Carlos Tenreiro

Notas do curso de Probabilidades

Coimbra, 2023

Junho de 2024
Versões anteriores: Jun. 2022, Jun. 2023

Nota prévia

As presentes notas foram escritas para servirem de texto de apoio às aulas de Probabilidades, disciplina do segundo ano, segundo semestre, da Licenciatura em Matemática da Faculdade de Ciências e Tecnologia da Universidade de Coimbra. Apesar dos assuntos aqui tratados, com exceção dos assinalados (*), corresponderem, no essencial, ao lecionado, as matérias completas, que incluem os exercícios que constam das folhas práticas da cadeira, foram expostas nas aulas.

Carlos Tenreiro
tenreiro@mat.uc.pt

Índice

1 Probabilidades: das origens ao Teorema do Limite Central **1**

1.1 Origens do cálculo de probabilidades 1
1.2 A Lei dos Grandes Números de Bernoulli 5
1.3 O Teorema do Limite Central de de Moivre 7
1.4 O Teorema do Limite Central de Laplace 8
1.5 O Teorema do Limite Central moderno 10
1.6 Bibliografia 11

2 Espaços de probabilidade **13**

2.1 Acontecimentos e σ-álgebra de acontecimentos 13
2.2 Axiomas da probabilidade 15
2.3 Propriedades da probabilidade 19
2.4 Probabilidade condicionada 25
2.4.1 Aplicação aos testes de diagnóstico 27
2.5 Independência de acontecimentos aleatórios 30
2.6 Bibliografia 33

3 Variáveis aleatórias **35**

3.1 Variáveis aleatórias e suas leis de probabilidade 35
3.2 Variáveis discretas e variáveis absolutamente contínuas 39
3.3 Função de distribuição 43
3.4 Transformação afim de variáveis aleatórias 50
3.5 Exemplos de leis de probabilidade 51
3.5.1 Leis discretas 51
3.5.2 Leis absolutamente contínuas 56
3.6 Transformação de variáveis com densidade* 62
3.7 Simulação de experiências e variáveis aleatórias 63

3.8 Bibliografia 65

4 Esperança matemática, variância e momentos **67**
4.1 Esperança matemática 67
4.1.1 Propriedades da esperança matemática 71
4.1.2 Cálculo da esperança matemática: alguns exemplos 76
4.2 Momentos e variância 79
4.2.1 Propriedades da variância 80
4.2.2 Cálculo da variância: alguns exemplos 81
4.2.3 Coeficientes de assimetria e de forma 84
4.3 Desigualdade de Markov e suas consequências 84
4.4 Localização e dispersão sem momentos 86
4.5 Bibliografia 87

5 Vetores aleatórios **89**
5.1 Introdução 89
5.2 Vetores aleatórios e suas leis de probabilidade 91
5.3 Vetores discretos e vetores absolutamente contínuos 96
5.4 Função de distribuição dum vetor aleatório 100
5.5 Independência de variáveis aleatórias 104
5.6 Distribuições condicionais 109
5.6.1 O caso discreto 109
5.6.2 O caso absolutamente contínuo 110
5.6.3 Esperança condicional 111
5.7 Covariância e correlação 114
5.8 Desigualdades de Hölder e de Minkowski* 118
5.9 Média e matriz de covariância dum vetor aleatório 119
5.10 Transformação de vetores com densidade* 121
5.11 Vetores aleatórios normais (parte I) 123
5.12 Bibliografia 127

6 Função caraterística **129**
6.1 Esperança matemática de variáveis complexas 129
6.2 Definição e primeiras propriedades 130
6.3 Cálculo da função caraterística: alguns exemplos 132
6.4 Função caraterística e momentos 134
6.5 Distribuição da soma de variáveis independentes 137
6.6 Função caraterística dum vetor aleatório* 139

6.7 Vetores aleatórios normais (parte II)* 140
6.8 Bibliografia 142

7 Leis dos Grandes Números **143**
7.1 Convergências funcionais 143
7.2 Primeiras leis dos grandes números 147
7.3 Leis de Khintchine e de Kolmogorov 151
7.4 Aplicações à Estatística 151
7.5 Duas ilustrações 153
7.6 Bibliografia 154

8 Teorema do Limite Central **157**
8.1 Introdução 157
8.2 Convergência em distribuição 158
8.3 O Teorema do Limite Central: dois exemplos 162
8.4 O Teorema do Limite Central clássico 165
8.5 Algumas consequências 167
8.6 Duas ilustrações 168
8.7 Bibliografia 172

Tabelas das distribuições Binomial, Poisson e Normal Standard **173**

Bibliografia **181**

Índice Remissivo **183**

1

Probabilidades: das origens ao Teorema do Limite Central

Origens do Cálculo de Probabilidades. A Lei dos Grandes Números de Bernoulli. O Teorema do Limite Central de de Moivre. O Teorema do Limite Central de Laplace. O Teorema do Limite Central moderno.

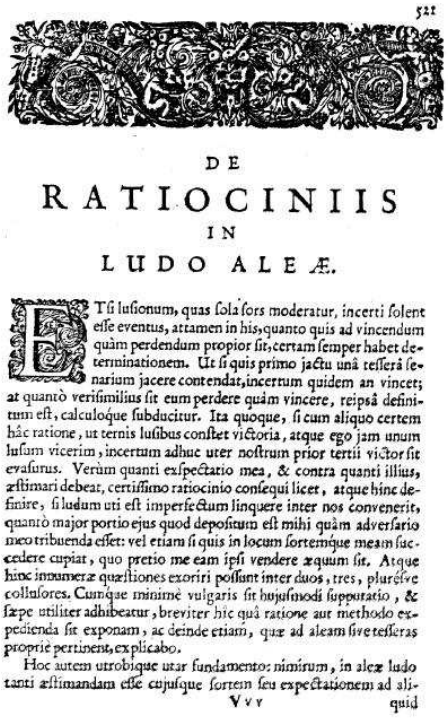
1.1 Origens do cálculo de probabilidades

A origem do cálculo de probabilidades é habitualmente situada em 1654, ano em que tem lugar uma célebre troca de correspondência entre Blaise Pascal (1623-1662) e Pierre de Fermat (1601-1665) sobre duas questões colocadas a Pascal por Antoine Gombaud (1607-1684), que ficará conhecido como chevalier de Méré ([1]). É, no entanto, verdade que muito antes desta data questões relativas à probabilidade de vitória em jogos de azar haviam sido analisadas por outros matemáticos como Luca Pacioli (1445-1517), Niccolò Tartaglia (1499-1557) ou Girolamo Cardano (1501-1576). Os dois problemas seguintes, razão principal da referida troca de correspondência, eram bem conhecidos quando, no verão de 1654, a mesma ocorre.

Problema 1.1.1. Quantos lançamentos de dois dados equilibrados devemos efetuar para que a "chance" (isto é, probabilidade) de obtermos pelo menos um par de 6 seja maior do que a de não obter nenhum par de 6?

Problema 1.1.2. Dois jogadores jogam uma série de partidas justas (isto é, ambos têm iguais possibilidade de ganhar cada partida) até que um deles obtenha 6 vitórias. Por motivos exteriores ao jogo, este é interrompido quando um dos jogadores soma 5 vitórias e o outro 3 vitórias. Como devemos dividir, de forma justa, o montante apostado por ambos os jogadores?

[1] Sobre a correspondência entre Pascal e Fermat, ver Hald (1990, pp. 54–63).



521

DE
RATIOCINIIS
IN
LUDO ALEÆ.

ETsi luſionum, quas ſola ſors moderatur, incerti ſolent eſſe eventus, attamen in his, quanto quis ad vincendum quàm perdendum propior ſit, certam ſemper habet determinationem. Ut ſi quis primo jactu unâ teſſerâ ſenarium jacere contendat, incertum quidem an vincet; at quantò veriſimilius ſit eum perdere quàm vincere, reipsâ definitum eſt, calculoque ſubducitur. Ita quoque, ſi cum aliquo certem hâc ratione, ut ternis luſibus conſtet victoria, atque ego jam unum luſum vicerim, incertum adhuc uter noſtrum prior tertii victor ſit evaſurus. Verùm quanti exſpectatio mea, & contra quanti illius, æſtimari debeat, certiſſimo ratiocinio conſequi licet, atque hinc definire, ſi ludum uti eſt imperfectum linquere inter nos convenerit, quantò major portio ejus quod depoſitum eſt mihi quàm adverſario meo tribuenda eſſet: vel etiam ſi quis in locum ſortemque meam ſuccedere cupiat, quo pretio me eam ipſi vendere æquum ſit. Atque hinc innumeræ quæſtiones exoriri poſſunt inter duos, tres, pluréſve colluſores. Cumque minimè vulgaris ſit hujuſmodi ſupputatio, & ſæpe utiliter adhibeatur, breviter hic quâ ratione aut methodo expedienda ſit exponam, ac deinde etiam, quæ ad aleam ſive teſſeras proprié pertinent, explicabo.

Hoc autem utrobique utar fundamento: nimirum, in aleæ ludo tanti æſtimandam eſſe cujuſque ſortem ſeu expectationem ad ali-

Figura 1.1: Primeira página da versão latina do livro de Christiaan Huygens (1629-1695) *De ratiociniis in aleae ludo* publicado em 1657.

Uma solução para este segundo problema, conhecido como problema da divisão das apostas, é pela primeira vez publicada, em 1657, por Christiaan Huygens (1629-1695), na obra *De ratiociniis in ludo aleae* (Sobre o raciocínio nos jogos de azar), primeiro livro publicado sobre probabilidades ([2]). A solução apresentada por Huygens é semelhante à de Pascal, mas esta só é publicada em 1665 na obra *Traité du triangle arithmétique*. Quase um século antes, por volta de 1564, também Cardano havia escrito o *Liber de ludo alea* (Livro sobre jogos de azar), mas este texto não é publicado antes de 1663 ([3]).

No final de *De ratiociniis in aleae ludo*, Huygens deixa ao leitor cinco problemas para os quais só indica a solução, não apresentando a respetiva resolução. Os dois problemas seguintes, são adaptações livres do primeiro e do último desses problemas ([4]). O segundo destes problemas ficou conhecido como problema da ruína do jogador.

[2]Sobre o livro de Huygens e a sua resolução do problema da divisão das apostas, ver Hald (1990, pp. 65–78). Sobre as primeiras tentativas para resolver o problema da divisão das apostas, ver Hald (1990, pp. 35–36).

[3]Sobre o texto de Cardano, ver Hald (1990, pp. 36–41).

[4]Sobre os cinco problema de Huygens, ver Hald (1990, pp. 74–78).


TRAITE
DV TRIANGLE
ARITHMETIQVE,
AVEC QVELQVES AVTRES
PETITS TRAITEZ SVR LA
MESME MATIERE.
Par Monſieur PASCAL.
A PARIS,
Chez GVILLAVME DESPREZ, ruë ſaint Iacques,
à Saint Proſper.
M. DC. LXV.

Figura 1.2: Página de título do *Traité du triangle arithmétique, avec quelques autres petits traitez sur la mesme matière* de Blaise Pascal (1623-1662), publicado em 1665.

Problema 1.1.3. Suponhamos que dois jogadores A e B lançam alternadamente dois dados nas condições seguintes: A ganha se tirar 6 pontos, B ganha se tirar 7 pontos, e é A que começa a lançar. Qual é a probabilidade de A ganhar?

Problema 1.1.4. Dois jogadores têm cada um 5 moedas e jogam com três dados. Se saem 11 pontos, o primeiro dá uma moeda ao segundo, e se saem 14 pontos, o segundo dá uma moeda ao primeiro. Ganha aquele que primeiro ficar com todas as moedas. Qual é a probabilidade de cada um deles ganhar?

Para todos os autores anteriores, a probabilidade de acontecimentos associados a uma experiência cujos resultados dependem do acaso, ditos, por isso, acontecimentos aleatórios, é igual ao quociente entre o número de casos favoráveis a esses acontecimentos e o número de casos possíveis da experiência aleatória em causa, em que se admite que o número de casos possíveis é finito e que todos têm igual possibilidade de ocorrer, ou seja, são equiprováveis. A esta forma de atribuir probabilidade associamos habitualmente o nome do matemático francês Pierre-Simon de Laplace (1749-1827). No entanto, no momento em que os problemas anteriores são colocados e resolvidos Laplace ainda não tinha nascido!

LIBER
DE LVDO
ALEÆ.

CAPVT PRIMVM.
De Ludorum generibus.

CAPVT II.
De Ludorum conditionibus.

CAPVT III
Quibus, & quando magis conueniat ludere.

Figura 1.3: Primeira página do *De ratiociniis in aleae ludo* de Girolamo Cardano (1501-1576) que é publicado após 1663.

Nesta altura inicial do cálculo de probabilidades era por todos assumido que a frequência relativa de um acontecimento aleatório, quando se repetia muitas vezes a experiência aleatória em causa, podia ser considerada uma boa aproximação para a probabilidade desse acontecimento. A propósito da interpretação frequencista do conceito de probabilidade, atentemos no problema seguinte que é colocado a Galileu Galilei (1564-1642), levando-o a escrever *Sopra le scoperte dei dadi* (Sobre uma descoberta acerca de dados) entre 1613 e 1623:

Problema 1.1.5. No lançamento de três dados equilibrados, 9 e 10 pontos podem ser obtidos de seis maneiras diferentes: 1 2 6, 1 3 5, 1 4 4, 2 2 5, 2 3 4, 3 3 3, e 1 3 6, 1 4 5, 2 2 6, 2 3 5, 2 4 4, 3 3 4, respetivamente. Como pode este facto ser compatível com a experiência que leva jogadores de dados a considerarem que a soma 9 ocorre menos vezes que a soma 10?


JACOBI BERNOULLI,
Profess. Basil. & utriusque Societ. Reg. Scientiar.
Gall. & Pruss. Sodal.
MATHEMATICI CELEBERRIMI,
ARS CONJECTANDI,
OPUS POSTHUMUM.
Accedit
TRACTATUS
DE SERIEBUS INFINITIS,
Et EPISTOLA Gallicè scripta
DE LUDO PILÆ
RETICULARIS.
BASILEÆ,
Impensis THURNISIORUM, Fratrum.
cIↃ IↃcc XIII.

Figura 1.4: Página de título da obra de Jakob Bernoulli (1645-1705) *Ars conjectandi* publicada em 1713.

1.2 A Lei dos Grandes Números de Bernoulli

Uma etapa maior na evolução do cálculo de probabilidades é dada pela publicação póstuma, em 1713, da obra *Ars conjectandi* (A arte da conjetura) de Jakob Bernoulli (1655-1705) ([5]). Neste livro generaliza-se o problema da divisão das apostas ao caso em que o jogo não é necessariamente justo. Denotando por S_n o número de vitórias (sucessos) obtidas em n partidas pelo jogador que tem uma probabilidade $p \in]0,1[$ de sucesso por partida, Bernoulli mostra que S_n possui uma distribuição binomial, isto é, que a probabilidade de obter exatamente $k \in \{0,1,\ldots,n\}$ vitórias é dada por

$$P\big(S_n = k\big) = C_k^n p^k (1-p)^{n-k}, \tag{1.2.1}$$

onde C_k^n representa as combinações de n elementos tomados k a k, isto é,

$$C_k^n = \frac{n!}{(n-k)!k!}.$$

[5] Sobre a *Ars conjectandi*, ver Hald (1990, pp. 223–225).

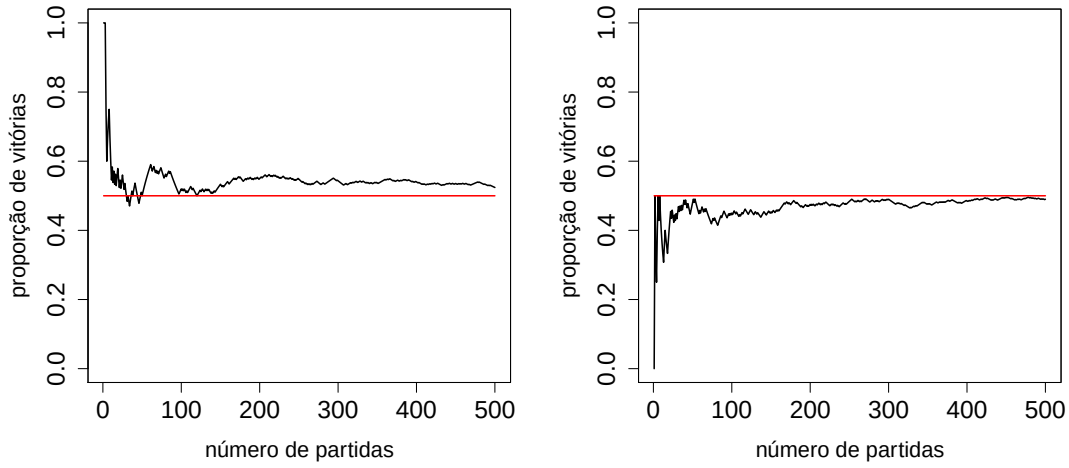


Figura 1.5: Evoluções da proporção de vitórias obtidas por um jogador que joga 500 partidas justas.

Além disso, Bernoulli estabelece o primeiro teorema limite (isto é, quando n tende para infinito) da teoria das probabilidades, provando que para todo o $\epsilon > 0$,

$$P\left(\left|\frac{1}{n}S_n - p\right| < \epsilon\right) \to 1,\ n \to \infty,$$

resultado este que é conhecido como lei dos grandes números de Bernoulli ([6]). A ele voltaremos no Capítulo 7. A expressão lei dos grandes números é introduzida muito mais tarde, em 1837, por Siméon Denis Poisson (1781-1840) ([7]). Verificando-se a convergência anterior, diremos também que S_n/n converge em probabilidade para p, quando $n \to \infty$. Por outras palavras, a probabilidade de sucesso por partida (isto é, p) pode ser aproximada, com a precisão que queiramos, pela frequência relativa (isto é, S_n/n) dos sucessos observados numa longa sucessão de partidas ($n \to \infty$).

Este resultado permite assim justificar a já referida interpretação frequencista para a noção de probabilidade (ver Figura 1.5), estabelecendo que a probabilidade de ocorrência de um acontecimento (neste caso, vitória na partida) pode ser aproximada pela frequência relativa da ocorrência desse mesmo acontecimento, quando repetimos, um grande número de vezes, a experiência aleatória em causa (ou seja, quando jogamos muitas partidas).

[6] Sobre a forma como o resultado é formulado e demonstrado por Bernoulli, ver Hald (1990, pp. 257–264).

[7] Poisson, S.D. (1837). *Recherches sur la probabilité des judgements en matière criminale et matière civile*, Paris.

1.3 O Teorema do Limite Central de de Moivre

Em 1718 é publicada a primeira edição da obra *Doctrine of chances* de Abraham de Moivre (1667-1754). Nas segunda e terceira edições da obra, publicadas em 1738 e 1756, surge um melhoramento da lei dos grandes números de Bernoulli obtido por de Moivre em 1733 ([8]). Continuando a denotar por S_n o número de vitórias (sucessos) obtidas em n partidas por um jogador que tem uma probabilidade $p \in]0,1[$ de sucesso por partida, de Moivre usa a fórmula de Stirling ([9])

$$n! \sim \sqrt{2\pi}\, n^{n+1/2} e^{-n},$$

onde o símbolo $\sim$ significa que o quociente entre as duas quantidades converge para 1, quando n tende para infinito, para mostrar que a probabilidade (1.2.1) pode ser aproximada por ([10])

$$P\big(S_n = k\big) \approx \frac{1}{\sqrt{2\pi np(1-p)}}\, e^{-(k-np)^2/(2np(1-p))},$$

quando n é grande, daqui deduzindo que

$$\begin{aligned} P\left(\left|\frac{1}{n}S_n - p\right| < \epsilon\right) &= \sum_{|k-np|<n\epsilon} P(S_n = k) \\ &\approx \int_{np-n\epsilon}^{np+n\epsilon} \frac{1}{\sqrt{2\pi np(1-p)}}\, e^{-(x-np)^2/(2np(1-p))}\, dx \\ &= \int_{-\gamma}^{\gamma} \frac{1}{\sqrt{2\pi}}\, e^{-t^2/2}\, dt, \end{aligned} \tag{1.3.1}$$

onde

$$\gamma = \epsilon\sqrt{\frac{n}{p(1-p)}}.$$

Este resultado limite, dito teorema do limite central de de Moivre, é a primeira versão de um conjunto de resultados que serão, muito mais tarde, conhecidos pela designação comum de Teoremas do Limite Central. Deste resultado podemos deduzir a lei dos grandes números de Bernoulli, mas também, e isso não decorre do resultado de Bernoulli, uma fórmula para aproximar a probabilidade (1.3.1) que envolve o integral, extendido ao intervalo $]-\gamma,\gamma[$, da função

$$f(t) = \frac{1}{\sqrt{2\pi}}\, e^{-t^2/2},\ t \in \mathbb{R},$$

[8] Moivre, A. de (1733). *Approximatio ad summam terminorum binomii* $(a+b)^n$ *in seriem expansi.*

[9] Stirling, J. (1730). *Methodus Differentialis*. London.

[10] Sobre a aproximação normal de de Moivre para a distribuição binomial, ver Hald (1990, pp. 485–495).

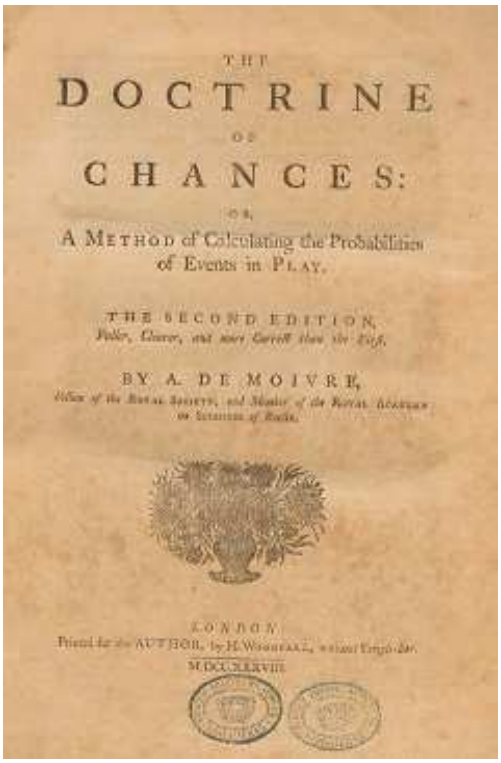

THE
DOCTRINE
OF
CHANCES:
OR,
A METHOD of Calculating the Probabilities
of Events in PLAY.
THE SECOND EDITION,
BY A. DE MOIVRE,
LONDON
M.DCC.XXXVIII

Figura 1.6: Página de título da segunda edição da obra de Abraham de Moivre (1667-1754) *Doctrine of chances*, publicada em 1738.

que agora designamos por densidade normal standard, ou curva de Gauss-Laplace, função esta que, por esta via, desempenha um papel central nas Probabilidades bem como nas suas aplicações à Estatística.

1.4 O Teorema do Limite Central de Laplace

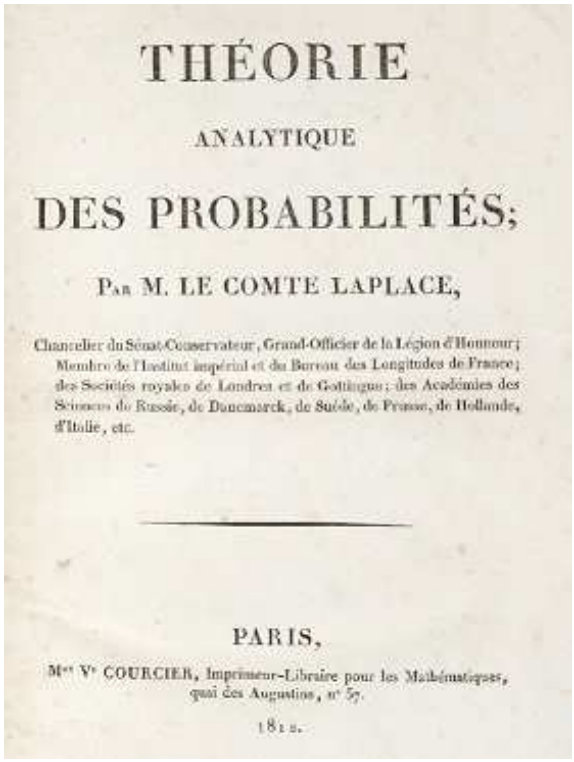
Nos resultados anteriores, o número obtido de vitórias quando realizamos n partidas pode ser escrito na forma

$$S_n = X_1 + X_2 + \cdots + X_n,$$

onde

$$X_i = \begin{cases} 1, & \text{se há sucesso na } i\text{-ésima partida} \\ 0, & \text{caso contrário} \end{cases}$$

A estas funções X_i, que ao resultado de uma experiência aleatória associam dois valores possíveis, 0 ou 1, chamamos variáveis aleatórias de Bernoulli de parâmetro p, sendo p a probabilidade de ocorrência do valor 1 (sucesso). Os resultados estabelecidos



THÉORIE

ANALYTIQUE

DES PROBABILITÉS;

PAR M. LE COMTE LAPLACE,

Chancelier du Sénat-Conservateur, Grand-Officier de la Légion d'Honneur; Membre de l'Institut impérial et du Bureau des Longitudes de France; des Sociétés royales de Londres et de Gottingue; des Académies des Sciences de Russie, de Danemarck, de Suède, de France, de Hollande, d'Italie, etc.

PARIS,

Mme Ve COURCIER, Imprimeur-Libraire pour les Mathématiques, quai des Augustins, n° 57.

1812.

Figura 1.7: Página de título da primeira edição da *Théorie analytique des probabilités* de Pierre-Simon de Laplace (1749-1827), publicada em 1812.

por Bernoulli e de Moivre valem assim para somas S_n muito particulares, em que as respetivas parcelas X_i são variáveis aleatórias de Bernoulli.

Generalizações dos teoremas de Bernoulli e de de Moivre ocorrem em todo o período de 1812 a 1935. No caso particular do teorema de de Moivre, uma primeira generalização para soma gerais é obtida por Pierre-Simon de Laplace (1749-1827) em 1810, surgindo pouco depois na obra *Théorie analytique des probabilités*, publicada em 1812 ([11]). Sendo $S_n = X_1 + X_2 + \cdots + X_n$, onde $X_1, X_2, \ldots, X_n$ representam observações "independentes" de uma variável quantitativa aleatória X (peso, altura, etc.), então para todo o $-\infty < a < b < \infty$ vale a convergência

$$P\left(a < \frac{S_n - n\mu}{\sqrt{n}\,\sigma} < b\right) \to \int_a^b \frac{1}{\sqrt{2\pi}}\, e^{-t^2/2} dt,\ n \to \infty,$$

onde μ e $\sigma > 0$ são números reais a que chamaremos média e desvio-padrão da variável observada X.

[11] Laplace, P.S. (1810). Mémoire sur les approximations des formules qui sont fonctions de très grands nombres d'observations, et sur leur applications aux probabilités. *Mém. Acad. Sci. Paris* 10, 353–415. Sobre o teorema do limite central de Laplace, ver Hald (1998, pp. 307-317).

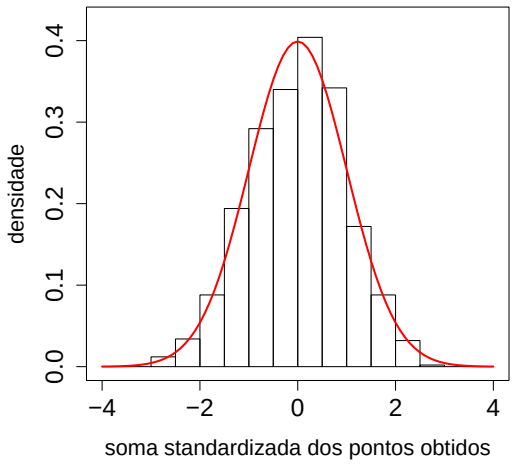


Figura 1.8: Distribuição da soma standardizada dos pontos obtidos em 1000 lançamentos de 5 dados equilibrados e densidade normal standard.

Este resultado assintótico, a que chamamos teorema do limite central de Laplace, é ilustrado na Figura 1.8 onde se apresenta a distribuição da soma standardizada dos pontos obtidos em 1000 lançamentos de cinco dados equilibrados, descrita pelo respetivo histograma, bem como a densidade da normal standard. Neste caso, a soma dos pontos ocorridos nos cinco dados é dada por $S_5 = X_1 + \cdots + X_5$, onde X_i representa o número de pontos obtidos no i-ésimo dado, e a soma standardizada é definida por $(S_5 - 5\mu)/(\sqrt{5}\,\sigma)$, com $\mu = 21/6$ e $\sigma = \sqrt{91}/6$ (mais à frente veremos a razão para a escolha destes valores de μ e σ). Mesmo para um valor pequeno de n, vemos que a distribuição normal standard dá uma razoável aproximação para a distribuição da soma standardizada dos pontos obtidos nos cinco dados equilibrados.

1.5 O Teorema do Limite Central moderno

A demonstração de Laplace do resultado anterior apresenta imprecisões, não sendo válida para variáveis quantitativas gerais. A primeira prova rigorosa do chamado teorema do limite central de Laplace-Liapounov é atribuída a Alexandre Liapounov (1857-1918) num trabalho publicado em 1900, que surge na sequência dos trabalhos de Pafnuty Tchebychev (1821-1894) e Andrei Markov (1856-1922), datados de 1887 e 1898, respetivamente ([12]). O teorema do limite central de Laplace-Liapounov será estudado no

[12] Tchebycheff, P.L. (1889). Sur les résidus intégraux qui donnent des valeurs approchées des intégrales. *Acta Mathematica* 12, 287–322 (tradução francesa do original russo). Markov, A.A. (1898). Sur les racines de l'équation $e^{x^2}\frac{d^m e^{-x^2}}{dx^m} = 0$. *Bull. Acad. Sci. St. Petersbourg* 9, 435–446. Liapounov, A. (1900). Sur une proposition de théorie des probabilités. *Bull. Acad. Sci. St. Petersbourg* 13,

Capítulo 8. Generalizações do teorema de Liapounov são obtidas pelo próprio Liapounov em 1901, e mais tarde por Jarl Lindeberg (1876-1932), em 1922 ([13]), e Paul Lévy (1886-1971) e William Feller (1906-1970), em 1935 ([14]). A partir de 1920 estes e outros resultados, que generalizam de diversas formas o teorema limite de de Moivre, passam a ser conhecidos pela designação comum, devida ao matemático George Pólya (1887-1985), de Teorema do Limite Central, onde a palavra "central" enfatiza o papel fulcral de tais resultados na teoria das Probabilidades e nas suas aplicações à Estatística ([15]).

1.6 Bibliografia

Adams, W.J. (2009). *The life and times of the central limit theorem.* American Mathematical Society, London Mathematical Society.

Alexandrov, A.D., Kolmogorov, A.N., Lavrentiev, M.A. (2020). *Mathématiques, leur contenu, leurs méthodes, leur signification.* Saint-Cyr-sur-Mer: Les Éditions du Bec de l'Aigle.

Deheuvels, P. (1990). *La probabilité de hasard et la certitude.* Paris: Presses Universitaires de France.

Feller, W. (1945). The fundamental limit theorems in probability. *Bull. Amer. Math. Soc.* 51, 800–832.

Hald, A. (1990). *A history of probability and statistics and their applications before 1750.* New York: Wiley.

Hald, A. (1998). *A history of mathematical statistics from 1750 to 1930.* New York: Wiley.

Le Cam, L. (1986). The central limit theorem around 1935. *Statist. Sci.* 1, 78–91.

359–386.

[13] Liapounov, A. (1901). Nouvelle forme du théorème sur la limite de probabilité. *Mémoires de l'Académie de St. Pétersbourg* 12, 1–24. Lindeberg, J.W. (1922). Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung. *Mathematische Zeitschrift* 15, 211–225.

[14] Lévy, P. (1935). Proprietés asymptotiques des sommes de variables indépendantes ou enchaínées. *J. Math. Pures Appli.* 14, 347–402. Feller, W. (1936). Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung. *Mathematische Zeitschrift* 40, 521–559.

[15] Sobre a história mais recente do teorema do limite central, ver Le Cam (1986).

2

Espaços de probabilidade

Acontecimentos e σ-álgebra de acontecimentos. Axiomas da probabilidade. σ-álgebra de Borel e medida de Lebesgue. Propriedades duma probabilidade. Probabilidade condicionada. Testes de diagnóstico. Independência de acontecimentos aleatórios.

2.1 Acontecimentos e σ-álgebra de acontecimentos

Em 1933, com a publicação de *Grundbegriffe der wahrscheinlichkeitrechnung* (*Fundamentos da teoria da probabilidade*), Andrei Kolmogorov (1903-1987) estabelece as bases axiomáticas do cálculo das probabilidades. O modelo proposto por Kolmogorov permitiu associar o cálculo das probabilidades à teoria da medida e da integração, possibilitando assim a utilização dos resultados e técnicas da análise matemática no desenvolvimento da teoria das probabilidades ([1]).

No modelo proposto por Kolmogorov, ao conjunto das realizações possíveis duma experiência aleatória associamos um conjunto Ω, a que chamamos espaço dos resultados ou espaço fundamental, em que cada elemento $\omega \in \Omega$ carateriza completamente uma realização possível da experiência aleatória. De uma forma genérica, entendemos por experiência aleatória, qualquer experiência que goza das seguintes propriedades: 1) pode repetir-se, mesmo que hipoteticamente, nas mesmas condições, ou em condições muito semelhantes; 2) o resultado observado em cada uma dessas repetições é um de entre um conjunto de resultados possíveis conhecidos antes de realizar a experiência; 3) esse resultado é consequência dum conjunto de fatores que não podemos, na totalidade, controlar, e que atribuímos ao acaso. Os acontecimentos aleatórios associados à experiência aleatória são identificados com subconjuntos do espaço fundamental, associando a cada acontecimento o conjunto dos pontos $\omega \in \Omega$ que correspondem a resultados da experiência aleatória favoráveis à realização desse acontecimento. Como casos extremos temos o acontecimento impossível e o acontecimento certo representados naturalmente

[1] Sobre o modelo proposto por Kolmogorov, ver Kolmogorov (1950, pp. 1–6).

pelos conjuntos $\emptyset$ e Ω, respetivamente. Os subconjuntos singulares de Ω dizem-se acontecimentos elementares.

Exemplo 2.1.1. Concretizemos este procedimento considerando a experiência aleatória que consiste no lançamento de um dado equilibrado. Representando por "i" a ocorrência da face com "i" pontos, o espaço dos resultados é $\Omega = \{1, 2, 3, 4, 5, 6\}$. Os acontecimentos aleatórios $A =$"saída de número par", $B =$"saída de número inferior a 3", etc., podem ser identificados com os subconjuntos do espaço dos resultados $A = \{2, 4, 6\}$, $B = \{1, 2\}$, etc., respectivamente.

As operações usuais entre conjuntos, reunião, interseção, complementação e diferença, entre outras, permitem exprimir ou construir acontecimentos a partir de outros acontecimentos. Alguns exemplos:

— $A \cup B \equiv$ representa o acontecimento que se realiza quando pelo menos um dos acontecimentos A ou B se realiza;

— $A \cap B \equiv$ acontecimento que se realiza quando A e B se realizam;

— $A^c \equiv$ acontecimento contrário ao acontecimento A, realizando-se quando A não se realiza;

— $A - B \equiv$ acontecimento que se realiza quando A se realiza e B não se realiza.

Também a inclusão, $A \subset B$, do conjunto A no conjunto B, é traduzida para o contexto dos acontecimentos significando que a realização do acontecimento A implica a realização do acontecimento B.

Por vezes é também útil lidar com acontecimentos definidos a partir de um número infinito de outros acontecimentos aleatórios:

— $\bigcup_{n=1}^{\infty} A_n \equiv$ acontecimento que se realiza quando pelo menos um dos acontecimentos A_n se realiza;

— $\bigcap_{n=1}^{\infty} A_n \equiv$ acontecimento que se realiza quando todos os acontecimentos A_n se realizam;

Exemplo 2.1.2. Suponhamos, por exemplo, que lançamos sucessivamente um dado equilibrado e que representamos por A_n a saída de número par no n-ésimo lançamento. Se lançarmos o dado m vezes e estivermos interessados nos acontecimentos "saída de número par em pelo menos um dos m lançamentos do dado" e "saída de número par em todos os m lançamentos do dado", a cada um destes acontecimentos podemos associar os conjuntos $\bigcup_{n=1}^{m} A_n$ e $\bigcap_{n=1}^{m} A_n$, respetivamente. Se imaginarmos que continuamos a lançar o dado e que estamos interessados nos acontecimentos "saída de número par em pelo menos um dos lançamentos do dado" e "saída de número par em todos os

lançamentos do dado", a estes acontecimentos deveremos associar os conjuntos $\bigcup_{n=1}^{\infty} A_n$ e $\bigcap_{n=1}^{\infty} A_n$, respetivamente.

2.2 Axiomas da probabilidade

Em resposta às perguntas "qual é a probabilidade de sair um número par no lançamento de um dado?" e "qual é a probabilidade de sair um número múltiplo de 3 no lançamento de um dado?", esperamos associar a cada um dos conjuntos $\{2,4,6\}$ e $\{3,6\}$, um número real que exprima a maior ou menor possibilidade de tais acontecimentos ocorrerem. Uma forma natural de o fazer, será associar a um acontecimento a proporção de vezes que esperamos que esse acontecimento ocorra em sucessivas repetições da experiência aleatória. Sendo o dado equilibrado, e atendendo a que em sucessivos lançamentos do mesmo esperamos que o acontecimento $\{2,4,6\}$ ocorra três vezes em cada seis lançamentos e que o acontecimento $\{3,6\}$ ocorra duas vezes em cada seis lançamentos, poderíamos ser levados a associar ao primeiro acontecimento o número $3/6$ e ao segundo o número $2/6$.

A definição de probabilidade de Kolmogorov que a seguir apresentamos, é motivada por considerações do tipo anterior relacionadas com o conceito frequencista de probabilidade, isto é, com as propriedades da frequência relativa de acontecimentos aleatórios em sucessivas repetições duma experiência aleatória. Em particular, se por $P(A)$ denotarmos a probabilidade do acontecimento A, $P(A)$ deverá ser um número real do intervalo $[0,1]$, com $P(\Omega) = 1$ e $P(A \cup B) = P(A) + P(B)$, se A e B são incompatíveis, isto é, se $A \cap B = \emptyset$. Estamos agora já muito perto de noção de probabilidade considerada por Kolmogorov. Além da propriedade de aditividade sobre P, Kolmogorov assume que P é σ-aditiva, isto é, que a propriedade de aditividade anterior é válida para uma qualquer família numerável de acontecimentos aleatórios dois a dois incompatíveis. O domínio natural de definição duma tal aplicação é uma classe $\mathscr{A}$ de partes de Ω a que chamamos σ-álgebra de partes de Ω.

Definição 2.2.1. *Uma classe $\mathscr{A}$ de partes de um conjunto Ω diz-se uma σ-álgebra de partes de Ω se contém o conjunto vazio e é fechada para a complementação e para a reunião numerável:*

a) $\emptyset \in \mathscr{A}$;

b) Para todo o $A \in \mathscr{A}$, então $A^c \in \mathscr{A}$;

c) Para toda a sucessão $A_n \in \mathscr{A}$, $n = 1, 2, \ldots$, então $\bigcup_{n=1}^{\infty} A_n \in \mathscr{A}$.

Das propriedades anteriores deduzimos facilmente que uma σ-álgebra contém Ω, e é fechada para a interseção numerável bem como para a interseção e reunião finitas.

A classe $\mathscr{P}(\Omega)$ de todos os subconjuntos do espaço dos resultados é uma σ-álgebra de partes de Ω. Quando Ω é finito ou numerável, é essa a σ-álgebra de acontecimentos que habitualmente se associa ao espaço dos resultados Ω. No entanto, quando o espaço dos resultados Ω é infinito não-numerável, por exemplo $\mathbb{R}$ ou um intervalo real, a classe $\mathscr{P}(\Omega)$ pode ser demasiado grande para nela podermos definir uma probabilidade no sentido da definição seguinte. Nesse caso os acontecimentos aleatórios associados ao espaço dos resultados Ω não podem ser identificados com toda a classe $\mathscr{P}(\Omega)$, mas apenas com uma subclasse desta que seja uma σ-álgebra.

Definição 2.2.2. *Uma probabilidade P sobre uma σ-álgebra $\mathscr{A}$ de partes de Ω é uma aplicação de $\mathscr{A}$ em $[0,1]$ tal que:*

a) $P(\Omega) = 1$;

b) Para toda a sucessão de conjuntos $A_n \in \mathscr{A}$, $n = 1, 2, \ldots$, disjuntos dois a dois,

$$P\Big(\bigcup_{n=1}^{\infty} A_n\Big) = \sum_{n=1}^{\infty} P(A_n) \ (\sigma\text{-}aditividade).$$

Ao terno $(\Omega, \mathscr{A}, P)$ chamamos espaço de probabilidade. Quando a uma experiência aleatória associamos o espaço de probabilidade $(\Omega, \mathscr{A}, P)$, dizemos também que este espaço é um modelo probabilístico para a experiência aleatória em causa. Os elementos de $\mathscr{A}$ dizem-se acontecimentos aleatórios. Notemos que quando o espaço dos resultados é finito, a condição b) é equivalente à aditividade de P, isto é, $P(A \cup B) = P(A) + P(B)$, sempre que os acontecimentos A e B sejam incompatíveis.

Como veremos na secção seguinte, a probabilidade do acontecimento impossível é igual a zero, isto é, $P(\emptyset) = 0$. Podemos assim dizer que uma probabilidade é uma função definida na σ-álgebra $\mathscr{A}$ de partes de um conjunto Ω, que é não negativa, σ-aditiva com $P(\emptyset) = 0$. Usando a linguagem da análise matemática, isto significa que uma probabilidade é uma medida em $\mathscr{A}$. Um exemplo muito importante de uma medida em $\mathbb{R}$ é-nos dado pela chamada medida de Lebesgue, que herda o nome do matemático Henri Lebesgue (1875-1941). Dado um intervalo $]a, b]$, com $-\infty < a \leq b < \infty$, a medida de Lebesgue de $]a, b]$, que denotamos por $\lambda(]a, b])$, é dada pelo comprimento do intervalo $]a, b]$, isto é,

$$\lambda\big(]a, b]\big) = b - a. \tag{2.2.3}$$

É possível provar que a medida de Lebesgue pode ser definida numa σ-álgebra de partes de $\mathbb{R}$, que denotamos por $\mathscr{B}(\mathbb{R})$, a que chamamos σ-álgebra de Borel em $\mathbb{R}$ em homenagem ao matemático Émile Borel (1871-1956). Esta σ-álgebra é, por definição, a mais pequena σ-álgebra de partes de $\mathbb{R}$ que contém os subconjuntos abertos de $\mathbb{R}$. Dizemos, por isso, que se trata da σ-álgebra gerada pelos abertos de $\mathbb{R}$. A σ-álgebra

de Borel é também a σ-álgebra gerada pelos conjuntos fechados de $\mathbb{R}$, ou ainda a σ-álgebra gerada por todos os intervalos da forma $]a, b]$, com $-\infty < a \leq b < \infty$. A esta σ-álgebra pertencem os conjuntos $\emptyset$ e $\mathbb{R}$, os conjuntos singulares, os conjuntos finitos ou numeráveis, todos os intervalos em $\mathbb{R}$ (finitos ou infinitos, abertos ou fechados, etc), bem como reuniões finitas ou numeráveis destes. Os subconjuntos de $\mathbb{R}$ pertencentes a esta σ-álgebra são chamados borelianos. Quando dissemos que é possível definir λ em $\mathscr{B}(\mathbb{R})$, estamos efetivamente a dizer que é possível provar que existe apenas uma medida definida em $\mathscr{B}(\mathbb{R})$ que satisfaz a propriedade (2.2.3). Assim, a todo o $A \in \mathscr{B}(\mathbb{R})$ é possível associar o número real $\lambda(A)$ a que chamamos medida de Lebesgue, ou comprimento, de A.

Podemos generalizar a definição de medida de Lebesgue ao espaço multidimensional $\mathbb{R}^d$, com $d \geq 2$. Dado um retângulo $]a, b]$ em $\mathbb{R}^d$, isto é,

$$]a, b] :=]a_1, b_1] \times \cdots \times]a_d, b_d],$$

com $-\infty < a_i \leq b_i < \infty$, a medida de Lebesgue de $]a, b]$ é dada pelo volume de $]a, b]$, isto é,

$$\lambda\big(]a, b]\big) = (b_1 - a_1) \times \cdots \times (b_d - a_d).$$

De forma análoga ao caso real, é possível demonstrar que a medida de Lebesgue em $\mathbb{R}^d$ está definida na σ-álgebra de Borel de $\mathbb{R}^d$, isto é, na σ-álgebra gerada pelos abertos de $\mathbb{R}^d$, que denotamos por $\mathscr{B}(\mathbb{R}^d)$. Uma vez que a cada conjunto $A \in \mathscr{B}(\mathbb{R}^d)$ podemos associar a sua medida (ou volume) $\lambda(A)$, dizemos que os conjuntos de $\mathscr{B}(\mathbb{R}^d)$ são mensuráveis.

A axiomatização da noção de probabilidade, não resolve o problema da atribuição de probabilidade aos acontecimentos de uma experiência aleatória particular; apenas fixa as regras gerais a que uma tal atribuição deve satisfazer. Nos exemplos que a seguir consideramos, a associação dum modelo probabilístico às experiências aleatórias que descrevemos pode ser feita de forma simples.

Exemplo 2.2.4. Retomando o exemplo do lançamento de um dado equilibrado, como todos os elementos de $\Omega = \{1, 2, 3, 4, 5, 6\}$ têm a mesma possibilidade de ocorrer, será natural tomar P definida em $\mathscr{A} = \mathscr{P}(\Omega)$ por $P(\{\omega\}) = 1/6$, para $\omega \in \Omega$. Duma forma geral, se o espaço Ω dos resultados duma experiência aleatória é finito e todos eles têm a mesma possibilidade de ocorrer, será natural tomar

$$P(A) = \frac{\#A}{\#\Omega}, \text{ para } A \subset \Omega, \tag{2.2.5}$$

isto é,

$$P(A) = \frac{\textit{número de resultados favoráveis a } A}{\textit{número de resultados possíveis}},$$

que não é mais do que a definição de Laplace de probabilidade.

Em consequência da propriedade de aditividade duma probabilidade, que estabeleceremos na próxima secção (Proposição 2.3.2), quando o espaço Ω dos resultados duma experiência aleatória é finito e todos esses resultados têm a mesma possibilidade de ocorrer, a expressão (2.2.5) é a única forma de atribuir probabilidade que é compatível com a Definição 2.2.2.

Exemplo 2.2.6. (cont. do Problema 1.1.5) Como já explicámos, a definição de probabilidade de Laplace só pode ser usada quando os resultados elementares da experiência aleatória em causa tiverem a mesma possibilidade de ocorrer (e forem em número finito). Tal não acontece com os resultados da experiência que são apresentados a Galileu. Os seis resultados, ditos favoráveis a cada uns dos acontecimentos em causa (A = "saída de soma 9" e B = "saída de soma 10"), não têm a mesma possibilidade de ocorrer. Por exemplo, "escondidos" no resultado 126 que é apresentado a Galilei estão seis resultados elementares da experiência, a saber, 126, 162, 216, 261, 612 e 621 (os primeiro, segundo e terceiro algarismos representam os resultados ocorridos no primeiro, segundo e terceiro dados), o segundo, estes sim com igual possibilidade de ocorrerem (pois estamos a assumir que os dados são equilibrados). De forma análoga, "escondidos" no resultado 144 que é apresentado a Galilei estão três resultados elementares da experiência com iguais possibilidades de ocorrerem, a saber, 144, 414 e 441. Em conclusão, para modelarmos convenientemente a experiência aleatória apresentada a Galileu devemos tomar para espaço fundamental

$$\Omega = \{ijk : i, j, k = 1, \ldots, 6\},$$

onde ijk representa a ocorrência de face i primeiro dado, j no segundo, e k no terceiro. Uma vez que $\#\Omega = 6^3 = 216$, $\#A = 25$ e $\#B = 27$, concluímos que $P(A) = 25/216 \approx 0{,}1157$ e $P(B) = 27/216 = 0{,}125$. À luz da lei dos grandes números de Bernoulli, a que aludimos em §1.2, fica assim justificado o facto da soma 9 ocorrer menos vezes do que a soma 10.

Exemplo 2.2.7. Suponhamos que extraímos ao acaso um ponto do intervalo real $[a, b]$. Neste caso $\Omega = [a, b]$. Sendo o número de resultados possíveis infinito, não podemos proceder como no exemplo anterior. No entanto, como intervalos com igual comprimento têm a mesma possibilidade de conter o ponto extraído, será natural tomar para probabilidade dum subintervalo $]c, d]$ de $[a, b]$, o quociente entre o seu comprimento e o comprimento de $[a, b]$, isto é,

$$P(]c, d]) = \frac{d - c}{b - a} = \frac{\lambda(]c, d])}{\lambda([a, b])},$$

para $a \leq c < d \leq b$. Mais geralmente, se Q é uma região mensurável de $\mathbb{R}^d$ com volume $\lambda(Q) \in]0, +\infty[$, onde λ é a medida de Lebesgue em $\mathbb{R}^d$, a extração ao acaso dum ponto de Q pode ser modelada pela probabilidade

$$P(A) = \frac{\textit{volume de } A}{\textit{volume de } Q} = \frac{\lambda(A)}{\lambda(Q)}, \text{ para } A \in \mathscr{B}(Q),$$

a que chamamos probabilidade geométrica.

Como veremos na próxima secção, decorre da Definição 2.2.2 que a probabilidade do acontecimento impossível é igual a zero (Proposição 2.3.1). Quando o número de resultados da experiência é finito e todos os resultados têm a mesma possibilidade de ocorrer, o recíproco é também verdadeiro. Com efeito, da igualdade (2.2.5) concluímos que o facto dum acontecimento A ter probabilidade zero implica naturalmente que $\#A = 0$, ou seja, o acontecimento A é necessariamente o acontecimento impossível. Como podemos verificar no caso do modelo de probabilidade descrito no Exemplo 2.2.7, esta conclusão não é válida em geral. Por exemplo, se tomarmos para Q o quadrado $[0,1] \times [0,1] \subset \mathbb{R}^2$, para o qual $\lambda(Q) = 1$, verificamos que o acontecimento $A = \{(x,y) \in Q : x = y\}$ tem probabilidade $P(A) = \lambda(A) = 0$ (visto ter volume, isto é, área, igual a zero), e A não é o acontecimento impossível.

2.3 Propriedades da probabilidade

As propriedades seguintes são válidas para uma qualquer probabilidade definida numa σ-álgebra $\mathscr{A}$ de partes do espaço dos resultados Ω. Como já vimos, os subconjuntos de Ω pertencentes a $\mathscr{A}$ são identificados com os acontecimentos aleatórios da experiência aleatória que é descrita pelo espaço de probabilidade $(\Omega, \mathscr{A}, P)$.

Proposição 2.3.1. *A probabilidade do acontecimento impossível é nula*

$$P(\emptyset) = 0.$$

Dem: Consideremos a sucessão (A_n) de acontecimentos, incompatíveis dois a dois, definida por $A_1 = \Omega$ e $A_n = \emptyset$, para $n \geq 2$. Sendo a probabilidade σ-aditiva, concluímos que

$$P(\Omega) = P\Big(\bigcup_{n=1}^{\infty} A_n\Big) = \sum_{n=1}^{\infty} P(A_n) = P(\Omega) + \sum_{n=2}^{\infty} P(\emptyset),$$

de onde deduzimos que $P(\emptyset) = 0$, pois $P(\Omega) < \infty$. ■

Proposição 2.3.2 (aditividade)**.** *Se $A_1, \ldots, A_k$, com $k \geq 2$, são acontecimentos aleatórios incompatíveis dois a dois, então*

$$P\Big(\bigcup_{n=1}^{k} A_n\Big) = \sum_{n=1}^{k} P(A_n).$$

Em particular, se A e B são acontecimentos incompatíveis, então

$$P(A \cup B) = P(A) + P(B).$$

Dem: Consideremos a sucessão (B_n) de acontecimentos, incompatíveis dois a dois, definida por $B_n = A_n$, para $n = 1, \ldots, k$, e $B_n = \emptyset$, para $n > k$. Assim, usando a σ-aditividade de P e a Proposição 2.3.1, concluímos que

$$P\Big(\bigcup_{n=1}^{k} A_n\Big) = P\Big(\bigcup_{n=1}^{\infty} B_n\Big) = \sum_{n=1}^{\infty} P(B_n) = \sum_{n=1}^{k} P(A_n). \qquad \blacksquare$$

Na proposição seguinte usaremos, pela primeira vez, o facto duma probabilidade verificar $P(\Omega) = 1$.

Proposição 2.3.3. *A probabilidade do acontecimento contrário ao acontecimento A é dada por*

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

Dem: Tomando na Proposição 2.3.2 $A_1 = A$, $A_2 = A^c$ e $k = 2$, obtemos

$$1 = P(\Omega) = P\big(A \cup A^c\big) = P(A) + P(A^c). \qquad \blacksquare$$

Proposição 2.3.4. *Se A e B são dois acontecimentos quaisquer então*

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

Além disso, se $A \subset B$ então

$$P(B - A) = P(B) - P(A).$$

Dem: Basta usar a aditividade de P e ter em conta que $(B - A) \cup (A \cap B) = B$, com $B - A$ e $A \cap B$ acontecimentos incompatíveis. $\blacksquare$

Proposição 2.3.5 (monotonia)**.** *Se A e B são acontecimentos com $A \subset B$ então*

$$P(A) \leq P(B).$$

Dem: Decorre imediatamente da segunda parte da proposição anterior. ■

Apesar de termos imposto na Definição 2.2.2 que uma probabilidade P é uma aplicação com valores no intervalo $[0,1]$, uma inspeção cuidadosa das demonstrações anteriores revela que, até ao momento, além das propriedades expressas nas alíneas a) e b) da definição de probabilidade, apenas usámos o facto de P tomar valores em $\mathbb{R}$. Isto significa que bastaria assumirmos isso para obter todas as propriedades anteriores. Uma vez que qualquer acontecimento aleatório A é tal que $\emptyset \subset A \subset \Omega$, da monotonia da probabilidade concluiríamos que $0 \leq P(A) \leq 1$. Isto é, mesmo sem impor, por definição, que uma probabilidade P é uma aplicação com valores no intervalo $[0,1]$, essa propriedade seria verdadeira desde que assumíssemos que uma probabilidade é uma aplicação real.

Proposição 2.3.6. *Se A e B são dois acontecimentos quaisquer então*

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

Dem: Tendo em conta que $A \cup B = (A - B) \cup (B - A) \cup (A \cap B)$, com $A - B$, $B - A$ e $A \cap B$ acontecimentos incompatíveis, então pela aditividade de P concluímos que

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

Usando agora a Proposição 2.3.4 concluímos que

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

A fórmula anterior pode ser generalizada à reunião de mais que dois acontecimentos aleatórios.

Proposição 2.3.7 (Fórmula de Daniel da Silva ou da Inclusão-Exclusão; Fórmula de Poincaré)**.** *Para quaisquer acontecimentos $A_1, \ldots, A_n$ temos*

$$\begin{aligned} P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) - \sum_{1 \leq i < j \leq n} P(A_i \cap A_j) \\ + \sum_{1 \leq i < j < k \leq n} P(A_i \cap A_j \cap A_k) + \cdots + (-1)^{n+1} P(A_1 \cap \cdots \cap A_n). \end{aligned}$$

Dem: Vamos demonstrar a fórmula para $n = 3$:

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

Comecemos por verificar que a reunião $A \cup B \cup C$ se pode escrever como reunião disjunta dos acontecimentos seguintes

$$\begin{aligned} A \cup B \cup C = &\big(A - (B \cup C)\big) \cup \big(B - (A \cup C)\big) \cup \big(C - (A \cup B)\big) \\ &\cup \big((A \cap B) - C\big) \cup \big((A \cap C) - B\big) \cup \big((B \cap C) - A\big) \\ &\cup \big(A \cap B \cap C\big). \end{aligned}$$

O resultado decorre agora da aditividade de P e das Proposições 2.3.4 e 2.3.6. ■

Proposição 2.3.8 (σ-subaditividade)**.** *Para qualquer sucessão de acontecimentos* (A_n) *temos*

$$P\Big(\bigcup_{n=1}^{\infty} A_n\Big) \leq \sum_{n=1}^{\infty} P(A_n).$$

Dem: Uma vez que

$$\bigcup_{n=1}^{\infty} A_n = A_1 \cup (A_2 - A_1) \cup (A_3 - (A_1 \cup A_2)) \cup \cdots = \bigcup_{n=1}^{\infty} (A_n - B_n),$$

onde $B_1 = \emptyset$ e $B_n = \bigcup_{k=1}^{n-1} A_k$, para $n \geq 2$, e onde os acontecimentos $(A_n - B_n)$ são dois a dois incompatíveis, então pela σ-aditividade e pela monotonia de P temos

$$\begin{aligned} P\Big(\bigcup_{n=1}^{\infty} A_n\Big) &= P\Big(\bigcup_{n=1}^{\infty} (A_n - B_n)\Big) \\ &= \sum_{n=1}^{\infty} P(A_n - B_n) \\ &\leq \sum_{n=1}^{\infty} P(A_n). \end{aligned}$$

■

As propriedades duma probabilidade que discutimos a seguir envolvem a noção de limite de uma sucessão (A_n) de acontecimentos aleatórios. Para formalizarmos uma tal noção vamos definir dois acontecimentos ditos limite inferior e limite superior da sucessão (A_n):

— $\liminf A_n := \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k \equiv$ acontecimento que se realiza quando se realizam todos os acontecimentos A_n com exceção dum número finito deles;

— $\limsup A_n := \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k \equiv$ acontecimento que se realiza quando se realiza uma infinidade de acontecimentos A_n.

Assim, dizemos que a sucessão (A_n) tem limite, que denotamos por $\lim A_n$, sempre que $\liminf A_n = \limsup A_n$. No caso da sucessão (A_n) ser crescente, isto é, $A_n \subset A_{n+1}$,

para todo o $n \in \mathbb{N}$, ou decrescente, isto é, $A_{n+1} \subset A_n$, para todo o $n \in \mathbb{N}$, a sucessão (A_n) tem sempre limite. No caso de (A_n) ser crescente temos

$$\lim A_n = \bigcup_{n=1}^{\infty} A_n,$$

e no caso de (A_n) ser decrescente temos

$$\lim A_n = \bigcap_{n=1}^{\infty} A_n$$

(ver Exercício 9).

Proposição 2.3.9 (continuidade monótona)**.** *Se (A_n) é uma sucessão monótona (crescente ou decrescente em sentido lato) de acontecimentos aleatórios então*

$$P(\lim A_n) = \lim P(A_n).$$

Dem: Comecemos por assumir que (A_n) é uma sucessão crescente de acontecimentos aleatórios, isto é, $A_n \subset A_{n+1}$, para todo o $n \in \mathbb{N}$. Sabemos que

$$\lim A_n = \bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} (A_n - A_{n-1}),$$

onde $A_0 = \emptyset$ e a sucessão $(A_n - A_{n-1})$ é formada por acontecimentos incompatíveis. Assim, da σ-aditividade de P e da Propriedade 2.3.4 temos

$$\begin{aligned} P(\lim A_n) &= P\Big(\bigcup_{n=1}^{\infty} (A_n - A_{n-1}) \Big) = \sum_{n=1}^{\infty} P(A_n - A_{n-1}) \\ &= \lim_{k\to\infty} \sum_{n=1}^{k} \big(P(A_n) - P(A_{n-1}) \big) = \lim_{k\to\infty} P(A_k). \end{aligned}$$

Seja agora (A_n) é uma sucessão decrescente de acontecimentos aleatórios, isto é, $A_{n+1} \subset A_n$, para todo o $n \in \mathbb{N}$. Sabemos que (ver Exercício 9)

$$\lim A_n = \bigcap_{n=1}^{\infty} A_n = \Big(\bigcup_{n=1}^{\infty} A_n^c \Big)^c = \big(\lim A_n^c \big)^c,$$

uma vez que (A_n^c) é uma sucessão crescente de acontecimentos aleatórios. Da primeira parte da demonstração concluímos que

$$P(\lim A_n) = 1 - P\big(\lim A_n^c \big) = 1 - \lim P(A_n^c) = 1 - \lim \big(1 - P(A_n)\big) = \lim P(A_n). \quad \blacksquare$$

Proposição 2.3.10 (continuidade)**.** *Se (A_n) é uma sucessão de acontecimentos aleatórios com limite, então*

$$P(\lim A_n) = \lim P(A_n).$$

Dem: Por hipótese, (A_n) é uma sucessão de acontecimentos aleatórios com limite, isto é, $\liminf A_n = \limsup A_n = \lim A_n$. Uma vez que $\left(\bigcap_{k=n}^{\infty} A_k\right)$ é crescente e $\left(\bigcup_{k=n}^{\infty} A_k\right)$ é decrescente, então

$$\lim A_n = \liminf A_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k = \lim \left(\bigcap_{k=n}^{\infty} A_k \right)$$

e

$$\lim A_n = \limsup A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k = \lim \left(\bigcup_{k=n}^{\infty} A_k \right).$$

Usando a Proposição 2.3.9 concluímos que

$$P(\lim A_n) = \lim P\Big(\bigcap_{k=n}^{\infty} A_k \Big) = \lim P\Big(\bigcup_{k=n}^{\infty} A_k \Big).$$

Finalmente, usando a monotonia de P, para $n \in \mathbb{N}$ temos

$$P\Big(\bigcap_{k=n}^{\infty} A_k \Big) \leq P(A_n) \leq P\Big(\bigcup_{k=n}^{\infty} A_k \Big),$$

o que permite, utilizando o teorema das sucessões enquadradas, concluir a demonstração. ■

Terminamos esta secção com um resultado conhecido como Lema de Borel-Cantelli que nos dá uma condição suficiente para que o acontecimento $\limsup A_n$, isto é, o acontecimento que se realiza quando se realiza uma infinidade de acontecimentos A_n, tenha probabilidade zero de ocorrer.

Teorema 2.3.11 (Lema de Borel-Cantelli ([2]))**.** *Se os acontecimentos aleatórios* (A_n) *satisfazem* $\sum_{n=1}^{\infty} P(A_n) < \infty$*, então* $P(\limsup A_n) = 0$.

Dem: Sabemos que a sucessão $\left(\bigcup_{k=n}^{\infty} A_k\right)$ é decrescente, e que

$$\limsup A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k = \lim \bigcup_{k=n}^{\infty} A_k.$$

Usando a Proposições 2.3.8 e 2.3.10 concluímos que

$$P\big(\limsup A_n\big) = \lim P\Big(\bigcup_{k=n}^{\infty} A_k \Big) \leq \lim \sum_{k=n}^{\infty} P(A_k) = 0,$$

uma vez que $\sum_{k=n}^{\infty} P(A_k)$ é o resto de ordem n de uma série convergente. ■

[2] Borel, E. (1909). Les probabilités dénombrables et leurs applications arithmétiques. *Rend. Circ. Mat. Palermo* 27, 247–271. Cantelli, F.P. (1917). Sulla probabilità come limite della frequenza. *Rend. Accad. Naz. Lincei* 26, 295–302.

2.4 Probabilidade condicionada

Retomemos o Exemplo 2.2.4 e suponhamos agora que lançamos o dado e que, apesar de não sabermos qual foi a face que ocorreu, é-nos dada a informação de que saiu face par, isto é, ocorreu o acontecimento $B = \{2, 4, 6\}$. Com esta nova informação sobre a experiência aleatória, o espaço de probabilidade inicialmente considerado não é mais o espaço adequado à descrição da mesma. Duma forma geral, se $(\Omega, \mathscr{A}, P)$ é o espaço de probabilidade associado a uma experiência aleatória, e se sabemos que se realizou o acontecimento $B \in \mathscr{A}$, com $P(B) > 0$, a probabilidade dum acontecimento $A \in \mathscr{A}$ depende naturalmente "da sua relação com B". Por exemplo, se $A \supset B$, A realizar-se-á, e se $A \cap B = \emptyset$, A não se realizará. Será natural substituir a probabilidade $P(A)$ pela probabilidade $P(A|B)$, a que chamamos probabilidade de A dado B ou probabilidade de A condicionada por B, que irá representar a probabilidade do acontecimento A após termos conhecimento de que o acontecimento B se realizou, enquanto que $P(A)$ representa a probabilidade de A sem termos informação adicional sobre a realização, ou não, do acontecimento B.

Fixemos a nossa atenção no caso em que estamos a utilizar a definição clássica de probabilidade. Neste caso será natural tomar para probabilidade de A condicionada por B o quociente

$$P(A|B) = \frac{\text{número de resultados favoráveis a } A \cap B}{\text{número de resultados favoráveis a } B} = \frac{\#A \cap B}{\#B},$$

uma vez que, como sabemos que B se realizou, o número de resultados possíveis da experiência reduz-se aos resultados que são favoráveis a B, enquanto que o número de resultados favoráveis a A não é agora mais do que o número de resultados favoráveis a $A \cap B$. Reescrevendo o quociente anterior na forma

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

verificamos que o numerador e o denominador não são mais do que as probabilidades de $A \cap B$ e de B, respetivamente, quando não temos informação adicional sobre a experiência aleatória. Isto leva-nos à definição seguinte de probabilidade condicionada para uma qualquer forma de atribuir probabilidade aos acontecimentos duma experiência aleatória.

Definição 2.4.1. *Para $B \in \mathscr{A}$, com $P(B) > 0$, e $A \in \mathscr{A}$, chamamos probabilidade condicionada de A dado B ou probabilidade de A condicionada por B, ao quociente*

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \quad \textit{para } A \in \mathscr{A}.$$

A probabilidade $P(A|B)$ denota-se também por $P_B(A)$.

Notemos que P_B é uma probabilidade sobre $\mathscr{A}$, isto é, P_B é um aplicação de $\mathscr{A}$ em $[0,1]$, com $P_B(\Omega) = 1$ e $P_B\big(\bigcup_{n=1}^{\infty} A_n\big) = \sum_{n=1}^{\infty} P_B(A_n)$, para toda a sucessão (A_n) de acontecimentos dois a dois incompatíveis. Sendo uma probabilidade, P_B satisfaz assim todas as propriedades estabelecidas em §2.3.

O conhecimento de $P(A)$ e de $P(B|A)$ permitem calcular a probabilidade da interseção $A \cap B$:

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

O resultado seguinte generaliza tal facto à interseção dum número finito de acontecimentos.

Teorema 2.4.2 (Fórmula da probabilidade composta)**.** *Se $A_1, \ldots, A_n$, com $n \geq 2$, são acontecimentos aleatórios com $P(A_1 \cap \cdots \cap A_{n-1}) > 0$, então*

$$P(A_1 \cap \cdots \cap A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \ldots P(A_n|A_1 \cap \cdots \cap A_{n-1}).$$

Dem: Para $n = 2$ o resultado é consequência imediata da definição de probabilidade condicionada. Para $n > 2$, se $A_1, \ldots, A_n$ são acontecimentos aleatórios com $P(A_1 \cap \cdots \cap A_{n-1}) > 0$, basta ter em conta que

$$P(A_1 \cap \cdots \cap A_n) = P(A_1 \cap \cdots \cap A_{n-1})P(A_n|A_1 \cap \cdots \cap A_{n-1}). \qquad \blacksquare$$

Consideremos agora um acontecimento B cuja realização está relacionada com a dos acontecimentos de uma família finita $A_1, \ldots, A_n$ de acontecimentos incompatíveis, e admitamos que conhecemos as probabilidades $P(B|A_i)$ de B na eventualidade do acontecimento A_i se realizar. O resultado seguinte mostra como calcular a probabilidade de B desde que conheçamos a probabilidade de cada um dos acontecimentos A_i.

Teorema 2.4.3 (Fórmula da probabilidade total)**.** *Sejam $A_1, \ldots, A_n$ acontecimentos incompatíveis de probabilidade positiva e B um acontecimento com $B \subset A_1 \cup \cdots \cup A_n$. Então*

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

Dem: Vamos demonstrar o resultado no caso $n = 2$ (o caso $n > 2$ é análogo). Atendendo a que $B \subset A_1 \cup A_2$, então

$$B = \big(B \cap A_1\big) \cup \big(B \cap A_2\big).$$

Uma vez que $B \cap A_1$ e $B \cap A_2$ são incompatíveis, obtemos

$$P(B) = P(B \cap A_1) + P(B \cap A_2) = P(A_1)P(B|A_1) + P(A_2)P(B|A_2). \qquad \blacksquare$$

Mostramos agora que conhecendo as probabilidades $P(A|B)$ e $P(B)$ é possível determinar as probabilidades $P(B|A)$.

Teorema 2.4.4 (Teorema de Bayes ([3])). *Se A e B são acontecimentos aleatórios de probabilidade positiva, então*

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

Dem: Decorre diretamente da definição de probabilidade condicionada. ■

2.4.1 Aplicação aos testes de diagnóstico

Uma aplicação interessante dos dois resultados anteriores surge no campo dos chamados testes de diagnóstico. Os testes de diagnóstico são procedimentos em que os indivíduos sujeitos a tais testes são classificados como saudáveis ou como tendo determinada doença (Teste TOTG (diabetes), Teste VIH (sida), Teste COVID-19, etc). Por serem procedimentos simples de usar e pouco dispendiosos, os testes de diagnóstico são imperfeitos, podendo indivíduos saudáveis ser classificados como doentes (ditos falsos positivos) e indivíduos doentes ser classificados como não doentes (ditos falsos negativos). Um teste de diagnóstico pode ser avaliado através da sua sensibilidade e da sua especificidade, definidas pelas seguintes probabilidades condicionais:

$$\begin{aligned}\text{sensibilidade} &= P(\text{teste positivo}|\text{doença presente}) = P(T^+|D^+)\\ \text{especificidade} &= P(\text{teste negativo}|\text{doença ausente}) = P(T^-|D^-).\end{aligned}$$

Um teste com grande sensibilidade identificará os doentes com grande precisão, isto é, fará um diagnóstico correto na população doente, enquanto que um teste com grande especificidade terá grande precisão na identificação dos não doentes, isto é, fará um diagnóstico correto na população não afetada pela doença em causa.

Em geral os testes de diagnóstico passam pela medição no indivíduo observado de determinada resposta quantitativa X, sendo o indivíduo classificado como saudável se a variável X não for superior a determinado limiar c, ou como doente, em caso contrário. Assim, se $X \leq c$ o indivíduo é classificado como não doente, e se $X > c$ o indivíduo é classificado como doente (ver Figura 2.1). Um teste será tanto melhor quanto maior discriminação induzir entre as populações de doentes e saudáveis.

Na prática, a sensibilidade e a especificidade de um teste de diagnóstico podem ser estimadas lançando mão da lei dos grandes números de Bernoulli a que fizemos referência em §1.2. Para tal, o teste de diagnóstico é aplicado a dois grupos de indivíduos, um deles constituído por indivíduos que sabemos doentes (no caso da sensibilidade), e o outro constituído por indivíduos que sabemos não-doentes (no caso da especificidade).

[3] Bayes, T. (1763). An essay towards solving a problem in the dictrine of chances. *Philosophical Transactions* 53, 370–418.

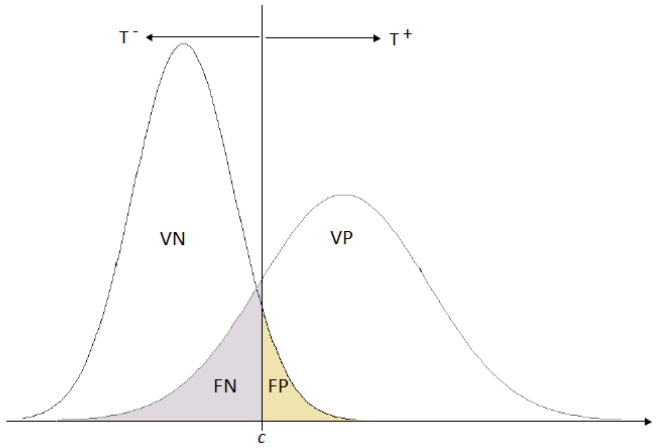


Figura 2.1: Distribuição dos valores da resposta X na população não doente (curva à esquerda) e doente (curva à direita) e classificação dos indivíduos para um determinado limiar c (VP="verdadeiros positivos"; FP="falsos positivos"; VN="verdadeiros negativos"; VP="falsos negativos").

Teste	Doença	
	Presente D^+	Ausente D^-
Positivo T^+	Verdadeiros Positivos (A)	Falsos Positivos (B)
Negativo T^-	Falsos Negativos (C)	Verdadeiros Negativos (D)
Total	Doentes (A+C)	Não Doentes (B+D)

Sendo A e C o número de indivíduos efetivamente doentes classificados como doentes (verdadeiros positivos) e como não-doentes (falsos negativos), respetivamente, a sensibilidade do teste de diagnóstico pode assim ser aproximada por

$$\text{sensibilidade} \approx \frac{\text{A}}{\text{A} + \text{C}}.$$

Por outro lado, sendo B e D o número de indivíduos saudáveis que são classificados como doentes (falsos positivos) e como não-doentes (verdadeiros negativos), respetivamente, a especificidade do teste de diagnóstico pode assim ser aproximada por

$$\text{especificidade} \approx \frac{\text{D}}{\text{B} + \text{D}}.$$

Duas caraterísticas importantes de um teste de diagnóstico são os seus valor preditivo

positivo (VPP) e valor preditivo negativo (VPN):

$$\begin{aligned}\text{VPP} &= P(\text{doença presente}|\text{teste positivo}) = P(D^+|T^+)\\ \text{VPN} &= P(\text{doença ausente}|\text{teste negativo}) = P(D^-|T^-).\end{aligned}$$

Estes valores preditivos são fundamentais para determinar o valor do teste de diagnóstico num indivíduo específico. Contrariamente à sensibilidade e à especificidade, que são atributos intrínsecos do teste de diagnóstico, veremos a seguir que os valores preditivos positivo e negativo dependem da sensibilidade e da especificidade do teste, mas também da prevalência, ou taxa de incidência, da doença na população.

Representando por Se e Es, a sensibilidade e a especificidade do teste, e por Pr a prevalência da doença na população em causa (isto é, a proporção de doentes na população em causa), a fórmula da probabilidade total e o teorema de Bayes permitem concluir que o valor preditivo positivo pode ser obtido a partir da fórmula

$$\begin{aligned}\text{VPP} &= \frac{P(T^+|D^+)P(D^+)}{P(T^+)}\\ &= \frac{P(T^+|D^+)P(D^+)}{P(T^+|D^+)P(D^+) + P(T^+|D^-)P(D^-)}\\ &= \frac{P(T^+|D^+)P(D^+)}{P(T^+|D^+)P(D^+) + (1-P(T^-|D^-))(1-P(D^+))}\\ &= \frac{\mathsf{Se}\times\mathsf{Pr}}{\mathsf{Se}\times\mathsf{Pr} + (1-\mathsf{Es})\times(1-\mathsf{Pr})},\end{aligned}$$

e que o valor preditivo negativo pode ser obtido a partir da fórmula

$$\begin{aligned}\text{VPN} &= \frac{P(T^-|D^-)P(D^-)}{P(T^-)}\\ &= \frac{P(T^-|D^-)(1-P(D^+))}{P(T^-|D^+)P(D^+) + P(T^-|D^-)P(D^-)}\\ &= \frac{P(T^-|D^-)(1-P(D^+))}{(1-P(T^+|D^+))P(D^+) + P(T^-|D^-)(1-P(D^+))}\\ &= \frac{\mathsf{Es}\times(1-\mathsf{Pr})}{(1-\mathsf{Se})\times\mathsf{Pr} + \mathsf{Es}\times(1-\mathsf{Pr})}.\end{aligned}$$

Como referimos atrás, as fórmulas anteriores mostram que os valores preditivos dependem da prevalência da doença na população. Isto significa que o mesmo teste de diagnóstico, quando utilizado em populações com prevalências distintas, pode conduzir a avaliações diferentes da probabilidade dum indivíduo ter a doença em causa quando o teste dá positivo. Na Figura 2.2 mostramos como evoluem os valores preditivos positivo e negativo de um teste de diagnóstico em função da prevalência da doença na população. Apesar de assumirmos que o teste de diagnóstico tem sensibilidade e especificidade

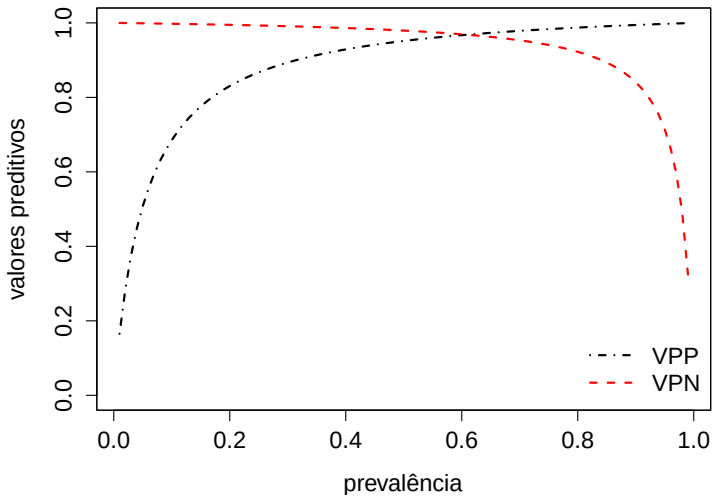


Figura 2.2: Valores preditivos de um teste de diagnóstico, com sensibilidade e especificidade iguais a 0,95, em função da prevalência da doença na população.

elevadas (ambas iguais a 0,95), é interessante constatar que o valor preditivo positivo pode ser baixo quando a prevalência da doença é muito baixa, e que o valor preditivo negativo pode ser reduzido quando a prevalência da doença é muito alta. Isto significa que numa população com muito baixa prevalência da doença, é pequena a probabilidade de estar efetivamente doente um indivíduo cujo teste deu positivo. Por outro lado, numa população com muito elevada prevalência da doença, é grande a probabilidade de estar efetivamente doente um indivíduo cujo teste deu negativo.

2.5 Independência de acontecimentos aleatórios

Dados dois acontecimentos A e B do espaço de probabilidade $(\Omega, \mathscr{A}, P)$ com probabilidades positivas, vimos que a probabilidade condicionada $P(A|B)$ pode ser interpretada como a probabilidade do acontecimento A quando temos a informação que o acontecimento B se realizou. A probabilidade $P(A|B)$ não coincide necessariamente com a probabilidade $P(A)$. No entanto, quando tal acontece, isto é, quando a probabilidade de A não se altera quando sabemos que B se realizou, isto é, $P(A|B) = P(A)$, diremos que os acontecimentos A e B são independentes. É simples ver que quando $P(A|B) = P(A)$ também é verdade que $P(B|A) = P(B)$, isto é, quando a ocorrência de B não altera a probabilidade de ocorrência de A, também a ocorrência de A não altera a probabilidade de ocorrência de B. Qualquer uma das igualdades anteriores é ainda equivalente à igualdade $P(A \cap B) = P(A)P(B)$. Uma vez que esta última

igualdade não necessita que os acontecimentos A e B tenham probabilidades positivas, vamos usá-la para formalizar a noção de independência dos acontecimentos A e B.

Definição 2.5.1. *Os acontecimentos A e B dizem-se* independentes *quando*

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

Se um acontecimento A tem probabilidade nula, então para qualquer acontecimento B tem-se $P(A \cap B) = 0$ (porquê?), verificando-se assim a igualdade $P(A \cap B) = P(A)P(B)$. Isto significa que dois acontecimentos A e B são independentes sempre que um deles tenha probabilidade nula. Como já vimos, se A e B são independentes e têm probabilidades positivas então

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

e também

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

Proposição 2.5.2. *Se A e B são acontecimentos independentes, também são independentes os pares de acontecimentos A e B^c, A^c e B, e A^c e B^c.*

Dem: Basta mostrar que A e B^c são independentes (porquê?). Pelas Proposições 2.3.3 e 2.3.4 temos

$$\begin{aligned} P\big(A \cap B^c\big) &= P(A - B) = P(A) - P(A \cap B) \\ &= P(A) - P(A)P(B) = P(A)(1 - P(B)) = P(A)P(B^c). \end{aligned}$$ ■

Generalizamos a seguir a noção de independência de acontecimentos aleatórios para uma qualquer família $A_t, t \in T$, onde o conjunto de índices T é um subconjunto, finito ou infinito, de $\mathbb{N}$.

Definição 2.5.3. *Os acontecimentos aleatórios $A_t, t \in T$, com $T \subset \mathbb{N}$ dizem-se* independentes *se para um qualquer subconjunto finito de índices $t_1, \ldots, t_n \in T$, com $n \in \mathbb{N}$, se tem*

$$P\big(A_{t_1} \cap \cdots \cap A_{t_n}\big) = P(A_{t_1}) \times \cdots \times P(A_{t_n}).$$

Decorre da definição anterior que se os acontecimentos aleatórios $A_t, t \in T$ são independentes então são dois a dois independentes, isto é, os acontecimentos A_s e A_t são independentes, para todo o $s, t \in T$ com $s \neq t$. No entanto, o recíproco não é verdadeiro (ver Exercício 22).

Como se ilustra a seguir, acontecimentos aleatórios independentes surgem naturalmente associados a experiências aleatórias que se repetem sempre não mesmas condições.

Exemplo 2.5.4. Suponhamos que lançamos um dado equilibrado N vezes. O espaço de probabilidade $(\Omega, \mathscr{A}, P)$ que podemos associar e esta experiência é

$$\Omega = \big\{(x_1, \ldots, x_N) : x_i \in \{1,2,3,4,5,6\}, i = 1, \ldots, N\big\},$$

com $\mathscr{A} = \mathscr{P}(\Omega)$ e

$$P\big(\{(x_1, \ldots, x_n)\}\big) = 1/6^N.$$

Consideremos agora dois acontecimentos aleatórios que dependem apenas do que ocorre no n-ésimo e no m-ésimo lançamento do dado, com $m \neq n$:

$A_n =$ "saiu número par no n-ésimo lançamento do dado"

$B_m =$ "saiu múltiplo de 3 no m-ésimo lançamento do dado"

Da noção intuitiva de independência de acontecimentos aleatório, somos levados a dizer que uma vez que o dado é lançado sempre nas mesma condições, saber que A_n ocorreu não altera a probabilidade de ocorrência de B_m, isto é, os acontecimentos A_n e B_m são independentes. Confirmemos esta ideia efetuando os cálculos das probabilidades associadas aos acontecimentos A_n, B_m e $A_n \cap B_m$:

$$P(A_n) = \frac{3 \times 6^{N-1}}{6^N} = \frac{1}{2},$$

$$P(B_m) = \frac{2 \times 6^{N-1}}{6^N} = \frac{1}{3},$$

$$P(A_n \cap B_m) = \frac{3 \times 2 \times 6^{N-2}}{6^N} = \frac{1}{6}.$$

Da mesma forma, se $m \neq n$ são independentes os acontecimentos A_n e A_m, bem com os acontecimentos B_n e B_m.

Proposição 2.5.5. *Se (A_n) são acontecimentos independentes, então*

$$P\Big(\bigcap_{n=1}^{\infty} A_n\Big) = \prod_{n=1}^{\infty} P(A_n).$$

Dem: Uma vez que a sucessão (B_k) definida por $B_k = \bigcap_{n=1}^{k} A_n$, para $k \in \mathbb{N}$, é decrescente, sabemos que

$$\lim B_k = \bigcap_{k=1}^{\infty} B_k = \bigcap_{k=1}^{\infty} \bigcap_{n=1}^{k} A_n = \bigcap_{n=1}^{\infty} A_n.$$

Para concluir basta agora usar a continuidade de P e a independência dos acontecimentos (A_n):

$$P\Big(\bigcap_{n=1}^{\infty} A_n\Big) = P(\lim_k B_k) = \lim_k P(B_k) = \lim_k \prod_{n=1}^{k} P(A_n) = \prod_{n=1}^{\infty} P(A_n). \qquad \blacksquare$$

Vimos anteriormente que quando a série $\sum_{n=1}^{\infty} P(A_n)$ é convergente o acontecimento $\limsup A_n$ tem probabilidade zero (Proposição 2.3.11). No resultado seguinte mostramos que no caso dos acontecimentos (A_n) serem independentes, uma tal probabilidade só pode tomar dois valores possíveis: zero ou um. Por outras palavras, a probabilidade de ocorrerem uma infinidade de acontecimentos A_n é zero ou um. Ou ainda, a probabilidade de ocorrência de quando muito um número finito desses acontecimentos é um ou zero, respetivamente.

Teorema 2.5.6 (Lei zero-um de Borel ([4])). *Se os acontecimentos (A_n) são independentes, então*

$$P\big(\limsup A_n\big) = \begin{cases} 0 & sse \quad \sum_{n=1}^{\infty} P(A_n) < \infty \\ 1 & sse \quad \sum_{n=1}^{\infty} P(A_n) = \infty. \end{cases}$$

Dem: Atendendo ao Teorema 2.3.11 basta provar que se $\sum_{n=1}^{\infty} P(A_n) = \infty$ então $P\big(\limsup A_n\big) = 1$. Tal é equivalente a provar que $P\big(\bigcup_{k=n}^{\infty} A_k\big) = 1$, para todo o $n \in \mathbb{N}$ (ver Exercício 8), ou ainda que, $P\big(\bigcap_{k=n}^{\infty} A_k^c\big) = 0$, para todo o $n \in \mathbb{N}$. Atendendo à independência dos acontecimentos (A_k^c), e à desigualdade $1 - x \leq \exp(-x)$, válida para todo o $x \geq 0$, da Proposição 2.5.5 obtemos

$$\begin{aligned} P\Big(\bigcap_{k=n}^{\infty} A_k^c\Big) &= \prod_{k=n}^{\infty} P(A_k^c) = \lim_m \prod_{k=n}^{m} P(A_k^c) \\ &= \lim_m \prod_{k=n}^{m} \big(1 - P(A_k)\big) \\ &\leq \lim_m \prod_{k=n}^{m} \exp\big(-P(A_k)\big) \\ &\leq \lim_m \exp\Big(-\sum_{k=n}^{m} P(A_k)\Big) = 0. \end{aligned}$$

■

2.6 Bibliografia

Kolmogorov, A.N. (1950). *Foundations of the theory of probability.* New York: Chelsea.

Gonçalves, E., Mendes Lopes, N. (2013). *Probabilidades: princípios teóricos.* Lisboa: Escolar Editora.

Haigh, J. (2002). *Probability models.* London: Springer.

[4] Borel, E. (1909). Les probabilités dénombrables et leurs applications arithmétiques. *Rend. Circ. Mat. Palermo* 27, 247–271.

James, B.R. (1981). *Probabilidade: um curso em nível intermediário.* Rio de Janeiro: IMPA.

Monfort, A. (1980). *Cours de probabilités.* Paris: Economica.

Resnick, S.I. (2005). *A probability path.* Boston: Birkhäuser.

Tenreiro, C. (2002). *Apontamentos de Teoria das Probabilidades.* Coimbra: DMUC. https://www.mat.uc.pt/∼tenreiro/apontamentos/ATP0204.pdf

3

Variáveis aleatórias

Variáveis aleatórias e suas leis de probabilidade. Variáveis discretas e variáveis absolutamente contínuas. Variáveis aleatórias mistas. Função de distribuição. Transformação afim de variáveis aleatórias. Exemplos de leis de probabilidade discretas e absolutamente contínuas. Transformação de variáveis absolutamente contínuas. Simulação de experiências e variáveis aleatórias.

3.1 Variáveis aleatórias e suas leis de probabilidade

Observado um resultado particular duma experiência aleatória, estamos por vezes interessados não no resultado em si mesmo, mas numa função desse resultado. Pensemos no que acontece quando jogamos ao Monopólio e lançamos os dados: interessa-nos a soma dos pontos obtidos e não os pontos ocorridos em cada um dos dados. Por outras palavras, sendo $(\Omega, \mathscr{A}, P)$ um modelo probabilístico para a experiência aleatória em causa, e observado um resultado $\omega \in \Omega$, interessamo-nos por uma função numérica $X(\omega)$ de ω a que chamamos variável aleatória. Uma variável aleatória X não é mais do que uma aplicação de Ω em $\mathbb{R}$, que a cada resultado elementar $\omega \in \Omega$ associa o número real $X(\omega)$. Do ponto de vista das probabilidades, sobre a variável X pretendemos conhecer a sua distribuição ou lei de probabilidade, ou seja, pretendemos conhecer os valores que a variável toma e a probabilidade com que tais valores são tomados. Mais geralmente, para um conjunto possível de valores $B \in \mathscr{B}(\mathbb{R})$, onde $\mathscr{B}(\mathbb{R})$ é a σ-álgebra de Borel de $\mathbb{R}$, isto é, a mais pequena σ-álgebra gerada pelos conjuntos abertos de $\mathbb{R}$, pretendemos conhecer a probabilidade com que X toma valores em B, ou seja, pretendemos conhecer $P(X \in B)$, onde

$$\{X \in B\} = \{\omega \in \Omega : X(\omega) \in B\}$$

(este conjunto, também denotado por $X^{-1}(B)$, é dito imagem recíproca de B por X). Tal como estabelecemos na definição seguinte, para podermos calcular as probabilidades dos conjuntos anteriores é preciso que tais conjuntos pertençam à σ-álgebra $\mathscr{A}$ onde

está definida a probabilidade P.

Definição 3.1.1. *Chamamos variável aleatória real (v.a.r.) a toda a aplicação $X : \Omega \to \mathbb{R}$ definida num espaço de probabilidade $(\Omega, \mathscr{A}, P)$ que satisfaz $\{X \in B\} \in \mathscr{A}$ para todo o $B \in \mathscr{B}(\mathbb{R})$.*

Sendo a σ-álgebra de Borel também gerada pelos intervalos da forma $]a, b]$, com $-\infty < a \leq b < \infty$, é possível provar que X é uma v.a.r. sse $\{X \in]a, b]\} = \{a < X \leq b\} \in \mathscr{A}$, para todo o $-\infty < a \leq b < \infty$. Quando $\mathscr{A} = \mathscr{P}(\Omega)$, qualquer aplicação X de Ω em $\mathbb{R}$, é uma variável aleatória. No caso em que $\mathscr{A}$ é a σ-álgebra de Borel de Ω, com Ω é um subconjunto de $\mathbb{R}^d$, para algum $d \geq 1$, é possível provar que qualquer aplicação contínua de Ω em $\mathbb{R}$ é uma variável aleatória. Outras classes mais vastas de funções são também variáveis aleatórias. Uma vez que as funções que encontramos em situações práticas são sempre variáveis aleatórias, não nos vamos preocupar com esta questão.

Como referimos, dada uma v.a.r. X estamos habitualmente interessados na probabilidade com que X toma valores em subconjuntos B de $\mathbb{R}$. No sentido da definição seguinte, estamos interessados na chamada distribuição de probabilidade de X.

Proposição 3.1.2. *Se X é uma v.a.r. definida num espaço de probabilidade $(\Omega, \mathscr{A}, P)$, então a aplicação P_X definida, para $B \in \mathscr{B}(\mathbb{R})$, por*

$$P_X(B) = P(X \in B),$$

é uma probabilidade em $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$, a que chamamos distribuição de probabilidade ou lei de probabilidade da variável X.

Dem: Para $B \in \mathscr{B}(\mathbb{R})$ temos $P_X(B) = P(X \in B) \in [0, 1]$, e $P_X(\mathbb{R}) = P(X \in \mathbb{R}) = P(\Omega) = 1$. Além disso, sendo (B_n) uma qualquer sucessão de acontecimentos em $\mathscr{B}(\mathbb{R})$ dois a dois incompatíveis, os acontecimentos $(\{X \in B_n\})$ em $\mathscr{A}$ são também dois a dois incompatíveis e assim

$$\begin{aligned} P_X\Big(\bigcup_{n=1}^{\infty} B_n\Big) &= P\Big(X \in \bigcup_{n=1}^{\infty} B_n\Big) = P\Big(\bigcup_{n=1}^{\infty} \{X \in B_n\}\Big) \\ &= \sum_{n=1}^{\infty} P\big(X \in B_n\big) = \sum_{n=1}^{\infty} P_X(B_n). \end{aligned}$$

■

Mais uma vez, atendendo ao facto da σ-álgebra de Borel $\mathscr{B}(\mathbb{R})$ ser gerada pelos intervalos da forma $]a, b]$, com $-\infty < a \leq b < \infty$, é possível provar que a distribuição de probabilidade de uma variável aleatória X é determinada pelos valores que P_X toma em tais intervalos reais. Assim, se X e Y são v.a.r. para as quais $P_X(]a, b]) = P_Y(]a, b])$,

para todo o $-\infty < a \leq b < \infty$, então $P_X(B) = P_Y(B)$, para todo o $B \in \mathscr{B}(\mathbb{R})$, isto é, X e Y têm a mesma distribuição de probabilidade. Quando tal acontece escrevemos $X \sim Y$. Semelhante propriedade é também válida para os intervalos da forma $]-\infty, b]$, com $-\infty < b < \infty$.

Exemplo 3.1.3. Consideremos a experiência aleatória que consiste no lançamento de dois dados equilibrados e seja X a variável que nos dá a soma dos pontos obtidos nos dois dados. Um modelo probabilístico $(\Omega, \mathscr{A}, P)$ para esta experiência aleatória pode ser definido por

$$\Omega = \{(i,j) : i,j \in \{1,\ldots,6\}\},$$

$$\mathscr{A} = \mathscr{P}(\Omega)$$

e

$$P(\{(i,j)\}) = 1/36, \quad (i,j) \in \Omega.$$

Atendendo à definição de $X(i,j) = i+j$, sabemos que X toma valores no conjunto $\{2,3,\ldots,12\} \subset \mathbb{R}$ com as seguintes probabilidades:

$$\begin{aligned}
P(X=2) &= P\big(\{\omega \in \Omega : X(\omega) = 2\}\big) = P(\{(1,1)\}) = \frac{1}{36},\\
P(X=3) &= P\big(\{\omega \in \Omega : X(\omega) = 3\}\big) = P(\{(1,2),(2,1)\}) = \frac{2}{36},\\
&\vdots\\
P(X=7) &= P\big(\{\omega \in \Omega : X(\omega) = 7\}\big) = P(\{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}) = \frac{6}{36},\\
&\vdots\\
P(X=12) &= P\big(\{\omega \in \Omega : X(\omega) = 12\}\big) = P(\{(6,6)\}) = \frac{1}{36}.
\end{aligned}$$

A partir das probabilidades anteriores podemos agora calcular a probabilidade $P_X(]a,b])$ para qualquer intervalo $]a,b]$. Com efeito,

$$P_X(]a,b]) = P\big(X \in]a,b]\big) = \sum_{x \in S \cap]a,b]} P(X = x).$$

Exemplo 3.1.4. Consideremos agora a experiência aleatória que consiste na extração ao acaso de um número do intervalo $[0,1]$. Como vimos no Exemplo 2.2.7, um modelo probabilístico $(\Omega, \mathscr{A}, P)$ para esta experiência aleatória é dado por

$$\Omega = [0,1],$$

$$\mathscr{A} = \mathscr{B}([0,1])$$

e

$$P(]a,b]) = b - a, \text{ com } 0 \leq a < b \leq 1.$$

Suponhamos que estamos interessados no quadrado do número extraído, isto é, estamos interessados na variável $X(\omega) = \omega^2$, para $\omega \in [0,1]$. Esta variável toma valores em todo o intervalo $[0,1]$. Pretendemos agora conhecer a forma como os valores desta variável se distribuem em $[0,1]$. Será que tais valores se distribuem aí uniformemente? Para responder a esta questão consideremos o subintervalo $]a,b]$ de $[0,1]$ e calculemos a probabilidade de X tomar valores neste intervalo:

$$\begin{aligned} P_X(]a,b]) &= P\big(X \in]a,b]\big) \\ &= P(\{\omega \in [0,1] : X(\omega) \in]a,b]\}) \\ &= P(\{\omega \in [0,1] : \omega^2 \in]a,b]\}) \\ &= P\{\omega \in [0,1] : \omega \in]\sqrt{a}, \sqrt{b}\,]\}) \\ &= P(]\sqrt{a}, \sqrt{b}\,]) = \sqrt{b} - \sqrt{a}. \end{aligned}$$

Tendo em conta este resultado, podemos concluir que a variável X não se distribui uniformemente no intervalo $[0,1]$ pois a probabilidade da variável tomar valores em $]a,b]$ não é proporcional ao tamanho do intervalo $]a,b]$. Por exemplo, a probabilidade de X "cair" no intervalo $]0,1/4]$ não é a mesma de "cair" no intervalo $]1/4,1]$. É também interessante verificar que a probabilidade anterior pode ser obtida a partir da fórmula

$$P_X(]a,b]) = \int_a^b \frac{1}{2\sqrt{x}}\, dx,$$

para $0 \leq a < b \leq 1$, isto é, a probabilidade de X pertencer ao intervalo $]a,b]$ é dada pela área da região limitada pelas retas $x = a$ e $x = b$ e compreendida entre o gráfico da função integranda e o eixo dos xx.

Exemplo 3.1.5. Reparemos que duas variáveis aleatórias distintas, quando entendidas como funções, podem ter a mesma distribuição de probabilidade. Um exemplo simples dessa situação é o da variável Y dada por

$$Y(\omega) = \begin{cases} X(\omega), & \omega \in [0,1] \setminus \{1/4\} \\ 0, & \omega = 1/4, \end{cases}$$

onde X é variável que considerámos no exemplo anterior. Uma vez que $P(X = Y) = P(\{\omega \in [0,1] : \omega \neq 1/4\}) = 1$, então $P(Y \in]a,b]) = P(Y \in]a,b], X = Y) = P(X \in]a,b], X = Y) = P(X \in]a,b])$, para todo intervalo $]a,b]$. Isto significa que X e Y possuem a mesma lei de probabilidade, ou seja, $X \sim Y$.

Neste último exemplo, vimos também que as variáveis X e Y eram tais que $P(X = Y) = 1$, ou de forma equivalente, $P(X \neq Y) = 0$. Sempre que tal acontece dizemos que X e Y são quase certamente iguais e escrevemos $X = Y$ *q.c.* Duas variáveis aleatórias quase certamente iguais possuem a mesma distribuição de probabilidade.

3.2 Variáveis discretas e variáveis absolutamente contínuas

Os dois exemplos anteriores motivam as definições seguintes.

Definição 3.2.1. *Dizemos que uma variável aleatória real (v.a.r.) X é discreta (ou que a sua lei de probabilidade P_X é discreta) se existe um conjunto finito ou infinito numerável S tal que $P_X(S) = 1$. Ao mais pequeno conjunto S_X (no sentido da inclusão de conjuntos) com a propriedade anterior chamamos suporte de X (ou de P_X), e à função f_X definida por*

$$f_X(x) = P_X(\{x\}) = P(X = x),\ x \in \mathbb{R},$$

chamamos função de probabilidade da variável X.

Proposição 3.2.2. *Se X é v.a.r. discreta então o seu suporte é dado por*

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

Dem: Se a variável é discreta, o suporte S_X é finito ou infinito numerável com $P_X(S_X) = 1$. Denotando por A o conjunto $\{x \in \mathbb{R} : P(X = x) > 0\}$, comecemos por provar que $A \subset S_X$. Seja $x \in A$, qualquer, e suponhamos, por absurdo, que $x \notin S_X$. Neste caso teríamos

$$P_X(S_X \cup \{x\}) = P_X(S_X) + P(X = x) = 1 + P(X = x) > 1,$$

o que é contradiz ser P_X uma probabilidade. Provemos agora que $S_X \subset A$. Seja $x \in S_X$ e suponhamos, por absurdo, que $x \notin A$, isto é, $P(X = x) = 0$. Então o conjunto $S_X - \{x\}$ é finito ou infinito numerável e

$$P_X(S_X - \{x\}) = P_X(S_X) - P_X(\{x\}) = P_X(S_X) = 1,$$

o que contradiz o facto de S_X ser o mais pequeno conjunto que satisfaz as condições da definição. ■

Da proposição anterior concluímos que a função de probabilidade da variável X é tal que $f_X(x) = 0$, para $x \notin S_X$. f_X é assim uma função não negativa com

$$\sum_{x\in\mathbb{R}} f_X(x) = \sum_{x\in S_X} P_X(\{x\}) = P_X\Big(\sum_{x\in S_X} \{x\}\Big) = P_X(S_X) = 1.$$

Exemplo 3.2.3. (cont. do Exemplo 3.1.3) De acordo com a definição anterior, a variável X que nos dá a soma dos pontos obtidos em dois dados equilibrados é dis-

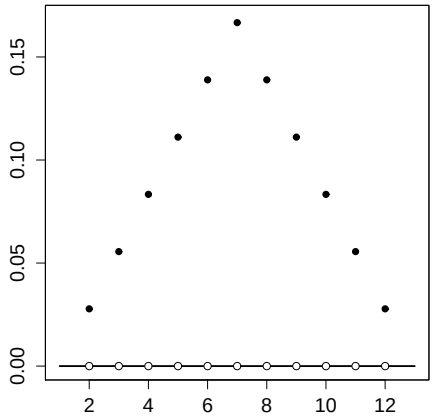


Figura 3.1: Função de probabilidade da v.a.r. X que nos dá a soma dos pontos obtidos no lançamento de dois dados equilibrados.

creta com suporte $S_X = \{2, 3, \ldots, 12\}$ e função de probabilidade dada por

$$f_X(x) = \begin{cases} 1/36, & x \in \{2, 12\} \\ 2/36, & x \in \{3, 11\} \\ 3/36, & x \in \{4, 10\} \\ 4/36, & x \in \{5, 9\} \\ 5/36, & x \in \{6, 8\} \\ 6/36, & x \in \{7\} \\ 0, & \text{caso contrário,} \end{cases}$$

Da definição anterior decorre que a função de probabilidade de uma variável discreta X carateriza a distribuição de probabilidade de X. Com efeito, para $-\infty < a \leq b < \infty$, temos

$$P_X\big(]a, b]\big) = P_X\big(]a, b] \cap S_X\big) = \sum_{x \in\,]a,b] \cap S_X} f_X(x) = \sum_{x \in\,]a,b]} f_X(x).$$

Esta igualdade motiva a definição seguinte.

Definição 3.2.4. *Dizemos que uma variável aleatória real (v.a.r.) X é absolutamente contínua (ou que a sua lei de probabilidade P_X é absolutamente contínua) se existe uma função não negativa e integrável $f_X : \mathbb{R} \to \mathbb{R}$ tal que*

$$P_X\big(]a, b]\big) = \int_a^b f_X(x)dx,$$

para todo o intervalo $]a, b]$, com $-\infty < a \leq b < \infty$. A função f_X é dita densidade de probabilidade da variável X.

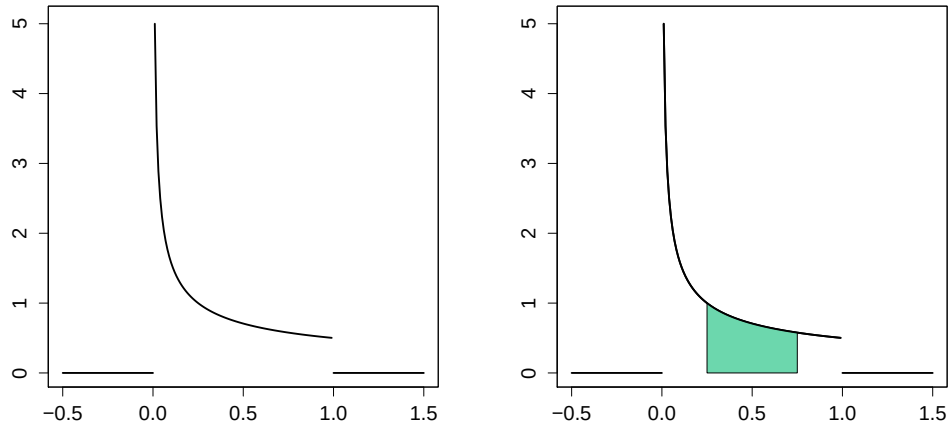


Figura 3.2: Densidade de probabilidade de X e região cuja área é a probabilidade do acontecimento $1/4 \leq X \leq 3/4$.

A densidade de probabilidade de uma variável absolutamente contínua X é assim uma função não negativa e integrável com

$$\begin{aligned}\int_{-\infty}^{\infty} f_X(x)dx = \lim \int_{-n}^{n} f_X(x)dx = \lim P_X\big(]-n,n]\big) \\ = P_X\big(\lim]-n,n]\big) = P_X(\mathbb{R}) = 1.\end{aligned}$$

Atendendo a que a σ-álgebra de Borel é gerada pelos intervalos da forma $]a,b]$, com $-\infty < a \leq b < \infty$, é possível provar que

$$P_X(B) = \int_B f_X(x)dx,$$

para todo o $B \in \mathscr{B}(\mathbb{R})$.

Exemplo 3.2.5 (cont. do Exemplo 3.1.4)**.** No caso da variável X que nos dá o quadrado de um número extraído ao acaso do intervalo $[0,1]$, podemos dizer que se trata de uma variável absolutamente contínua uma vez que para todo intervalo $]a,b]$, com $-\infty < a \leq b < \infty$, se tem

$$P_X\big(]a,b]\big) = \int_a^b f_X(x)dx,$$

onde

$$f_X(x) = \begin{cases} \dfrac{1}{2\sqrt{x}}, & 0 < x < 1 \\ 0, & \text{caso contrário.} \end{cases}$$

Em abono da verdade, é preciso dizer que os integrais que surgem nas expressões anteriores devem ser interpretados como integrais à Lebesgue e não como os habituais

integrais à Riemann. Usando a notação habitual para o integral de Lebesgue, deveríamos ter escrito $\int_a^b f_X(x)d\lambda(x)$ em vez de $\int_a^b f_X(x)dx$. O integral de Lebesgue, que não vamos aqui desenvolver por ultrapassar os objetivos deste curso, tem propriedades mais ricas que as do integral de Riemann permitindo que a exposição das matérias se possa desenvolver a um nível que integral de Riemann não permite. No entanto, em termos puramente calculatórios é possível usar o integral de Riemann, bem nosso conhecido, para calcular o integral de Lebesgue e, neste aspeto, não se perderá muito ao usarmos o integral de Riemann em vez do integral de Lebesgue. Para explicarmos completamente esta questão precisamos da noção de conjunto de medida nula (à Lebesgue).

Definição 3.2.6. *Dado um subconjunto A de $\mathbb{R}$, dizemos que A tem medida nula (ou é de medida nula) se para todo o $\epsilon > 0$ existir uma sucessão de intervalos $\left(]a_n, b_n[\right)$ tais que*

$$A \subset \bigcup_{n=1}^{\infty}]a_n, b_n[\quad \text{e} \quad \sum_{n=1}^{\infty}(b_n - a_n) \leq \epsilon.$$

É fácil provar que: 1) o conjunto vazio tem medida nula; 2) todo o subconjunto de um conjunto de medida nula tem medida nula; 3) um conjunto com um número finito de pontos tem medida nula; 4) um conjunto com uma infinidade numerável de pontos tem medida nula; 5) a reunião finita ou infinita numerável de conjuntos com medida nula tem medida nula.

Retomando agora a questão do cálculo do integral no sentido de Lebesgue, verifica-se que se uma função limitada f for integrável à Riemann no intervalo $[a, b]$, o que acontece se o conjunto dos seus pontos de descontinuidade tiver medida nula (esta importante caraterização das funções integráveis no sentido de Riemann, foi estabelecida por Henri Lebesgue (1875-1941) em 1902) — o que acontece, por exemplo, se f é contínua em $[a, b]$ ou tiver aí um conjunto finito ou infinito numerável de pontos de descontinuidade —, então f também é integrável à Lebesgue e tem-se

$$\int_a^b f_X(x)d\lambda(x) = \int_a^b f_X(x)dx.$$

Reparemos também que não existe apenas uma função f_X que satisfaz a Definição 3.2.4. Se g for uma função integrável e não negativa que coincide com f_X exceto num conjunto de medida nula (dizemos neste caso que $g = f_X$ quase em todo o ponto), tem-se

$$\int_a^b f_X(x)dx = \int_a^b g(x)dx,$$

para todo o intervalo $]a, b]$, com $-\infty < a \leq b < \infty$, e, de acordo com a definição dada, g

é também densidade de probabilidade da variável aleatória X. Por esta razão, dizemos muitas vezes que f_X (ou g) é uma versão da densidade de probabilidade de X.

Provamos a seguir que a distribuição de probabilidade P_X de uma v.a.r. X absolutamente contínua é difusa, isto é, $P_X(\{x\}) = 0$, para todo o $x \in \mathbb{R}$.

Proposição 3.2.7. *Se X é v.a.r. absolutamente contínua então*

$$P(X = x) = 0, \text{ para todo o } x \in \mathbb{R}.$$

Dem: Apesar do resultado ser válido para uma qualquer variável aleatória absolutamente contínua, vamos limitar-nos a demonstrá-lo no caso em que f_X é limitada e integrável à Riemann em pelo menos um dos intervalos $[x-\epsilon, x]$ e $[x, x+\epsilon]$, para algum $\epsilon > 0$. No primeiro caso, tendo em conta que

$$\{x\} = \lim]x - 1/n, x],$$

pela continuidade de P_X (ver Proposição 2.3.10) temos

$$P(X = x) = P_X(\{x\}) = \lim P_X\big(]x - 1/n, x]\big) = \lim \int_{x-1/n}^{x} f_X(y)dy.$$

Para concluir a demonstração basta agora ter em conta que sendo $M > 0$ tal que $f_X(y) \leq M$, para todo o $y \in [x-\epsilon, x]$, então para n suficientemente grande temos

$$\int_{x-1/n}^{x} f_X(y)dy \leq M \int_{x-1/n}^{x} dy = M\frac{1}{n}.$$

De forma análoga se procederia quando f_X é limitada e integrável à Riemann em $[x, x+\epsilon]$, para algum $\epsilon > 0$. ■

3.3 Função de distribuição

Na secção anterior definimos duas funções que caraterizam a distribuição de probabilidade duma v.a.r. X. No caso discreto falámos da função de probabilidade de X e no caso absolutamente contínuo falámos da densidade de probabilidade de X. Nesta secção estudamos uma nova função que, independentemente do tipo de distribuição da v.a.r. X, carateriza a sua distribuição de probabilidade.

Definição 3.3.1. *Chamamos função de distribuição da v.a.r. X, a função real de variável real definida por*

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

Sendo X uma v.a.r. discreta com suporte S_X, decorre da definição anterior que, para $x \in \mathbb{R}$, a sua função de distribuição é dada por

$$\begin{aligned} F_X(x) &= P_X\big(]-\infty, x]\big) = P_X\big(]-\infty, x] \cap S_X\big) \\ &= \sum_{y \in S_X : y \leq x} P_X(\{y\}) = \sum_{y \in S_X : y \leq x} P(X = y) \\ &= \sum_{y \in S_X : y \leq x} f_X(y) = \sum_{y \leq x} f_X(y). \end{aligned}$$

No caso em que X é absolutamente contínua, para $x \in \mathbb{R}$ temos

$$\begin{aligned} F_X(x) &= P_X(]-\infty, x]) = \lim P_X(]-n, x]) \\ &= \lim \int_{-n}^{x} f_X(y)dy = \int_{-\infty}^{x} f_X(y)dy. \end{aligned}$$

Retomemos agora os Exemplos 3.2.3 e 3.2.5 e calculemos a função de distribuição das variáveis aleatórias aí definidas.

Exemplo 3.3.2 (cont. do Exemplo 3.2.3)**.** Para a variável discreta X que nos dá a soma dos pontos obtidos em dois dados equilibrados, cujo suporte é $S_X = \{2, 3, \ldots, 12\}$, a sua função de distribuição é dada por

$$F_X(x) = \begin{cases} 0, & x < 2 \\ 1/36, & 2 \leq x < 3 \\ 3/36, & 3 \leq x < 4 \\ 6/36, & 4 \leq x < 5 \\ 10/36, & 5 \leq x < 6 \\ 15/36, & 6 \leq x < 7 \\ 21/36, & 7 \leq x < 8 \\ 26/36, & 8 \leq x < 9 \\ 30/36, & 9 \leq x < 10 \\ 33/36, & 10 \leq x < 11 \\ 35/36, & 11 \leq x < 12 \\ 1, & x \geq 12. \end{cases}$$

Exemplo 3.3.3 (cont. do Exemplo 3.2.5)**.** Para a variável absolutamente contínua X que nos dá o quadrado de um número extraído ao acaso do intervalo $[0, 1]$, a sua função de distribuição é dada por

$$F_X(x) = \begin{cases} 0, & x \leq 0 \\ \sqrt{x}, & 0 < x < 1 \\ 1, & x \geq 1. \end{cases}$$

A função de distribuição F_X goza de um conjunto importante de propriedades que enunciamos no resultado seguinte. A última dessas propriedades permite-nos dizer que F_X carateriza a distribuição de probabilidade de X.

Proposição 3.3.4. *F_X satisfaz as seguintes propriedades:*

a) F_X é não negativa, limitada e crescente (por crescente entendemos sempre, crescente em sentido lato, isto é, $F_X(x) \leq F_X(y)$, para todo o $x < y$);

b) $\lim_{x\to-\infty} F_X(x) = 0$ e $\lim_{x\to+\infty} F_X(x) = 1$;

c) F_X é contínua à direita, isto é, para todo o $x \in \mathbb{R}$

$$F_X(x^+) := \lim_{y\to x, y\geq x} F_X(y) = F_X(x);$$

d) Para $x \in \mathbb{R}$

$$F_X(x^-) := \lim_{y\to x, y\leq x} F_X(y) = F_X(x) - P(X = x);$$

e) F_X é contínua em $x \in \mathbb{R}$ sse $P(X = x) = 0$;

f) O conjunto dos pontos de descontinuidade de F_X é finito ou infinito numerável;

g) Para todo o intervalo $]a, b]$, com $-\infty < a \leq b < \infty$, temos

$$P_X\big(]a, b]\big) = F_X(b) - F_X(a);$$

h) F_X carateriza P_X (isto é, $F_X = F_Y$ sse $X \sim Y$).

Dem: a) Para todo o $x \in \mathbb{R}$ temos $F_X(x) \in [0, 1]$, sendo assim F_X não negativa e limitada. Por outro lado, pela monotonia de P_X, para $x < y$ temos $F_X(x) = P_X\big(]-\infty, x]\big) \leq P_X\big(]-\infty, y]\big) = F_X(y)$.

b) Sendo F_X monótona e limitada, existem os dois limites indicados. Assim, pela continuidade de P_X (ver Proposição 2.3.10) temos

$$\begin{aligned}\lim_{x\to-\infty} F_X(x) = \lim_{n\to\infty} F_X(-n) &= \lim_{n\to\infty} P_X\big(]-\infty, -n]\big) \\ &= P_X\big(\lim_{n\to\infty}]-\infty, -n]\big) = P_X\big(\emptyset\big) = 1,\end{aligned}$$

e

$$\begin{aligned}\lim_{x\to+\infty} F_X(x) = \lim_{n\to+\infty} F_X(n) &= \lim_{n\to+\infty} P_X\big(]-\infty, n]\big) \\ &= P_X\big(\lim_{n\to+\infty}]-\infty, n]\big) = P_X(\mathbb{R}) = 1,\end{aligned}$$

uma vez que

$$\lim_{n\to\infty}]-\infty, -n] = \bigcap_{n=1}^{\infty}]-\infty, -n] = \emptyset$$

e

$$\lim_{n\to\infty}]-\infty, n] = \bigcup_{n=1}^{\infty}]-\infty, n] = \mathbb{R}.$$

c) e d) Sendo F_X monótona e limitada, para todo o $x \in \mathbb{R}$ os limites laterais $\lim_{y\to x, y\leq x} F_X(y)$ e $\lim_{y\to x, y\geq x} F_X(y)$ existem. Assim, pela continuidade de P_X temos

$$\begin{aligned}\lim_{y\to x, y\leq x} F_X(y) &= \lim_{n\to+\infty} F_X(x-1/n) = \lim_{n\to+\infty} P_X(]-\infty, x-1/n])\\ &= P_X\big(]-\infty, x[\big) = P_X\big(]-\infty, x] - \{x\}\big) = F_X(x) - P(X=x)\end{aligned}$$

e

$$\begin{aligned}\lim_{y\to x, y\geq x} F_X(y) &= \lim_{n\to+\infty} F_X(x+1/n) = \lim_{n\to+\infty} P_X(]-\infty, x+1/n])\\ &= P_X\big(]-\infty, x]\big) = F_X(x),\end{aligned}$$

uma vez que

$$\lim_{n\to+\infty}]-\infty, x-1/n] = \bigcup_{n=1}^{\infty}]-\infty, x-1/n] =]-\infty, x[$$

e

$$\lim_{n\to+\infty}]-\infty, x+1/n] = \bigcap_{n=1}^{\infty}]-\infty, x+1/n] =]-\infty, x].$$

e) A propriedade enunciada é consequência de c) e d).

f) Pela alínea anterior, sabemos que o conjunto D dos pontos de descontinuidade de F_X é dado por $D = \{x \in \mathbb{R} : P(X=x) > 0\}$. Reparemos que $D = \bigcup_{n=1}^{\infty} D_n$ com $D_n = \{x \in \mathbb{R} : P(X=x) \geq 1/n\}$. Para concluirmos que F_X admite quando muito uma infinidade numerável de pontos de descontinuidade basta provar que cada um dos conjuntos D_n é finito. Mais precisamente, vamos mostrar que $\#D_n \leq n$. Com efeito, se fosse $\#D_n > n$, para $D'_n \subset D_n$ com $\#D'_n = n+1$, teríamos

$$P_X(D_n) \geq P_X(D'_n) = P_X\Big(\bigcup_{x\in D'_n} \{x\}\Big) = \sum_{x\in D'_n} P(X=x) \geq \sum_{x\in D'_n} \frac{1}{n} = \frac{\#D'_n}{n} > 1,$$

o que é falso.

g) Da Proposição 2.3.4 tiramos

$$P_X(]a,b]) = P_X(]-\infty, b]-]-\infty, a]) = P_X(]-\infty, b]) - P_X(]-\infty, a]) = F_X(b) - F_X(a).$$

h) Se $X \sim Y$, então $P_X = P_Y$ e assim $F_X(x) = P_X(]-\infty, x]) = P_Y(]-\infty, x]) = F_Y(x)$, para $x \in \mathbb{R}$. Reciprocamente, se $F_X = F_Y$, da alínea anterior concluímos $P_X(]a,b]) = P_Y(]a,b])$, para todo o intervalo $]a,b]$ com $-\infty < a \leq b < \infty$. Podemos então concluir que $P_X(B) = P_Y(B)$, para todo o $B \in \mathscr{B}(\mathbb{R})$, ou seja que $X \sim Y$. ■

Proposição 3.3.5. *Se X é uma v.a.r. discreta de suporte S_X então*

$$S_X = \{x \in \mathbb{R} : F_X \text{ é descontínua em } x\},$$

e

$$f_X(x) = F_X(x) - F_X(x^-), \quad x \in \mathbb{R}.$$

Dem: Sendo X discreta, pelas Proposições 3.2.2 e 3.3.4.d) sabemos que o suporte de X é dado por

$$S_X = \{x \in \mathbb{R} : P(X = x) > 0\} = \{x \in \mathbb{R} : F_X \text{ é descontínua em } x\}.$$

Utilizando novamente a Proposição 3.3.4.d), para $x \in \mathbb{R}$ temos

$$f_X(x) = P(X = x) = F_X(x) - F_X(x^-). \qquad \blacksquare$$

Sendo X é discreta, a função F_X é assim uma função em escada, tendo descontinuidades ("saltos") nos pontos de S_X.

Proposição 3.3.6. *Sejam X é uma v.a.r. e $S = \{x \in \mathbb{R} : F_X \text{ é descontínua em } x\}$. Então X é discreta sse $P_X(S) = 1$.*

Dem: Se S é tal que $P_X(S) = 1$, uma vez que S é finito ou infinito numerável (Proposição 3.3.4.e), concluímos que X é discreta. Suponhamos agora que X é discreta (ver Definição 3.2.1). Pela Proposição 3.3.5 sabemos que o seu suporte S_X é dado por $S_X = S$ e portanto $P_X(S) = P_X(S_X) = 1$. $\blacksquare$

Proposição 3.3.7. *Se X é uma v.a.r. com*

$$F_X(x) = \int_{-\infty}^{x} g(y)dy, \ x \in \mathbb{R},$$

para alguma função não negativa e integrável $g : \mathbb{R} \to \mathbb{R}$, então X é absolutamente contínua e g é uma versão da densidade de probabilidade de X.

Dem: Para $-\infty \leq a < b \leq +\infty$ temos

$$P_X(\,]a, b]) = F_X(b) - F_X(a) = \int_{-\infty}^{b} g(y)dy - \int_{-\infty}^{a} g(y)dy = \int_{a}^{b} g(y)dy.$$

Uma vez que, por hipótese, g é uma função não negativa e integrável, da Definição 3.2.4 concluímos que X é absolutamente contínua e que g uma versão da densidade de probabilidade de X. $\blacksquare$

Proposição 3.3.8. *Se X é absolutamente contínua então $f_X = F'_X$ em quase todo o ponto de $\mathbb{R}$ (isto é, o conjunto dos pontos $x \in \mathbb{R}$ para os quais $f_X(x) \neq F'_X(x)$, tem medida de Lebesgue nula).*

Dem: Vamos limitar-nos a demonstrar este resultado no caso em que f_X é contínua em $\mathbb{R}$. Para $x \in \mathbb{R}$ e $a < x$ temos

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt = \int_{-\infty}^{a} f_X(t)dt + \int_{a}^{x} f_X(t)dt = F_X(a) + \int_{a}^{x} f_X(t)dt.$$

Sendo f_X contínua em $\mathbb{R}$, pelo Teorema Fundamental do Cálculo sabemos que F_X é derivável em $\mathbb{R}$ e para todo o $x \in \mathbb{R}$ temos que

$$F'_X(x) = \frac{d}{dx}\int_{a}^{x} f_X(t)dt = f_X(x). \qquad \blacksquare$$

No caso de X ser absolutamente contínua podemos deduzir do resultado anterior que F_X é diferenciável em $\mathbb{R}$ a menos de um conjunto de medida nula. Um tal resultado é, no entanto, válido para uma qualquer variável aleatória X. Tal facto decorre de um teorema de Lebesgue (de 1904) que estabelece que qualquer função monótona — propriedade que, como já vimos, F_X possui —, é diferenciável em todo o ponto de $\mathbb{R}$ com exceção de um conjunto de pontos de medida nula. Tal é óbvio no caso em que X é uma variável discreta de suporte S_X. Sendo F_X descontínua para $x \in S_X$, sabemos que F_X não é diferenciável nestes ponto. Para os restantes pontos $x \notin S_X$, sendo F_X constante numa vizinhança desses pontos concluímos que $F'_X(x) = 0$, para $x \notin S_X$. Atendendo a que S_X tem medida de Lebesgue nula (pois S_X é finito ou infinito numerável), isto significa que $F'_X(x) = 0$ em quase todo o ponto $x \in \mathbb{R}$. Assim, quando falamos da derivada F'_X da função de distribuição de uma variável aleatória X, não estamos a afirmar que tal derivada existe em todo o ponto de $\mathbb{R}$. O que sabemos do referido teorema de Lebesgue é que F'_X existe a menos de um conjunto de pontos de medida nula.

Proposição 3.3.9. *A v.a.r. X é absolutamente contínua sse $\int_{-\infty}^{\infty} F'_X(x)dx = 1$.*

Dem: Comecemos por supor que X é absolutamente contínua. Da proposição anterior sabemos que $f_X = F'_X$ em quase todo o ponto de $\mathbb{R}$, e portanto

$$\int_{-\infty}^{\infty} F'_X(x)dx = \int_{-\infty}^{\infty} f_X(x)dx = 1.$$

Relativamente à implicação recíproca, vamos demonstrá-la apenas no caso em que F_X é continuamente diferenciável em $\mathbb{R}$. Pelo Teorema Fundamental do Cálculo deduzimos que

$$\int_{a}^{b} F'_X(x)dx = F_X(b) - F_X(a) = P_X(]a, b]),$$

para todo o intervalo real $]a, b]$. Uma vez que F'_X é não negativa (porquê?) e integrável, concluímos que X é absolutamente contínua. ■

Há casos de variáveis aleatórias que não pertencem a nenhuma das duas classes que temos vindo a analisar. A tais variáveis chamamos variáveis aleatórias mistas.

Exemplo 3.3.10. Consideremos novamente a experiência aleatória que consiste na extração ao acaso de um número do intervalo $[0, 1]$, cujo modelo probabilístico associado é $(\Omega, \mathscr{A}, P)$ com $\Omega = [0, 1]$, $\mathscr{A} = \mathscr{B}([0, 1])$ e $P(]a, b]) = b - a$, com $0 \leq a < b \leq 1$ (ver Exemplo 2.2.7). Suponhamos agora que estamos interessados na v.a.r. definida por $X(\omega) = \min(1/2, \omega)$, isto é, $X(\omega) = \omega$ se $0 \leq \omega \leq 1/2$ e $X(\omega) = 1/2$ se $1/2 < \omega \leq 1$. A função de distribuição de X é dada por

$$F_X(x) = \begin{cases} 0, & x < 0 \\ x, & 0 \leq x < 1/2 \\ 1, & x \geq 1/2. \end{cases}$$

Neste caso $S = \{x \in \mathbb{R} : F_X \text{ é descontínua em } x\} = \{1/2\}$ e $P_X(S) = P(X = 1/2) = 1 - 1/2 = 1/2 \neq 1$. Da Proposição 3.3.6 concluímos que X não é discreta. X também não é absolutamente contínua uma vez que F_X não é contínua em $\mathbb{R}$. Outra forma de verificar que X não é absolutamente contínua é calcular

$$F'_X(x) = \begin{cases} 0, & x < 0 \vee x > 1/2 \\ 1, & 0 < x < 1/2, \end{cases}$$

(não interessa o que se passa em $x = 0$ e $x = 1/2$) e verificar que $\int F'_X(x)dx = \int_0^{1/2} 1dx = 1/2 \neq 1$.

É fácil verificar que F_X se pode escrever na forma

$$F_X(x) = \frac{1}{2}F_d(x) + \frac{1}{2}F_{ac}(x),$$

chamada decomposição de Lebesgue de F_X, onde

$$F_d(x) = \begin{cases} 0, & x < 1/2 \\ 1, & x \geq 1/2 \end{cases} \quad \text{e} \quad F_{ac}(x) = \begin{cases} 0, & x < 0 \\ 2x, & 0 \leq x < 1/2 \\ 1, & x \geq 1/2. \end{cases}$$

Notemos que F_d é a função de distribuição da variável aleatória discreta $Y(\omega) = 1/2$ e F_{ac} é a função de distribuição da variável absolutamente contínua $Z(\omega) = \omega/2$, para $\omega \in [0, 1]$.

3.4 Transformação afim de variáveis aleatórias

Uma primeira aplicação dos resultados anteriores pode ser encontrada para provar que uma transformação afim de uma variável discreta é ainda discreta, e que uma transformação afim de uma variável absolutamente contínua é ainda absolutamente contínua.

Proposição 3.4.1. *Sejam X uma v.a.r. e $Y = aX + b$ com $a > 0$ e $b \in \mathbb{R}$. Então*

$$F_Y(y) = F_X\big((y-b)/a\big),\ y \in \mathbb{R}.$$

Além disso,

a) se X é discreta com suporte S_X, então Y é discreta com suporte $aS_X + b$ e função de probabilidade

$$f_Y(y) = f_X\big((y-b)/a\big),\ y \in \mathbb{R};$$

b) se X é absolutamente contínua, então Y é absolutamente contínua com densidade de probabilidade

$$f_Y(y) = f_X\big((y-b)/a\big)/a,\ y \in \mathbb{R}.$$

Dem: Para $y \in \mathbb{R}$ temos

$$F_Y(y) = P(Y \le y) = P(aX + b \le y) = P\big(X \le (y-b)/a\big) = F_X\big((y-b)/a\big).$$

a) Suponhamos que X é discreta com suporte S_X, que sabemos ser finito ou infinito numerável. Assim, $aS_X + b = \{ax + b : x \in S_X\}$ é também finito ou infinito numerável e

$$P_Y(aS_X + b) = P(Y \in aS_X + b) = P(aX + b \in aS_X + b) = P(X \in S_X) = 1.$$

Por definição de v.a.r. discreta, concluímos que Y é discreta e que $S_Y \subset aS_X + b$. Para estabelecer que $S_Y = aS_X + b$ tomemos $y \in aS_X + b$, isto é, $y = ax + b$ com $P(X = x) > 0$. Daqui, concluímos que $P(Y = y) = P(aX + b = ax + b) = P(X = x) > 0$, ou seja, $y \in S_Y$. Para concluir a demonstração da parte a), basta utilizar a Proposição 3.3.5 para concluirmos que

$$f_Y(y) = F_Y(y) - F_Y(y^-) = F_X\big((y-b)/a\big) - F_X\big(((y-b)/a)^-\big) = f_X\big((y-b)/a\big).$$

b) Suponhamos agora que X é absolutamente contínua. Para quase todo o ponto $y \in \mathbb{R}$ temos

$$F_Y'(y) = F_X'\big((y-b)/a\big)/a = f_X\big((y-b)/a\big)/a$$

e

$$\int_{-\infty}^{\infty} F_Y'(y)\,dy = \int_{-\infty}^{\infty} f_X\big((y-b)/a\big)/a\,dy = \int_{-\infty}^{\infty} f_X(x)dx = 1.$$

Pela Proposição 3.3.9 concluímos que Y é absolutamente contínua, e pela Proposição 3.3.8 concluímos que $f_Y(y) = f_X\big((y-b)/a\big)/a$ é a sua densidade de probabilidade. ■

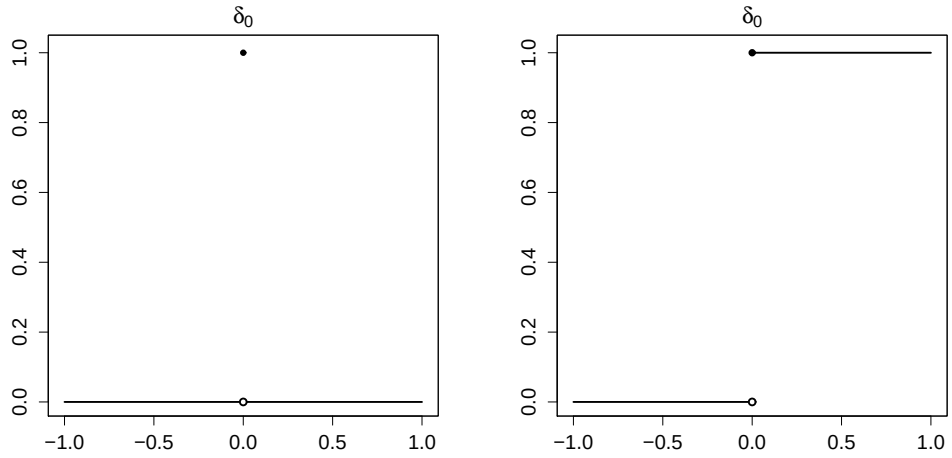


Figura 3.3: Funções de probabilidade e de distribuição da variável de Dirac no ponto 0.

3.5 Exemplos de leis de probabilidade

3.5.1 Leis discretas

Lei de Dirac

Consideremos uma experiência aleatória descrita pelo espaço de probabilidade $(\Omega, \mathscr{A}, P)$ e seja A um acontecimento com $P(A) = 1$. Seja X uma v.a.r. que satisfaz $X(\omega) = a$ para $\omega \in A$, para $a \in \mathbb{R}$ fixo, a que chamamos variável de Dirac no ponto a. A sua lei de probabilidade, dita lei de Dirac no ponto a, é dada por

$$P_X(\{x\}) = P(X = x) = \begin{cases} P(A) = 1, & x = a \\ 0, & x \neq a \end{cases} =: \delta_a(x),$$

onde δ_a é dita função de Dirac no ponto a. A sua função de distribuição é dada por

$$F_X(x) = \begin{cases} 0, & x < a \\ 1, & x \geq a \end{cases} = \mathbb{1}_{[a,+\infty[}(x),$$

onde denotamos por $\mathbb{1}_A$ a função indicatriz do conjunto A, isto é,

$$\mathbb{1}_A(x) = \begin{cases} 1, & x \in A \\ 0, & x \notin A. \end{cases}$$

Uma vez que $P(X = a) = 1$ dizemos também que X é igual a a quase certamente, e escrevemos $X = a$ *q.c.* Mais geralmente, uma propriedade $\mathscr{P}(\omega)$ relativa aos pontos $\omega \in \Omega$ diz-se quase certamente verdadeira se $P\big(\{\omega : \mathscr{P}(\omega) \text{ é verdadeira}\}\big) = 1$.

Lei de Bernoulli

Sejam $(\Omega, \mathscr{A}, P)$ um espaço de probabilidade e A um acontecimento cuja probabilidade é $p \in]0,1[$. Considere a v.a.r. definida por $X(\omega) = 1$, se $\omega \in A$ e $X(\omega) = 0$, se $\omega \notin A$, a que chamamos X variável de Bernoulli de parâmetro p. A lei de probabilidade de X, dita lei de Bernoulli de parâmetro p, é discreta de suporte $S_X = \{0,1\}$. A função de probabilidade de X é dada por

$$f_X(x) = \begin{cases} 1-p, & x = 0 \\ p, & x = 1 \\ 0, & x \notin \{0,1\}, \end{cases}$$

e a função de distribuição de X é dada por

$$F_X(x) = \begin{cases} 0, & x < 0 \\ 1-p, & 0 \leq x < 1 \\ 1, & x \geq 1. \end{cases}$$

Quando pretendemos indicar que uma determinada variável X possui ou segue uma lei de Bernoulli de parâmetro p, escrevemos $X \sim B(p)$. Notemos que nos casos limite em que $p = 0$ e $p = 1$, a lei de Bernoulli reduz-se à lei de Dirac nos pontos 0 e 1, respetivamente.

Lei uniforme discreta

Dado um espaço de probabilidade $(\Omega, \mathscr{A}, P)$, suponha que existem n acontecimentos aleatórios incompatíveis A_k, $k = 1, \ldots, n$, com $\Omega = A_1 \cup \cdots \cup A_n$ e $P(A_k) = 1/n$. A variável aleatória definida por $X(\omega) = k$, se $\omega \in A_k$, tem suporte $S_X = \{1, \ldots, n\}$ e lei de probabilidade

$$P_X(\{k\}) = P(X = k) = P(A_k) = 1/n,$$

para $k = 1, \ldots, n$, a que chamamos lei uniforme discreta sobre $\{1, \ldots, n\}$. À v.r.a. chamamos variável uniforme discreta sobre $\{1, \ldots, n\}$.

Lei binomial

Consideremos um acontecimento A de probabilidade $p \in]0,1[$, associado a determinada experiência aleatória, e suponhamos que a experiência aleatória é repetida n vezes, sempre nas mesmas condições. Representemos por X a v.a.r. que nos dá o número vezes em que ocorreu o acontecimento A nas n repetições da experiência. Designando por sucesso a ocorrência do acontecimento A e por insucesso a não ocorrência do acontecimento A, X dá-nos o número de sucessos ocorridos nas n repetições da experiência inicial. X é

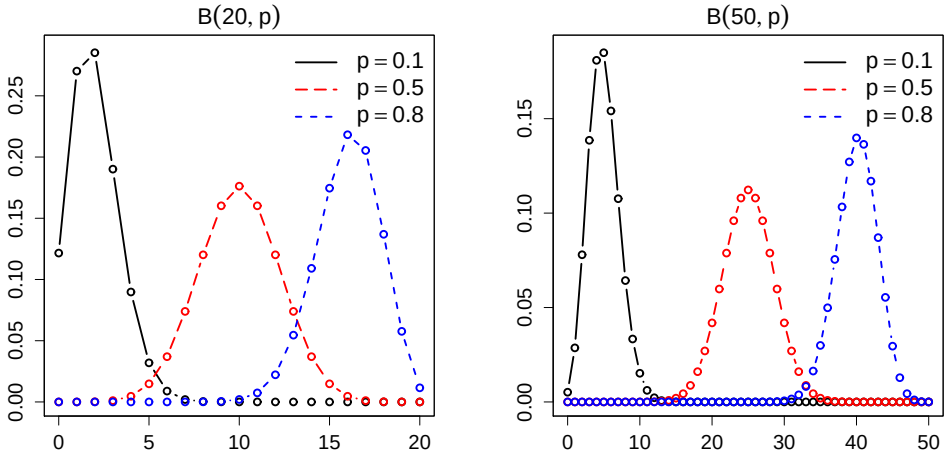


Figura 3.4: Funções de probabilidade da variável $B(n, p)$.

assim uma variável discreta com suporte $S_X = \{0, 1, \ldots, n\}$ a que chamamos variável binomial de parâmetros n e p, e indicamos $X \sim B(n, p)$.

Para determinar a distribuição de probabilidade de X, temos de calcular $P(X = k)$, para $k \in S_X$. O acontecimento $X = k$ realiza-se quando ocorrerem k sucessos e $n - k$ insucessos nas n repetições. Suponhamos que ocorreram sucessos nas k primeiras repetições da experiência e insucessos nas restantes $n - k$ repetições, ou seja, ocorreram os acontecimentos A_i para $i = 1, \ldots, k$ e A_i^c para $i = k + 1, \ldots, n$, onde A_i representa a ocorrência do acontecimento A na i-ésima repetição da experiência. Por outras palavras, ocorreu o acontecimento

$$A_1 \cap \cdots \cap A_k \cap A_{k+1}^c \cap \cdots \cap A_n^c.$$

Sendo os vários acontecimentos independentes (uma vez que a experiência é repetida nas mesmas condições), a probabilidade deste acontecimento é dada por

$$p^k (1 - p)^{n-k}.$$

Atendendo a que os resultados favoráveis ao acontecimento $X = k$ são sempre do tipo anterior, surgindo os k sucessos em diferentes posições, o número total de tais resultados é igual ao número de subconjuntos de $\{1, \ldots, n\}$ com k elementos, ou seja,

$$C_k^n = \frac{n!}{(n - k)!k!},$$

onde C_k^n representa as combinações de n elementos tomados k a k. Podemos então concluir que

$$P(X = k) = C_k^n p^k (1 - p)^{n-k},$$

para $k = 0, 1, \ldots, n$.

Notemos que se X representa o número de sucessos obtidos nas n repetições, então $n - X$ dá-nos o número de insucessos obtidos nas n repetições. Uma vez que a probabilidade de insucesso é $1 - p$, deduzimos que $n - X \sim B(n, 1 - p)$.

Lei de Poisson

Consideremos agora uma sucessão de experiências aleatórias binomiais, em que a probabilidade de sucesso $p_n \in]0, 1[$ não se mantém constante, dependendo do número n de repetições da experiência com $\lim np_n = \lambda \in]0, +\infty[$ (o que implica que $\lim p_n = 0$). Se representarmos por X_n a variável aleatória binomial de parâmetros n e p_n, vimos que X_n tem suporte $S_{X_n} = \{0, 1, \ldots, n\}$ e que

$$P(X_n = k) = C_k^n p_n^k (1 - p_n)^{n-k},$$

para $k = 0, 1, \ldots, n$. Se mantivermos k fixo e fizermos n tender para $+\infty$, para $\epsilon_n = np_n$ obtemos

$$\begin{aligned} P(X_n = k) &= C_k^n \Big(\frac{\epsilon_n}{n}\Big)^k \Big(1 - \frac{\epsilon_n}{n}\Big)^{n-k} \\ &= \frac{n(n-1)(n-(k-1))}{n^k} \frac{\epsilon_n^k}{k!} \Big(1 - \frac{\epsilon_n}{n}\Big)^n \Big(1 - \frac{\epsilon_n}{n}\Big)^{-k} \\ &\to \frac{\lambda^k}{k!} e^{-\lambda},\ n \to +\infty. \end{aligned}$$

É interessante verificar que como limite obtemos uma distribuição de probabilidade com suporte $\mathbb{N}_0$ uma vez que

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

Representando por X a v.a.r. com suporte $\mathbb{N}_0$ tal que

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

dizemos que X é uma variável aleatória de Poisson de parâmetro $\lambda > 0$, e escrevemos $X \sim \mathscr{P}(\lambda)$. Esta distribuição de probabilidade poderá ser adequada para modelar o número de ocorrências de determinado acontecimento de probabilidade pequena (pois $\lim p_n = 0$) quando não limitamos, à partida, o número de repetições da experiência. Por exemplo, suponhamos que determinado sistema de produção complexo depende de um grande número (n) de componentes, cuja respetiva probabilidade de falha num determinado período de tempo (p_n) é muito pequena. A variável X representará o número de componentes que falham nesse período de tempo.

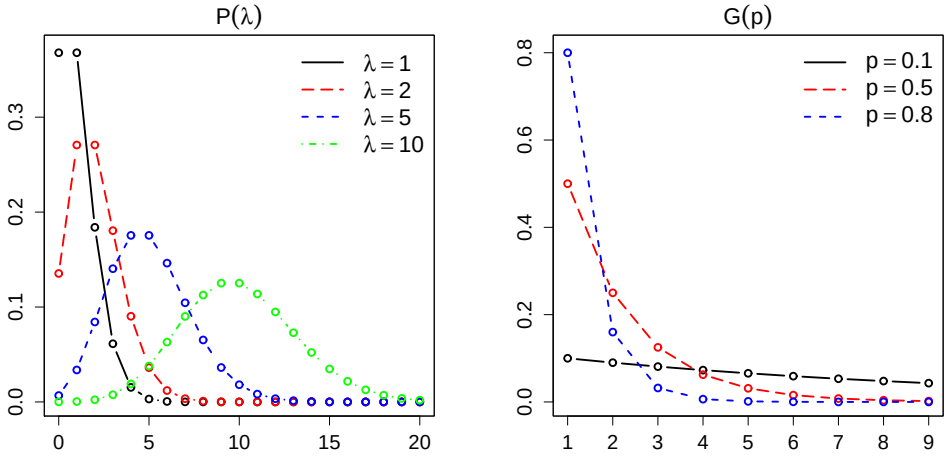


Figura 3.5: Funções de probabilidade das variáveis $\mathscr{P}(\lambda)$ e $G(p)$.

A distribuição de Poisson de parâmetro $\lambda > 0$ pode também ser interpretada como a distribuição de probabilidade da variável aleatória X que nos dá o número de ocorrências (sucessos) de determinado fenómeno aleatório (por exemplo, chegada de uma chamada a uma central telefónica) durante um intervalo de tempo $]a, b]$, em que o número médio de ocorrências do fenómeno em estudo em tal intervalo é λ. Para justificar esta interpretação, vamos decompor o intervalo $]a, b]$ em n subintervalos de amplitude $(b-a)/n$,

$$]a, b] = \bigcup_{k=1}^{n} \left]a + (k-1)(b-a)/n, a + k(b-a)/n\right],$$

em que n é suficientemente grande para que o fenómeno em causa não ocorra mais do que uma vez em cada subintervalo. Denotando por $A_{n,k}$ o acontecimento que ocorre se o fenómeno em estudo ocorre no subintervalo $]a+(k-1)(b-a)/n, a+k(b-a)/n]$, assumimos que os acontecimentos $A_{n,1}, \ldots, A_{n,n}$ são independentes e que a probabilidade de $A_{n,k}$ é $p_n \in]0, 1[$. Nestas condições, se X_n representar o número de ocorrências (sucessos) do fenómeno aleatório durante o intervalo $]a, b]$, podemos dizer que $X_n \sim B(n, p_n)$. Como veremos no próximo capítulo, o número médio de sucessos associados a esta variável binomial é np_n, o que significa que devemos ter $\lim_{n\to\infty} np_n = \lambda \in]0, +\infty[$, uma vez que λ é o número médio de ocorrências do fenómeno em estudo no intervalo $]a, b]$. Assim sendo, para $k \in \mathbb{N}_0$, será natural tomar

$$P(X = k) = \lim_{n\to\infty} P(X_n = k) = \frac{\lambda^k}{k!} e^{-\lambda},$$

ou seja, X possui uma distribuição de Poisson de parâmetro $\lambda > 0$.

Lei geométrica

Consideremos um acontecimento A de probabilidade $p \in]0,1[$, associado a determinada experiência aleatória. Suponhamos que a experiência aleatória é sucessivamente repetida, sempre nas mesmas condições, e denotemos por X a v.a.r. que nos dá o número repetições da experiência até se realizar, pela primeira vez, o acontecimento A. X toma assim valores em $\mathbb{N}$ e

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

Dizemos que X possui uma distribuição geométrica de parâmetro p e escrevemos $X \sim G(p)$. A distribuição geométrica é a assim a distribuição do tempo de espera até ao primeiro sucesso de uma sucessão de experiências de Bernoulli com probabilidade de sucesso p. A função de distribuição de X é dada por $F_X(x) = 0$ para $x < 1$, e por

$$F_X(x) = \sum_{i=1}^{k}(1-p)^{i-1}p = \frac{p(1-(1-p)^{k-1})}{1-(1-p)} = 1-(1-p)^{k-1},$$

para $x \in [k, k+1[$, com $k \in \mathbb{N}$.

Lei hipergeométrica

Consideremos uma população de tamanho $N \in \mathbb{N}$ que contém K elementos de tipo 1 (sucessos) e $N-K$ elementos de tipo 0 (insucessos). A distribuição hipergeométrica é a distribuição de probabilidade da v.a.r. X que nos dá o número de sucessos obtidos em n extrações sem reposição realizadas na população em causa. Para $k \in \{\max(0, n+K-N), \ldots, \min(n,K)\}$ temos

$$P(X = k) = C_k^K C_{n-k}^{N-K} / C_n^N.$$

A distribuição hipergeométrica está naturalmente relacionada com a distribuição binomial pois caso as n extrações fossem realizadas com reposição, a variável X possuiria uma distribuição binomial de parâmetros n e $p = K/N$ (probabilidade de sucesso).

3.5.2 Leis absolutamente contínuas

Lei uniforme

A lei uniforme sobre um intervalo $[a,b]$ é a lei de probabilidade da variável aleatória X que nos dá a extração ao acaso de um ponto do intervalo $[a,b]$, a que chamamos variável uniforme sobre o intervalo $[a,b]$ e escrevemos $X \sim U[a,b]$. X toma assim valores no intervalo $[a,b]$ e vimos já (ver Exemplo 2.2.7) que se $]c,d]$ é um subintervalo de $[a,b]$, então

$$P_X(]c,d]) = \frac{d-c}{b-a}.$$

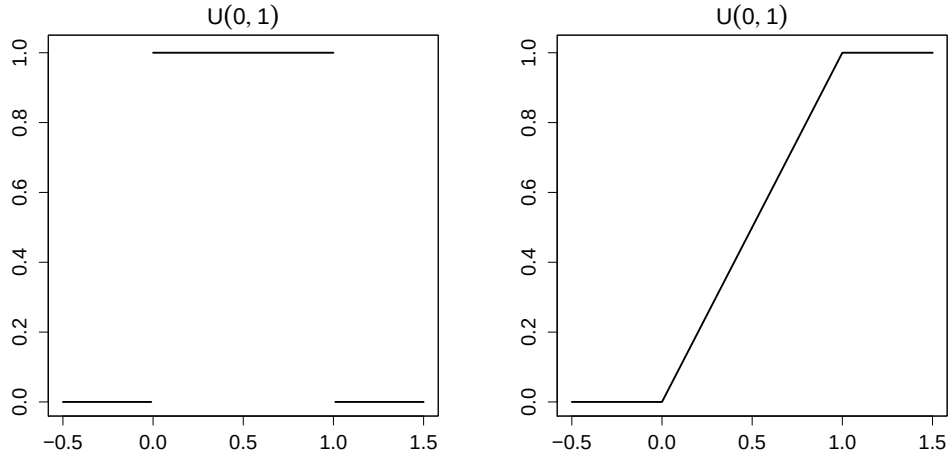


Figura 3.6: Densidade de probabilidade e função de distribuição da variável $U[0,1]$.

Uma vez que

$$P_X(]c,d]) = \int_c^d \frac{1}{b-a}\,dx,$$

para todo o subintervalo $]c,d]$ de $[a,b]$, então X é absolutamente contínua com densidade de probabilidade dada por

$$f_X(x) = \begin{cases} \dfrac{1}{b-a}, & a \leq x \leq b \\ 0, & x < a \text{ ou } x > b. \end{cases}$$

A função de distribuição de X é dada por

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

Lei exponencial

Uma v.a.r. X diz-se exponencial de parâmetro $\lambda > 0$, e escrevemos $X \sim \text{Exp}(\lambda)$, se a sua densidade de probabilidade é dada por

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

A função de distribuição de X é dada por

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

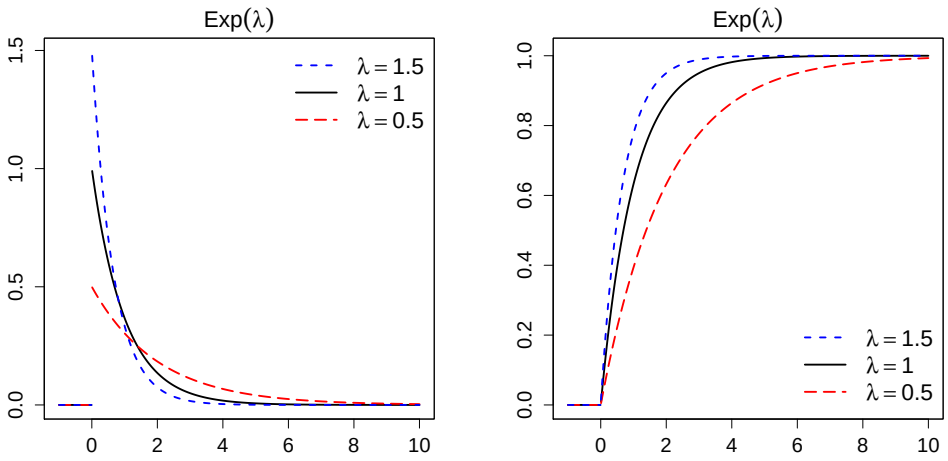


Figura 3.7: Densidades de probabilidade e funções de distribuição da variável Exp(λ).

Uma variável aleatória exponencial está associada à descrição de diversos fenómenos aleatórios, como são os casos tempo entre chegadas consecutivas de clientes a um posto de atendimento, quando assumimos que o número de chegadas por unidade de tempo é constante, ou tempo de funcionamento duma componente ou sistema, quando assumimos que o número de falhas por unidade de tempo é constante.

Mais precisamente, suponhamos que estamos interessados na variável aleatória X que nos dá o primeiro instante t, após o instante inicial $t = 0$, em que ocorre determinado fenómeno aleatório, fenómeno esse que assumimos ter um número médio de ocorrência por unidade de tempo igual a $\lambda > 0$. Se representarmos por Y o número de ocorrências do fenómeno no intervalo $]0, t]$, é natural assumir que Y possui uma distribuição de Poisson, o que significa que $Y \sim \mathscr{P}(\lambda t)$, uma vez que se λ é o número médio de ocorrências por unidade de tempo, o número médio de ocorrência no intervalo $]0, t]$ será igual a λt. Nestas condições,

$$P(X > t) = P(Y = 0) = e^{-\lambda t},$$

o que significa que, para $t > 0$, a função de distribuição de X é dada por

$$F_X(t) = 1 - P(X > t) = 1 - e^{-\lambda t},$$

e $F_X(t) = 0$, para $t \leq 0$. Concluímos assim que X possui uma distribuição exponencial de parâmetro λ.

Lei gama

A distribuição exponencial é um caso particular de uma família mais vasta de distribuições de probabilidade. Dizemos que uma v.a.r. X possui uma distribuição gama de

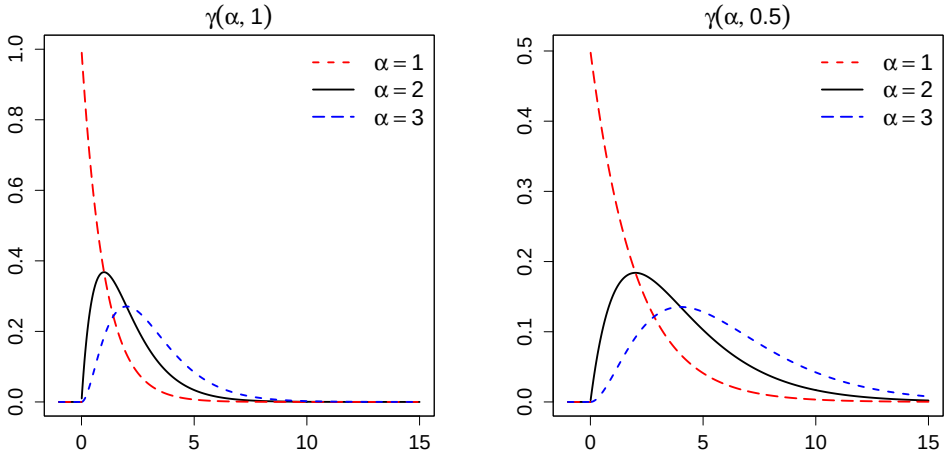


Figura 3.8: Densidades de probabilidade das variáveis $\gamma(\alpha, 1)$ e $\gamma(\alpha, 0{,}5)$.

parâmetros $\alpha > 0$ e $\beta > 0$, e escrevemos $X \sim \gamma(\alpha, \beta)$, se for absolutamente contínua com densidade de probabilidade dada por

$$f_X(x) = \begin{cases} 0, & x \leq 0 \\ \dfrac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, & x > 0, \end{cases}$$

onde

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

é a função gama. Para todo o $\alpha > 0$, $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$, e para $\alpha \in \mathbb{N}$, $\Gamma(\alpha) = (\alpha - 1)!$. A lei exponencial $\text{Exp}(\lambda)$ é assim a lei gama $\gamma(1, \lambda)$.

Lei normal ou gaussiana

Dizemos que uma v.a.r. X possui uma distribuição normal (ou gaussiana) de parâmetros $\mu \in \mathbb{R}$ e $\sigma > 0$ (ou σ^2), e escrevemos $X \sim N(\mu, \sigma^2)$, se X for absolutamente contínua com densidade de probabilidade dada por

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right),$$

para $x \in \mathbb{R}$.

A distribuição normal é a mais usada das distribuições de probabilidade, descrevendo, por exemplo, o efeito global aditivo de um número elevado de pequenos efeitos independentes, como é o caso dos erros de instrumentação. A justificação teórica para o papel de relevo que esta distribuição assume na modelação deste tipo de fenómenos aleatórios, é o já referido Teorema do Limite Central que estudaremos mais à frente.

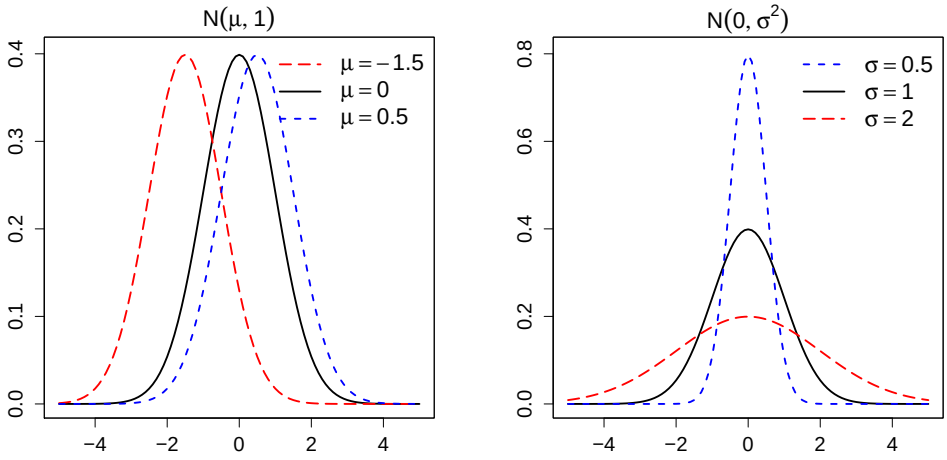


Figura 3.9: Densidades de probabilidade da variável $N(\mu, \sigma^2)$.

Se $\mu = 0$ e $\sigma = 1$, dizemos que X é uma variável normal standard, ou, por razões que veremos mais à frente, normal centrada e reduzida. Como podemos ver pelos gráficos seguintes, os parâmetros μ e σ estão relacionados com o "centro" e a "dispersão" da distribuição de X. Voltaremos a este assunto mais à frente.

Comecemos por verificar que a função f_X dada pela expressão anterior é efetivamente uma densidade de probabilidade. Sendo claramente não negativa, basta mostrar que $\int_{-\infty}^{\infty} f_X(x)dx = 1$. Comecemos por notar que efetuando a mudança de variável $(x-\mu)/\sigma = y$, obtemos

$$I = \int_{-\infty}^{\infty} f_X(x)dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-y^2/2} dy,$$

e então, pelo Teorema de Fubini,

$$I^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-(x^2+y^2)/2} dxdy.$$

Efetuando agora a mudança para coordenadas polares $x = r\cos\theta$ e $y = r\sin\theta$, obtemos

$$I^2 = \frac{1}{2\pi} \int_0^{2\pi} \int_0^{+\infty} e^{-r^2/2} r\, drd\theta = \Big(-e^{-r^2/2}\Big]_0^{+\infty} = 1,$$

de onde concluímos que $I = 1$.

Vimos atrás que uma transformação afim de uma variável absolutamente contínua é ainda absolutamente contínua. Na proposição seguinte mostramos que uma transformação afim de uma variável normal é ainda normal.

Proposição 3.5.1. *Se $X \sim N(\mu, \sigma^2)$ então $aX + b \sim N(a\mu + b, a^2\sigma^2)$, para $a \neq 0$ e $b \in \mathbb{R}$.*

Dem: Para $a > 0$, sabemos da Proposição 3.4.1 que $Y = aX + b$ é absolutamente contínua com densidade de probabilidade

$$\begin{aligned} f_Y(y) &= f_X\big((y-b)/a\big)/a \\ &= \frac{1}{\sqrt{2\pi a^2\sigma^2}} \exp\Big(-\frac{((y-b)/a-\mu)^2}{2\sigma^2}\Big) \\ &= \frac{1}{\sqrt{2\pi a^2\sigma^2}} \exp\Big(-\frac{((y-(a\mu+b))^2}{2a^2\sigma^2}\Big), \end{aligned}$$

ou seja, Y é uma variável normal de parâmetros $a\mu + b$ e $a^2\sigma^2$.

Para $a < 0$, usando os argumentos utilizados na demonstração da Proposição 3.4.1 podemos concluir que $Y = aX + b$ é absolutamente contínua com densidade de probabilidade

$$\begin{aligned} f_Y(y) &= -f_X\big((y-b)/a\big)/a \\ &= \frac{1}{\sqrt{2\pi a^2\sigma^2}} \exp\Big(-\frac{((y-b)/a-\mu)^2}{2\sigma^2}\Big) \\ &= \frac{1}{\sqrt{2\pi a^2\sigma^2}} \exp\Big(-\frac{((y-(a\mu+b))^2}{2a^2\sigma^2}\Big), \end{aligned}$$

ou seja, Y é uma variável normal de parâmetros $a\mu + b$ e $a^2\sigma^2$. ■

Corolário 3.5.2. *Sejam $\mu \in \mathbb{R}$ e $\sigma > 0$.*

a) Se $X \sim N(\mu, \sigma^2)$ então $Z = (X-\mu)/\sigma \sim N(0,1)$.

b) Se $Z \sim N(0,1)$ então $\sigma Z + \mu \sim N(\mu, \sigma^2)$.

A primeira parte deste resultado é particularmente útil pois permite que o cálculo de probabilidades associado a uma variável normal de parâmetros μ e σ possa ser feito a partir da distribuição normal standard. Por exemplo, se para um dado valor $x \in \mathbb{R}$, pretendermos calcular a probabilidade $P(X \leq x)$, onde $X \sim N(\mu, \sigma^2)$, bastará passar da variável X para a variável $(X-\mu)/\sigma$, que sabemos possuir uma distribuição normal standard. Assim,

$$P(X \leq x) = P\big((X-\mu)/\sigma \leq (x-\mu)/\sigma\big) = P(Z \leq (X-\mu)/\sigma),$$

podendo esta última probabilidade ser obtida da tabela da distribuição normal standard. Por razões que serão claras mais à frente, a variável normal standard Z é também dita variável normal centrada e reduzida. Quando a X subtraímos μ e dividimos a diferença $X - \mu$ por σ, dizemos que centrámos e reduzimos a variável X.

3.6 Transformação de variáveis com densidade*

Vimos na Proposição 3.4.1 que sempre que a variável X é absolutamente contínua, também o é a variável $Y = aX + b = h(X)$, variável que se obtém de X através da transformação afim $h(x) = ax + b$, para $x \in \mathbb{R}$, onde $a > 0$ e $b \in \mathbb{R}$. No Exemplo 3.1.4 determinámos a densidade de probabilidade da variável X que nos dá o quadrado de um número extraído ao acaso no intervalo $]0, 1[$. Uma vez que a extração ao acaso de um número no intervalo $]0, 1[$ pode ser vista como a realização de uma variável U com distribuição uniforme sobre o intervalo $]0, 1[$, a variável X pode ser vista como uma transformação da variável U, isto é, $X = U^2 = h(U)$, onde $h :]0, 1[\to]0, 1[$ é definida por $h(x) = x^2$. Neste caso particular chegámos à conclusão que uma versão da densidade de probabilidade de X é dada por

$$f_X(x) = \begin{cases} \dfrac{1}{2\sqrt{x}}, & x \in]0, 1[\\ 0, & x \notin]0, 1[. \end{cases}$$

Sob certas condições na transformação h, é possível concluir que se X é uma v.a.r. absolutamente contínua então $Y = h(X)$ é também absolutamente contínua, e que podemos obter a densidade de probabilidade de Y a partir da de X.

Proposição 3.6.1. *Sejam X uma v.a.r. absolutamente contínua com densidade f_X e $h :]a, b[\to]c, d[$ uma aplicação bijetiva com h e h^{-1} continuamente diferenciáveis, com $-\infty \leq a < b \leq +\infty$ e $-\infty \leq c < d \leq +\infty$. Se $f_X \mathbb{1}_{]a,b[}$ é uma versão de f_X então $Y = h(X)$ é absolutamente contínua com densidade f_Y dada por*

$$f_Y(y) = f_X(h^{-1}(y))\, |(h^{-1})'(y)| \mathbb{1}_{]c,d[}(y).$$

Dem: Se $f_X \mathbb{1}_{]a,b[}$ é uma versão de f_X concluímos que $P(X \in]a, b[) = 1$ e portanto

$$P(Y \in]c, d[) = P(h(X) \in]c, d[) = P(X \in h^{-1}(]c, d[)) = P(X \in]a, b[) = 1.$$

Supondo que h é estritamente crescente (o que implica que h^{-1} também seja estritamente crescente), para $y \in]c, d[$ temos

$$F_Y(y) = P(Y \leq y) = P\big(h(X) \leq y\big) = P\big(X \leq h^{-1}(y)\big) = F_X\big(h^{-1}(y)\big).$$

Supondo que h é estritamente decrescente (o que implica que h^{-1} também seja estritamente decrescente), para $y \in]c, d[$ temos

$$F_Y(y) = P(Y \leq y) = P\big(h(X) \leq y\big) = P\big(X \geq h^{-1}(y)\big) = 1 - F_X\big(h^{-1}(y)\big).$$

No primeiro caso

$$F'_Y(y) = f_X\big(h^{-1}(y)\big)(h^{-1})'(y) = f_X\big(h^{-1}(y)\big)(h^{-1})'(y)$$

e segundo caso

$$F'_Y(y) = -f_X\big(h^{-1}(y)\big)(h^{-1})'(y),$$

o que dá origem à expressão unificada

$$F'_Y(y) = f_X\big(h^{-1}(y)\big)|(h^{-1})'(y)|,$$

para $y \in]c, d[$. Para concluir a demonstração basta agora provar que

$$\int_c^d F'_Y(y)dy = 1.$$

Com efeito, efetuando a mudança de variável $x = h^{-1}(y)$ obtemos

$$\int_c^d F'_Y(y)dy = \int_c^d f_X\big(h^{-1}(y)\big)|(h^{-1})'(y)|dy = \int_a^b f_X(x)dx = 1. \qquad \blacksquare$$

Exemplo 3.6.2. Retomando a Proposição 3.4.1 vamos admitir que $Y = aX + b$, com $a \neq 0$ e $b \in \mathbb{R}$, isto é, $Y = h(X)$ onde $h : \mathbb{R} \to \mathbb{R}$ é definida por $h(x) = ax + b$, para $x \in \mathbb{R}$. Usando o resultado anterior, uma vez que

$$h^{-1}(y) = (y - b)/a,$$

concluímos que, para $y \in \mathbb{R}$, f_Y é dada por

$$f_Y(y) = f_X(h^{-1}(y))\,|(h^{-1})'(y)| = f_X((y - b)/a)/|a|.$$

Exemplo 3.6.3. Retomemos o Exemplo 3.1.4 e calculemos a densidade de $X = h(U)$ com $h(u) = u^2$ e $U \sim U[0, 1]$. Sabemos que uma versão de f_U é dada por $f_U(u) = 1$ se $u \in]0, 1[$ e $f_U(u) = 0$ se $u \notin]0, 1[$. $h :]0, 1[\to]0, 1[$ é uma aplicação diferenciável com inversa também diferenciável dada por $h^{-1}(x) = \sqrt{x}$. Uma vez que $(h^{-1})'(x) = \frac{1}{2\sqrt{x}}$, para $x \in]0, 1[$, da proposição anterior concluímos que X é uma variável absolutamente contínua com densidade dada por

$$f_X(x) = f_U(h^{-1}(x))|(h^{-1})'(x)| = \frac{1}{2\sqrt{x}},$$

para $x \in]0, 1[$ e $f_X(x) = 0$ para $x \notin]0, 1[$.

3.7 Simulação de experiências e variáveis aleatórias

Muitas das experiências aleatórias que até agora descrevemos podem ser simuladas num computador. Na base de todo o processo de simulação está a extração ao acaso de

pontos do intervalo $]0,1[$, ou seja, a simulação duma variável aleatória uniforme sobre o intervalo $]0,1[$ (ver Exemplo 2.2.7). No caso do software estatístico R, a função

$$\mathsf{runif}(n,\ \mathsf{min} = 0,\ \mathsf{max} = 1)$$

simula n valores de uma variável $U \sim U]0,1[$, a que chamamos habitualmente números aleatórios (mais precisamente, números pseudoaleatórios). A partir desta variável é simples simular o lançamento de um dado. Admitindo que o dado é equilibrado, e sabendo que cada um dos acontecimentos

$$\big\{(i-1)/6 \leq U < i/6\big\},\ i = 1, \ldots, 6,$$

tem probabilidade $1/6$, então a variável definida por

$$X = i \text{ sse } (i-1)/6 \leq U < i/6,$$

possui uma distribuição uniforme discreta sobre o conjunto $\{1, 2, \ldots, 6\}$.

A função seguinte, escrita em linguagem R, permite simular n lançamentos de um dado equilibrado:

```
rdadoequi = function(n)
{
 u <- runif(n)
 1*(u<1/6)+2*(u>=1/6)*(u<2/6)+3*(u>=2/6)*(u<3/6)+
 4*(u>=3/6)*(u<4/6)+5*(u>=4/6)*(u<5/6)+6*(u>=5/6)
}
```

A partir da variável $U \sim U]0,1[$ podemos também definir facilmente uma variável com distribuição $U]a,b[$. Basta tomar

$$X = a(1-U) + bU.$$

A variável X assim definida simula a extração ao acaso de um ponto do intervalo $]a,b[$. Em alternativa, podemos usar a função

$$\mathsf{runif}(n,\ \mathsf{min} = \mathsf{a},\ \mathsf{max} = \mathsf{b})$$

para gerar n valores de uma variável $U]a,b[$.

Suponhamos agora que X é uma v.a.r. da qual conhecemos a sua função de distribuição F_X e que pretendemos simular valores de X a partir de valores de $U \sim U]0,1[$. O resultado seguinte indica como podemos fazê-lo quando F_X é estritamente crescente.

Proposição 3.7.1. *Se a função de distribuição F_X da v.a.r. X é estritamente crescente em $\{x \in \mathbb{R} : 0 < F_X(x) < 1\}$, então a variável Y definida por $Y = F_X^{-1}(U)$, onde $U \sim U]0,1[$, é tal que $Y \sim X$.*

Dem: Sendo F_X estritamente crescente, F_X possui inversa F_X^{-1}. Para provar que a variável Y definida por $Y = F_X^{-1}(U)$ possui a mesma distribuição de probabilidade de X, basta provar que $F_X = F_Y$ uma vez que a função de distribuição carateriza a distribuição de probabilidade. Para $x \in \mathbb{R}$, temos

$$F_Y(x) = P(Y \leq x) = P\big(F_X^{-1}(U) \leq x\big) = P\big(U \leq F_X(x)\big) = F_X(x). \qquad \blacksquare$$

Exemplo 3.7.2. Um exemplo simples de utilização do resultado anterior, é o da simulação de uma variável exponencial de parâmetro $\lambda > 0$ (ver Exercício 43d). Se $X \sim \text{Exp}(\lambda)$, sabemos que

$$F_X(x) = (1 - e^{-\lambda x})\mathbb{1}_{]0,+\infty[}(x),$$

que é uma função estritamente crescente em $]0, +\infty[$. Uma vez que

$$F_X^{-1}(y) = -\frac{1}{\lambda}\log(1-y),\ y \in]0,1[,$$

então

$$Y = -\frac{1}{\lambda}\log(1-U) \sim \text{Exp}(\lambda),$$

onde $U \sim U]0,1[$. A função seguinte, escrita em linguagem R, permite simular n valores de uma variável exponencial de parâmetro $\lambda > 0$:

```
rexponencial = function(n, lambda=1)
{
 -log(1-runif(n))/lambda
}
```

3.8 Bibliografia

Chae, S.B. (1980). *Lebesgue integration.* New York: Marcel Dekker.

Gonçalves, E., Mendes Lopes, N. (2013). *Probabilidades: princípios teóricos.* Lisboa: Escolar Editora.

Jacod, J., Protter, P. (2000). *Probability essentials.* Berlin: Springer.

Lima, E.L. (1995). *Curso de análise, Vol. 1.* Rio de Janeiro: IMPA.

Monfort, A. (1980). *Cours de probabilités.* Paris: Economica.

Natanson, I.P. (1961). *Theory of functions of a real variable.* New York: Frederick Unger.

Resnick, S.I. (2005). *A probability path.* Boston: Birkhäuser.

Tenreiro, C. (2002). *Apontamentos de Teoria das Probabilidades.* Coimbra: DMUC. https://www.mat.uc.pt/∼tenreiro/apontamentos/ATP0204.pdf

4

Esperança matemática, variância e momentos

Esperança matemática. Propriedades da esperança matemática e cálculo respetivo. Momentos e variância. Propriedades da variância e cálculo respetivo. Coeficientes de assimetria e de forma. Desigualdade de Markov e suas consequências. Localização e dispersão sem momentos.

4.1 Esperança matemática

Neste capítulo vamos definir parâmetros associados a uma v.a.r. X que não são mais que resumos numéricos da sua distribuição de probabilidade. Vamos centrar a nossa atenção em medidas do centro e da dispersão, ou variabilidade, da distribuição de probabilidade de X, ditos parâmetros de localização e de dispersão.

Começamos pela noção de esperança matemática de uma v.a.r. X. Não podendo lançar mão da teoria da integração à Lebesgue, o que permitiria definir de forma unificada a noção esperança matemática de X como o integral de X relativamente à probabilidade P, vamos abordar separadamente os casos em que X é discreta e em que X é absolutamente contínua.

Definição 4.1.1. *a) Se X é uma v.a.r. discreta com suporte S_X, chamamos esperança matemática de X ou média da variável X, que denotamos por $E(X)$, a*

$$E(X) = \sum_{x \in \mathbb{R}} x f_X(x),$$

sempre que

$$\sum_{x \in \mathbb{R}} |x| f_X(x) < \infty.$$

b) Se X é uma v.a.r. absolutamente contínua, chamamos esperança matemática de

X ou média da variável X, que denotamos por $E(X)$, a

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

sempre que

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

No caso em que X é uma v.a.r. discreta com suporte S_X, sabemos que $f_X(x) = P(X = x)$ para $x \notin S_X$, e portanto a soma $\sum_{x\in\mathbb{R}} x f_X(x)$ reduz-se a $\sum_{x\in S_X} xP(X = x)$, sendo esta uma soma com um número finito de parcelas se o suporte S_X é finito, ou uma série numérica se S_X é infinito numerável.

A interpretação de $E(X)$ como medida de localização do centro da distribuição de probabilidade de X pode ser reforçada se recordarmos que o centro de massa, ou ≪ponto de equilíbrio≫, dum sistema de pontos materiais $x_1, \ldots, x_N$, com massas $m_1, \ldots, m_N$ é o ponto

$$x_{\text{CM}} = \frac{\sum_{i=1}^{N} m_i x_i}{\sum_{i=1}^{N} m_i}.$$

Se a massa total do sistema de pontos for 1, isto é, $\sum_{i=1}^{N} m_i = 1$, caso em que as massas m_i podem ser interpretadas como probabilidades associadas a cada um dos pontos x_i, o centro de massa x_{CM} não é mais do que a esperança matemática da variável discreta X que toma os valores x_i com probabilidades m_i.

No caso do suporte de X ser finito, $E(X)$ existe sempre uma vez que a soma anterior possui um número finito de parcelas. No caso do suporte de X ser infinito numerável, $E(X)$ não está definida quando a série $\sum_{x\in S_X} |x| f_X(x)$ for divergente. Dizemos nesse caso que a esperança matemática não existe. Tal como no caso discreto, no caso em que X é absolutamente contínua a esperança $E(X)$ não está definida (dizemos que não existe) quando $\int_{-\infty}^{\infty} |x| f_X(x)\, dx = \infty$. No caso de X ser uma v.a.r. absolutamente contínua a definição dada não é ainda matematicamente satisfatória a menos que o integral $\int_{-\infty}^{\infty} x f_X(x)\, dx$ seja interpretado no sentido de Lebesgue e não no sentido de Riemann. Tal facto leva a que algumas das propriedades que enunciamos não possam ser demonstradas na sua generalidade plena.

Reparemos que a esperança matemática depende apenas da distribuição de probabilidade P_X da variável X, o que implica que $E(X) = E(Y)$ sempre que $X \sim Y$.

Exemplo 4.1.2 (cont. do Exemplo 3.2.3)**.** Para a variável discreta X que nos dá a soma dos pontos obtidos em dois dados equilibrados, calculemos $E(X)$. Assim,

$$E(X) = 2 \times \frac{1}{36} + 3 \times \frac{2}{36} + 4 \times \frac{3}{36} + \cdots + 12 \times \frac{1}{36} = 7.$$

Devemos interpretar este valor como o número médio de pontos que se obtêm por lançamento quando jogamos com dois dados equilibrados.

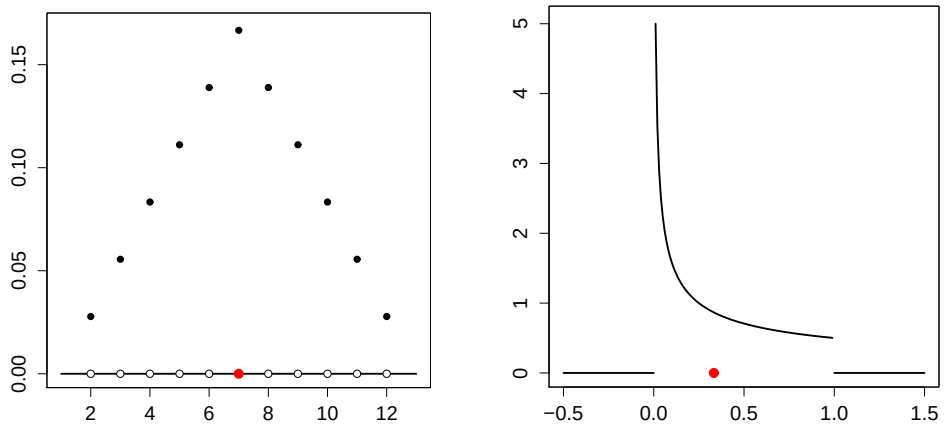


Figura 4.1: Esperanças matemáticas das variáveis aleatórias que nos dão a soma dos pontos obtidos no lançamento de dois dados equilibrados (esquerda) e o quadrado de um número extraído ao acaso do intervalo $[0,1]$ (direita).

Exemplo 4.1.3 (cont. do Exemplo 3.2.5)**.** No caso da variável X que nos dá o quadrado de um número extraído ao acaso do intervalo $[0,1]$, temos

$$\int_{-\infty}^{\infty} |x| f_X(x)dx = \int_0^1 x f_X(x)dx = \int_0^1 \frac{1}{2}x^{1/2}dx = \frac{1}{3}\Big[x^{3/2}\Big]_0^1 = \frac{1}{3} < \infty,$$

e portanto

$$E(X) = \int_{-\infty}^{\infty} x f_X(x)dx = \int_0^1 x f_X(x)dx = \frac{1}{3}.$$

Como já referimos, a noção de esperança matemática pode ser dada mesmo quando a variável X não é discreta nem contínua.

Exemplo 4.1.4. Vimos no Exemplo 3.3.10 que a função de distribuição da variável X aí considerada se podia escrever na forma

$$F_X(x) = \frac{1}{2}F_d(x) + \frac{1}{2}F_{ac}(x),$$

onde

$$F_d(x) = \begin{cases} 0, & x < 1/2 \\ 1, & x \geq 1/2 \end{cases}$$

é a função de distribuição de uma variável aleatória discreta Y e

$$F_{ac}(x) = \begin{cases} 0, & x < 0 \\ 2x, & 0 \leq x < 1/2 \\ 1, & x \geq 1/2. \end{cases}$$

é função de distribuição da variável absolutamente contínua Z. Neste caso será natural definir

$$\begin{aligned} E(X) &= \frac{1}{2}E(Y) + \frac{1}{2}E(Z) \\ &= \frac{1}{2}\sum_{y\in S_Y} y f_Y(y) + \frac{1}{2}\int_{-\infty}^{\infty} y f_Z(y)dy \\ &= \frac{1}{2}\frac{1}{2}P(Y = 1/2) + \frac{1}{2}\int_0^{1/2} 2y dy = \frac{3}{8}. \end{aligned}$$

No Exemplo 4.1.2 a distribuição de X é simétrica relativamente ao ponto $x = 7$, isto é, $f_X(7 - x) = f_X(7 + x)$, para $x \in \mathbb{R}$, e a esperança matemática de X coincide com esse ponto de simetria. Este resultado pode ser generalizado da forma seguinte que reforça a interpretação da esperança matemática, ou média de uma variável, como medida de resumo do centro da distribuição de X.

Proposição 4.1.5. *Seja X uma v.a.r. (discreta ou absolutamente contínua) simétrica relativamente a $x = a$, isto é,*

$$f_X(a - x) = f_X(a + x),\ x \in \mathbb{R}.$$

Se $E(X)$ existe, então $E(X) = a$.

Dem: No caso em que X é discreta temos

$$\begin{aligned} E(X) &= \sum_{x\in\mathbb{R}} x f_X(x) = \sum_{y\in\mathbb{R}} (a - y) f_X(a - y) = \sum_{y\in\mathbb{R}} (a - y) f_X(a + y) \\ &= \sum_{x\in\mathbb{R}} (2a - x) f_X(x) = 2a \sum_{x\in\mathbb{R}} f_X(x) - \sum_{x\in\mathbb{R}} x f_X(x) = 2a - E(X). \end{aligned}$$

No caso em que X é absolutamente contínua temos

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x f_X(x)dx = \int_{-\infty}^{\infty} (a - y) f_X(a - y)dy = \int_{-\infty}^{\infty} (a - y) f_X(a + y)dy \\ &= \int_{-\infty}^{\infty} (2a - x) f_X(x)dx = 2a \int_{-\infty}^{\infty} f_X(x)dx - \int_{-\infty}^{\infty} x f_X(x)dx = 2a - E(X). \end{aligned}$$

Num e noutro caso, concluímos que $E(X) = a$. ■

Atendendo à Proposição 3.4.1, sabemos que $f_X(a-x) = f_{-(X-a)}(x)$ e $f_X(a+x) = f_{X-a}(x)$. Assim, dizer que X é simétrica relativamente a $x = a$ é dizer $X-a \sim -(X-a)$. Quando, $a = 0$, isto é, quando $X \sim -X$, dizemos apenas que X é simétrica (ou que a distribuição de X é simétrica).

A existência da esperança matemática da variável X é fundamental para a validade do resultado anterior. Tal é ilustrado no exemplo seguinte.

Exemplo 4.1.6. Dizemos que uma v.a.r. X possui uma distribuição de Cauchy (standard) se

$$f_X(x) = \frac{1}{\pi(1+x^2)},\ x \in \mathbb{R}.$$

Apesar de f_X ser simétrica relativamente a $x = 0$, a esperança matemática de X não existe uma vez que

$$\int_{-n}^{n} |x| f_X(x)dx = \frac{1}{\pi}\int_0^n \frac{2x}{1+x^2}\,dx = \frac{1}{\pi}\Big[\log(1+x^2)\Big]_0^n = \frac{1}{\pi}\log(1+n^2),$$

e portanto

$$\int_{-\infty}^{\infty} |x| f_X(x)dx = \lim_{n\to+\infty}\frac{1}{\pi}\log(1+n^2) = +\infty.$$

Exemplo 4.1.7. Como aplicação direta da Proposição 4.1.5, concluímos que se X representa o número de pontos que ocorrem no lançamento de um dado equilibrado então $E(X) = 3{,}5$ (porquê?). Isto significa que, em média, ocorrem 3,5 pontos por lançamento quando lançamos um dado equilibrado vulgar.

4.1.1 Propriedades da esperança matemática

Apresentam-se no resultado seguintes propriedades fundamentais da esperança matemática ou média de uma variável aleatória. A demonstração de algumas delas no caso não discreto, ultrapassa em muito os objetivos do curso.

Proposição 4.1.8. *Sejam X e Y v.a.r. definidas no espaço de probabilidade $(\Omega, \mathscr{A}, P)$, e $a, b \in \mathbb{R}$.*

a) Se $X = a$ q.c., então $E(X) = a$;

b) Se $X = \mathbb{1}_A$ q.c., então $E(X) = P(A)$;

c) Se $X \geq 0$ q.c. e $E(X)$ existe, então $E(X) \geq 0$. Além disso, $E(X) = 0$ sse $X = 0$ q.c.;

d) Se $|X| \leq Y$ q.c. e $E(Y)$ existe, então $E(X)$ existe;

e) Se $|X| \leq M$ q.c. com $M > 0$, então $E(X)$ existe;

f) Se $E(X)$ e $E(Y)$ existem, então $E(aX + bY)$ existe e

$$E(aX + bY) = aE(X) + bE(Y)\ (\textit{linearidade});$$

g) Se $E(X)$ e $E(Y)$ existem e $X \leq Y$ q.c., então $E(X) \leq E(Y)$ (monotonia).

Dem: a) Por hipótese $P(X = a) = 1$, ou seja, $f_X(a) = 1$ e $f_X(x) = 0$, se $x \neq a$ (X é uma variável de Dirac no ponto a). Assim $S_X = \{a\}$, e

$$E(X) = \sum_{x\in S_X} x f_X(x) = a \times f_X(a) = a.$$

b) $X = \mathbb{1}_A$ é uma v.a.r. discreta com suporte $S_X = \{0, 1\}$, $P(X = 1) = P(A)$ e $P(X = 0) = 1 - P(A)$ (X é uma variável de Bernoulli de parâmetro $p = P(A)$). Assim

$$E(X) = 0 \times (1 - P(A)) + 1 \times P(A) = P(A).$$

c) Suponhamos que X é discreta. Se $E(X)$ existe então $E(X) = \sum_{x \in S_X} xP(X = x)$. Como $X \geq 0$ *q.c.*, concluímos que $S_X \subset [0, +\infty[$. Assim $x \geq 0$ para todo o $x \in S_X$ de onde tiramos que $E(X) \geq 0$.

Se $X = 0$ *q.c.*, sabemos da alínea a) que $E(X) = 0$. Reciprocamente, suponhamos que $E(X) = 0$, ou seja, $\sum_{x \in S_X} xP(X = x) = 0$. Como $P(X = x) > 0$ para todo o $x \in S_X$, então temos de ter $x = 0$ para todo o $x \in S_X$, ou seja, $S_X = \{0\}$. Assim $P(X = 0) = P_X(\{0\}) = P_X(S_X) = 1$, ou seja, $X = 0$ *q.c.*

d) Suponhamos que X e Y são discretas. Por hipótese $P(|X| \leq Y) = 1$, ou ainda, $P(|X| > Y) = 0$. Mas

$$\begin{aligned} P(|X| > Y) &= P\Big(|X| > Y, \bigcup_{x \in S_X} \{X = x\}, \bigcup_{y \in S_Y} \{Y = y\}\Big) \\ &= \sum_{x \in S_X} \sum_{y \in S_Y} P\big(|X| > Y, X = x, Y = y\big) \\ &= \sum_{x \in S_X, y \in S_Y} P\big(|x| > y, X = x, Y = y\big) \\ &= \sum_{x \in S_X, y \in S_Y : |x| > y} P\big(X = x, Y = y\big), \end{aligned}$$

o que permite concluir que $P\big(X = x, Y = y\big) = 0$, para todo o $x \in S_X$ e $y \in S_Y$ com $|x| > y$. Provemos então que $E(X)$ existe. Com efeito,

$$\begin{aligned} \sum_{x \in S_X} |x| P(X = x) &= \sum_{x \in S_X} |x| \sum_{y \in S_Y} P(X = x, Y = y) \\ &= \sum_{x \in S_X, y \in S_Y : |x| \leq y} |x| P(X = x, Y = y) \\ &\leq \sum_{x \in S_X, y \in S_Y : |x| \leq y} y P(X = x, Y = y) \\ &\leq \sum_{x \in S_X, y \in S_Y} y P(X = x, Y = y) \\ &\leq \sum_{y \in S_Y} y \sum_{x \in S_X} P(X = x, Y = y) \\ &\leq \sum_{y \in S_Y} y P(Y = y) < \infty. \end{aligned}$$

e) Consequência de a) e d) tomando $Y = a = M$.

f) Suponhamos que X e Y são discretas. Vamos provar que $E(X+Y)$ e $E(aX)$ existem e que $E(X+Y) = E(X)+E(Y)$ e $E(aX) = aE(X)$. Provemos que $E(X+Y)$ existe. Com efeito,

$$\begin{aligned}
&\sum_{s\in S_{X+Y}} |s|P(X+Y=s)\\
&= \sum_{s\in S_{X+Y}} |s| \sum_{x\in S_X, y\in S_Y : x+y=s} P(X=x, Y=y)\\
&= \sum_{s\in S_{X+Y}} \sum_{x\in S_X, y\in S_Y : x+y=s} |x+y|P(X=x, Y=y)\\
&= \sum_{x\in S_X, y\in S_Y} |x+y|P(X=x, Y=y)\\
&\leq \sum_{x\in S_X, y\in S_Y} (|x|+|y|)P(X=x, Y=y)\\
&= \sum_{x\in S_X, y\in S_Y} |x|P(X=x, Y=y) + \sum_{x\in S_X, y\in S_Y} |y|P(X=x, Y=y)\\
&= \sum_{x\in S_X} |x| \sum_{y\in S_Y} P(X=x, Y=y) + \sum_{y\in S_Y} |y| \sum_{x\in S_X} P(X=x, Y=y)\\
&= \sum_{x\in S_X} |x|P(X=x) + \sum_{y\in S_Y} |y|P(Y=y) < \infty.
\end{aligned}$$

Repetindo os passos anteriores temos

$$\begin{aligned}
E(X+Y) &= \sum_{s\in S_{X+Y}} sP(X+Y=s)\\
&= \sum_{s\in S_{X+Y}} s \sum_{x\in S_X, y\in S_Y : x+y=s} P(X=x, Y=y)\\
&= \sum_{x\in S_X, y\in S_Y} (x+y)P(X=x, Y=y)\\
&= \sum_{x\in S_X, y\in S_Y} xP(X=x, Y=y) + \sum_{x\in S_X, y\in S_Y} yP(X=x, Y=y)\\
&= \sum_{x\in S_X} x \sum_{y\in S_Y} P(X=x, Y=y) + \sum_{y\in S_Y} y \sum_{x\in S_X} P(X=x, Y=y)\\
&= \sum_{x\in S_X} xP(X=x) + \sum_{y\in S_Y} yP(Y=y)\\
&= E(X)+E(Y).
\end{aligned}$$

Provemos agora que existe a esperança matemática de aX. Com efeito,

$$\sum_{x\in S_X} |ax|f_X(x) \leq |a| \sum_{x\in S_X} |x|f_X(x) < \infty.$$

Além disso,

$$E(aX) = \sum_{x \in S_X} (ax) f_X(x) = a \sum_{x \in S_X} x f_X(x) = aE(X).$$

g) Se $X \leq Y$ *q.c.* então $Y - X \geq 0$ *q.c.*. Pela alíneas c) e f) temos $E(Y - X) \geq 0$, e $E(Y) - E(X) \geq 0$, ou ainda $E(X) \leq E(Y)$. ■

Da alínea g) podemos também concluir que se X é uma variável com média μ então a variável $Y = X - \mu$ tem média zero

$$E(Y) = E(X) - \mu = \mu - \mu = 0.$$

Por ter média zero dizemos que a variável Y é centrada. Sempre que subtraímos a uma variável X a sua média, dizemos que centrámos a variável X.

Exemplo 4.1.9 (cont. do Exemplo 4.1.2)**.** Consideremos a variável discreta X que nos dá a soma dos pontos obtidos em dois dados equilibrados. Reparemos que a linearidade da esperança matemática, estabelecida na alínea f) da proposição anterior, pode ser usada para simplificar o cálculo de $E(X)$. Para tal basta notar que $X = X_1 + X_2$, onde X_1 e X_2 representam os pontos obtidos no primeiro e no segundo dado, respetivamente. Uma vez que X_1 e X_2 têm a mesma distribuição de probabilidade com $E(X_1) = E(X_2) = 3{,}5$, concluímos que $E(X) = E(X_1) + E(X_2) = 3{,}5 + 3{,}5 = 7$.

Mostramos a seguinte que a esperança matemática de uma função $h(X)$ de uma v.a.r. X, onde h é uma função real de variável real, pode ser calculada a partir do conhecimento da distribuição de X. Não necessitamos de conhecer a distribuição de $h(X)$.

Proposição 4.1.10. *Seja X uma v.a.r. discreta ou absolutamente contínua e h uma função real de variável real.*

a) Se X é discreta então

$$E(h(X)) = \sum_{x \in \mathbb{R}} h(x) f_X(x),$$

desde que $\sum_{x \in \mathbb{R}} |h(x)| f_X(x) < \infty$.

b) Se X é absolutamente contínua então

$$E(h(X)) = \int_{\mathbb{R}} h(x) f_X(x)\, dx,$$

desde que $\int_{\mathbb{R}} |h(x)| f_X(x)\, dx < \infty$.

Dem: a) Sendo S_X o suporte de X, o suporte de $Y = h(X)$ é dado por $S_Y = \{h(x) : x \in S_X\}$. Para todo o $y \in S_Y$, vamos representar por E_y o conjunto

$$E_y = \{x \in S_X : h(x) = y\}.$$

Reparemos que Y toma o valor $y \in S_Y$ sse X toma valores em E_y, isto é,

$$\{Y = y\} = \{X \in E_y\}$$

e portanto

$$P(Y = y) = P(X \in E_y) = \sum_{x \in E_y} P(X = x).$$

Assim

$$\begin{aligned}\sum_{x \in S_X} h(x) f_X(x) &= \sum_{y \in S_Y} \sum_{x \in E_y} h(x) P(X = x) = \sum_{y \in S_Y} \sum_{x \in E_y} y P(X = x) \\ &= \sum_{y \in S_Y} y \sum_{x \in E_y} P(X = x) = \sum_{y \in S_Y} y P(Y = y) \\ &= E(Y) = E(h(X)).\end{aligned}$$

b) A demonstração para o caso em que X é absolutamente contínua não pode ser feita em geral uma vez que a distribuição de $h(X)$ pode ser discreta, absolutamente contínua ou mista, o que obrigaria a uma demonstração diferente para cada caso. Vamos analisar o caso em que $Y = h(X)$ é discreta com suporte S_Y. Sem perda de generalidade, vamos admitir que h é uma aplicação com contradomínio S_Y, e para todo o $y \in S_Y$ vamos representar por E_y o conjunto $E_y = \{x \in \mathbb{R} : h(x) = y\}$. Assim, para $y \in S_Y$,

$$P(Y = y) = P(X \in E_y) = \int_{E_y} f_X(x) dx,$$

e

$$\begin{aligned}\int_{-\infty}^{\infty} h(x) f_X(x) dx &= \sum_{y \in S_Y} \int_{E_y} h(x) f_X(x) dx = \sum_{y \in S_Y} \int_{E_y} y f_X(x) dx \\ &= \sum_{y \in S_Y} y \int_{E_y} f_X(x) dx = \sum_{y \in S_Y} y P(Y = y) \\ &= E(Y) = E(h(X)).\end{aligned}$$

■

Reparemos que o resultado anterior pode ser usado para calcular a esperança matemática de uma variável $h(X)$ que seja função duma variável discreta ou duma variável absolutamente contínua, mesmo que $h(X)$ não seja discreta nem absolutamente contínua.

Exemplo 4.1.11. Retomemos novamente o Exemplo 3.3.10. Vimos no Exemplo 4.1.4 que $E(X) = 3/8$. A variável X pode ser escrita na forma $X = \min(1/2, U) = h(U)$, onde $h(u) = \min(1/2, u)$ e $U \sim U[0,1]$. Usando a Proposição 4.1.10 podemos escrever

$$E(X) = \int_{-\infty}^{\infty} h(u) f_U(u) du = \int_0^1 \min(1/2, u) du = \int_0^{1/2} u du + \int_{1/2}^1 \frac{1}{2} du = \frac{3}{8},$$

o que confirma o valor anteriormente calculado.

Proposição 4.1.12. *Se X é uma v.a.r. então $X \in \mathscr{L}^1$ sse $|X| \in \mathscr{L}^1$ e*

$$|E(X)| \leq E(|X|).$$

Dem: Quer no caso discreto quer no caso absolutamente contínuo, se usarmos a proposição anterior com $h(X) = |X|$, verificamos que a condição que garante que $|X|$ possui esperança matemática coincide com a condição que garante que X possui esperança matemática. Concluímos assim que $E(X)$ existe sse $E(|X|)$ existe. Além disso, uma vez que

$$-|X| \leq X \leq |X|,$$

da monotonia e da linearidade da esperança matemática obtemos

$$-E(|X|) \leq E(X) \leq E(|X|),$$

ou seja

$$|E(X)| \leq E(|X|).$$ ■

4.1.2 Cálculo da esperança matemática: alguns exemplos

Para cada uma das variáveis definidas na Secção 3.5, calculemos a respetiva esperança matemática.

Lei de Dirac

Uma variável de Dirac no ponto $a \in \mathbb{R}$ é tal que $P(X = a) = 1$. Já vimos na Proposição 4.1.8 que

$$E(X) = a \times P(X = a) = a.$$

Lei de Bernoulli

Uma variável de Bernoulli X toma valores 1 e 0 com probabilidades p e $1 - p$, respetivamente. Assim,

$$E(X) = 0 \times P(X = 0) + 1 \times P(X = 1) = 0 \times (1 - p) + 1 \times p = p.$$

Lei uniforme discreta

Uma variável aleatória discreta sobre o conjunto $\{1, \ldots, n\}$ é uma variável simétrica relativamente ao ponto $(n+1)/2$. Usando a Proposição 4.1.5 concluímos que

$$E(X) = \frac{n+1}{2}.$$

Vamos confirmar este valor efetuando o cálculo usando a definição:

$$E(X) = 1 \times \frac{1}{n} + \cdots + n \times \frac{1}{n} = \frac{(1+n)n}{2}\frac{1}{n} = \frac{n+1}{2}.$$

Lei binomial

Uma forma simples de calcular $E(X)$ quando $X \sim B(n,p)$, começa por constatar que uma variável aleatória binomial não é mais do que uma soma de variáveis de Bernoulli de parâmetro p: $X = X_1 + \cdots + X_n$, com $X_i \sim B(p)$, onde X_i toma os valores 1 ou 0 de acordo com ter ou não ocorrido um sucesso na i-ésima repetição da experiência. Uma vez que a probabilidade de sucesso é p então $X_i \sim B(p)$. Assim usando a alínea f) da Proposição 4.1.8 temos

$$E(X) = E(X_1) + \cdots + E(X_n) = np.$$

Recorrendo diretamente à definição de esperança matemática, temos:

$$\begin{aligned} E(X) &= \sum_{k=0}^{n} kP(X=k) = \sum_{k=1}^{n} kC_k^n p^k (1-p)^{n-k} \\ &= \sum_{k=1}^{n} nC_{k-1}^{n-1} p^k (1-p)^{n-k} = np \sum_{k=1}^{n} C_{k-1}^{n-1} p^{k-1} (1-p)^{(n-1)-(k-1)} \\ &= np \sum_{\ell=0}^{n-1} C_\ell^{n-1} p^\ell (1-p)^{(n-1)-\ell} = np\big(p + (1-p)\big)^{n-1} = np. \end{aligned}$$

Concluímos assim que o número médio de sucessos que ocorrem numa experiência aleatória binomial é np.

Lei de Poisson

No caso da variável de Poisson de parâmetro $\lambda > 0$ temos

$$\begin{aligned} \sum_{k=0}^{\infty} |k| P(X=k) &= \sum_{k=1}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = \lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} e^{-\lambda} \\ &= \lambda \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} = \lambda e^{\lambda} e^{-\lambda} = \lambda < \infty. \end{aligned}$$

Assim

$$E(X) = \sum_{k=0}^{\infty} kP(X = k) = \lambda,$$

ou seja, o parâmetro λ representa o número médio de sucessos (ocorrências) numa experiência aleatória de Poisson.

Lei geométrica

No caso da variável geométrica de parâmetro $p \in]0, 1[$ temos

$$\sum_{k=1}^{\infty} |k|P(X = k) = \sum_{k=1}^{\infty} k(1-p)^{k-1}p = \frac{p}{(1-(1-p))^2} = \frac{1}{p},$$

e assim

$$E(X) = \frac{1}{p}.$$

Lei uniforme

Se a variável X possui uma distribuição uniforme sobre o intervalo $[a, b]$ ($X \sim U[a, b]$) então a sua densidade de probabilidade é dada por

$$f_X(x) = \begin{cases} \dfrac{1}{b-a}, & a \leq x \leq b \\ 0, & x < a \text{ ou } x > b, \end{cases}$$

densidade esta que é simétrica relativamente ao ponto $(a + b)/2$. Além disso, uma vez que $|X| \leq \max(|a|, |b|)$, pela alínea c) da Proposição 4.1.8 sabemos em $E(X)$ existe. Usando agora a Proposição 4.1.5 temos então

$$E(X) = \frac{a+b}{2}.$$

Querendo fazer o cálculo usando diretamente a definição de esperança matemática temos:

$$E(X) = \int_a^b x f_X(x)dx = \frac{1}{b-a}\int_a^b x\,dx = \frac{b^2-a^2}{2(b-a)} = \frac{a+b}{2}.$$

Leis exponencial e gama

Se $X \sim \text{Exp}(\lambda)$, $\lambda > 0$, a esperança matemática existe pois

$$\int_{-\infty}^{\infty} |x| f_X(x)dx = \int_0^{\infty} x\lambda e^{-\lambda x}dx = \frac{1}{\lambda} < \infty.$$

Assim

$$E(X) = \int_0^\infty x f_X(x) dx = \frac{1}{\lambda}.$$

No caso mais geral em que $X \sim \gamma(\alpha, \beta)$, é possível provar que $E(X)$ existe e que

$$E(X) = \frac{\alpha}{\beta}.$$

Lei normal ou gaussiana

Consideremos agora o caso em que $X \sim N(\mu, \sigma^2)$, isto é, X possui uma distribuição normal de parâmetros $\mu \in \mathbb{R}$ e $\sigma > 0$. Comecemos por analisar o caso da variável aleatória normal standard, isto é, $Z \sim N(0, 1)$. A densidade de probabilidade de Z é dada por

$$f_Z(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2},$$

que é uma densidade simétrica relativamente ao ponto $x = 0$. Para concluirmos que $E(Z) = 0$, basta mostrar que $\int_{-\infty}^{\infty} |x| f_Z(x)\, dx < \infty$, o que é verdade uma vez que

$$\begin{aligned}\int_{-n}^{n} |x| f_Z(x)\, dx &= \frac{2}{\sqrt{2\pi}} \int_0^n x\, e^{-x^2/2}\, dx = \frac{2}{\sqrt{2\pi}} \Big(- e^{-x^2/2} \Big]_0^n \\ &= \frac{2}{\sqrt{2\pi}} \Big(1 - e^{-n^2/2} \Big) \to \frac{2}{\sqrt{2\pi}}, n \to \infty.\end{aligned}$$

Retomemos agora o caso geral em que $X \sim N(\mu, \sigma^2)$. Pelo Corolário 3.5.2 sabemos que

$$Z = (X - \mu)/\sigma \sim N(0, 1),$$

e assim $E(Z) = 0$. Uma vez que $X = \sigma Z + \mu$, pela alínea g) da Proposição 4.1.8 concluímos que a média de X existe sendo dada por

$$E(X) = \sigma E(Z) + \mu = \mu.$$

Concluímos assim que o parâmetro μ da distribuição $N(\mu, \sigma^2)$ é a esperança matemática ou média da distribuição.

4.2 Momentos e variância

A par da esperança matemática que estudámos no parágrafo anterior, definimos neste parágrafo outros parâmetros de resumo da distribuição de probabilidade duma variável aleatória que têm um papel importante no seu estudo.

Definição 4.2.1. *Sejam $p \in \mathbb{N}$ e X uma v.a.r. Chamamos* momento de ordem p de X *a $m_p = E(X^p)$, sempre que tal esperança matemática exista, e* momento centrado de ordem p de X *a $\mu_p = E(X - E(X))^p$.*

Vamos denotar por $\mathscr{L}^p$ o conjunto das v.a.r. para as quais o momento de ordem p existe. $\mathscr{L}^1$ é assim o conjunto das v.a.r. que possuem esperança matemática, também designadas por variáveis aleatórias integráveis. Uma vez que $X \in \mathscr{L}^p$ sse $X^p \in \mathscr{L}^1$, dizemos também que $\mathscr{L}^p$ é o conjunto das variáveis aleatórias de potência p integrável.

$\mathscr{L}^p$ é um espaço vetorial com a soma e produto escalar definidos da forma habitual. Com efeito, se $X, Y \in \mathscr{L}^p$ e $\alpha \in \mathbb{R}$, então $X + Y \in \mathscr{L}^p$ e $\alpha X \in \mathscr{L}^p$, uma vez que $|X + Y|^p \leq 2^p(|X|^p + |Y|^p) \in \mathscr{L}^1$, e $|\alpha X|^p = |\alpha|^p |X|^p \in \mathscr{L}^1$.

Por outro lado, $\mathscr{L}^p \subset \mathscr{L}^q$ para $p \geq q$. Tal é consequência da desigualdade $|X|^q \leq 1 + |X|^p$, e da alínea d) da Proposição 4.1.8. Em particular $X \in \mathscr{L}^1$ sempre que $X \in \mathscr{L}^p$, para algum $p \geq 2$.

Como parâmetros de resumo da distribuição de probabilidade duma variável aleatória, particular interesse têm para nós o momento de primeira ordem (média), já estudado nos parágrafos anteriores, e o momento centrado de segunda ordem. Este último, por razões que decorrem da sua definição, é um parâmetro de dispersão da distribuição da variável X, medindo essa dispersão, ou variabilidade, relativamente à média $E(X)$ da distribuição.

Definição 4.2.2. *Se $X \in \mathscr{L}^2$, chamamos* variância de X, *que denotamos por* $\mathrm{Var}(X)$, *ao seu momento centrado de segunda ordem, isto é,*

$$\mathrm{Var}(X) = E(X - E(X))^2.$$

A $\sigma(X) = \sqrt{\mathrm{Var}(X)}$, chamamos desvio padrão de X.

Notemos que o desvio padrão $\sigma(X)$ é expresso nas mesmas unidades da variável X, enquanto que a variância $\mathrm{Var}(X)$ está expressa no quadrado dessas unidades.

4.2.1 Propriedades da variância

Proposição 4.2.3. *Se $X \in \mathscr{L}^2$, então* $\mathrm{Var}(X) = 0$ *sse X é constante quase certamente.*

Dem: Pela alínea c) da Proposição 4.1.8 temos $\mathrm{Var}(X) = 0$ sse $E(X - E(X))^2 = 0$ sse $(X - E(X))^2 = 0$ *q.c.* sse $X = E(X)$ *q.c.* ■

Proposição 4.2.4. *Se $X \in \mathscr{L}^2$, então:*
a) $\mathrm{Var}(X) = E(X^2) - (E(X))^2$*;*
b) $\mathrm{Var}(aX + b) = a^2 \mathrm{Var}(X)$.

Dem: Usando as propriedades da esperança matemática temos

$$\begin{aligned}\mathrm{Var}(X) &= E(X - E(X))^2 = E\big(X^2 - 2XE(X) + (E(X))^2\big) \\ &= E(X^2) - 2(E(X))^2 + (E(X))^2 = E(X^2) - (E(X))^2,\end{aligned}$$

e

$$\begin{aligned}\mathrm{Var}(aX+b) &= E\big((aX+b)-E(aX+b)\big)^2 = E\big(aX+b-aE(X)-b\big)^2\\ &= a^2E(X-E(X))^2 = a^2\mathrm{Var}(X).\end{aligned}$$ ■

4.2.2 Cálculo da variância: alguns exemplos

No que se segue vamos limitar-nos ao cálculo da variância de algumas das distribuições de probabilidades consideradas anteriormente.

Lei de Dirac

Uma variável de Dirac no ponto a é tal que $P(X=a)=1$. Já vimos na Proposição 4.2.3 que

$$\mathrm{Var}(X)=0.$$

Lei de Bernoulli

Se $X\sim B(p)$ vimos que $E(X)=p$. Uma vez que $X^2\sim X$ (porquê?) temos

$$E(X^2)=p,$$

e assim

$$\mathrm{Var}(X)=p-p^2=p(1-p).$$

Lei uniforme discreta

Se X é uma variável uniforme discreta sobre o conjunto $\{1,\dots,n\}$, vimos que $E(X)=(n+1)/2$. O seu momento de segunda ordem é dado por

$$E(X^2)=1^2\times\frac{1}{n}+\cdots+n^2\times\frac{1}{n}=\frac{n(n+1)(2n+1)}{6}\frac{1}{n}=\frac{(n+1)(2n+1)}{6}$$

e

$$\mathrm{Var}(X)=\frac{(n+1)(2n+1)}{6}-\Big(\frac{n+1}{2}\Big)^2=\frac{(n+1)(2n-1)}{12}.$$

Lei binomial

Se $X\sim B(n,p)$ vimos que $E(X)=np$. Como

$$\begin{aligned}E(X(X-1)) &= \sum_{k=0}^{n}k(k-1)P(X=k)=\sum_{k=2}^{n}k(k-1)C_k^n p^k(1-p)^{n-k}\\ &= n(n-1)p^2\sum_{k=2}^{n}C_{k-2}^{n-2}p^{k-2}(1-p)^{(n-2)-(k-2)}\end{aligned}$$

$$= n(n-1)p^2 \sum_{l=0}^{n-2} C_\ell^{n-2} p^\ell (1-p)^{(n-2)-\ell}$$
$$= n(n-1)p^2 \big(p + (1-p)\big)^{n-2} = n(n-1)p^2,$$

então

$$E(X^2) = n(n-1)p^2 + E(X) = n(n-1)p^2 + np = np(np - p + 1).$$

Assim

$$\mathrm{Var}(X) = np(np - p + 1) - (np)^2 = np(p-1).$$

Lei de Poisson

Se $X \sim \mathscr{P}(\lambda)$ vimos que $E(X) = \lambda$. Temos também

$$\sum_{k=0}^{\infty} k(k-1)P(X=k) = \sum_{k=2}^{\infty} k(k-1)\frac{\lambda^k}{k!}e^{-\lambda} = \lambda^2 \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!}e^{-\lambda}$$
$$= \lambda^2 \sum_{k=0}^{\infty} \frac{\lambda^k}{k!}e^{-\lambda} = \lambda^2 < \infty,$$

de onde concluímos que $E(X(X-1)) = \lambda^2$. Assim, também $E(X^2)$ existe e

$$E(X^2) = \lambda^2 + \lambda = \lambda(\lambda + 1).$$

Finalmente temos

$$\mathrm{Var}(X) = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

Lei geométrica

No caso da variável geométrica de parâmetro $p \in]0,1[$, temos $E(X) = 1/p$ e

$$\sum_{k=1}^{\infty} k(k-1)P(X=k) = p(1-p)\sum_{k=2}^{\infty} k(k-1)(1-p)^{k-2}$$
$$= p(1-p)\frac{2}{(1-(1-p))^3} = \frac{2(1-p)}{p^2},$$

de onde concluímos que $E(X^2)$ existe e

$$E(X^2) = \frac{2(1-p)}{p^2} + \frac{1}{p} = \frac{2-p}{p^2}.$$

Assim,

$$\mathrm{Var}(X) = \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2}.$$

Lei uniforme

Se $X \sim U[a,b]$ vimos que $E(X) = (a+b)/2$. Temos também

$$E(X^2) = \int_a^b x^2 f_X(x)dx = \frac{1}{b-a}\int_a^b x^2 dx = \frac{b^3-a^3}{3(b-a)}$$

e

$$\mathrm{Var}(X) = \frac{b^3-a^3}{3(b-a)} - \Big(\frac{a+b}{2}\Big)^2 = \frac{(b-a)^2}{12}.$$

Leis exponencial e gama

Se $X \sim \mathrm{Exp}(\lambda)$, vimos que $E(X) = 1/\lambda$. Temos também

$$\int_{-\infty}^{\infty} x^2 f_X(x)dx = \int_0^{\infty} x^2 \lambda e^{-\lambda x}dx = 2\int_0^{\infty} xe^{-\lambda x}dx = \frac{2}{\lambda^2} < \infty.$$

Assim

$$E(X^2) = \frac{2}{\lambda^2}$$

e

$$\mathrm{Var}(X) = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

A distribuição $\mathrm{Exp}(\lambda)$ é um caso particular da distribuição gama $\gamma(\alpha,\beta)$ em que $\alpha = 1$ e $\beta = \lambda$. Se $X \sim \gamma(\alpha,\beta)$, vimos que $E(X) = \alpha/\beta$. É possível provar que

$$E(X^2) = \frac{\alpha(\alpha+1)}{\beta^2}.$$

Assim

$$\mathrm{Var}(X) = \frac{\alpha(\alpha+1)}{\beta^2} - \frac{\alpha^2}{\beta^2} = \frac{\alpha}{\beta^2}.$$

Lei normal ou gaussiana

Se $X \sim N(\mu,\sigma^2)$, com $\mu \in \mathbb{R}$ e $\sigma > 0$, sabemos que $X = \sigma Z + \mu$ com $Z \sim N(0,1)$. Existindo o momento de segunda ordem de Z, pela alínea b) da Proposição 4.2.4 temos

$$\mathrm{Var}(X) = \sigma^2 \mathrm{Var}(Z) = \sigma^2 E(Z^2),$$

e portanto para calcular a variância de X basta calcular o momento de segunda ordem de Z. Comecemos por mostrar que o momento de segunda ordem de Z existe, isto é, que $\int_{-\infty}^{\infty} x^2 f_Z(x)\,dx < \infty$. Integrando por partes temos

$$\int_{-\infty}^{\infty} x^2 f_Z(x)\,dx = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} x^2\, e^{-x^2/2}\,dx = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-x^2/2}\,dx = 1 < \infty.$$

Assim, $E(Z^2) = 1$ e portanto

$$\mathrm{Var}(X) = \sigma^2.$$

O parâmetro σ^2 (resp. σ) da distribuição $N(\mu, \sigma^2)$ é assim a variância (resp. o desvio padrão) da distribuição.

4.2.3 Coeficientes de assimetria e de forma

Terminamos esta secção fazendo referência a outros dois parâmetros de resumo da distribuição de probabilidade duma variável aleatória que nos dão indicação sobre a forma da distribuição de X. São, por isso, ditos parâmetros de forma.

Definição 4.2.5. *Se $X \in \mathscr{L}^3$ chamamos coeficiente de assimetria de X a*

$$\beta_1 = \mu_3/\mu_2^{3/2}.$$

Se $X \in \mathscr{L}^4$ chamamos coeficiente de achatamento ou coeficiente de forma de X a

$$\beta_2 = \mu_4/\mu_2^2.$$

Notemos que se X é simétrica relativamente a $a \in \mathbb{R}$, então $\beta_1 = 0$. Se $\beta_1 > 0$ dizemos que X tem assimetria positiva, e se $\beta_1 < 0$ dizemos que X tem assimetria negativa. O coeficiente de achatamento, também chamado de curtose, que traduz "o peso das caudas" da distribuição de X, é habitualmente comparado com o da distribuição normal para a qual $\beta_2 = 3$. Quando o $\beta_2 = 3$ dizemos que a distribuição é mesocúrtica. Se $\beta_2 > 3$ dizemos que a distribuição de X é leptocúrtica pois possui caudas mais pesadas que as da normal. Se $\beta_2 < 3$ dizemos que a distribuição de X é platicúrtica possuindo caudas mais leves que as da normal.

4.3 Desigualdade de Markov e suas consequências

As desigualdades seguintes permitem obter limites superiores para as "probabilidades de cauda" da distribuição de uma v.a.r. X a partir dos momentos dessa variável. A primeira dessas desigualdades é habitualmente referida como desigualdade de Markov, mas é-o também como desigualdade de Tchebychev.

Proposição 4.3.1 (Desigualdade de Markov ([1])). *Se $X \in \mathscr{L}^1$ é não negativa, então*

$$P(X \geq \alpha) \leq \frac{E(X)}{\alpha},$$

para todo o $\alpha > 0$.

[1] Markov, A.A. (1913). *Ischislenie Veroiatnostei*, 3rd ed. Gosizdat. Moscow.

Dem: Para todo o $\alpha > 0$, é válida a desigualdade

$$\mathbb{1}_{\{X \geq \alpha\}} \leq X/\alpha.$$

Assim, usando as alíneas b), g) e h) da Proposição 4.1.8 obtemos

$$P(X \geq \alpha) = E\big(\mathbb{1}_{\{X \geq \alpha\}}\big) \leq E\big(X/\alpha\big) \leq E(X)/\alpha. \qquad \blacksquare$$

Da desigualdade anterior deduzimos que

$$P\big(X > rE(X)\big) \leq \frac{1}{r}, \quad r > 0,$$

que traduz o facto da média ser também uma medida de dispersão no caso caso das variáveis aleatórias não negativas. De facto, neste caso qualquer um dos momentos o é uma vez que para todo o $p \in \mathbb{N}$ se tem

$$P\big(X > r(E(X^p))^{1/p}\big) \leq \frac{1}{r^p}, \quad r > 0.$$

Proposição 4.3.2 (Desigualdade de Tchebychev-Markov)**.** *Sejam $p \in \mathbb{N}$ e $X \in \mathscr{L}^p$. Então*

$$P\big(|X| \geq \alpha\big) \leq \frac{E\big(|X|^p\big)}{\alpha^p},$$

para todo o $\alpha > 0$.

Dem: Para todo o $\alpha > 0$ temos

$$P(|X| \geq \alpha) = P(|X|^p \geq \alpha^p).$$

Basta agora usar a desigualdade de Markov com $|X|^p$ no lugar de X e α^p no lugar de α. $\blacksquare$

Proposição 4.3.3 (Desigualdade de Bienaymé-Tchebychev ([2]))**.** *Se $X \in \mathscr{L}^2$, então*

$$P\big(|X - E(X)| \geq \alpha\big) \leq \frac{\mathrm{Var}(X)}{\alpha^2},$$

para todo o $\alpha > 0$.

Dem: Basta aplicar a desigualdade de Tchebychev-Markov com $p = 2$ à variável $Y = X - E(X)$. $\blacksquare$

[2] Bienaymé, I.-J. (1853). Considérations à l'appui de la découverte de Laplace sur la loi de probabilité dans la méthode des moindres carré. *C. R. Acad. Sci. Paris* 37, 309–324; *J. Math. Pures Appl. (2^e série)* 12, 158–176, 1867; Tchébychef, P.L. (1867). Des valeurs moyennes. *J. Math. Pures Appl. (2^e série)* 12, 177–184.

Esta última desigualdade permite-nos naturalmente obter limites inferiores para a probabilidade da variável X pertencer a um intervalo centrado na sua média com semi-amplitude igual a α:

$$P\big(|X - E(X)| < \alpha\big) \geq 1 - \frac{\mathrm{Var}(X)}{\alpha^2}.$$

Podemos assim concluir que para qualquer variável aleatória com momento de segunda ordem, a probabilidade do seu desvio absoluto relativamente à média ser superior ou igual a r vezes o seu desvio-padrão, não é superior a $1/r^2$. Com efeito, usando esta última desigualdade com $\alpha = r\sigma(X)$, obtemos

$$P\big(|X - E(X)| < r\sigma(X)\big) \geq 1 - \frac{\mathrm{Var}(X)}{r^2\sigma(X)^2} = 1 - \frac{1}{r^2},$$

isto é, a probabilidade de X pertencer ao intervalo

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

é sempre maior ou igual a $1 - 1/r^2$. Se $r = 2$ obtemos $1 - 1/r^2 = 0{,}75$; se $r = 3$ obtemos $1 - 1/r^2 = 0{,}888\ldots$; se $r = 5$ obtemos $1 - 1/r^2 = 0{,}96$; e para $r = 18$ temos $1 - 1/r^2 = 0{,}997$ (ver Exercício 46).

4.4 Localização e dispersão sem momentos

A média e o desvio padrão são os parâmetros mais usados de localização e de dispersão da distribuição de uma variável aleatória real X. Nesta secção referimos brevemente outras duas medidas de localização e de dispersão que não são baseados na noção de momento da variável X mas sim na noção de quantil da variável X.

Definição 4.4.1. *Dado $p \in]0,1[$, chamamos quantil de ordem p da distribuição da v.a.r. X (ou de X) a todo o número real Q_p tal que*

$$P_X(]-\infty, Q_p[) \leq p \quad e \quad P_X(]-\infty, Q_p]) \geq p.$$

Atendendo a que $P_X(]-\infty, Q_p[) = F_X(Q_p^-)$ e $P_X(]-\infty, Q_p]) = F_X(Q_p)$, se F_X é contínua então o quantil de ordem p de X é tal que

$$F_X(Q_p) = p.$$

Além disso, se X é absolutamente contínua o quantil de ordem p de X é tal que

$$\int_{-\infty}^{Q_p} f_X(x)dx = p,$$

onde f_X é a densidade de probabilidade de X.

No que se segue vamos restringir-nos ao caso em que F_X é contínua e estritamente crescente, caso em que o quantil de ordem p de X é único, sendo dado

$$Q_p = F_X^{-1}(p).$$

Ao quantil de ordem $1/2$ chamamos mediana de X. A mediana divide a distribuição de X em duas partes com iguais probabilidades, sendo, por isso, uma medida muito utilizada do centro da distribuição de X. Se X é simétrica relativamente a $x = a$, isto é, $P_X(]-\infty, a-x]) = P_X([a+x, +\infty[)$, para todo o $x \in \mathbb{R}$, então $Q_{1/2} = a$. Neste caso a mediana coincide com a média $E(X)$, caso esta exista (ver Proposição 4.1.5).

Outros dois casos importantes são os dos quantis de ordens $1/4$ e $3/4$, a que chamamos primeiro e terceiro quartis de X, respetivamente. A mediana é também denominada segundo quartil de X. Os quartis dividem a distribuição de X em quatro partes com iguais probabilidades. A distância entre o primeiro e o terceiro quartil, $Q_{3/4} - Q_{1/4}$, é chamada amplitude interquartil e é uma medida muito utilizada da dispersão ou variabilidade de X.

4.5 Bibliografia

Gonçalves, E., Mendes Lopes, N. (2013). *Probabilidades: princípios teóricos*. Lisboa: Escolar Editora.

Haigh, J. (2002). *Probability models*. London: Springer.

Jacod, J., Protter, P. (2000). *Probability essentials*. Berlin: Springer.

James, B.R. (1981). *Probabilidade: um curso em nível intermediário*. Rio de Janeiro: IMPA.

Monfort, A. (1980). *Cours de probabilités*. Paris: Economica.

Resnick, S.I. (2005). *A probability path*. Boston: Birkhäuser.

Tenreiro, C. (2002). *Apontamentos de Teoria das Probabilidades*. Coimbra: DMUC. https://www.mat.uc.pt/∼tenreiro/apontamentos/ATP0204.pdf

5

Vetores aleatórios

Vetores aleatórios e suas leis de probabilidade. Leis de probabilidade marginais. Vetores discretos e vetores absolutamente contínuos. Função de distribuição dum vetor aleatório. Independência de variáveis aleatórias. Distribuições condicionais. Covariância e correlação. Desigualdades de Hölder e de Minkowski. Média e matriz de covariância dum vetor aleatório. Vetores aleatórios normais (parte I).

5.1 Introdução

Para 200 homens registaram-se a altura (X) e o peso (Y) respetivos, que estão representados no gráfico de dispersão seguinte, onde cada ponto (x, y) aí marcado corresponde à altura x e ao peso y observados num mesmo indivíduo. Como podemos constatar, o gráfico de dispersão permite uma representação simples da distribuição do vetor bidimensional (X, Y) com valores em $\mathbb{R}^2$.

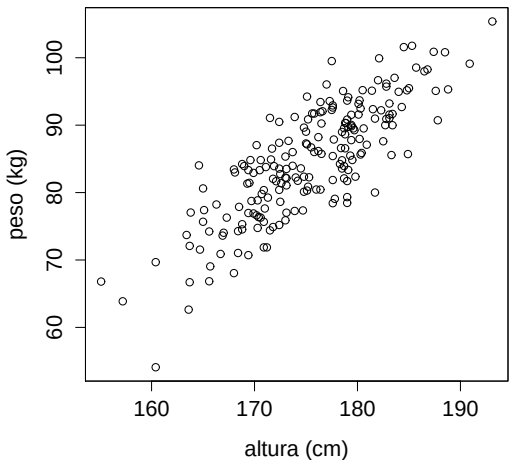


Figura 5.1: Gráfico de dispersão relativo às alturas e pesos observados.

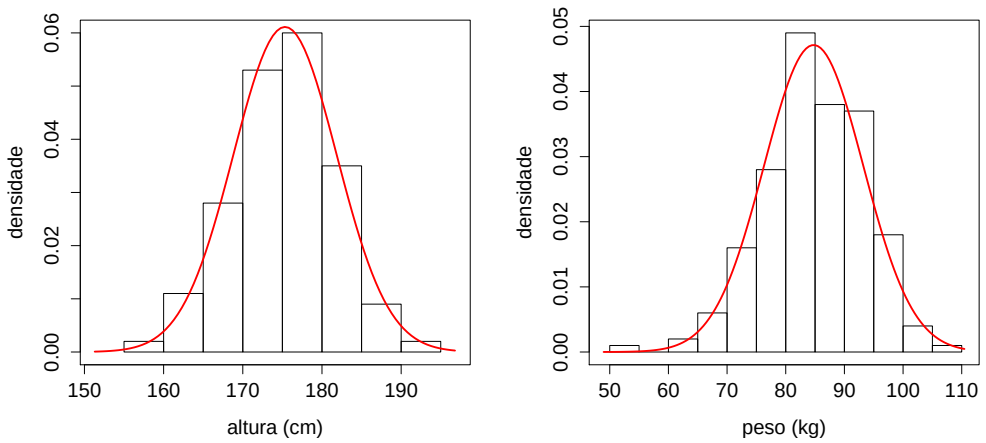


Figura 5.2: Histogramas de frequência relativos às alturas e pesos observados, e respetivas densidades normais com parâmetros iguais às médias e aos desvios padrão das observações.

O gráfico revela que pesos mais baixos e mais altos estão predominantemente associados a alturas mais baixas e mais elevadas, respetivamente. Esta caraterística da distribuição bidimensional, ou conjunta, das variáveis X e Y não é revelada quando analisamos a distribuição das variáveis X e Y isoladamente, o que pode ser feito usando histogramas de frequência (ver Figura 5.2). Tal é ainda mais claro se representarmos novamente as observações através dum gráfico de dispersão em que apesar de possuirmos os valores dos pesos e das alturas, perdemos a informação sobre a correspondência entre uns e outros valores. Nesse caso, apesar da distribuição dos pesos e das alturas revelada pelos histogramas se manter, a informação dada pelo gráfico de dispersão não

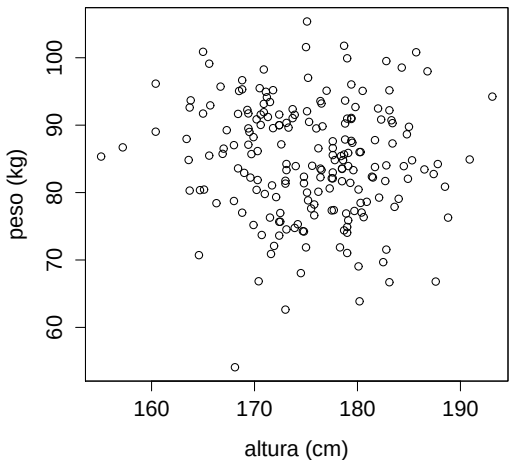


Figura 5.3: Gráfico de dispersão relativo às alturas e a uma permutação dos pesos observados.

é mais reveladora da distribuição conjunta do vetor (X, Y). Isso é ilustrado na Figura 5.3 onde se representam os pontos $(x_i, y_{p(i)})$, com p uma dada permutação do conjunto $\{1, \ldots, 200\}$.

O estudo que fizemos até aqui possibilita a descrição da distribuição de probabilidade de variáveis aleatórias reais X, mas não permite, em geral, a descrição da distribuição de probabilidade de vetores aleatórios bidimensionais (X, Y) para os quais existe, tal como no exemplo anterior, uma associação forte entre os valores das variáveis X e Y. Por essa razão, estudamos neste capítulo distribuições multidimensionais que nos permitem modelar experiências aleatórias em que a cada resultado da experiência $\omega \in \Omega$ associamos dois valores numéricos $(X(\omega), Y(\omega))$, ou, mais geralmente, d valores numéricos $(X_1(\omega), \ldots, X_d(\omega))$.

5.2 Vetores aleatórios e suas leis de probabilidade

Sendo $(\Omega, \mathscr{A}, P)$ um modelo probabilístico para a experiência aleatória que estudamos, e observado um resultado da experiência, $\omega \in \Omega$, interessamo-nos agora por uma função multivariada de ω. Representando por $X = (X_1, \ldots, X_d)$ uma tal função, X não é mais do que uma aplicação de Ω em $\mathbb{R}^d$, que a $\omega \in \Omega$ associa o vetor $X(\omega) = (X_1(\omega), \ldots, X_d(\omega))$. Vamos denotar por $\mathscr{B}(\mathbb{R}^d)$ a σ-álgebra de Borel em $\mathbb{R}^d$, isto é, a σ-álgebra gerada pelos abertos de $\mathbb{R}^d$.

Definição 5.2.1. *Chamamos* vetor aleatório real de dimensão d *(ve.a.r.) a toda a aplicação $X = (X_1, \ldots, X_d)$ de Ω em $\mathbb{R}^d$ que satisfaz $\{X \in B\} \in \mathscr{A}$, para todo o $B \in \mathscr{B}(\mathbb{R}^d)$.*

Atendendo a que $\mathscr{B}(\mathbb{R}^d)$ é também a σ-álgebra gerada pelos conjuntos da forma $B_1 \times \cdots \times B_d$, com $B_i \in \mathscr{B}(\mathbb{R})$, para $i = 1, \ldots, d$, $X = (X_1, \ldots, X_d)$ é um vetor aleatório sse $\{X \in B\} \in \mathscr{A}$ para todo o $B = B_1 \times \cdots \times B_d \in \mathscr{B}(\mathbb{R}^d)$. Uma vez que para $B = B_1 \times \cdots \times B_d \in \mathscr{B}(\mathbb{R}^d)$ se tem

$$\{X \in B\} = \{X_1 \in B_1, \ldots, X_d \in B_d\} = \bigcap_{i=1}^{d} \{X_i \in B_i\},$$

concluímos que X é um ve.a.r. sse X_i é uma v.a.r. para todo o $i = 1, \ldots, d$. As v.a.r. $X_1, \ldots, X_d$ são chamadas margens de X ou variáveis aleatórias marginais de X.

Proposição 5.2.2. *Se X é um ve.a.r. definido num espaço de probabilidade $(\Omega, \mathscr{A}, P)$ então a aplicação P_X definida para $B \in \mathscr{B}(\mathbb{R}^d)$, por*

$$P_X(B) = P(X \in B),$$

é uma probabilidade em $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$*, a que chamamos* distribuição de probabilidade *ou* lei de probabilidade *do vetor* X*.*

Dem: Análoga à da Proposição 3.1.2. ■

Para $x = (x_1, \ldots, x_d)$ e $y = (y_1, \ldots, y_d)$ em $\mathbb{R}^d$, escrevemos $x \leq y$ (resp. $x < y$) sempre que $x_i \leq y_i$ (resp. $x_i < y_i$), para todo o $i = 1, \ldots, d$. Tal com em $\mathbb{R}$, os conjuntos dos pontos x tais que $a < x \leq b$, ou dos pontos x tais que $x \leq b$, serão denotados por $]a, b]$ ou $]-\infty, b]$, respetivamente.

Atendendo ao facto da σ-álgebra de Borel $\mathscr{B}(\mathbb{R}^d)$ ser também gerada pelos retângulos da forma $]a, b]$, com $a \leq b$, é possível provar que a distribuição de probabilidade de um ve.a.r. X é determinada pelos valores que P_X toma nos rectângulos $]a, b]$, com $a \leq b$. Assim, se X e Y são ve.a.r. para os quais $P_X(]a, b]) = P_Y(]a, b])$, para todo o $a \leq b$, então $P_X(B) = P_Y(B)$, para todo o $B \in \mathscr{B}(\mathbb{R}^d)$, isto é, X e Y têm a mesma distribuição de probabilidade. Quando tal acontece escrevemos $X \sim Y$. Comentário análogo é também válido para os conjuntos da forma $]-\infty, b]$, com $b \in \mathbb{R}^d$.

As distribuições de probabilidade das margens do ve.a.r. $X = (X_1, \ldots, X_d)$, ditas distribuições marginais, ou, mais geralmente, as distribuições de probabilidade de qualquer subvetor de X, podem ser obtidas a partir da lei do vetor X. Para $\{i_1, \ldots, i_m\} \subset \{1, \ldots, d\}$ vamos representar por $\pi_{i_1,\ldots,i_m}$ a aplicação de $\mathbb{R}^d$ em $\mathbb{R}^m$, dita projeção, definida por

$$\pi_{i_1,\ldots,i_m}(x_1, \ldots, x_d) = (x_{i_1}, \ldots, x_{i_m}).$$

Para $B = B_{i_1} \times \ldots \times B_{i_m} \in \mathscr{B}(\mathbb{R}^m)$, vamos representar por $\pi^{-1}_{i_1,\ldots,i_m}(B)$ a imagem recíproca de B por $\pi_{i_1,\ldots,i_m}$, isto é,

$$\pi^{-1}_{i_1,\ldots,i_m}(B) = \{(x_1, \ldots, x_d) \in \mathbb{R}^d : \pi_{i_1,\ldots,i_m}(x_1, \ldots, x_d) \in B\} = \prod_{i=1}^{d} C_i,$$

onde $C_i = B_i$, sempre que $i \in \{i_1, \ldots, i_m\}$, e $C_i = \mathbb{R}$, caso contrário.

Proposição 5.2.3. *Se* $X = (X_1, \ldots, X_d)$ *é um ve.a.r., então para todo o* $\{i_1, \ldots, i_m\} \subset \{1, \ldots, d\}$*, temos*

$$P_{(X_{i_1},\ldots,X_{i_m})}\big(B_{i_1} \times \ldots \times B_{i_m}\big) = P_X\big(\pi^{-1}_{i_1,\ldots,i_m}(B_{i_1} \times \ldots \times B_{i_m})\big),$$

para $B_{i_1} \times \ldots \times B_{i_m} \in \mathscr{B}(\mathbb{R}^m)$*. Em particular,*

$$P_{X_i}(B_i) = P_X\big(\pi_i^{-1}(B_i)\big),$$

para $B_i \in \mathscr{B}(\mathbb{R})$ *e* $i = 1, \ldots, d$*.*

Dem: Para $B_{i_1} \times \ldots \times B_{i_m} \in \mathscr{B}(\mathbb{R}^m)$ temos

$$\begin{aligned} P_{(X_{i_1},\ldots,X_{i_m})}\big(B_{i_1} \times \ldots \times B_{i_m}\big) &= P\big(X_{i_1} \in B_{i_1}, \ldots, X_{i_m} \in B_{i_m}\big) \\ &= P\big(X_1 \in C_1, \ldots, X_d \in C_d\big) \\ &= P_X\big(C_1 \times \cdots \times C_d\big), \end{aligned}$$

onde $C_i = B_i$, quando $i \in \{i_1, \ldots, i_m\}$, e $C_i = \mathbb{R}$, caso contrário. Isto conclui a demonstração uma vez que $C_1 \times \cdots \times C_d = \pi^{-1}_{i_1,\ldots,i_m}(B_{i_1} \times \ldots \times B_{i_m})$. ■

Como se ilustra no exemplo seguinte, o conhecimento das distribuições marginais de um ve.a. não permite, em geral, caraterizar a sua distribuição de probabilidade.

Exemplo 5.2.4. Consideremos os vectores $X = (X_1, X_2)$ e $Y = (Y_1, Y_2)$ com distribuições distintas definidas, para $(i, j) \in \{0, 1\}^2$, por $P_X(\{(i, j)\}) = 1/8$, se $i = j$, $P_X(\{(i, j)\}) = 3/8$, se $i \neq j$, e $P_Y(\{(i, j)\}) = 1/4$, para todo o (i, j). As distribuições marginais destes vetores coincidem, sendo distribuições de Bernoulli de parâmetro $1/2$.

Nos exemplos seguintes retomamos experiências aleatórias consideradas nos capítulos anteriores para exemplificar diferentes tipos de distribuições que podemos observar em vetores aleatórios.

Exemplo 5.2.5. Consideremos novamente a experiência aleatória que consiste no lançamento de dois dados equilibrados e seja (X, Y) o ve.a.r. em que X nos dá o número de pontos que sai no primeiro dado e Y nos dá a soma dos pontos obtidos nos dois dados. Como vimos no Exemplo 3.1.3, o modelo probabilístico para esta experiência aleatória é dado por $(\Omega, \mathscr{A}, P)$, onde $\Omega = \{(i, j) : i, j \in \{1, \ldots, 6\}\}$, $\mathscr{A} = \mathscr{P}(\Omega)$ e

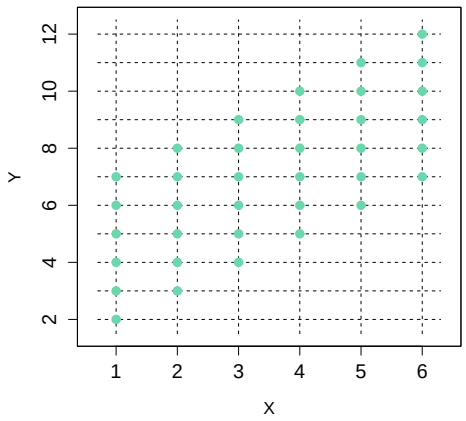


Figura 5.4: Suporte da distribuição do vetor (X, Y), onde Y é a soma dos pontos obtidos no lançamento de dois dados equilibrados e X os pontos obtidos no primeiro dos dados.

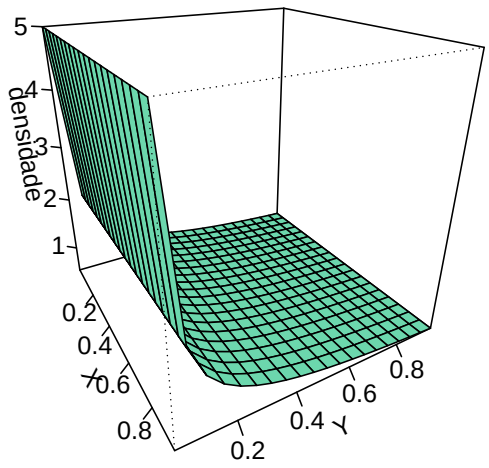


Figura 5.5: Densidade de probabilidade do vetor (X, Y) onde X é a abcissa e Y o quadrado da ordenada de um ponto extraído ao acaso no quadrado $[0, 1]^2$.

$P(\{\omega\}) = 1/36, \quad \omega \in \Omega$. As distribuições das margens do vetor (X, Y) são já nossas conhecidas. Relativamente à distribuição de (X, Y), este vetor aleatório toma valores no subconjunto S de $\{1, \ldots, 6\} \times \{2, 3, \ldots, 12\} \subset \mathbb{R}^2$ dado por

$$S = \{(x, y) \in \mathbb{N}^2 : 1 \leq x \leq 6, x + 1 \leq y \leq x + 6\},$$

com probabilidades $P((X, Y) = (x, y)) = 1/36$, para $(x, y) \in S$.

Por analogia com o caso das variáveis aleatórias reais, existindo um conjunto finito ou numerável S onde se tem $P((X, Y) \in S) = 1$, diremos que o vetor (X, Y) é discreto (ver Definição 5.3.1).

Exemplo 5.2.6. Consideremos agora a experiência aleatória que consiste na extração ao acaso de um ponto do quadrado $[0, 1]^2$. O modelo probabilístico $(\Omega, \mathscr{A}, P)$ para esta experiência aleatória é dado por $\Omega = [0, 1]^2$, $\mathscr{A} = \mathscr{B}([0, 1]^2)$ e $P(]a, b]\times]c, d]) = (b - a) \times (d - c) = \lambda(]a, b]\times]c, d])$, com $0 \leq a < b \leq 1$ e $0 \leq c < d \leq 1$ (ver Exemplo 2.2.7). Sendo $\omega = (\omega_1, \omega_2)$ o ponto extraído ao acaso do quadrado $[0, 1]^2$, consideremos o vetor aleatório (X, Y), onde $X(\omega) = \omega_1$ é a abcissa do ponto extraído e $Y(\omega) = \omega_2^2$ é o quadrado da ordenada desse ponto. Também neste caso, as distribuições das margens do vetor (X, Y) são já nossas conhecidas. Relativamente à distribuição conjunta, (X, Y)

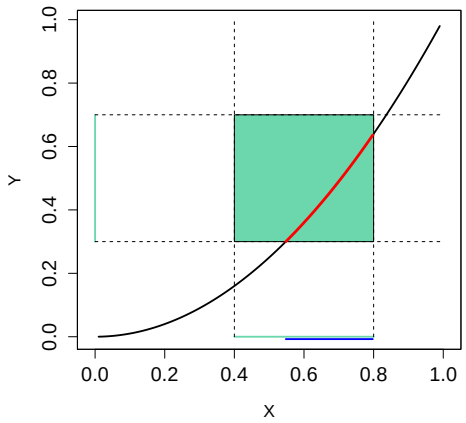


Figura 5.6: Representação da probabilidade $P_{(X,Y)}(]0{,}4;0{,}8]\times]0{,}3;0{,}7])$, onde X é o número extraído do intervalo $[0,1]$ e Y é o quadrado desse número.

toma valores no quadrado $[0,1]^2$ e para $]a,b]\times]c,d] \subset [0,1]^2$ temos

$$\begin{aligned} P_{(X,Y)}(]a,b]\times]c,d]) &= P\big(X \in]a,b], Y \in]c,d]\big) \\ &= P\big(\{(\omega_1,\omega_2) \in [0,1]^2 : \omega_1 \in]a,b], \omega_2^2 \in]c,d]\}\big) \\ &= P\big(\{(\omega_1,\omega_2) \in [0,1]^2 : \omega_1 \in]a,b], \omega_2 \in]\sqrt{c},\sqrt{d}]\}\big) \\ &= P\big(]a,b]\times]\sqrt{c},\sqrt{d}]\big) \\ &= (b-a)(\sqrt{d}-\sqrt{c}) \\ &= \int_{]a,b]\times]c,d]} \frac{1}{2\sqrt{y}}\, dxdy. \end{aligned}$$

Por analogia com o caso das variáveis aleatórias reais, podendo a probabilidade anterior ser identificada com o volume da região sob o gráfico de determinada função não negativa na região $]a,b]\times]c,d]$, diremos que (X,Y) é absolutamente contínuo (ver Definição 5.3.4).

O exemplo seguinte dá-nos conta de um vetor que não é discreto nem absolutamente contínuo.

Exemplo 5.2.7. Consideremos agora a experiência aleatória que consiste na extração ao acaso de um número do intervalo $[0,1]$. Como vimos no Exemplo 3.1.4, o modelo probabilístico $(\Omega,\mathscr{A},P)$ para esta experiência aleatória é dado por $\Omega = [0,1]$, $\mathscr{A} = \mathscr{B}([0,1])$ e $P(]a,b]) = b-a = \lambda(]a,b])$, com $0 \leq a < b \leq 1$. Consideremos o vetor aleatório (X,Y) onde X é o número extraído do intervalo $[0,1]$ e Y é o quadrado desse número. As distribuições das margens do vetor (X,Y) são já nossas conhecidas. Relativamente à distribuição conjunta, (X,Y) toma valores no subconjunto de $[0,1]^2$

dado por $\{(x,y) \in [0,1]^2 : y = x^2\}$ e para $]a,b]\times]c,d] \subset [0,1]^2$,

$$\begin{aligned} P_{(X,Y)}(]a,b]\times]c,d]) &= P\big(X \in]a,b], Y \in]c,d]\big) \\ &= P(\{\omega \in [0,1] : \omega \in]a,b], \omega^2 \in]c,d]\}) \\ &= P(\{\omega \in [0,1] : \omega \in]a,b] \cap]\sqrt{c}, \sqrt{d}]\}) \\ &= \lambda(]a,b] \cap]\sqrt{c}, \sqrt{d}]). \end{aligned}$$

5.3 Vetores discretos e vetores absolutamente contínuos

Tal como no caso da variáveis aleatórias reais, vamos centrar a nossa atenção nos vetores discretos e nos vetores absolutamente contínuos.

Definição 5.3.1. *Dizemos que um ve.a.r. $X = (X_1, \ldots, X_d)$ é* discreto *(ou que a sua lei de probabilidade P_X é discreta) se existe um conjunto finito ou infinito numerável S tal que $P_X(S) = 1$. Ao mais pequeno conjunto S_X com a propriedade anterior chamamos* suporte de X *(ou de P_X), e à função f_X definida por*

$$f_X(x) = P(X = x),\ x \in \mathbb{R}^d,$$

chamamos função de probabilidade *do vetor X.*

Tal como no caso real, se X é ve.a.r. discreto então o seu suporte é dado por

$$S_X = \big\{x \in \mathbb{R}^d : P(X = x) > 0\big\},$$

e a função de probabilidade de X é uma função não negativa com $f_X(x) = 0$, para $x \notin S_X$, e

$$\sum_{x\in\mathbb{R}^d} f_X(x) = \sum_{x\in S_X} P(X = x) = P(X \in S_X) = 1.$$

A função de probabilidade do vetor aleatório discreto X, carateriza a distribuição de probabilidade de X. Com efeito, para qualquer subconjunto $B \in \mathscr{B}(\mathbb{R}^d)$ temos

$$P_X\big(B\big) = P\big(X \in B\big) = \sum_{x\in B\cap S_X} f_X(x) = \sum_{x\in B} f_X(x).$$

Proposição 5.3.2. *Se $X = (X_1, \ldots, X_d)$ é um ve.a.r. discreto com suporte S_X e função de probabilidade f_X, então, para todo o $\{i_1, \ldots, i_m\} \subset \{1, \ldots, d\}$, $(X_{i_1}, \ldots, X_{i_m})$ é discreto com suporte $\pi_{i_1,\ldots,i_m}(S_X)$ e função de probabilidade*

$$h(x_{i_1}, \ldots, x_{i_m}) = \sum_{(x_1,\ldots,x_{i_1-1},x_{i_1+1},\ldots,x_{i_m-1},x_{i_m+1},\ldots,x_d)\in\mathbb{R}^{d-m}} f_X(x_1, \ldots, x_d),$$

para $(x_{i_1}, \ldots, x_{i_m}) \in \mathbb{R}^m$. *Em particular, as margens* X_i *são v.a.r. discretas com função de probabilidade*

$$f_{X_i}(x_i) = \sum_{(x_1,\ldots,x_{i_1-1},x_{i_1+1},\ldots,x_d)\in\mathbb{R}^{d-1}} f_X(x_1, \ldots, x_d),$$

para $x_i \in \mathbb{R}$.

Dem: Vamos provar o resultado no caso $d = 2$. Os argumentos utilizados podem ser facilmente adaptados ao caso $d > 2$. Seja então (X, Y) um ve.a.r. discreto com função de probabilidade $f_{(X,Y)}$ e suporte $S_{(X,Y)}$, que sabemos ser finito ou infinito numerável. Comecemos por provar que a v.a.r. X é discreta. Para tal vamos mostrar que o conjunto finito ou infinito numerável $\pi_1(S_{(X,Y)})$ é tal que $P_X(\pi_1(S_{(X,Y)})) = 1$ (ver Definição 3.2.1). Com efeito, uma vez que $\{(X, Y) \in S_{(X,Y)}\} \subset \{X \in \pi_1(S_{(X,Y)})\}$, temos

$$P_X\big(\pi_1(S_{(X,Y)})\big) = P\big(X \in \pi_1(S_{(X,Y)})\big) \geq P\big((X, Y) \in S_{(X,Y)}\big) = P_{(X,Y)}\big(S_{(X,Y)}\big) = 1.$$

Atendendo à definição de suporte S_X de X, para concluirmos que $S_X = \pi_1(S_{(X,Y)})$, basta provar que $\pi_1(S_{(X,Y)}) \subset S_X$. Dado então $x \in \pi_1(S_{(X,Y)})$, sabemos existir $y \in \mathbb{R}$ tal que $(x, y) \in S_{(X,Y)}$. Assim,

$$P(X = x) \geq P(X = x, Y = y) > 0,$$

ou seja $x \in S_X$ (ver Proposição 3.2.2). De igual forma se provaria que Y é discreta de suporte $\pi_2(S_{(X,Y)})$.

Finalmente, para $x \in \mathbb{R}$, temos

$$\begin{aligned} f_X(x) &= P\big(X = x, Y \in S_Y\big) \\ &= \sum_{y\in S_Y} P(X = x, Y = y) \\ &= \sum_{y\in\mathbb{R}} P(X = x, Y = y) \\ &= \sum_{y\in\mathbb{R}} f_{(X,Y)}(x, y). \end{aligned}$$

De igual forma se provaria que $f_Y(y) = \sum_{x\in\mathbb{R}} f_{(X,Y)}(x, y)$, para $y \in \mathbb{R}$. ■

Exemplo 5.3.3. No caso vetor (X, Y) considerado no Exemplo 5.2.5, (X, Y) é assim um vetor aleatório discreto com suporte

$$S_{(X,Y)} = \{(x, y) \in \mathbb{N}^2 : 1 \leq x \leq 6, x + 1 \leq y \leq x + 6\},$$

(ver Figura 5.4) e função de probabilidade

$$f_{(X,Y)}(x,y) = 1/36, \text{ para } (x,y) \in S_{(X,Y)}.$$

De acordo com a Proposição 5.3.2, X e Y são v.a.r. discretas com funções de probabilidade dadas por

$$f_X(x) = \sum_{y\in\mathbb{R}} f_{(X,Y)}(x,y) = \sum_{y=x+1}^{x+6} \frac{1}{36} = \frac{1}{6},$$

para $x \in \pi_1(S_{(X,Y)}) = \{1,\ldots,6\}$, e $f_X(x) = 0$ para $x \notin \{1,\ldots,6\}$, e

$$f_Y(y) = \sum_{x\in\mathbb{R}} f_{(X,Y)}(x,y) = \sum_{x=1}^{6} f_{(X,Y)}(x,y),$$

para $y \in \pi_2(S_{(X,Y)}) = \{2,\ldots,12\}$ e $f_Y(y) = 0$, para $y \notin \{2,\ldots,12\}$, ou seja,

$$f_Y(y) = \begin{cases} 1/36, & y \in \{2,12\} \\ 2/36, & y \in \{3,11\} \\ 3/36, & y \in \{4,10\} \\ 4/36, & y \in \{5,9\} \\ 5/36, & y \in \{6,8\} \\ 6/36, & y \in \{7\} \\ 0, & \text{caso contrário,} \end{cases}$$

Definição 5.3.4. *Dizemos que um vetor aleatório real (ve.a.r.) X é absolutamente contínuo (ou que a sua lei de probabilidade P_X é absolutamente contínua) se existe uma função não negativa e integrável $f_X : \mathbb{R}^d \to \mathbb{R}$ tal que*

$$P_X\big(]a,b]\big) = \int_{]a,b]} f_X(x_1,\ldots,x_d)dx_1\ldots dx_d,$$

para todo o retângulo $]a,b]$, com $a \le b$. A função f_X é dita densidade de probabilidade do vetor X.

A densidade de probabilidade de um vetor absolutamente contínua X é assim uma função não negativa e integrável com

$$\int_{\mathbb{R}^d} f_X(x)dx = \lim \int_{]-n,n]^d} f_X(x)dx = \lim P_X\big(]-n,n]^d\big) = P_X(\mathbb{R}^d) = 1.$$

Como já referimos, a distribuição de probabilidade P_X é determinada pelos valores que toma nos retângulos da forma $]a,b]$, com $a \le b$. Da Definição 5.3.4 decorre que sendo X absolutamente contínuo, para um qualquer subconjunto $B \in \mathscr{B}(\mathbb{R}^d)$ temos

$$P_X\big(B\big) = \int_B f_X(x_1,\ldots,x_d)dx_1\ldots dx_d.$$

Proposição 5.3.5. *Se $X = (X_1, \ldots, X_d)$ é um ve.a.r. absolutamente contínuo com densidade de probabilidade f_X, então, para todo o $\{i_1, \ldots, i_m\} \subset \{1, \ldots, d\}$, o ve.a.r. $(X_{i_1}, \ldots, X_{i_m})$ é absolutamente contínuo com densidade de probabilidade*

$$f(x_{i_1}, \ldots, x_{i_m}) = \int_{\mathbb{R}^{d-m}} f_X(x_1, \ldots, x_d) dx_1, \ldots, dx_{i_1-1}, dx_{i_1+1}, \ldots, dx_d,$$

para $(x_{i_1}, \ldots, x_{i_m}) \in \mathbb{R}^m$. Em particular, as margens X_i são v.a.r. absolutamente contínuas com densidade de probabilidade

$$f_{X_i}(x_i) = \int_{\mathbb{R}^{d-1}} f_X(x_1, \ldots, x_d), dx_1, \ldots, dx_{i_1-1}, dx_{i_1+1}, \ldots, dx_d,$$

para $x_i \in \mathbb{R}$.

Dem: Vamos provar o resultado no caso $d = 2$. Seja então (X, Y) um ve.a.r. absolutamente contínuo com densidade de probabilidade $f_{(X,Y)}$. Para todo o intervalo real $]a, b]$, temos

$$\begin{aligned} P\big(X \in]a, b]\big) = P\big(X \in]a, b], Y \in \mathbb{R}\big) &= P_{(X,Y)}\big(]a, b] \times \mathbb{R}\big) \\ &= \int_{]a,b]\times\mathbb{R}} f_{(X,Y)}(x, y) dx dy \\ &= \int_a^b \int_{\mathbb{R}} f_{(X,Y)}(x, y) dy dx = \int_a^b f(x) dx, \end{aligned}$$

onde a função $f(x) = \int_{\mathbb{R}} f_{(X,Y)}(x, y) dy$ é não negativa e, pelo teorema de Fubini,

$$\int_{\mathbb{R}} f(x) dx = \int_{\mathbb{R}^2} f_{(X,Y)}(x, y) dx dy = 1.$$

Assim, X é uma v.a.r. absolutamente contínua com densidade de probabilidade dada por $f_X(x) = \int_{\mathbb{R}} f_{(X,Y)}(x, y) dy$.

De modo análogo se prova que Y é uma v.a.r. absolutamente contínua com densidade de probabilidade dada por $f_Y(y) = \int_{\mathbb{R}} f_{(X,Y)}(x, y) dx$. ■

Exemplo 5.3.6. No caso vetor (X, Y) considerado no Exemplo 5.2.6, concluímos que se trata dum vetor absolutamente contínuo com densidade de probabilidade

$$f_{(X,Y)}(x, y) = \begin{cases} \dfrac{1}{2\sqrt{y}}, & 0 < x, y < 1 \\ 0, & \text{caso contrário,} \end{cases} = \frac{1}{2\sqrt{y}} \mathbb{1}_{]0,1[}(x) \mathbb{1}_{]0,1[}(y)$$

(ver Figura 5.5). De acordo com a Proposição 5.3.5 as v.a.r. X e Y são absolutamente contínuas com densidades de probabilidade

$$f_X(x) = \int_{\mathbb{R}} f_{(X,Y)}(x, y) dy = \int_0^1 \frac{1}{2\sqrt{y}} dy \mathbb{1}_{]0,1[}(x) = \mathbb{1}_{]0,1[}(x)$$

e

$$f_Y(y) = \int_{\mathbb{R}} f_{(X,Y)}(x, y) dx = \int_0^1 dx \frac{1}{2\sqrt{y}} \mathbb{1}_{]0,1[}(y) = \frac{1}{2\sqrt{y}} \mathbb{1}_{]0,1[}(y).$$

5.4 Função de distribuição dum vetor aleatório

Neste parágrafo generalizamos a noção de função de distribuição ao caso multivariado.

Definição 5.4.1. *Chamamos função de distribuição do vetor aleatório $X = (X_1, \ldots, X_d)$, e denotamo-la por F_X, a função de $\mathbb{R}^d$ em $[0,1]$ definida por*

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

Proposição 5.4.2. *F_X goza das seguintes propriedades:*

a) F_X é não negativa, limitada e crescente coordenada a coordenada;

b) Para $i = 1, \ldots, d$, $\lim_{x_i \to -\infty} F_X(x) = 0$, e $\lim_{x_1,\ldots,x_d \to +\infty} F_X(x) = 1$;

c) F_X é contínua à direita, isto é, para todo o $x \in \mathbb{R}^d$

$$\lim_{y \to x, y \geq x} F_X(y) = F_X(x);$$

d) Para $x \in \mathbb{R}^d$

$$\lim_{y \to x, y \leq x} F_X(y) = F_X(x) - P_X\big(\mathrm{fr}(]-\infty, x])\big),$$

onde $\mathrm{fr}(A)$ denota a fronteira de $A \subset \mathbb{R}^d$;

e) F_X é contínua em $x \in \mathbb{R}^d$ sse $P_X\big(\mathrm{fr}(]-\infty, x])\big) = 0$;

f) Para $a \leq b$,

$$P_X\big(]a, b]\big) = \sum_{x \in V} \mathrm{sgn}(x) F_X(x),$$

onde V é o conjunto dos vértices de $]a, b]$, isto é, V o conjunto dos pontos da forma $(x_1, \ldots, x_d)$ com $x_i = a_i$ ou $x_i = b_i$, para $i = 1, \ldots, d$, e $sgn(x)$ representa o sinal do vértice x definido por $sgn(x) = (-1)^{\#\{i : x_i = a_i\}}$.

g) F_X carateriza P_X (isto é, $F_X = F_Y$ sse $X \sim Y$).

Dem: a) Para todo o $x \in \mathbb{R}^d$ temos $F_X(x) \in [0,1]$, sendo assim F_X não negativa e limitada. Por outro lado, F_X é crescente coordenada a coordenada uma vez que para $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_d \in \mathbb{R}$ e $x, y \in \mathbb{R}$ com $x < y$, temos

$$\begin{aligned} F_X(x_1, \ldots, x_{i-1}, x, x_{i+1}, \ldots, x_d) &= P(X_1 \leq x_1, \ldots, X_i \leq x, \ldots, X_d \leq x_d) \\ &\leq P(X_1 \leq x_1, \ldots, X_i \leq y, \ldots, X_d \leq x_d) \\ &= F_X(x_1, \ldots, x_{i-1}, y, x_{i+1}, \ldots, x_d). \end{aligned}$$

b) O limite $\lim_{x_i \to -\infty} F_X(x)$ pode ser obtido como no caso real. Relativamente ao segundo limite, para quaisquer sucessões $x_{1,n}, \ldots, x_{d,n} \to +\infty$, podemos provar que

$$\lim\,]-\infty, x_{1,n}] \times \cdots \times]-\infty, x_{d,n}] = \mathbb{R}^d.$$

Assim, pela continuidade da probabilidade concluímos que

$$\begin{aligned}\lim F_X(x_{1,n},\ldots,x_{d,n}) &= \lim P_X\big(]-\infty,x_{1,n}]\times\cdots\times]-\infty,x_{d,n}]\big)\\ &= P_X\big(\lim]-\infty,x_{1,n}]\times\cdots\times]-\infty,x_{d,n}]\big)\\ &= P_X(\mathbb{R}^d)=1.\end{aligned}$$

c) Sendo (y_n) uma qualquer sucessão em $\mathbb{R}^d$ com $y_n \to x$, $y_n \geq x$ e $y_n \neq x$, podemos provar que $\lim\,]-\infty,y_n] =]-\infty,x]$. Da continuidade de P_X obtemos

$$\begin{aligned}\lim F_X(y_n) &= \lim P_X\big(]-\infty,y_n]\big)\\ &= P_X\big(\lim\,]-\infty,y_n]\big)\\ &= P_X\big(]-\infty,x]\big)=F_X(x).\end{aligned}$$

d) Sendo (y_n) uma qualquer sucessão em $\mathbb{R}^d$ com $y_n \to x$, $y_n \leq x$ e $y_n \neq x$, podemos provar que $\lim\,]-\infty,y_n] =]-\infty,x[=]-\infty,x] - \mathrm{fr}(]-\infty,x])$. Basta agora usar a continuidade de P_X para concluir.

e) Sendo F_X contínua à direita e crescente coordenada a coordenada, a continuidade de F_X no ponto $x \in \mathbb{R}^d$ é equivalente à continuidade à esquerda nesse ponto, isto é, F_X é contínua em x sse $\lim_{y\to x, y\leq x} F_X(y) = F_X(x)$. Assim, o resultado enunciado decorre da alínea d).

f) Limitamo-nos a demonstrar o caso $d = 2$. Para $a = (a_1,a_2), b = (b_1,b_2) \in \mathbb{R}^2$, com $a_1 \leq b_1$ e $a_2 \leq b_2$, temos

$$]a_1,b_1]\times]a_2,b_2] =]-\infty,b_1]\times]-\infty,b_2] - \big(]-\infty,a_1]\times]-\infty,b_2]\cup]-\infty,b_1]\times]-\infty,a_2]\big),$$

Assim

$$\begin{aligned}&P_X\big(]a_1,b_1]\times]a_2,b_2]\big)\\ &= P_X\big(]-\infty,b_1]\times]-\infty,b_2]\big) - P_X\big(]-\infty,a_1]\times]-\infty,b_2]\cup]-\infty,b_1]\times]-\infty,a_2]\big)\\ &= P_X\big(]-\infty,b_1]\times]-\infty,b_2]\big) - P_X\big(]-\infty,a_1]\times]-\infty,b_2]\big)\\ &\quad - P_X\big(]-\infty,b_1]\times]-\infty,a_2]\big) + P_X\big(]-\infty,a_1]\times]-\infty,a_2]\big)\\ &= F_X(b_1,b_2) - F_X(a_1,b_2) - F_X(b_1,a_2) + F_X(a_1,a_2).\end{aligned}$$

g) Se $X \sim Y$, então $P_X = P_Y$ e assim $F_X(x) = P_X(]-\infty,x]) = P_Y(]-\infty,x]) = F_Y(x)$, para $x \in \mathbb{R}^d$. Reciprocamente, se $F_X = F_Y$, da alínea anterior concluímos $P_X(]a,b]) = P_Y(]a,b])$ para todo o retângulo $]a,b]$, com $a \leq b$. Podemos então concluir que $P_X(B) = P_Y(B)$, para todo o $B \in \mathscr{B}(\mathbb{R}^d)$, ou seja que $X \sim Y$. ■

Tal como podemos concluir da alínea e) do resultado anterior, no caso em que X é discreto, e contrariamente ao caso real, o suporte de X não pode ser identificado com os pontos de descontinuidade de F_X.

Mostramos a seguir que conhecida a função de distribuição dum vetor X, podemos facilmente obter a função de distribuição dum seu subvector.

Proposição 5.4.3. *Se $X = (X_1, \ldots, X_d)$ é um ve.a.r. com função de distribuição F_X, então, para todo o $\{i_1, \ldots, i_m\} \subset \{1, \ldots, d\}$, a função de distribuição de $(X_{i_1}, \ldots, X_{i_m})$ é dada por*

$$F(x_{i_1}, \ldots, x_{i_m}) = \lim F_X(x_1, \ldots, x_d),$$

para $(x_{i_1}, \ldots, x_{i_m}) \in \mathbb{R}^m$, onde o limite é tomado quando $x_j \to +\infty$, para todo o $j \in \{1, \ldots, d\} \setminus \{i_1, \ldots, i_m\}$. Em particular, a função de distribuição da margem X_i é dada por

$$F_{X_i}(x_i) = \lim F_X(x_1, \ldots, x_d),$$

para $x_i \in \mathbb{R}$, onde o limite é tomado quando $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_d \to +\infty$.

Dem: Para todo o $j \in \{1, \ldots, d\} \setminus \{i_1, \ldots, i_m\}$, seja $(x_{j,n})$ uma qualquer sucessão com $x_{j,n} \to +\infty$. Uma vez que

$$\begin{aligned}
&\{X_{i_1} \leq x_{i_1}, \ldots, X_{i_m} \leq x_{i_m}\} \\
&= \lim_{n\to\infty} \big\{X_1 \leq x_{1,n}, \ldots, X_{i_1} \leq x_{i_1}, \ldots, X_{i_m} \leq x_{i_m}, \ldots, X_d \leq x_{d,n}\big\},
\end{aligned}$$

pela continuidade de P_X concluímos que

$$\begin{aligned}
&F_{(X_{i_1},\ldots,X_{i_m})}(x_{i_1}, \ldots, x_{i_m}) \\
&= P_X\big(\{X_{i_1} \leq x_{i_1}, \ldots, X_{i_m} \leq x_{i_m}\}\big) \\
&= \lim_{n\to\infty} P_X\big(\{X_1 \leq x_{1,n}, \ldots, X_{i_1} \leq x_{i_1}, \ldots, X_{i_m} \leq x_{i_m}, \ldots, X_d \leq x_{d,n}\}\big) \\
&= \lim_{n\to\infty} F_X(x_{1,n}, \ldots, x_{i_1}, \ldots, x_{i_m}, \ldots, x_{d,n}).
\end{aligned}$$

■

Proposição 5.4.4. *Sejam X um ve.a.r. e $x \in \mathbb{R}^d$.*

a) Se X é discreto de suporte S_X então

$$F_X(x) = \sum_{y \in S_X : y \leq x} f_X(y);$$

b) Se X é absolutamente contínuo então F_X é contínua em $\mathbb{R}^d$ e

$$F_X(x) = \int_{]-\infty, x]} f_X(y)dy = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_d} f_X(y_1, \ldots, y_d)dy_1 \ldots dy_d.$$

Dem: Vimos já que para qualquer subconjunto $B \in \mathscr{B}(\mathbb{R}^d)$ se tem

$$P_X\big(B\big) = \sum_{y \in B} f_X(y),$$

quando X é discreto, e

$$P_X\big(B\big) = \int_B f_X(y)dy,$$

quando X é absolutamente contínuo. Tomando $B =]-\infty, x]$ e atendendo a que $P_X\big(]-\infty, x]\big) = F(x)$, obtemos o pretendido. Limitando-nos agora ao caso $d = 2$, vamos estabelecer a continuidade de F_X em (x_1, x_2) apenas no caso em que f_X é limitada, ou seja, quando $f_X(x, y) \leq M$, para todo o $(x, y) \in \mathbb{R}^2$. Para tal vamos usar a alínea e) da Proposição 5.4.2. Pela continuidade e monotonia de P_X temos

$$\begin{aligned}
&P_X\big(\mathrm{fr}(]-\infty, x_1]\times]-\infty, x_2])\big)\\
&= P_X\big(]-\infty, x_1]\times]-\infty, x_2]-]-\infty, x_1[\times]-\infty, x_2[\big)\\
&= P_X\big(]-\infty, x_1]\times]-\infty, x_2]\big) - P_X\big(]-\infty, x_1[\times]-\infty, x_2[\big)\\
&= \lim\Big(P_X\big(]x_1-n, x_1]\times]x_2-n, x_2]\big) - P_X\big(]x_1-n, x_1-1/n^2]\times]x_2-n, x_2-1/n^2]\big)\Big)\\
&= \lim\Big(P_X\big(]x_1-n, x_1]\times]x_2-n, x_2]-]x_1-n, x_1-1/n^2]\times]x_2-n, x_2-1/n^2]\big)\Big)\\
&\leq \lim\Big(P_X\big(]x_1-1/n^2, x_1]\times]x_2-n, x_2]\big) + P_X\big(]x_1-n, x_1-1/n^2]\times]x_2-n, x_2-1/n^2]\big)\Big),
\end{aligned}$$

onde

$$\begin{aligned}
&P_X\big(]x_1-1/n^2, x_1]\times]x_2-n, x_2]\big)\\
&= \int_{]x_1-n, x_1-1/n^2]\times]x_2-n, x_2-1/n^2]} f_X(x, y)dxdy \leq M\frac{1}{n^2}n = M\frac{1}{n}
\end{aligned}$$

e

$$\begin{aligned}
&P_X\big(]x_1-n, x_1-1/n^2]\times]x_2-n, x_2-1/n^2]\big)\\
&= \int_{]x_1-n, x_1]\times]x_2-1/n^2, x_2]} f_X(x, y)dxdy \leq M\frac{1}{n^2}n = M\frac{1}{n}.
\end{aligned}$$

■

Vamos agora generalizar a Proposição 3.3.8 ao caso multidimensional.

Proposição 5.4.5. *Se X é um ve.a.r. absolutamente contínuo então*

$$f_X(x_1, \ldots, x_d) = \frac{\partial^d F_X}{\partial x_{p(1)} \ldots \partial x_{p(d)}}(x_1, \ldots, x_d),$$

em quase todo o ponto $(x_1, \ldots, x_d) \in \mathbb{R}^d$ (isto é, o conjunto dos pontos $(x_1, \ldots, x_d) \in \mathbb{R}^d$ para os quais a igualdade anterior é falsa, tem medida de Lebesgue nula), para qualquer permutação p definida em $\{1, \ldots, d\}$.

Dem: Vamos considerar apenas o caso $d = 2$, e vamos assumir que $f_{(X_1,X_2)}$ é contínua em $\mathbb{R}^2$. Para $(x_1, x_2) \in \mathbb{R}^2$, tomemos $a < x_1$ e consideremos a decomposição

$$\begin{aligned} F_{(X_1,X_2)}(x_1, x_2) &= \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} f_{(X_1,X_2)}(u_1, u_2) du_2 du_1 \\ &= \int_{-\infty}^{a} \int_{-\infty}^{x_2} f_{(X_1,X_2)}(u_1, u_2) du_2 du_1 + \int_{a}^{x_1} \int_{-\infty}^{x_2} f_{(X_1,X_2)}(u_1, u_2) du_2 du_1. \end{aligned}$$

Usando o Teorema Fundamental do Cálculo, sendo $f_{(X_1,X_2)}$ é contínua em $\mathbb{R}^2$ sabemos que $x_1 \mapsto F_{(X_1,X_2)}(x_1, x_2)$ é derivável em $\mathbb{R}$ e

$$\frac{\partial F_X}{\partial x_1}(x_1, x_2) = \frac{d}{dx_1} \int_{a}^{x_1} \int_{-\infty}^{x_2} f_{(X_1,X_2)}(u_1, u_2) du_2 du_1 = \int_{-\infty}^{x_2} f_{(X_1,X_2)}(x_1, u_2) du_2.$$

Tomando $b < x_2$ e usando novamente o Teorema Fundamental do Cálculo, agora relativamente à variável x_2, temos

$$\frac{\partial F_X}{\partial x_1}(x_1, x_2) = \int_{-\infty}^{b} f_{(X_1,X_2)}(x_1, u_2) du_2 + \int_{b}^{x_2} f_{(X_1,X_2)}(x_1, u_2) du_2,$$

e

$$\frac{\partial^2 F_X}{\partial x_1 \partial x_2}(x_1, x_2) = \frac{d}{dx_2} \int_{b}^{x_2} f_{(X_1,X_2)}(x_1, u_2) du_2 = f_{(X_1,X_2)}(x_1, x_2).$$

De forma análoga se prova que

$$\frac{\partial^2 F_X}{\partial x_2 \partial x_1}(x_1, x_2) = f_{(X_1,X_2)}(x_1, x_2). \qquad \blacksquare$$

5.5 Independência de variáveis aleatórias

Já dissemos que as leis das margens de um vetor aleatório real $X = (X_1, \ldots, X_d)$ não caraterizam a distribuição de probabilidade desse vetor. No entanto, no caso que estudamos neste parágrafo a distribuição de probabilidade de X depende exclusivamente das distribuições marginais.

Definição 5.5.1. *As variáveis aleatórias reais $X_1, \ldots, X_d$ definidas sobre o espaço de probabilidade $(\Omega, \mathscr{A}, P)$ dizem-se* independentes, *se forem independentes os acontecimentos $\{X_1 \in B_1\}, \ldots, \{X_d \in B_d\}$ de $\mathscr{A}$, para todo o $B_1, \ldots, B_d \in \mathscr{B}(\mathbb{R})$.*

Apesar de nos limitar aqui à noção de independência de variáveis aleatórias reais, reparemos que de igual forma poderíamos definir o conceito de independência para vetores aleatórios reais.

Dada uma v.a.r. X interessamo-nos a seguir por uma função desta variável, ou seja, pela composição $g \circ X$, onde g é uma aplicação real de variável real. De acordo com a Definição 3.1.1, uma tal função será ela própria uma variável aleatória real quando

$$\{g \circ X \in B\} = \{\omega \in \Omega : g(X(\omega)) \in B\} \in \mathscr{A},$$

para todo o $B \in \mathscr{B}(\mathbb{R})$. Uma vez que

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

onde o conjunto $g^{-1}(B) = \{x \in \mathbb{R} : g(x) \in B\}$ é dito imagem recíproca de B por g, a composição $g \circ X$ será uma variável aleatória real no sentido da Definição 3.1.1, quando a aplicação g for tal que $g^{-1}(B) \in \mathscr{B}(\mathbb{R})$, para todo o $B \in \mathscr{B}(\mathbb{R})$. Dizemos neste caso que g é mensurável (relativamente às σ-álgebras de Borel). É possível provar que as aplicações contínuas de $\mathbb{R}$ em $\mathbb{R}$ são mensuráveis. Sendo as funções que habitualmente consideramos mensuráveis, não vamos preocupar-nos com esta questão.

Provamos no resultado seguinte que quaisquer funções mensuráveis de v.a.r. independentes são ainda v.a.r. independentes.

Proposição 5.5.2. *Se $X_1, \ldots, X_d$ são v.a.r. independentes e $g_1, \ldots, g_d$ são aplicações reais de variável real mensuráveis, então $g_1 \circ X_1, \ldots, g_d \circ X_d$ são v.a.r. independentes.*

Dem: Pretendemos provar que os acontecimentos $\{g_1 \circ X_1 \in B_1\}, \ldots, \{g_d \circ X_d \in B_d\}$ são independentes, para todo o $B_1, \ldots, B_d \in \mathscr{B}(\mathbb{R})$. Uma vez que

$$\{g_i \circ X_i \in B_i\} = \{\omega \in \Omega : X_i(\omega) \in g_i^{-1}(B_i)\} = \{X_i \in g_i^{-1}(B_i)\},$$

com $g_i^{-1}(B_i) \in \mathscr{B}(\mathbb{R})$, então a independência dos acontecimentos $\{g_1 \circ X_1 \in B_1\}, \ldots, \{g_d \circ X_d \in B_d\}$ decorre imediatamente da independência das v.a.r. $X_1, \ldots, X_d$. ■

Proposição 5.5.3. *As v.a.r. $X_1, \ldots, X_d$ são v.a.r. independentes sse*

$$P_{(X_1,\ldots,X_d)}(B_1 \times \cdots \times B_d) = P_{X_1}(B_1) \times \cdots \times P_{X_d}(B_d),$$

para todo o $B_1, \ldots, B_d \in \mathscr{B}(\mathbb{R})$. O resultado permanece válido se a igualdade anterior se verificar para todos os borelianos da forma $B_i =]a_i, b_i]$, ou da forma $B_i =]-\infty, b_i]$, com $a_i < b_i$, $i = 1, \ldots, d$.

Dem: Se $X_1, \ldots, X_d$ são v.a.r. independentes, então para $B_1, \ldots, B_d \in \mathscr{B}(\mathbb{R})$, temos

$$\begin{aligned} P_{(X_1,\ldots,X_d)}(B_1 \times \cdots \times B_d) &= P\big(X_1 \in B_1, \ldots, X_d \in B_d\big) \\ &= P\big(X_1 \in B_1\big) \times \cdots \times P\big(X_d \in B_d\big) \\ &= P_{X_1}(B_1) \times \cdots \times P_{X_d}(B_d). \end{aligned}$$

Para estabelecer a implicação recíproca, necessitamos de provar que, para todo o $B_1, \ldots, B_d \in \mathscr{B}(\mathbb{R})$, os acontecimentos $\{X_1 \in B_1\}, \ldots, \{X_d \in B_d\}$ são independentes. Para tal, vamos considerar um qualquer subconjunto finito de índices $t_1, \ldots, t_n \in \{1, \ldots, d\}$, com $n \in \mathbb{N}$. Por hipótese temos então

$$\begin{aligned} & P\big(X_{t_1} \in B_{t_1}, \ldots X_{t_n}, \in B_{t_n}\big) \\ & = P\big(X_1 \in \mathbb{R}, \ldots, X_{t_1} \in B_{t_1}, \ldots, X_{t_n}, \in B_{t_n}, \ldots, X_d \in \mathbb{R}\big) \end{aligned}$$

$$\begin{aligned} &= P_{(X_1,\ldots,X_d)}\big(\mathbb{R} \times \cdots \times B_{t_1} \times \cdots \times B_{t_n} \times \cdots \times \mathbb{R}\big) \\ &= P_{X_1}(\mathbb{R}) \times \cdots \times P_{X_{t_1}}(B_{t_1}) \times \cdots \times P_{X_{t_n}}(B_{t_n}) \times \cdots \times P_{X_d}(\mathbb{R}) \\ &= P_{X_{t_1}}(B_{t_1}) \times \cdots \times P_{X_{t_n}}(B_{t_n}) \\ &= P(X_{t_1} \in B_{t_1}) \times \cdots \times P(X_{t_n} \in B_{t_n}). \end{aligned}$$

De acordo com a Definição 2.5.3 podemos então concluir que os $\{X_1 \in B_1\}, \ldots, \{X_d \in B_d\}$ são independentes.

A parte final do resultado enunciado é consequência de $P_{(X_1,\ldots,X_d)}$ ser determinada pelos valores que toma nos borelianos da forma $]a_1, b_1] \times \cdots \times]a_d, b_d]$ ou $]-\infty, b_1] \times \cdots \times]-\infty, b_d]$. ■

Nos resultados seguintes apresentamos caraterizações da independência das margens dum vetor aleatório em termos da sua função de distribuição e, no caso deste ser discreto ou absolutamente contínuo, em termos da sua função de probabilidade ou da sua densidade de probabilidade.

Proposição 5.5.4. *Seja $(X_1, \ldots, X_d)$ um ve.a.r. com função de distribuição $F_{(X_1,\ldots,X_d)}$. As v.a.r $X_1, \ldots, X_n$ são independentes sse*

$$F_{(X_1,\ldots,X_d)}(x_1, \ldots, x_d) = F_{X_1}(x_1) \times \cdots \times F_{X_d}(x_d),$$

onde F_{X_i} denota a função de distribuição da variável aleatória X_i. Além disso, se $F_{(X_1,\ldots,X_d)}$ admite a fatorização

$$F_{(X_1,\ldots,X_d)}(x_1, \ldots, x_d) = G_1(x_1) \times \cdots \times G_d(x_d),$$

onde cada G_i é uma distribuição de probabilidade em $\mathbb{R}$ (isto é, G_i é crescente, contínua à direita, com $\lim_{x\to-\infty} G_i(x) = 0$ e $\lim_{x\to+\infty} G_i(x) = 1$), então $G_i = F_{X_i}$, para $i = 1, \ldots, d$, e as variáveis aleatórias $X_1, \ldots, X_d$ são independentes.

Dem: Se $X_1, \ldots, X_d$ são independentes, temos

$$\begin{aligned} F_{(X_1,\ldots,X_d)}(x_1, \ldots, x_d) &= P_{(X_1,\ldots,X_n)}\big(]-\infty, x_1] \times \cdots \times]-\infty, x_d]\big) \\ &= P_{X_1}(]-\infty, x_1]) \times \cdots \times P_{X_d}(]-\infty, x_d]) \\ &= F_{X_d}(x_d) \times \cdots \times F_{X_d}(x_d), \end{aligned}$$

para $(x_1, \ldots, x_d) \in \mathbb{R}^d$. Reciprocamente, se $F_{(X_1,\ldots,X_d)}(x_1, \ldots, x_d) = F_{X_d}(x_d) \times \cdots \times F_{X_d}(x_d)$, então $P_{(X_1,\ldots,X_n)}$ e $P_{X_1} \times \cdots \times P_{X_d}$ coincidem sobre os borelianos da forma $]-\infty, x_1] \times \cdots \times]-\infty, x_d]$. Atendendo ao teorema anterior, concluímos que as variáveis $X_1, \ldots, X_d$ são independentes. Suponhamos agora que $F_{(X_1,\ldots,X_d)} = \prod_{i=1}^{d} G_i$, onde cada G_i é uma distribuição de probabilidade em $\mathbb{R}$. Assim, para $i = 1, \ldots, d$, e $x_i \in \mathbb{R}$,

$$F_{X_i}(x_i) = \lim_{\substack{x_j \to +\infty \\ j \neq i}} F_{(X_1,\ldots,X_d)}(x_1, \ldots, x_d) = \lim_{\substack{x_j \to +\infty \\ j \neq i}} \prod_{k=1}^{d} G_k(x_k) = G_i(x_i).$$

Além disso, $F_{(X_1,\ldots,X_d)} = \prod_{i=1}^{d} F_i$, o que pela primeira parte da demonstração é equivalente à independência de $X_1, \ldots, X_d$. ■

Proposição 5.5.5. *Seja $(X_1, \ldots, X_d)$ um ve.a.r. discreto com função de probabilidade $f_{(X_1,\ldots,X_d)}$. As v.a.r $X_1, \ldots, X_n$ são independentes sse*

$$f_{(X_1,\ldots,X_d)}(x_1, \ldots, x_d) = f_{X_1}(x_1) \times \cdots \times f_{X_d}(x_d),$$

onde f_{X_i} denota a função de probabilidade da variável aleatória X_i. Além disso, se $f_{(X_1,\ldots,X_d)}$ admite a fatorização

$$f_{(X_1,\ldots,X_d)}(x_1, \ldots, x_d) = g_1(x_1) \times \cdots \times g_d(x_d),$$

onde cada g_i é uma função de probabilidade em $\mathbb{R}$ (isto é, g_i é não negativa, com $\sum_{x \in S_i} g_i(x) = 1$, para algum conjunto finito ou infinito numerável S_i), então $g_i = f_{X_i}$, para $i = 1, \ldots, d$, e as variáveis aleatórias $X_1, \ldots, X_d$ são independentes.

Dem: Se $X_1, \ldots, X_d$ são independentes, temos

$$\begin{aligned} f_{(X_1,\ldots,X_d)}(x_1, \ldots, x_d) &= P\big(X_1 = x_1, \ldots, X_d = x_d\big) \\ &= P(X_1 = x_1) \times \cdots \times P(X_d = x_d) \\ &= f_{X_d}(x_d) \times \cdots \times f_{X_d}(x_d), \end{aligned}$$

para $(x_1, \ldots, x_d) \in \mathbb{R}^d$. Reciprocamente, se $f_{(X_1,\ldots,X_d)}(x_1, \ldots, x_d) = f_{X_1}(x_1) \ldots f_{X_d}(x_d)$, concluímos que

$$\begin{aligned} P_{(X_1,\ldots,X_d)}\big(]a_1, b_1] \times \cdots \times]a_d, b_d]\big) &= \sum_{(x_1,\ldots,x_d) \in]a,b]} f_{(X_1,\ldots,X_d)}(x_1, \ldots, x_d) \\ &= \sum_{(x_1,\ldots,x_d) \in]a,b]} f_{X_1}(x_1) \times \cdots \times f_{X_d}(x_d) \\ &= \sum_{x_1 \in\,]a_1,b_1]} f_{X_1}(x_1) \times \cdots \times \sum_{x_d \in\,]a_d,b_d]} f_{X_d}(x_d) \\ &= P_{X_1}(]a_1, b_1]) \times \cdots \times P_{X_d}(]a_d, b_d]). \end{aligned}$$

Da Proposição 5.5.3 concluímos que $X_1, \ldots, X_d$ são independentes. Suponhamos agora que $f_{(X_1,\ldots,X_d)} = \prod_{i=1}^{d} g_i$, onde cada g_i é uma função de probabilidade em $\mathbb{R}$. Então para $i = 1, \ldots, d$, e $x_i \in \mathbb{R}$,

$$\begin{aligned} f_{X_i}(x_i) &= \sum_{(x_1,\ldots,x_{i_1-1},x_{i_1+1},\ldots,x_d)\in\mathbb{R}^{d-1}} f_{(X_1,\ldots,X_d)}(x_1,\ldots,x_d) \\ &= \sum_{(x_1,\ldots,x_{i_1-1},x_{i_1+1},\ldots,x_d)\in\mathbb{R}^{d-1}} g_1(x_1)\times\cdots\times g_d(x_d) \\ &= g_i(x_i), \end{aligned}$$

e $f_{(X_1,\ldots,X_d)} = \prod_{i=1}^{d} f_{X_i}$, o que pela primeira parte da demonstração é equivalente à independência de $X_1, \ldots, X_d$. ■

Proposição 5.5.6. *Seja $(X_1, \ldots, X_d)$ um ve.a.r. absolutamente contínuo com densidade de probabilidade $f_{(X_1,\ldots,X_d)}$. As v.a.r. $X_1, \ldots, X_n$ são independentes sse*

$$f_{(X_1,\ldots,X_d)}(x_1,\ldots,x_d) = f_{X_1}(x_1)\times\cdots\times f_{X_d}(x_d),$$

onde f_{X_i} denota a densidade de probabilidade da variável aleatória X_i. Além disso, se $f_{(X_1,\ldots,X_d)}$ admite a fatorização

$$f_{(X_1,\ldots,X_d)}(x_1,\ldots,x_d) = g_1(x_1)\times\cdots\times g_d(x_d),$$

onde cada g_i é uma densidade de probabilidade em $\mathbb{R}$ (isto é, g_i é não negativa, com $\int_{\mathbb{R}} g_i(x)dx = 1$), então $g_i = f_{X_i}$, para $i = 1, \ldots, d$, e as variáveis aleatórias $X_1, \ldots, X_d$ são independentes.

Dem: Se $X_1, \ldots, X_n$ são independentes então para $a \le b$ temos

$$\begin{aligned} P_{(X_1,\ldots,X_d)}\big(]a,b]\big) &= P_{X_1}(]a_1,b_1]) \times\cdots\times P_{X_d}(]a_1,b_1]) \\ &= \int_{a_1}^{b_1} f_{X_1}(y_1)dy_1 \times\cdots\times \int_{a_d}^{b_d} f_{X_d}(y_d)dy_d \\ &= \int_{]a,b]} f_{X_1}(y_1)\ldots f_{X_d}(y_d)dy_1\ldots dy_d. \end{aligned}$$

Assim $f_{X_1}(x_1)\ldots f_{X_d}(x_d)$ é uma versão da densidade de probabilidade de $(X_1, \ldots, X_d)$. Reciprocamente, se $f_{(X_1,\ldots,X_d)}(x_1,\ldots,x_d) = f_{X_1}(x_1)\ldots f_{X_d}(x_d)$, concluímos que

$$\begin{aligned} P_{(X_1,\ldots,X_d)}\big(]a_1,b_1]\times\cdots\times]a_d,b_d]\big) &= \int_{]a,b]} f_{(X_1,\ldots,X_d)}(x_1,\ldots,x_d)dx_1\ldots dx_d \\ &= \int_{a_1}^{b_1} f_{X_1}(x_1)dy_1 \times\cdots\times \int_{a_d}^{b_d} f_{X_d}(x_d)dx_d \\ &= P_{X_1}(]a_1,b_1]) \times\cdots\times P_{X_d}(]a_d,b_d]). \end{aligned}$$

Da Proposição 5.5.3 concluímos que $X_1, \ldots, X_d$ são independentes. Suponhamos agora que $f_{(X_1,\ldots,X_d)} = \prod_{i=1}^{d} g_i$, onde cada g_i é uma densidade de probabilidade em $\mathbb{R}$. Então para $i = 1, \ldots, d$, e $x_i \in \mathbb{R}$,

$$\begin{aligned} f_{X_i}(x_i) &= \int f_{(X_1,\ldots,X_d)}(x_1, \ldots, x_d) dx_1 \ldots dx_{i-1} dx_{i+1} dx_d \\ &= \int g_1(x_1) \ldots g_d(x_d) dx_1 \ldots dx_{i-1} dx_{i+1} dx_d \\ &= g_i(x_i), \end{aligned}$$

e $f_{(X_1,\ldots,X_d)} = \prod_{i=1}^{d} f_{X_i}$, o que pela primeira parte da demonstração é equivalente à independência de $X_1, \ldots, X_d$. ■

Exemplo 5.5.7. Vimos no Exemplo 5.3.6 que o vetor (X, Y) em $\mathbb{R}^2$ aí considerado tinha por densidade

$$f_{(X,Y)}(x, y) = \frac{1}{2\sqrt{y}} \mathbb{1}_{]0,1[}(x) \mathbb{1}_{]0,1[}(y).$$

Uma vez que $g(x) = \mathbb{1}_{]0,1[}(x)$ e $h(y) = \frac{1}{2\sqrt{y}} \mathbb{1}_{]0,1[}(y)$ são densidades de probabilidade em $\mathbb{R}$, da Proposição 5.5.6 podemos imediatamente concluir que X e Y são independentes com $f_X = g$ e $f_Y = h$.

5.6 Distribuições condicionais

Na Secção 2.4 estudámos a noção de probabilidade condicionada. Aí vimos que se os acontecimentos aleatórios A e B não são independentes, então a probabilidade $P(B)$ do acontecimento B é alterada se tivermos informação adicional de que o acontecimento A se realizou. Essa probabilidade foi denotada por $P(B|A)$ para a distinguir da probabilidade $P(B)$ do acontecimento B quando nada sabemos sobre a ocorrência, ou não, do acontecimento A. Esse contexto é agora extendido ao caso em que dispomos de um vetor aleatório (X, Y) com valores em $\mathbb{R}^{p+q}$ e sabemos que foi observado o valor x de X, isto é, sabemos que ocorreu o acontecimento $\{X = x\}$. Pretendemos agora saber de que forma esta informação altera a distribuição de probabilidade de Y.

5.6.1 O caso discreto

Para responder à questão anterior, começamos por considerar o caso em que o ve.a.r. (X, Y) é discreto e vamos denotar por $f_X(x) = P(X = x)$, $f_Y(y) = P(Y = y)$ e $f_{(X,Y)}(x, y) = P(X = x, Y = y)$ as funções de probabilidade de X, Y e (X, Y), respetivamente. Denotando por $f_{Y|X=x}$ a função de probabilidade do vetor Y quando

sabemos que ocorreu o acontecimento $\{X = x\}$, onde $x \in \mathbb{R}^p$ é tal que $f_X(x) > 0$, para $y \in \mathbb{R}^q$ temos então

$$f_{Y|X=x}(y) = P(Y = y|X = x) = \frac{P(X = x, Y = y)}{P(X = x)} = \frac{f_{(X,Y)}(x, y)}{f_X(x)}. \tag{5.6.1}$$

Para $x \in \mathbb{R}^p$ com $f_X(x) > 0$, é simples verificar que a função $f_{Y|X=x}$ é uma função de probabilidade em $\mathbb{R}^q$. Com efeito, $f_{Y|X=x}$ é não negativa e, pela Proposição 5.3.2,

$$\sum_{y\in\mathbb{R}^q} f_{Y|X=x}(y) = \frac{\sum_{y\in\mathbb{R}^q} f_{(X,Y)}(x, y)}{f_X(x)} = \frac{f_X(x)}{f_X(x)} = 1.$$

A função de probabilidade $f_{Y|X=x}$ é dita função de probabilidade condicional de Y dado $X = x$. Sendo uma função de probabilidade, pode ser interpretada como a função de probabilidade dum ve.a.r. discreto que vamos denotar por $Y|X = x$. Para $B \in \mathscr{B}(\mathbb{R}^q)$, a distribuição condicional de Y dado $X = x$ é naturalmente definida por

$$P_{Y|X=x}(B) = \sum_{y\in B} f_{Y|X=x}(y).$$

Da igualdade (5.6.1) e da Proposição 5.5.5 concluímos que se a função de probabilidade condicional de Y dado $X = x$ é igual a f_Y quando e só quando os ve.a. X e Y são independentes.

5.6.2 O caso absolutamente contínuo

Suponhamos agora que o vetor (X, Y) com valores em $\mathbb{R}^{p+q}$ é absolutamente contínuo e denotemos por f_X, f_Y e $f_{(X,Y)}$ as densidades de probabilidade de X, Y e (X, Y), respetivamente. Para $x \in \mathbb{R}^p$ com $f_X(x) > 0$, é simples verificar que a função definida, para $y \in \mathbb{R}^q$, por

$$f_{Y|X=x}(y) = \frac{f_{(X,Y)}(x, y)}{f_X(x)}, \tag{5.6.2}$$

é uma densidade de probabilidade em $\mathbb{R}^q$. Com efeito, $f_{Y|X=x}$ é não negativa e, pela Proposição 5.3.5,

$$\int_{\mathbb{R}^q} f_{Y|X=x}(y)dy = \frac{\int_{\mathbb{R}^q} f_{(X,Y)}(x, y)dy}{f_X(x)} = \frac{f_X(x)}{f_X(x)} = 1.$$

A densidade de probabilidade $f_{Y|X=x}$ é dita densidade condicional de Y dado $X = x$. Sendo uma densidade de probabilidade, pode ser interpretada como a densidade de probabilidade dum ve.a.r. absolutamente contínuo que vamos também denotar por $Y|X = x$. Para $B \in \mathscr{B}(\mathbb{R}^q)$, a distribuição condicional de Y dado $X = x$ é naturalmente definida por

$$P_{Y|X=x}(B) = \int_B f_{Y|X=x}(y)dy.$$

Tal como no caso discreto, da igualdade (5.6.2) e da Proposição 5.5.6 concluímos que a densidade condicional de Y dado $X = x$ é igual a f_Y quando e só quando os ve.a. X e Y são independentes.

A noção de distribuição condicional pode ser também usada para obter a distribuição de probabilidade dum vetor $(X, Y) \in \mathbb{R}^{p+q}$ à custa da distribuição de X e da distibuição de $Y|X = x$. Com efeito, no caso discreto ou absolutamente contínuo sabemos que

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

De forma análoga, a distribuição de (X, Y) pode ser obtida a partir da distribuição de Y e da distibuição de $X|Y = y$:

$$f_{(X,Y)}(x, y) = f_Y(y) f_{X|Y=y}(x).$$

Exemplo 5.6.3. Suponhamos que X é uma v.a.r. com uma distribuição normal standard, $X \sim N(0, 1)$, e que observado um valor x da variável X, a variável Y é definida como sendo uma v.a.r. normal com média x e variância 1, ou seja, $Y|X = x \sim N(x, 1)$. Determinemos a distribuição de probabilidade do ve.a. (X, Y). Atendendo a que

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-x^2/2}$$

e

$$f_{Y|X=x}(y) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-(y-x)^2/2},$$

temos

$$f_{(X,Y)}(x, y) = f_X(x) f_{Y|X=x}(y) = \frac{1}{2\pi} \exp\big(-(2x^2 - 2xy + y^2)/2\big).$$

A partir da distribuição de (X, Y) podemos determinar a distribuição (não condicional) de Y. Com efeito,

$$f_Y(y) = \int_{\mathbb{R}} f_{(X,Y)}(x, y) dx = \frac{1}{\sqrt{4\pi}} \mathrm{e}^{-y^2/4}.$$

Concluímos assim que $Y \sim N(0, 2)$.

5.6.3 Esperança condicional

Da mesma maneira que definimos a esperança de uma variável aleatória real Y, podemos também considerar a noção de esperança condicional de Y dado $X = x$.

Definição 5.6.4. *Sejam (X, Y) um ve.a.r. discreto ou absolutamente contínuo com valores em $\mathbb{R}^{d+1}$ onde $Y \in \mathscr{L}^1$.*

a) Se (X,Y) é discreto, a esperança condicional de Y dado $X=x$, *com $f_X(x)>0$, é dada por*

$$E(Y|X=x)=\sum_{y\in\mathbb{R}} y f_{Y|X=x}(y).$$

b) Se (X,Y) é absolutamente contínuo, a esperança condicional de Y dado $X=x$, *com $f_X(x)>0$, é dada por*

$$E(Y|X=x)=\int_{\mathbb{R}} y f_{Y|X=x}(y)dy.$$

Todos os resultados obtidos para a esperança matemática não condicional (ver Secção 4.1.1) são também válidos para a esperança condicional uma vez que as definições respetivas são essencialmente as mesmas, apenas sendo substituídas as funções e as densidades de probabilidade não condicionais pelas correspondentes funções e densidades de probabilidade condicionais.

Decorre da respetiva definição que esperança condicional $E(Y|X=x)$ é uma função de x. A composição desta função, que denotamos por $h(x)=E(Y|X=x)$ com a variável aleatória X, isto é, $h(X)$, é uma variável aleatória que a que chamamos esperança condicional de Y dado X. Uma tal variável é denotada por $E(Y|X)$. Os resultados seguintes estabelecem que esta variável é constantemente igual a $E(Y)$ quando X e Y são independentes, e que a esperança matemática de $E(Y|X)$ é igual à esperança matemática de Y, o que permite uma fórmula alternativa de cálculo de $E(Y)$ a partir de $E(Y|X=x)$ e da distribuição de probabilidade de X.

Proposição 5.6.5. *Se (X,Y) é um ve.a.r. discreto ou absolutamente contínuo com valores em $\mathbb{R}^{d+1}$ onde $Y\in\mathscr{L}^1$, e X e Y são independentes, então*

$$E(Y|X)=E(Y).$$

Dem: Basta ter em conta que se X e Y são independentes então $f_{Y|X=x}=f_Y$, para todo o $x\in\mathbb{R}^d$ com $f_X(x)>0$, de onde resulta

$$E(Y|X=x)=\sum_{y\in\mathbb{R}} y f_{Y|X=x}(y)=\sum_{y\in\mathbb{R}} y f_Y(y)=E(Y).$$

De forma análoga se procede no caso do vetor (X,Y) ser absolutamente contínuo. ■

Proposição 5.6.6. *Se (X,Y) é um ve.a.r. discreto ou absolutamente contínuo com valores em $\mathbb{R}^{d+1}$ onde $Y\in\mathscr{L}^1$, então*

$$E(Y)=E\big(E(Y|X)\big).$$

Dem: Supondo que (X,Y) é discreto e usando a Proposição 4.1.10 com $h(x) = E(Y|X=x)$ temos

$$\begin{aligned} E\big(E(Y|X)\big) = E(h(X)) &= \sum_{x\in\mathbb{R}^d} h(x) f_X(x) \\ &= \sum_{x\in\mathbb{R}^d} E(Y|X=x) f_X(x) \\ &= \sum_{x\in\mathbb{R}^d} \sum_{y\in\mathbb{R}} y f_{Y|X=x}(y) f_X(x) \\ &= \sum_{y\in\mathbb{R}} y \sum_{x\in\mathbb{R}^d} f_{(X,Y)}(x,y) \\ &= \sum_{y\in\mathbb{R}} y f_Y(y) = E(Y). \end{aligned}$$

A demonstração é análoga quando (X,Y) é absolutamente contínuo, bastando substituir os somatórios $\sum_{x\in\mathbb{R}^d}$ e $\sum_{y\in\mathbb{R}}$ pelos integrais $\int_{\mathbb{R}^d} dx$ e $\sum_{\mathbb{R}} dy$, respetivamente. ■

Exemplo 5.6.7. Retomamos agora o Exemplo 5.6.3 para ilustrar o processo alternativo para o cálculo da esperança matemática da variável Y com base na proposição anterior. Uma vez que $Y|X=x \sim N(x,1)$, temos naturalmente

$$E(Y|X=x) = x.$$

Usando agora a Proposição 5.6.6

$$E(Y) = \int_{\mathbb{R}} E(Y|X=x) f_X(x) dx = \int_{\mathbb{R}} x f_X(x) dx = E(X) = 0,$$

pois $X \sim N(0,1)$. A Proposição 5.6.6 pode também ser usada para calcular $E(Y^2)$ (com Y^2 no lugar de Y). Atendendo a que $Y|X=x \sim N(x,1)$, temos

$$E(Y^2|X=x) = 1 + x^2,$$

e assim

$$\begin{aligned} E(Y^2) = E(E(Y^2|X)) &= \int_{\mathbb{R}} E(Y^2|X=x) f_X(x) dx \\ &= \int_{\mathbb{R}} (1+x^2) f_X(x) dx = 1 + E(X^2) = 2. \end{aligned}$$

Assumindo que $Y \in \mathscr{L}^2$, vamos agora estabelecer uma interessante propriedade da esperança condicional. Vamos mostrar que a esperança condicional $E(Y|X)$ é a melhor aproximação de Y que é função de X, no sentido da norma $||\cdot||_2$. Dito de outro modo, $E(Y|X)$ é a projeção ortogonal de $Y \in \mathscr{L}^2$ no subespaço vetorial de $\mathscr{L}^2$ das v.a.r. que são função de X.

Proposição 5.6.8. *Seja (X,Y) um ve.a. discreto ou absolutamente contínuo com valores em $\mathbb{R}^{d+1}$. Se $Y \in \mathscr{L}^2$ então*

$$||Y - E(Y|X)||_2 \leq ||Y - h(X)||_2,$$

para toda a função $h : \mathbb{R}^d \to \mathbb{R}$ com $h(X) \in \mathscr{L}^2$.

Dem: Vamos admitir que (X,Y) é um vetor discreto, mas os mesmos argumentos são válidos no caso contínuo. Temos

$$\begin{aligned}
||Y - h(X)||_2^2 &= E\big(Y - h(X)\big)^2 \\
&= \sum_{x\in\mathbb{R}^d}\sum_{y\in\mathbb{R}} (y - h(x))^2 f_{(X,Y)}(x,y) \\
&= \sum_{x\in\mathbb{R}^d}\sum_{y\in\mathbb{R}} (y - h(x))^2 f_X(x) f_{Y|X=x}(y) \\
&= \sum_{x\in\mathbb{R}^d} f_X(x) \sum_{y\in\mathbb{R}} (y - h(x))^2 f_{Y|X=x}(y),
\end{aligned}$$

onde, para todo o $x \in \mathbb{R}^d$, a função $\mu \mapsto \sum_{y\in\mathbb{R}} (y-\mu)^2 f_{Y|X=x}(y)$ tem mínimo absoluto no ponto $\mu = \sum_{y\in\mathbb{R}} y f_{Y|X=x}(y)dy = E(Y|X=x)$, ou seja,

$$\sum_{y\in\mathbb{R}} (y-\mu)^2 f_{Y|X=x}(y) \geq \sum_{y\in\mathbb{R}} (y - E(Y|X=x))^2 f_{Y|X=x}(y),$$

para todo o $\mu \in \mathbb{R}$. Assim,

$$\begin{aligned}
||Y - h(X)||_2^2 &\geq \sum_{x\in\mathbb{R}^d} f_X(x) \sum_{y\in\mathbb{R}} (y - E(Y|X=x))^2 f_{Y|X=x}(y) \\
&= \sum_{x\in\mathbb{R}^d}\sum_{y\in\mathbb{R}} (y - E(Y|X=x))^2 f_{(X,Y)}(x,y) \\
&= E\big(Y - E(Y|X)\big)^2 \\
&= ||Y - E(Y|X)||_2^2.
\end{aligned}$$

■

5.7 Covariância e correlação

Se (X,Y) é um vetor aleatório em $\mathbb{R}^2$, e variáveis X e Y possuem momentos de segunda ordem, isto é, $X, Y \in \mathscr{L}^2$, os parâmetros de resumo das distribuições de X e de Y que estudámos até ao momento, $E(X)$, $E(Y)$, $\mathrm{Var}(X)$ e $\mathrm{Var}(Y)$, são também parâmetros de resumo da distribuição de (X,Y). Contrariamente a tais parâmetros que incidem unicamente sobre as distribuições marginais do vetor (X,Y), vamos neste parágrafo

estudar um outro parâmetro de resumo da distribuição de (X, Y) que, como veremos, nos dá uma medida da dependência afim entre as variáveis X e Y.

Sabemos já que o espaço $\mathscr{L}^2$ das v.a.r. com momento de segunda ordem finito é um espaço vetorial. Identificando variáveis aleatórias que são quase certamente iguais, $\mathscr{L}^2$ é um espaço vetorial com produto interno definido por $\langle X, Y\rangle = E(XY)$. Com efeito, das propriedades da esperança matemática, para quaisquer $X, Y, Z \in \mathscr{L}^2$ e $\alpha, \beta \in \mathbb{R}$, temos:

PI.1) $\langle X, X\rangle = 0$ sse $X = 0$ q.c.;
PI.2) $\langle X, Y\rangle = \langle Y, X\rangle$;
PI.3) $\langle \alpha X + \beta Y, Z\rangle = \alpha\langle X, Z\rangle + \beta\langle Y, Z\rangle$.

Reparemos que a esperança matemática $E(XY)$ existe para $X, Y \in \mathscr{L}^2$, uma vez que $|XY| \leq (X^2+Y^2)/2$ (ver alínea d) da Proposição 4.1.8). Sendo $\mathscr{L}^2$ um espaço vetorial com produto interno, é válida a propriedade seguinte conhecida como desigualdade de Cauchy-Schwarz.

Proposição 5.7.1 (Desigualdade de Cauchy-Schwarz)**.** *Se $X, Y \in \mathscr{L}^2$ então $XY \in \mathscr{L}^1$ e*

$$|E(XY)| \leq \sqrt{E(X^2)}\,\sqrt{E(Y^2)}.$$

Além disso, tem-se a igualdade sse X é quase certamente um múltiplo escalar de Y.

Dem: A desigualdade é verdadeira se $E(Y^2) = 0$. Suponhamos então que $E(Y^2) \neq 0$ e seja $\alpha = E(XY)/E(Y^2)$. Consideremos a v.a.r. definida por $Z = X - \alpha Y$ e comecemos por provar que $E(ZY) = 0$ (isto é, Z e Y são ortogonais). Com efeito,

$$E(ZY) = E\big((X - \alpha Y)Y\big) = E(XY) - \alpha E(Y^2) = 0.$$

Assim,

$$\begin{aligned} E(X^2) &= E\big((Z + \alpha Y)^2\big) = E(Z^2) + 2\alpha E(ZY) + \alpha^2 E(Y^2) \\ &= E(Z^2) + \frac{\big(E(XY)\big)^2}{E(Y^2)} \geq \frac{\big(E(XY)\big)^2}{E(Y^2)}, \end{aligned}$$

de onde concluímos que

$$|E(XY)| \leq \sqrt{E(X^2)}\,\sqrt{E(Y^2)}.$$

Além disso, sendo válida a igualdade $|E(XY)| = \sqrt{E(X^2)}\,\sqrt{E(Y^2)}$, temos $E(Z^2) = 0$, ou seja, $Z = 0$ q.c., isto é, $X = \alpha Y$ q.c., com $\alpha = E(XY)/E(Y^2)$. Reciprocamente, suponhamos que $X = \lambda Y$ q.c. para algum $\lambda \in \mathbb{R}$. Neste caso $|E(XY)| = |\lambda| E(Y^2)$ e $\sqrt{E(X^2)}\,\sqrt{E(Y^2)} = |\lambda| E(Y^2)$, sendo válida a igualdade. ■

Sempre que X e Y não sejam nulas, da Proposição 5.7.1 concluímos que o quociente $E(XY)/\sqrt{E(X^2)}\sqrt{E(Y^2)} \in [-1,1]$ surge como uma medida natural da dependência linear entre X e Y. Se pretendemos avaliar não só a dependência linear mas também a dependência afim entre X e Y, o coeficiente anterior deixa de ser indicado para o efeito.

Definição 5.7.2. *Se $X, Y \in \mathscr{L}^2$, chamamos* covariância de *(X,Y) ao número real*

$$\mathrm{Cov}(X,Y) = E\big((X - E(X))(Y - E(Y))\big).$$

Se além disso X e Y são de variância não nula, chamamos coeficiente de correlação (linear) de *(X,Y) ao número real*

$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}} = \frac{\mathrm{Cov}(X,Y)}{\sigma(X)\sigma(Y)}.$$

Contrariamente à covariância, o coeficiente de correlação não depende das unidades em que as variáveis X e Y são expressas.

Listam-se na proposição seguinte as principais propriedades da covariância de (X,Y).

Proposição 5.7.3. *Para $X, Y, Z \in \mathscr{L}^2$ e $a, b \in \mathbb{R}$ temos:*

a) $\mathrm{Cov}(X,Y) = E(XY) - E(X)E(Y)$;

b) $\mathrm{Var}(X) = \mathrm{Cov}(X,X)$;

c) $\mathrm{Cov}(aX + bY, Z) = a\mathrm{Cov}(X,Z) + b\mathrm{Cov}(Y,Z)$.

As propriedades seguintes reforçam a interpretação do coeficiente de correlação de (X,Y) como uma medida da dependência afim entre as variáveis X e Y. Nos casos limite em que $\rho(X,Y) = \pm 1$ a distribuição de (X,Y) está concentrada numa reta.

Proposição 5.7.4. *Se $X, Y \in \mathscr{L}^2$ são de variância não nula, então*

$$|\rho(X,Y)| \leq 1.$$

Além disso, $\rho(X,Y) = \pm 1$ sse existem $a, b, c \in \mathbb{R}$, com $ab \neq 0$, tais que

$$aX + bY + c = 0, \text{ q.c.}$$

Dem: A primeira afirmação decorre da desigualdade de Cauchy-Schwarz uma vez que

$$\begin{aligned} |\mathrm{Cov}(X,Y)| &= \big|E\big((X - E(X))(Y - E(Y))\big)\big| \\ &\leq \sqrt{E(X - E(X))^2}\sqrt{E(Y - E(Y))^2} \\ &= \sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}. \end{aligned}$$

Se $\rho(X,Y) = \pm 1$, também da desigualdade de Cauchy-Schwarz sabemos que existe $\lambda \neq 0$ tal que $X - E(X) = \lambda(Y - E(Y))$ *q.c.* (se fosse $\lambda = 0$, teríamos $X = E(X)$ q.c., tendo X variância nula). Assim, $aX + bY + c = 0$ q.c. com $a = 1$, $b = -\lambda$ e $c = \lambda E(Y) - E(X)$. Reciprocamente, se $Y = -b^{-1}aX - b^{-1}c$ q.c. temos

$$\rho(X,Y) = \rho\big(X, -b^{-1}aX - b^{-1}c\big) = \frac{-b^{-1}a}{|b^{-1}a|} = \pm 1. \qquad \blacksquare$$

Sempre que $\mathrm{Cov}(X,Y) = 0$ diremos que as variáveis X e Y são não correlacionadas. Vamos provar a seguir que variáveis independentes em $\mathscr{L}^2$ são necessariamente não correlacionadas. Antes de enunciarmos um tal resultado, começamos por generalizar a Proposição 4.1.10 ao caso multidimensional.

Proposição 5.7.5. *Seja $(X_1, \ldots, X_d)$ um ve.a.r. discreto ou absolutamente contínuo e h uma função real definida em $\mathbb{R}^d$.*

a) Se $(X_1, \ldots, X_d)$ é discreto então

$$E(h(X_1, \ldots, X_d)) = \sum_{(x_1,\ldots,x_d)\in\mathbb{R}^d} h(x_1, \ldots, x_d) f_{(X_1,\ldots,X_d)}(x_1, \ldots, x_d),$$

desde que $\sum_{(x_1,\ldots,x_d)\in\mathbb{R}^d} |h(x_1, \ldots, x_d)| f_{(X_1,\ldots,X_d)}(x_1, \ldots, x_d) < \infty$.

b) Se $(X_1, \ldots, X_d)$ é absolutamente contínua então

$$E(h(X_1, \ldots, X_d)) = \int_{\mathbb{R}^d} h(x_1, \ldots, x_d) f_{(X_1,\ldots,X_d)}(x_1, \ldots, x_d)\, dxdy,$$

desde que $\int_{\mathbb{R}^d} |h(x_1, \ldots, x_d)| f_{(X_1,\ldots,X_d)}(x_1, \ldots, x_d)\, dx_1 \ldots dx_d < \infty$.

Proposição 5.7.6. *Se $X, Y \in \mathscr{L}^1$ são v.a.r. independentes, então $XY \in \mathscr{L}^1$ e*

$$E(XY) = E(X)E(Y).$$

Dem: Analisemos em primeiro lugar o caso em que (X,Y) é discreto. Comecemos por demonstrar que se $X, Y \in \mathscr{L}^1$ (isto é, X e Y possuem esperança matemática), então $XY \in \mathscr{L}^1$ (isto é, XY possui esperança matemática). Com efeito

$$\sum_{(x,y)\in\mathbb{R}^2} |xy| f_{(X,Y)}(x,y) = \sum_{(x,y)\in\mathbb{R}^2} |xy| f_X(x) f_Y(y) = \sum_{x\in\mathbb{R}} |x| f_X(x) \sum_{y\in\mathbb{R}} |y| f_Y(y) < \infty.$$

Além disso,

$$E(XY) = \sum_{(x,y)\in\mathbb{R}^2} xy f_{(X,Y)}(x,y) = \sum_{x\in\mathbb{R}} x f_X(x) \sum_{y\in\mathbb{R}} y f_Y(y) = E(X)E(Y).$$

No caso em que (X,Y) é absolutamente contínuo, a demonstração é muito semelhante: $E(XY)$ existe pois

$$\int_{\mathbb{R}^2}|xy|f_{(X,Y)}(x,y)dxdy=\int_{\mathbb{R}^2}|xy|f_X(x)f_Y(y)dxdy=\int_{\mathbb{R}}|x|f_X(x)dxdy\int_{\mathbb{R}}|y|f_Y(y)dy<\infty,$$

e

$$E(XY)=\int_{\mathbb{R}^2}xyf_{(X,Y)}(x,y)dxdy=\int_{\mathbb{R}}xf_X(x)dx\int_{\mathbb{R}}yf_Y(y)dy=E(X)E(Y). \quad \blacksquare$$

Corolário 5.7.7. *Se $X,Y\in\mathscr{L}^2$ são v.a.r. independentes, então X e Y são não correlacionadas.*

No entanto, podemos ter variáveis não correlacionadas que não sejam independentes. Esse é o caso, por exemplo, das variáveis X e $Y=X^2$, onde X é uma v.a.r. com suporte $\{-1,0,1\}$ e $P(X=-1)=P(X=1)=1/4$ e $P(X=0)=1/2$ (ver Exercício 67).

5.8 Desigualdades de Hölder e de Minkowski*

Generalizando a notação introduzida em §4.2, para qualquer $p>0$, vamos denotar $\mathscr{L}^p$ o conjunto das v.a.r. para as quais $|X|^p\in\mathscr{L}^1$. Vimos na Proposição 5.7.1 que se X e Y são v.a.r. em $\mathscr{L}^2$, então o seu produto XY é integrável, verificando-se a desigualdade de Cauchy-Schwarz: $|E(XY)|\leq\sqrt{E(X^2)}\sqrt{E(Y^2)}$. Começamos por generalizar esta desigualdade ao caso em que $X\in\mathscr{L}^p$ e $Y\in\mathscr{L}^q$, com p e q não necessariamente iguais a 2.

Proposição 5.8.1 (Desigualdade de Hölder)**.** *Se $X\in\mathscr{L}^p$ e $Y\in\mathscr{L}^q$, com $p,q>1$ e $\frac{1}{p}+\frac{1}{q}=1$, então $XY\in\mathscr{L}^1$ e*

$$|E(XY)|\leq\big(E(|X|^p)\big)^{1/p}\big(E(|Y|^q)\big)^{1/q}.$$

Dem: Se $E(|X|^p)=0$ ou $E(|Y|^q=0$ então $XY=0$ *q.c.* e a desigualdade é válida. Suponhamos então que $E(|X|^p),E(|Y|^q>0$. Uma vez que

$$xy\leq\frac{1}{p}x+\frac{1}{q}y,$$

para todo o $x,y\geq0$, tomando $x=|X|/\big(E(|X|^p)\big)^{1/p}$ e $y=|Y|/\big(E(|Y|^q)\big)^{1/q}$ obtemos

$$\frac{|XY|}{\big(E(|X|^p)\big)^{1/p}\big(E(|Y|^q)\big)^{1/q}}\leq\frac{1}{p}\frac{|X|^p}{E(|X|^p)}+\frac{1}{q}\frac{|Y|^q}{E(|Y|^q)}.$$

Pela monotonia da esperança matemática concluímos o pretendido. $\blacksquare$

Da desigualdade de Cauchy-Schwarz podemos também concluir que a aplicação definida, para $X \in \mathscr{L}^2$, por $||X||_2 = \sqrt{\langle X, X\rangle} = \sqrt{E(X^2)}$, satisfaz a chamada desigualdade triangular, isto é, $||X+Y||_2 \leq ||X||_2 + ||Y||_2$, para todo o $X, Y \in \mathscr{L}^2$, sendo uma semi-norma em $\mathscr{L}^2$ (é uma norma se identificarmos variáveis aleatórias que são quase certamente iguais), isto é, para quaisquer $X, Y \in \mathscr{L}^2$ e $\alpha \in \mathbb{R}$ temos:

SN.1) $||X||_2 = 0$ sse $X = 0$ q.c.;
SN.2) $||\alpha X||_2 = |\alpha| ||X||_2$;
SN.3) $||X+Y||_2 \leq ||X||_2 + ||Y||_2$.

Da desigualdade seguinte concluímos que a aplicação $||X||_p = \big(E(|X|^p)\big)^{1/p}$ definida para $X \in \mathscr{L}^p$, é uma semi-norma em $\mathscr{L}^p$ quando $p \geq 1$ (é uma norma se identificarmos variáveis aleatórias que são iguais quase certamente).

Proposição 5.8.2 (Desigualdade de Minkowski)**.** *Se $X, Y \in \mathscr{L}^p$, com $p \geq 1$, então $X + Y \in \mathscr{L}^p$ e*

$$||X+Y||_p \leq ||X||_p + ||Y||_p.$$

Dem: Para $p = 1$ o resultado é consequência da monotonia e linearidade da esperança matemática. Para $p > 1$, da desigualdade de Hölder temos

$$\begin{aligned} E(|X+Y|^p) &= E\big(|X+Y||X+Y|^{p-1}\big) \\ &\leq E\big(|X||X+Y|^{p-1}\big) + E\big(|Y||X+Y|^{p-1}\big) \\ &\leq \big(E(|X|^p)\big)^{1/p} \big(E(|X+Y|^{(p-1)q})\big)^{1/q} \\ &\quad + \big(E(|Y|^p)\big)^{1/p} \big(E(|X+Y|^{(p-1)q})\big)^{1/q} \\ &= \Big(\big(E(|X|^p)\big)^{1/p} + \big(E(|Y|^p)\big)^{1/p}\Big) \big(E(|X+Y|^p)\big)^{1/q}. \end{aligned}$$

■

5.9 Média e matriz de covariância dum vetor aleatório

A noção de esperança matemática ou média de uma v.a.r. pode ser extendida de forma natural ao caso dos vetores aleatórios em $\mathbb{R}^d$.

Definição 5.9.1. *Se $X = (X_1, \ldots, X_d)$ é um ve.a.r. com $X_1, \ldots, X_d \in \mathscr{L}^1$, chamamos esperança matemática de X ou média de X ao vetor de $\mathbb{R}^d$ dado por*

$$E(X) = \big(E(X_1), \ldots, E(X_d)\big).$$

Tal como no caso real, a interpretação de $E(X)$ como medida de localização do centro da distribuição de probabilidade de X pode ser reforçada recordando que o centro de massa, ou «ponto de equilíbrio», dum sistema de pontos materiais $x_1, \ldots, x_N \in \mathbb{R}^d$,

com massas $m_1, \ldots, m_N$, onde $\sum_{i=1}^{N} m_i = 1$, não é mais do que a esperança matemática do vetor discreto X que toma os valores x_i com probabilidades m_i.

A par da esperança matemática, a noção de matriz de covariâncias que a seguir introduzimos é um dos parâmetros de resumo duma distribuição de probabilidade mais utilizados no caso multidimensional. É a generalização natural a este contexto, da noção real de variância, permitindo caraterizar a dispersão dum ve.a.r. em torno da sua média.

Definição 5.9.2. *Se $X = (X_1, \ldots, X_d)$ é um ve.a.r. com $X_1, \ldots, X_d \in \mathscr{L}^2$, chamamos matriz de covariância de X à matriz*

$$C_X = [\mathrm{Cov}(X_i, X_j)]_{1 \leq i,j \leq d}.$$

Atendendo a que o operador de covariância Cov é simétrico e bilinear, isto é, é linear nos dois argumentos (ver Proposição 5.7.3), concluímos que a matriz de covariância é simétrica e semidefinida positiva, pois

$$0 \leq \mathrm{Var}(\lambda^T X) = \mathrm{Var}\left(\sum_{i=1}^{d} \lambda_i X_i \right) = \lambda^T C_X \lambda,$$

para todo o $\lambda \in \mathbb{R}^d$.

Da alínea c) da Proposição 5.7.4 sabemos que a matriz de covariância $C_{(X,Y)}$ dum vetor aleatório em $\mathbb{R}^2$ nos dá informação sobre o tipo de distribuição de (X, Y). Mais precisamente, sabemos que se $C_{(X,Y)}$ possui caraterística 1 então a distribuição de (X, Y) está concentrada numa reta, não sendo, por isso, absolutamente contínua. Generalizamos a seguir este resultado ao caso dum vetor aleatório em $\mathbb{R}^d$:

Proposição 5.9.3. *Sejam $X = (X_1, \ldots, X_d)$ um ve.a.r. com $X_1, \ldots, X_d \in \mathscr{L}^2$, e C_X a sua matriz de covariância. Se $\mathrm{car}(C_X) = r$, então existe um subespaço afim de $\mathbb{R}^d$ de dimensão r, S, tal que $P(X \in S) = 1$.*

Dem: Tendo o espaço das colunas de C_X dimensão r, o espaço nulo $\mathscr{N}(C_X)$ de C_X, que não é mais do que o conjunto das soluções $y \in \mathbb{R}^d$ do sistema $C_X y = 0$, tem dimensão $d - r$. Sendo $y_1, \ldots, y_{d-r}$ uma base de $\mathscr{N}(C_X)$, temos que $y_i^T C_X y_i = 0$, ou seja $\mathrm{Var}(y_i^T X) = 0$, para $i = 1, \ldots, d - r$. Assim, cada variável aleatória $y_i^T X$ é degenerada, sendo quase certamente igual à sua média, ou seja, $y_i^T (X - E(X)) = 0\, q.c.$, para $i = 1, \ldots, d - r$ (ver Proposição 4.2.3). Isto significa que, com probabilidade 1, $X - E(X)$ pertence ao complemento ortogonal $\mathscr{N}(C_X)^\perp$ de $\mathscr{N}(C_X)$, ou seja, X pertence ao subespaço afim de $\mathbb{R}^d$ de dimensão r dado por $S = E(X) + \mathscr{N}(C_X)^\perp$. ■

Um vetor aleatório com margens independentes tem uma estrutura distribucional muito simples. Como vimos na Proposição 5.5.6, no caso do vetor ser absolutamente

contínuo a sua densidade de probabilidade é dada pelo produto das densidades de probabilidade marginais. Uma forma simples de obter um vetor aleatório Y de margens não necessariamente independentes a partir de um vetor aleatório X com margens independentes, é definir Y como uma transformação afim do vetor X.

Proposição 5.9.4. *Sejam A uma matriz real de tipo $p \times d$ e b um vector em $\mathbb{R}^p$. Se X é um ve.a. em $\mathbb{R}^d$ com $X_1, \ldots, X_d \in \mathscr{L}^2$, e Y é o ve.a. em $\mathbb{R}^p$ definido por*

$$Y = AX + b,$$

então a esperança matemática e a matriz de covariância de Y são dadas por

$$E(Y) = AE(X) + b \quad e \quad C_Y = AC_X A^T.$$

Dem: Sendo A uma matriz com elemento genérico $a_{i,j}$, e b um vetor de coordenadas b_i, para $i, j = 1, \ldots, p$, temos

$$Y_i = \sum_{k=1}^{d} a_{ik} X_k + b_i.$$

Assim,

$$E(Y_i) = \sum_{k=1}^{d} a_{ik} E(X_k) + b_i,$$

e

$$\mathrm{Cov}(Y_i, Y_j) = \mathrm{Cov}\Big(\sum_{k=1}^{d} a_{ik} X_k + b_i, \sum_{l=1}^{d} a_{jl} X_l + b_j \Big) = \sum_{k,l=1}^{d} a_{ik} \mathrm{Cov}(X_k, X_l) a_{jl},$$

o que prova o pretendido. ■

Reparemos que se as margens de X são independentes, de acordo com o corolário 5.7.7 a matriz C_X é diagonal. No entanto, o mesmo já não acontece necessariamente com a matriz de covariância de Y. De acordo com o Corolário 5.7.7, se C_Y não é diagonal, então há pelo menos duas margens de Y que não são independentes.

5.10 Transformação de vetores com densidade*

Vimos na Proposição 3.6.1 que sob certas condições na transformação h, é possível concluir que se X é uma v.a.r. absolutamente contínua então a variável definida por $Y = h(X)$ é também absolutamente contínua. Além disso, podemos obter a densidade de probabilidade de Y a partir da de X. Na proposição seguinte generalizamos o resultado anterior ao contexto dos vetores aleatórios.

Proposição 5.10.1. *Suponhamos que X e Y são ve.a. em $\mathbb{R}^d$ tais que $Y = h(X)$ com $h : U \to V$, bijetiva entre os abertos U e V, e h e h^{-1} de classe C^1. Se X é absolutamente contínuo com densidade $f_X \mathbb{1}_U$, então Y é absolutamente contínuo com densidade de probabilidade dada por*

$$f_Y(y) = f_X(h^{-1}(y))|\det(J_{h^{-1}}(y))|\mathbb{1}_V(y).$$

Dem: Para $B \in \mathscr{B}(\mathbb{R}^d)$ temos

$$\begin{aligned} P(Y \in B) &= P(h(X) \in B) = P(X \in h^{-1}(B)) \\ &= \int_{h^{-1}(B)} f_X(x)\mathbb{1}_U(x)dx = \int_U f_X(x)\mathbb{1}_B(h(x))dx. \end{aligned}$$

Efetuando a mudança de variável $y = h(x)$, obtemos

$$\begin{aligned} P(Y \in B) &= \int_V f_X(h^{-1}(y))\mathbb{1}_B(y)|\det(J_{h^{-1}}(y))|dy \\ &= \int_B f_X(h^{-1}(y))|\det(J_{h^{-1}}(y))|\mathbb{1}_V(y)dy, \end{aligned}$$

o que permite concluir que Y é absolutamente contínuo com densidade de probabilidade dada por $f_Y(y) = f_X(h^{-1}(y))|\det(J_{h^{-1}}(y))|\mathbb{1}_V(y)$. ■

Exemplo 5.10.2. Quando Y é uma transformação afim de X, isto é, $Y = AX + b$, com A uma matriz invertível de dimensão d e $b \in \mathbb{R}^d$, usando a proposição anterior com $h(x) = Ax+b$, concluímos que se X é absolutamente contínuo, então Y é absolutamente contínuo com densidade de probabilidade

$$f_Y(y) = \frac{1}{|\det(A)|} f_X(A^{-1}(y - b)),\ y \in \mathbb{R}^d.$$

Exemplo 5.10.3. Uma aplicação interessante do resultado anterior surge na determinação da densidade de probabilidade da soma $X + Y$ de duas v.a.r. independentes e absolutamente contínuas X e Y. Uma vez que (X, Y) é absolutamente contínuo com densidade de probabilidade $f_{(X,Y)}(x, y) = f_X(x)f_Y(y)$ (ver Proposição 5.5.5) e $(X + Y, Y) = h(X, Y)$, onde $h : \mathbb{R}^2 \to \mathbb{R}^2$, definida por $h(x, y) = (x + y, y)$, satisfaz as condições do resultado anterior, então $(X + Y, Y)$ é absolutamente contínuo com densidade

$$f_{(X+Y,Y)}(u, v) = f_{(X,Y)}(u - v, v) = f_X(u - v)f_Y(v),$$

pois $h^{-1}(u, v) = (u - v, v)$ e $\det(J_{h^{-1}}(u, v)) = 1$. Finalmente, pela Proposição 5.3.5, $X + Y$ é absolutamente contínua com densidade de probabilidade

$$f_{X+Y}(u) = \int_{\mathbb{R}} f_X(u - v)f_Y(v)dv,$$

que é dita produto de convolução das densidades f_X e f_Y.

5.11 Vetores aleatórios normais (parte I)

Uma família muito importante de vetores aleatórios, cuja distribuição depende exclusivamente das respetivas média e matriz de covariâncias, é a família dos vetores aleatórios normais ou gaussianos. No caso real, sabemos que uma variável normal $X \sim N(\mu, \sigma^2)$, com $\sigma > 0$, pode ser obtida por transformação afim duma variável normal standard $Z \sim N(0,1)$:

$$X = \sigma Z + \mu.$$

Vamos seguir esta mesma estratégia para definir a classe dos vetores aleatórios normais em $\mathbb{R}^d$. Começamos assim por considerar o vetor aleatório $Z = (Z_1, \ldots, Z_d)$ em $\mathbb{R}^d$ cujas margens são independentes com distribuições normais standard. Dizemos que Z é um vetor normal standard em $\mathbb{R}^d$. Pela Proposição 5.5.6 sabemos que Z é um vetor absolutamente contínuo com densidade de probabilidade

$$\begin{aligned} f_Z(z_1, \ldots, z_d) &= f_{Z_1}(z_1) \times \cdots \times f_{Z_d}(z_d) = \prod_{i=1}^{d} \frac{1}{\sqrt{2\pi}} e^{-z_i^2/2} \\ &= \frac{1}{(2\pi)^{d/2}} \exp\left(-\sum_{i=1}^{d} z_i^2/2 \right) = \frac{1}{(2\pi)^{d/2}} \exp\left(-||z||^2/2 \right), \end{aligned}$$

com $z = (z_1, \ldots, z_d) \in \mathbb{R}^d$ e $||z||$ a norma euclideana em $\mathbb{R}^d$. A esta densidade chamamos densidade normal standard em $\mathbb{R}^d$. Além disso,

$$E(Z) = 0 \quad \text{e} \quad C_Z = I_d,$$

onde I_d é a matriz identidade de dimensão d.

Definição 5.11.1. *Um vetor aleatório $X = (X_1, \ldots, X_d)$ em $\mathbb{R}^d$ diz-se normal ou gaussiano se*

$$X = AZ + b,$$

onde Z é um vetor normal standard em $\mathbb{R}^d$, A uma matriz real quadrada de dimensão d e b um vetor em $\mathbb{R}^d$.

Da Proposição 5.9.4 sabemos que

$$E(X) = AE(Z) + b = b$$

e

$$C_X = AC_Z A^T = AA^T.$$

Uma forma simples de exibir um vetor aleatório normal em $\mathbb{R}^d$ é considerar um vetor cujas margens sejam normais e independentes.

Proposição 5.11.2. *Se $X = (X_1, \ldots, X_d)$ é um vetor aleatório em $\mathbb{R}^d$ cujas margens são v.a.r. normais independentes, então X é um vetor aleatório normal.*

Dem: Por hipótese, as v.a.r. $X_1, \ldots, X_n$ são independentes com $X_i \sim N(\mu_i, \sigma_i^2)$, para $i = 1, \ldots, d$. Definindo

$$Z_i = (X_i - \mu_i)/\sigma_i,$$

temos

$$X_i = \sigma_i Z_i + \mu_i,$$

para $i = 1, \ldots, d$, onde as v.a.r. $Z_1, \ldots, Z_d$ são independentes com distribuições normais standard. Assim, X é um vetor normal em $\mathbb{R}^d$ uma vez que

$$X = AZ + \mu,$$

onde o vetor $Z = (Z_1, \ldots, Z_d)$ é normal standard em $\mathbb{R}^d$, A é a matriz diagonal

$$A = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_d \end{bmatrix}$$

e $\mu = (\mu_1, \ldots, \mu_d)$. ■

Proposição 5.11.3. *Se $X = (X_1, \ldots, X_d)$ é um vetor aleatório normal em $\mathbb{R}^d$ com $\mathrm{car}(C_X) = d$, então X é absolutamente contínuo com densidade de probabilidade dada por*

$$f_X(x) = \frac{1}{\sqrt{(2\pi)^d \det(C_X)}} \exp\Big(-(x - E(X))^T C_X^{-1} (x - E(X))/2 \Big),$$

para $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$.

Dem: Se $X = (X_1, \ldots, X_d)$ é um vetor aleatório normal em $\mathbb{R}^d$, então $X = AZ + b$, onde Z é um vetor normal standard em $\mathbb{R}^d$, A uma matriz real quadrada de dimensão d e b um vetor em $\mathbb{R}^d$. Quando $\mathrm{car}(C_X) = d$, a matriz A é invertível pois $\mathrm{car}(C_X) = \mathrm{car}(AA^T) = \mathrm{car}(A)$, e pela Proposição 5.10.1 (ver Exemplo 5.10.2) sabemos que X é absolutamente contínuo com densidade de probabilidade dada por

$$\begin{aligned} f_X(x) &= \frac{1}{|\det(A)|} f_Z(A^{-1}(x - b)) \\ &= \frac{1}{(\det(C_X))^{1/2}} \frac{1}{(2\pi)^{d/2}} \exp\big(-||A^{-1}(x - E(X))||^2/2 \big) \\ &= \frac{1}{\sqrt{(2\pi)^d \det(C_X)}} \exp\Big(-(x - E(X))^T C_X^{-1} (x - E(X))/2 \Big), \end{aligned}$$

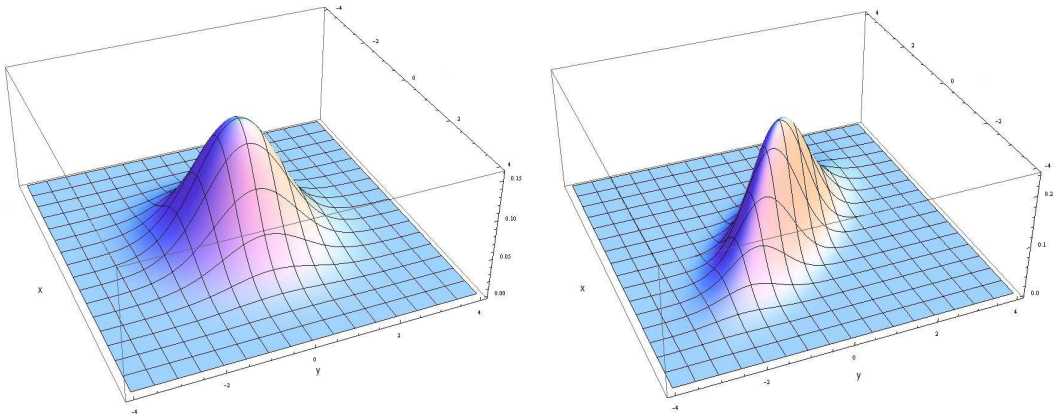


Figura 5.7: Densidade de probabilidade do vetor aleatório normal (X, Y), onde $X, Y \sim N(0, 1)$ e $\rho = 0$ (esquerda) e $\rho = 0{,}75$ (direita).

para $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$. ■

Da proposição anterior concluímos que quando $\text{car}(C_X) = d$ a distribuição de X fica determinada pelo conhecimento da média e da matriz de covariâncias de X. O mesmo acontece quando $\text{car}(C_X) < d$, mas deixamos a justificação deste facto para o próximo capítulo. Para já, sabemos da Proposição 5.9.3 que, quando $\text{car}(C_X) = r < d$, o vetor X não é absolutamente contínuo, estando a sua distribuição concentrada num subespaço afim de dimensão r.

No caso dos vetores em $\mathbb{R}^2$, a matriz de covariâncias pode ser escrita na forma

$$C_{(X_1,X_2)} = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix},$$

com $\sigma_1 = \sigma(X_1)$, $\sigma_2 = \sigma(X_2)$ e $\rho = \rho(X_1, X_2)$, sendo invertível sse $\rho \neq \pm 1$. Neste caso temos

$$C^{-1}_{(X_1,X_2)} = \frac{1}{1-\rho^2} \begin{bmatrix} \frac{1}{\sigma_1^2} & -\frac{\rho}{\sigma_1\sigma_2} \\ -\frac{\rho}{\sigma_1\sigma_2} & \frac{1}{\sigma_2^2} \end{bmatrix}.$$

e a densidade de probabilidade de (X_1, X_2) toma a forma

$$f_{(X_1,X_2)}(x_1, x_2)$$
$$= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\frac{x_1-\mu_1}{\sigma_1}\frac{x_2-\mu_2}{\sigma_2} + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right]\right),$$

para $(x_1, x_2) \in \mathbb{R}^2$, onde $\mu_1 = E(X_1)$ e $\mu_2 = E(X_2)$ (ver Figura 5.7). Com algum trabalho de cálculo integral, podemos provar que a densidade de probabilidade de X_1

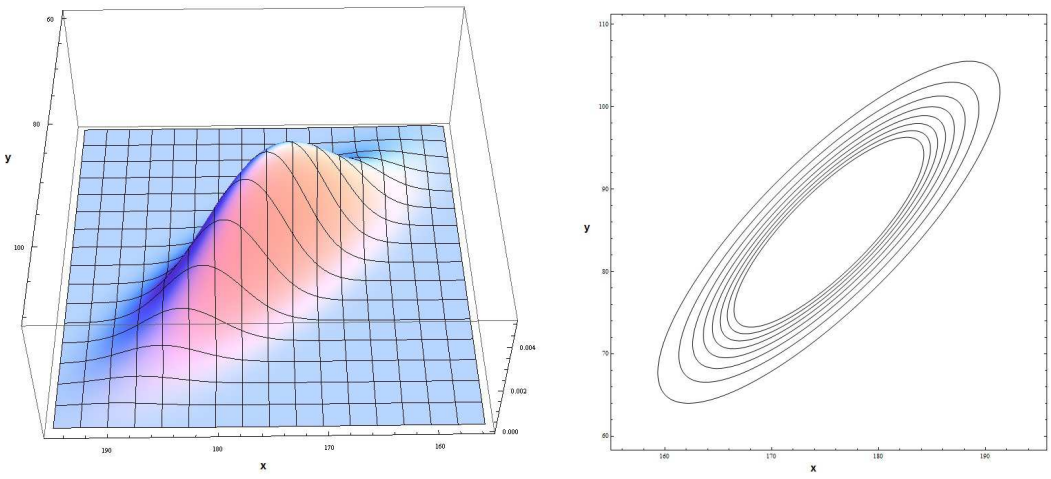


Figura 5.8: Densidade de probabilidade do vetor aleatório normal (X, Y), onde $X \sim N(175{,}33; 6{,}52)$, $Y \sim N(84{,}75; 8{,}46)$ e $\rho = 0{,}82$, e respetivas curvas de nível.

é dada por

$$f_{X_1}(x_1) = \int_{\mathbb{R}} f_{(X_1,X_2)}(x_1, x_2)dx_2 = \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left(-\frac{(x_1-\mu_1)^2}{2\sigma_1^2}\right),$$

o que permite concluir que $X_1 \sim N(\mu_1, \sigma_1^2)$. Com mais algum cálculo pode provar-se que a distribuição condicional $X_2|X_1 = x_1$ é também normal:

$$X_2|X_1 = x_1 \sim N\Big(\mu_2 + \rho\sigma_1^{-1}\sigma_2(x_1-\mu_1), \sigma_2^2(1-\rho^2)\Big).$$

De forma análoga se conclui que X_2 e $X_1|X_2 = x_2$ são também normais.

Outra caraterística interessante dos vetores aleatórios normais tem a ver com a possibilidade de caraterizarmos a independência das suas margens de forma muito simples. Vejamos o que se passa no caso $d = 2$. O caso geral é estudado no próximo capítulo. Quando $\rho = 0$, isto é, quando $\mathrm{Cov}(X_1, X_2) = 0$, a densidade de probabilidade de (X_1, X_2) escreve-se na forma

$$\begin{aligned} f_{(X_1,X_2)}(x_1, x_2) &= \frac{1}{2\pi\sigma_1\sigma_2} \exp\left(-\frac{1}{2}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right]\right) \\ &= \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left(-\frac{(x_1-\mu_1)^2}{2\sigma_1^2}\right) \frac{1}{\sqrt{2\pi\sigma_2^2}} \exp\left(-\frac{(x_2-\mu_2)^2}{2\sigma_2^2}\right) \\ &= f_{X_1}(x_1) f_{X_2}(x_2), \end{aligned}$$

para todo o $(x_1, x_2) \in \mathbb{R}^2$, o que permite concluir que X_1 e X_2 são independentes (ver Proposição 5.5.6). Esta propriedade é também válida para qualquer vetor aleatório

normal em $\mathbb{R}^d$. As margens de um vetor aleatório normal são independentes sse forem duas a duas não correlacionadas (ver Proposição 6.7.3).

Retomando o exemplo com que começámos este capítulo, é razoável admitir que a distribuição do vetor (X, Y), onde X representa a altura e Y o peso dos indivíduos observados, possa ser normal. Na Figura 5.8 apresentamos a densidade normal bivariada com parâmetros μ_1, μ_2, σ_1, σ_2 e ρ aproximados a partir das observações feitas de pesos e alturas nos 200 indivíduos considerados (no próximo capítulo detalhamos este procedimento). Reparemos, em particular, na proximidade entre a forma das curvas de nível da densidade de probabilidade teórica e a distribuição revelada pelo gráfico de dispersão da Figura 5.1.

5.12 Bibliografia

Gonçalves, E., Mendes Lopes, N. (2013). *Probabilidades: princípios teóricos.* Lisboa: Escolar Editora.

Jacod, J., Protter, P. (2000). *Probability essentials.* Berlin: Springer.

James, B.R. (1981). *Probabilidade: um curso em nível intermediário.* Rio de Janeiro: IMPA.

Monfort, A. (1980). *Cours de probabilités.* Paris: Economica.

Resnick, S.I. (2005). *A probability path.* Boston: Birkhäuser.

Tenreiro, C. (2002). *Apontamentos de Teoria das Probabilidades.* Coimbra: DMUC. https://www.mat.uc.pt/∼tenreiro/apontamentos/ATP0204.pdf

6

Função caraterística

Esperança matemática de variáveis aleatórias complexas. Definição de função característica e primeiras propriedades. Função caraterística e o cálculo dos momentos. Cálculo da função caraterística: alguns exemplos. Distribuição da soma de variáveis aleatórias independentes. Função caraterística dum vetor aleatório. Vetores aleatórios normais (parte II).

6.1 Esperança matemática de variáveis complexas

Como bem sabemos, o conjunto dos números complexos pode ser identificado com o conjunto $\mathbb{R}^2$ dos pontos do plano, associando-se a cada complexo $z = x + \mathrm{i}y$ o par ordenado (x, y). A x chamamos parte real de z, e escrevemos $x = Re(z)$ e a y parte imaginária de z que denotamos por $y = Im(z)$. Considerando em $\mathbb{R}^2$ a norma euclideana e em $\mathbb{C}$ a norma do módulo ($|z| = \sqrt{x^2 + y^2}$), concluímos facilmente que os abertos de cada um dos conjuntos podem ser também identificados, o mesmo acontecendo relativamente às σ-álgebras de Borel $\mathscr{B}(\mathbb{C})$ e $\mathscr{B}(\mathbb{R}^2)$.

Toda a função complexa Z definida num conjunto Ω pode escrever-se na forma $Z = Re(Z) + \mathrm{i}Im(Z)$, onde $Re(Z)$ e $Im(Z)$ são funções reais definidas, para $\omega \in \Omega$, por $Re(Z)(\omega) = Re(Z(\omega))$ e $Im(Z)(\omega) = Im(Z(\omega))$.

Definição 6.1.1. *Chamamos variável aleatória complexa (v.a.c.) a toda a aplicação $Z = Re(Z) + \mathrm{i}Im(Z) : \Omega \to \mathbb{C}$ definida num espaço de probabilidade $(\Omega, \mathscr{A}, P)$ que satisfaz $\{Z \in B\} \in \mathscr{A}$ para todo o $B \in \mathscr{B}(\mathbb{C})$.*

As observações preliminares anteriores implicam que uma função Z definida num espaço de probabilidade $(\Omega, \mathscr{A}, P)$ com valores em $(\mathbb{C}, \mathscr{B}(\mathbb{C}))$ é uma variável aleatória sse a função de $(\Omega, \mathscr{A}, P)$ em $(\mathbb{R}^2, \mathscr{B}(\mathbb{R}^2))$ definida por $(Re(Z), Im(Z))$ é um vetor aleatório em $\mathbb{R}^2$, ou ainda, sse $Re(Z)$ e $Im(Z)$ são variáveis aleatórias reais.

Tendo em conta o que foi dito, a definição de esperança matemática duma variável aleatória complexa surge agora de forma natural.

Definição 6.1.2. *Dizemos que uma variável aleatória complexa Z possui esperança matemática sempre que $Re(Z), Im(Z) \in \mathscr{L}^1$, e nesse caso, a* esperança matemática *de Z é dada por*

$$E(Z) = E(Re(Z)) + \mathrm{i}E(Im(Z)).$$

Vamos representar por $\mathscr{L}^1$ o espaço vetorial complexo (com a soma e produto escalar definidos da forma habitual) das variáveis aleatórias complexas que possuem esperança matemática. Sempre que haja lugar a possível confusão escreveremos $\mathscr{L}^1(\mathbb{R})$ ou $\mathscr{L}^1(\mathbb{C})$ para representar o conjunto das variáveis reais ou complexas com esperança matemática. Atendendo à linearidade da esperança matemática para variáveis reais podemos concluir que a aplicação $Z \to E(Z)$, de $\mathscr{L}^1(\mathbb{C})$ em $\mathbb{C}$, é linear.

Proposição 6.1.3. *Se Z é uma v.a.c. então $Z \in \mathscr{L}^1(\mathbb{C})$ sse $|Z| \in \mathscr{L}^1(\mathbb{R})$, e nesse caso*

$$|E(Z)| \leq E(|Z|).$$

Dem: A primeira afirmação resulta das desigualdades $|Re(Z)| \leq |Z|$, $|Im(Z)| \leq |Z|$ e $|Z| \leq |Re(Z)| + |Im(Z)|$ e da Proposição 4.1.8. A desigualdade $|E(Z)| \leq E(|Z|)$ é válida se $E(Z) = 0$. Se $E(Z) \neq 0$, seja $w = E(Z)/|E(Z)|$. Então

$$|E(Z)| = w^{-1}E(Z) = E(w^{-1}Z) = E(Re(w^{-1}Z)) \leq E(|w^{-1}Z|) = E(|Z|). \qquad \blacksquare$$

Observemos que outros resultados que enunciámos relativos à esperança matemática de variáveis aleatórias reais, são também válidos para variáveis aleatórias complexas. Tais resultados podem ser estabelecidos a partir dos correspondentes resultados para variáveis aleatórias reais, considerando separadamente as partes reais e imaginárias das variáveis aleatórias intervenientes.

6.2 Definição e primeiras propriedades

A noção de função característica que introduzimos a seguir é, como veremos ao longo deste capítulo, um instrumento essencial no estudo da distribuição duma variável aleatória e na convergência em distribuição de uma sucessão de variáveis aleatórias que estudaremos mais à frente.

Definição 6.2.1. *Chamamos* função caraterística *duma v.a.r. X (ou função caraterística de P_X), à função de $\mathbb{R}$ em $\mathbb{C}$ definida por*

$$\phi_X(t) = E\big(e^{\,\mathrm{i}tX}\big) = E(\cos(tX)) + \mathrm{i}E(\sin(tX)), \ \textit{para } t \in \mathbb{R}.$$

Notemos que como $|e^{\,\mathrm{i}tX}| = 1$, a esperança matemática anterior está bem definida.

Proposição 6.2.2. *Se ϕ_X é a função caraterística da v.a.r. X então:*

a) $\phi_X(0) = 1$;

b) $|\phi_X(t)| \leq 1$, *para todo o* $t \in \mathbb{R}$;

c) $\phi_{-X}(t) = \overline{\phi_X(t)}$, *para todo o* $t \in \mathbb{R}$;

d) ϕ_X *é uma função contínua em* $\mathbb{R}$.

e) ϕ_X *carateriza* P_X *(isto é,* $\phi_X = \phi_Y$ *sse* $X \sim Y$*).*

Dem: As alíneas a), b) e c) são consequência imediata da definição de função caraterística. A demonstração da última alínea não será apresentada visto ultrapassar largamente os objetivos do curso. Apresentamos apenas a demonstração da alínea d) para a qual precisamos de um resultado auxiliar conhecido por teorema da convergência dominada de Lebesgue. Enunciamos para já uma versão fraca desse resultado, ao qual voltaremos no próximo capítulo (Teorema 7.1.10):

Teorema 6.2.3 (da convergência dominada de Lebesgue (versão fraca))**.** *Se a sucessão* (X_n) *de v.a.r. é tal que*

a) $\lim_{n\to\infty} X_n = X$;

b) $|X_n| \leq Y$ *para todo o* n, *com* $Y \in \mathscr{L}^1$;

então $X \in \mathscr{L}^1$ *e* $E(X_n) \to E(X)$.

Sendo $t \in \mathbb{R}$ e (t_n) uma sucessão de números reais que converge para t, pretendemos provar que $\phi_X(t_n) = E\big(e^{\,\mathrm{i}t_nX}\big)$ converge para $\phi_X(t) = E\big(e^{\,\mathrm{i}tX}\big)$. Reparemos que sabemos que

$$e^{\,\mathrm{i}t_nX} = \cos(t_nX) + \,\mathrm{i}\sin(t_nX) \to \cos(tX) + \,\mathrm{i}\sin(tX) = e^{\,\mathrm{i}tX}$$

(para todo o $\omega \in \Omega$), e que

$$|e^{\,\mathrm{i}t_nX}| \leq 1, \text{ para todo o } n.$$

Estamos assim nas condições do teorema da convergência dominada, o que permite concluir que

$$\phi_X(t_n) = E\big(e^{\,\mathrm{i}t_nX}\big) \to E\big(e^{\,\mathrm{i}tX}\big) = \phi_X(t). \qquad \blacksquare$$

Proposição 6.2.4. *Se X é uma v.a.r. simétrica relativamente à origem então*

$$\phi_X(t) = E(\cos(tX)),\ t \in \mathbb{R}.$$

Dem: Se X é simétrica relativamente à origem então $X \sim -X$. Atendendo à alínea c) da proposição anterior, concluímos que

$$\phi_X(t) = \phi_{-X}(t) = \overline{\phi_X(t)}.$$

Isto significa que ϕ_X é uma função real, o que prova o pretendido. $\blacksquare$

Proposição 6.2.5. *Para $a, b \in \mathbb{R}$, temos $\phi_{aX+b}(t) = e^{\,\mathrm{i}tb}\phi_X(at)$, para $t \in \mathbb{R}$.*

Dem: Para $t \in \mathbb{R}$ temos

$$\phi_{aX+b}(t) = E\big(e^{\,\mathrm{i}t(aX+b)}\big) = E\big(e^{\,\mathrm{i}taX}e^{\,\mathrm{i}b}\big) = e^{\,\mathrm{i}b}E\big(e^{\,\mathrm{i}atX}\big) = e^{\,\mathrm{i}tb}\phi_X(at). \qquad \blacksquare$$

6.3 Cálculo da função caraterística: alguns exemplos

O cálculo da função caraterística duma variável aleatória pode revelar-se um trabalho árduo. Nos exemplos seguintes um tal cálculo não coloca questões de maior.

Lei de Dirac

Uma variável de Dirac no ponto a é tal que $P(X = a) = 1$. Assim, para $t \in \mathbb{R}$,

$$\phi_X(t) = E\big(e^{\,\mathrm{i}tX}\big) = e^{\,\mathrm{i}ta}.$$

Lei de Bernoulli

Uma variável de Bernoulli X toma valores 1 e 0 com probabilidades p e $1-p$, respetivamente. Assim, para $t \in \mathbb{R}$,

$$\phi_X(t) = E\big(e^{\,\mathrm{i}tX}\big) = e^{\,\mathrm{i}t1}p + e^{\,\mathrm{i}t0}(1-p) = 1 - p(1 - e^{\,\mathrm{i}t}).$$

Lei uniforme discreta

Se X é uma variável aleatória discreta sobre o conjunto $\{1, \ldots, n\}$, então, para $t \in \mathbb{R}$,

$$\phi_X(t) = \sum_{k=1}^{n} e^{\,\mathrm{i}tk}\frac{1}{n} = \frac{e^{\,\mathrm{i}t}}{n}\sum_{k=0}^{n-1} e^{\,\mathrm{i}tk} = \frac{e^{\,\mathrm{i}t}}{n}\frac{1 - e^{\,\mathrm{i}tn}}{1 - e^{\,\mathrm{i}t}}.$$

Lei binomial

Se $X \sim B(n, p)$, então, para $t \in \mathbb{R}$,

$$\begin{aligned}\phi_X(t) &= \sum_{k=0}^{n} e^{\,\mathrm{i}tk}C_k^n p^k(1-p)^{n-k} = \sum_{k=0}^{n} C_k^n (pe^{\,\mathrm{i}t})^k(1-p)^{n-k} \\ &= \big(pe^{\,\mathrm{i}t} + 1 - p\big)^n = \big(1 - p(1 - e^{\,\mathrm{i}t})\big)^n.\end{aligned}$$

Lei de Poisson

No caso da variável de Poisson de parâmetro $\lambda > 0$, para $t \in \mathbb{R}$ temos

$$\phi_X(t) = \sum_{k=0}^{\infty} e^{\,\mathrm{i}tk}\frac{\lambda^k}{k!}e^{-\lambda} = \sum_{k=0}^{\infty} \frac{(\lambda e^{\,\mathrm{i}t})^k}{k!}e^{-\lambda} = \exp\big(\lambda(e^{\,\mathrm{i}t} - 1)\big).$$

Lei uniforme

Se $U \sim U[0,1]$, para $t \in \mathbb{R} \setminus \{0\}$ temos

$$\phi_U(t) = \int_0^1 e^{\,\mathrm{i}tx} dx = \int_0^1 \cos(tx)dx + \,\mathrm{i}\int_0^1 \sin(tx)dx = \frac{e^{\,\mathrm{i}t} - 1}{\mathrm{i}t}.$$

Usando agora o facto de $X = (b-a)U + a \sim U[a,b]$, pela Proposição 6.2.5 e para $t \in \mathbb{R} \setminus \{0\}$ temos

$$\phi_X(t) = e^{\,\mathrm{i}ta}\phi_U((b-a)t) = e^{\,\mathrm{i}ta}\frac{e^{\,\mathrm{i}t(b-a)} - 1}{\mathrm{i}t(b-a)} = \frac{e^{\,\mathrm{i}tb} - e^{\,\mathrm{i}ta}}{\mathrm{i}t(b-a)}.$$

Leis exponencial e gama

Se $X \sim \mathrm{Exp}(\lambda)$, com $\lambda > 0$, para $t \in \mathbb{R}$ temos

$$\phi_X(t) = \int_0^\infty e^{\,\mathrm{i}tx}\lambda \mathrm{e}^{-\lambda x} dx = \int_0^\infty \cos(tx)\lambda \mathrm{e}^{-\lambda x} dx + \,\mathrm{i}\int_0^\infty \sin(tx)\lambda \mathrm{e}^{-\lambda x} dx.$$

Integrando por partes, temos

$$\int_0^\infty \cos(tx)\lambda \mathrm{e}^{-\lambda x} dx = \frac{\lambda^2}{\lambda^2 + t^2} \quad \text{e} \quad \int_0^\infty \sin(tx)\lambda \mathrm{e}^{-\lambda x} dx = \frac{t\lambda}{\lambda^2 + t^2},$$

e assim

$$\phi_X(t) = \frac{\lambda(\lambda + \mathrm{i}t)}{\lambda^2 + t^2} = \frac{\lambda}{\lambda - \mathrm{i}t},\ t \in \mathbb{R}.$$

Como sabemos, a distribuição $\mathrm{Exp}(\lambda)$ é um caso particular da distribuição gama de parâmetros $\alpha = 1$ e $\beta = \lambda$. Se $X \sim \gamma(\alpha, \beta)$, com $\alpha, \beta > 0$, é possível provar que

$$\phi_X(t) = \left(\frac{\beta}{\beta - \mathrm{i}t}\right)^\alpha,\ t \in \mathbb{R}.$$

Lei normal

Sabemos que se $X \sim N(\mu, \sigma^2)$ então X pode ser escrita na forma $X = \sigma Z + \mu$, onde $Z \sim N(0,1)$. Assim, pela Proposição 6.2.5 temos

$$\phi_X(t) = e^{\,\mathrm{i}t\mu}\phi_Z(\sigma t),\ t \in \mathbb{R},$$

o que significa que precisamos apenas de calcular a função caraterística da distribuição normal standard. Atendendo a que esta distribuição é simétrica relativamente à origem então, de acordo com a Proposição 6.2.4,

$$\phi_Z(t) = E(\cos(tZ)).$$

Procedendo como na demonstração da Proposição 6.4.1, concluiríamos que é possível derivar sob o sinal de esperança matemática e assim

$$\begin{aligned}\phi'_Z(t) &= E((\cos(tZ))') \\ &= E(-Z\sin(tZ)) \\ &= \int_{\mathbb{R}}(-x)\sin(tx)\frac{1}{\sqrt{2\pi}}e^{-x^2/2}dx \\ &= -t\int_{\mathbb{R}}\cos(tx)\frac{1}{\sqrt{2\pi}}e^{-x^2/2}dx \\ &= -tE(\cos(tZ)) = -t\phi_Z(t).\end{aligned}$$

Obtemos assim a equação diferencial $\phi'_Z(t)/\phi_Z(t) = -t$, que possui como solução $\phi_Z(t) = e^c e^{-t^2/2}$, ou ainda, $\phi_Z(t) = e^{-t^2/2}$, uma vez que $\phi_Z(0) = 1$. Finalmente, concluímos que

$$\phi_X(t) = e^{\,\mathrm{i}t\mu}e^{-\sigma^2t^2/2},\ t\in\mathbb{R}.$$

6.4 Função caraterística e momentos

Uma das aplicações interessantes da função caraterística é a de permitir uma forma alternativa para calcular os momentos de uma variável aleatória.

Teorema 6.4.1. *Se X é uma v.a.r. em $\mathscr{L}^m$ para algum $m\in\mathbb{N}$, então ϕ_X possui derivada de ordem m e, para $t\in\mathbb{R}$,*

$$\phi_X^{(m)}(t) = \mathrm{i}^m E\big(X^m e^{\,\mathrm{i}tX}\big).$$

Em particular,

$$\phi_X^{(k)}(0) = \mathrm{i}^k E(X^k),\ \textit{para } k = 1,\dots,m.$$

Dem: Vamos limitar-nos a provar o resultado para $m = 1$. Usando o método de indução matemática e a técnica do caso $m = 1$, podemos estabelecer o resultado enunciado. Sendo $X\in\mathscr{L}^1$ e $t\in\mathbb{R}$, pretendemos provar que

$$\lim_{h\to 0}\frac{\phi_X(t+h)-\phi(t)}{h} = \mathrm{i}E\big(X\mathrm{e}^{\,\mathrm{i}tX}\big),$$

o que é equivalente a provar que

$$\lim_{n\to\infty}\frac{\phi_X(t+h_n)-\phi(t)}{h_n} = \mathrm{i}E\big(X\mathrm{e}^{\,\mathrm{i}tX}\big),$$

para qualquer sucessão de números reais (h_n), com $h_n\to 0$, quando n tende para infinito. Reparemos que

$$\frac{\phi_X(t+h_n)-\phi(t)}{h_n} = E\left(\frac{e^{\,\mathrm{i}(t+h_n)X}-e^{\,\mathrm{i}tX}}{h_n}\right) = E\left(e^{\,\mathrm{i}tX}\frac{e^{\,\mathrm{i}h_nX}-1}{h_n}\right)$$

onde

$$e^{\,\mathrm{i}tX}\frac{e^{\,\mathrm{i}h_nX}-1}{h_n}\to e^{\,\mathrm{i}tX}\,\mathrm{i}X,$$

e

$$\left|e^{\,\mathrm{i}tX}\frac{e^{\,\mathrm{i}h_nX}-1}{h_n}\right|=|e^{\,\mathrm{i}tX}|\frac{|e^{\,\mathrm{i}h_nX}-1|}{h_n}\le|X|,$$

com $X\in\mathscr{L}^1$ (verifique que $|e^{\,\mathrm{i}x}-1|\le|x|$, para todo o $x\in\mathbb{R}$).

Pelo teorema da convergência dominada (Teorema 6.2.3), concluímos que

$$\frac{\phi_X(t+h_n)-\phi(t)}{h_n}\to E\big(e^{\,\mathrm{i}tX}\,\mathrm{i}X\big)=\,\mathrm{i}E\big(Xe^{\,\mathrm{i}tX}\big),$$

o que prova o pretendido. ■

A não existência da derivada de ordem k de ϕ_X na origem, implica assim que $X\notin\mathscr{L}^k$. De acordo com o resultado seguinte, cuja demonstração omitimos (ver Métivier, 1972, p. 157 e seguintes), é possível utilizar a função caraterística no estudo da existência dos momentos de X.

Teorema 6.4.2. *Se a função caraterística da v.a.r. X possui derivada de ordem $m\in\mathbb{N}$ na origem, então $X\in\mathscr{L}^m$ se m é par, e $X\in\mathscr{L}^{m-1}$ se m é ímpar.*

Para algumas das distribuições que considerámos na secção anterior, vamos efetuar o cálculo dos respetivos momentos de primeira e segunda ordens usando para o efeito o Teorema 6.4.1.

Lei de Dirac

Atendendo a que

$$\phi_X(t)=\mathrm{e}^{\,\mathrm{i}ta},\ t\in\mathbb{R},$$

então

$$\phi'_X(t)=\,\mathrm{i}a\phi_X(t)\ \text{ e }\ \phi''_X(t)=\,\mathrm{i}a\phi'_X(t)=-a^2\phi_X(t).$$

Assim,

$$\phi'_X(0)=\,\mathrm{i}a\ \text{ e }\ \phi''_X(0)=-a^2,$$

de onde concluímos que

$$E(X)=a\ \text{ e }\ E(X^2)=a^2.$$

Lei de Bernoulli

Atendendo a que

$$\phi_X(t) = 1 - p(1 - \mathrm{e}^{\mathrm{i}t}),\ t \in \mathbb{R},$$

então

$$\phi'_X(0) = \mathrm{i}p \ \text{ e } \ \phi''_X(0) = -p,$$

de onde concluímos que

$$E(X) = p \text{ e } E(X^2) = p.$$

Lei binomial

Uma vez que

$$\phi_X(t) = \big(1 - p(1 - \mathrm{e}^{\mathrm{i}t})\big)^n,\ t \in \mathbb{R},$$

então

$$\phi'_X(0) = \mathrm{i}np \ \text{ e } \ \phi''_X(0) = -np(1 - p + np),$$

de onde concluímos que

$$E(X) = np \text{ e } E(X^2) = np(1 - p + np).$$

Lei de Poisson

Atendendo a que

$$\phi_X(t) = \exp\big(\lambda(\mathrm{e}^{\mathrm{i}t} - 1)\big),\ t \in \mathbb{R},$$

temos

$$\phi'_X(t) = \mathrm{i}\lambda\mathrm{e}^{\mathrm{i}t}\phi_X(t) \ \text{ e } \ \phi''_X(t) = -\lambda\mathrm{e}^{\mathrm{i}t}\phi_X(t) + \mathrm{i}\lambda\mathrm{e}^{\mathrm{i}t}\phi'_X(t).$$

Assim,

$$\phi'_X(0) = \mathrm{i}\lambda \ \text{ e } \ \phi''_X(0) = -\lambda - \lambda^2,$$

de onde concluímos que

$$E(X) = \lambda \text{ e } E(X^2) = \lambda(\lambda + 1).$$

Lei exponencial

Atendendo a que

$$\phi_X(t) = \frac{\lambda}{\lambda - \mathrm{i}t},\ t \in \mathbb{R},$$

temos

$$\phi'_X(t) = \mathrm{i}\frac{1}{\lambda}\phi_X(t)^2 \ \text{ e } \ \phi''_X(t) = \mathrm{i}\frac{2}{\lambda}\phi_X(t)\phi'_X(t).$$

Assim,

$$\phi'_X(0) = \mathrm{i}\frac{1}{\lambda} \text{ e } \phi''_X(0) = -\frac{2}{\lambda},$$

e portanto

$$E(X) = \frac{1}{\lambda} \text{ e } E(X^2) = \frac{2}{\lambda}.$$

Lei normal

Uma vez

$$\phi_X(t) = \mathrm{e}^{\mathrm{i}t\mu}\mathrm{e}^{-\sigma^2 t^2/2}, \ t \in \mathbb{R},$$

temos

$$\phi'_X(t) = (\mathrm{i}\mu - t\sigma^2)\phi_X(t) \text{ e } \phi''_X(t) = \big((\mathrm{i}\mu - t\sigma^2)^2 - \sigma^2\big)\phi_X(t).$$

Assim,

$$\phi'_X(0) = \mathrm{i}\mu \text{ e } \phi''_X(0) = -(\mu^2 + \sigma^2),$$

e portanto

$$E(X) = \mu \text{ e } E(X^2) = \mu^2 + \sigma^2.$$

6.5 Distribuição da soma de variáveis independentes

A função caraterística tem um papel importante no estudo da distribuição da soma S_n de v.a.r. independentes onde

$$S_n = X_1 + \cdots + X_n.$$

Proposição 6.5.1. *Se $X_1, \ldots, X_n$ são v.a.r. independentes então*

$$\phi_{S_n}(t) = \prod_{i=1}^{n} \phi_{X_i}(t), \ t \in \mathbb{R}.$$

Dem: Para $t \in \mathbb{R}$ temos

$$\begin{aligned}\phi_{S_n}(t) &= E\big(e^{\mathrm{i}t(X_1+\cdots+X_n)}\big) \\ &= E\big(e^{\mathrm{i}tX_1}\big) \times \cdots \times E\big(e^{\mathrm{i}tX_n}\big) \\ &= \phi_{X_1}(t) \times \cdots \times \phi_{X_n}(t).\end{aligned}$$

■

A recíproca do resultado anterior não é válida (tome $X_i = X$, com X a variável de Cauchy standard definida no Exemplo 4.1.6).

Os resultados seguintes usam as expressões da função caraterística de algumas das distribuições de probabilidade consideradas em §6.3.

Corolário 6.5.2. *Se $X_1, \ldots, X_n$ são v.a.r. independentes com distribuições de Bernoulli de parâmetro $p \in]0,1[$, então*

$$S_n \sim B(n,p).$$

Dem: Para $t \in \mathbb{R}$ temos

$$\phi_{S_n}(t) = \big(1 - p(1 - e^{\,\mathrm{i}t})\big)^n,$$

que não é mais do que a função caraterística duma distribuição $B(n,p)$. Como a função caraterística carateriza a distribuição, concluímos que $X_1 + \cdots + X_n \sim B(n,p)$. ■

Corolário 6.5.3 (estabilidade da lei de Poisson)**.** *Se $X_1, \ldots, X_n$ são v.a.r. independentes com distribuições de Poisson $\mathscr{P}(\lambda_i)$, $i = 1, \ldots, n$, então*

$$S_n \sim \mathscr{P}\big(\lambda_1 + \cdots + \lambda_n\big).$$

Dem: Para $t \in \mathbb{R}$ temos

$$\begin{aligned}\phi_{S_n}(t) &= \exp\big(\lambda_1(e^{\,\mathrm{i}t} - 1)\big) \times \cdots \times \exp\big(\lambda_n(e^{\,\mathrm{i}t} - 1)\big) \\ &= \exp\big((\lambda_1 + \cdots + \lambda_n)(e^{\,\mathrm{i}t} - 1)\big),\end{aligned}$$

que é a função caraterística duma variável de Poisson de parâmetro $\lambda_1 + \cdots + \lambda_n$. ■

Corolário 6.5.4. *Se $X_1, \ldots, X_n$ são v.a.r. independentes com distribuições gama $\gamma(\alpha_i, \lambda)$, $i = 1, \ldots, n$, então*

$$S_n \sim \gamma\big(\alpha_1 + \cdots + \alpha_n, \lambda\big).$$

Dem: Para $t \in \mathbb{R}$ temos

$$\begin{aligned}\phi_{S_n}(t) &= \Big(\frac{\lambda}{\lambda - \,\mathrm{i}t}\Big)^{\alpha_1} \times \cdots \times \Big(\frac{\lambda}{\lambda - \,\mathrm{i}t}\Big)^{\alpha_n} \\ &= \Big(\frac{\lambda}{\lambda - \,\mathrm{i}t}\Big)^{\alpha_1 + \cdots + \alpha_n},\end{aligned}$$

que é a função caraterística duma variável gama de parâmetros $\alpha_1 + \cdots + \alpha_n$ e λ. ■

Uma vez que a distribuição exponencial $\mathrm{Exp}(\lambda)$ é a distribuição $\gamma(1, \lambda)$, concluímos que se $X_1, \ldots, X_n$ são v.a.r. independentes com distribuições $\mathrm{Exp}(\lambda)$ então

$$S_n \sim \gamma\big(n, \lambda\big).$$

Corolário 6.5.5 (estabilidade da lei normal)**.** *Se $X_1, \ldots, X_n$ são v.a.r. independentes com distribuições normais $N(\mu_i, \sigma_i^2)$, $i = 1, \ldots, n$, então*

$$S_n \sim N\big(\mu_1 + \cdots + \mu_n, \sigma_1^2 + \cdots + \sigma_n^2\big).$$

Dem: Para $t \in \mathbb{R}$ temos

$$\begin{aligned}\phi_{S_n}(t) &= e^{\,\mathrm{i}t\mu_1}e^{-\sigma_1^2t^2/2} \times \cdots \times e^{\,\mathrm{i}t\mu_n}e^{-\sigma_n^2t^2/2} \\ &= e^{\,\mathrm{i}t(\mu_1+\cdots+\mu_n)}e^{-(\sigma_1^2+\cdots+\sigma_n^2)t^2/2},\end{aligned}$$

que é a função caraterística duma variável normal de média $\mu_1 + \cdots + \mu_n$ e variância $\sigma_1^2 + \cdots + \sigma_n^2$. ■

No caso particular em que as variáveis $X_1, \ldots, X_n$ são independentes e identicamente distribuídas com distribuição comum normal de média μ e desvio padrão σ, concluímos que

$$\frac{1}{n}S_n \sim N\left(\mu, \frac{\sigma^2}{n}\right),$$

ou ainda

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

6.6 Função caraterística dum vetor aleatório*

A noção de função caraterística que temos vindo a estudar pode ser facilmente extendida ao caso dos vetores aleatórios.

Definição 6.6.1. *Chamamos função caraterística do ve.a. X em $\mathbb{R}^d$ (ou função caraterística de P_X), à função ϕ_X de $\mathbb{R}^d$ em $\mathbb{C}$ definida, para $t \in \mathbb{R}^d$, por*

$$\phi_X(t) = E\big(e^{\,\mathrm{i}t^TX}\big) = E(\cos(t^TX)) + \mathrm{i}E(\sin(t^TX)).$$

Tal como no caso real, notemos que como $|e^{\,\mathrm{i}\langle t,X\rangle}| = 1$, a esperança matemática anterior está bem definida. Enunciamos a seguir algumas propriedades da função caraterística dum vetor aleatório. As primeiras são em tudo semelhantes às da função caraterística duma variável aleatória real.

Proposição 6.6.2. *Se ϕ_X é a função característica do ve.a. X então:*

a) $\phi_X(0) = 1$;

b) $|\phi_X(t)| \leq 1$, para todo o $t \in \mathbb{R}^d$;

c) $\phi_{-X}(t) = \overline{\phi_X(t)}$, para todo o $t \in \mathbb{R}^d$;

d) ϕ_X é uma função contínua.

e) ϕ_X carateriza P_X (isto é, $\phi_X = \phi_Y$ sse $X \sim Y$).

As funções caraterísticas de subvetores dum vetor aleatório X podem ser obtidas facilmente a partir de ϕ_X.

Proposição 6.6.3. *Se $X = (X_1, \ldots, X_d)$ é um ve.a.r. com função caraterística ϕ_X, então, para todo o $\{i_1, \ldots, i_m\} \subset \{1, \ldots, d\}$, o ve.a.r. $(X_{i_1}, \ldots, X_{i_m})$ tem função característica*

$$\phi(t_{i_1}, \ldots, t_{i_m}) = \phi_X(0, \ldots, 0, t_{i_1}, 0, \ldots, 0, t_{i_m}, 0, \ldots, 0)$$

para $(t_{i_1}, \ldots, t_{i_m}) \in \mathbb{R}^m$. Em particular, as margens X_i têm funções caraterísticas

$$\phi_{X_i}(t_i) = \phi_X(0, \ldots, 0, t_i, 0, \ldots, 0)$$

para $t_i \in \mathbb{R}$.

Tal com já fizemos para a função de distribuição, função de probabilidade e densidade de probabilidade, apresentamos no resultado seguinte uma caraterização da independência das margens dum vetor aleatório em termos da sua função caraterística.

Proposição 6.6.4. *Seja $X = (X_1, \ldots, X_d)$ um ve.a. em $\mathbb{R}^d$. $X_1, \ldots, X_d$ são v.a.r. independentes sse*

$$\phi_X(t) = \phi_{X_1}(t_1) \times \cdots \times \phi_{X_d}(t_d),$$

para todo o $t = (t_1, \ldots, t_d) \in \mathbb{R}^d$.

6.7 Vetores aleatórios normais (parte II)*

Uma propriedade importante dos vetores normais (efetivamente é uma caraterização) é que qualquer combinação linear não degenerada (isto é, que não é quase certamente constante) das suas margens é uma variável aleatória normal. Em particular, as margens não degeneradas (isto é, que não são quase certamente constantes) de um vetor aleatório normal são v.a.r. normais. Estamos agora em condições de demonstrar esta propriedade.

Proposição 6.7.1. *Qualquer combinação linear não degenerada das margens de um ve.a. normal é uma v.a.r. normal.*

Dem: Sendo X é um ve.a. normal em $\mathbb{R}^d$ então $X = AZ + b$, com Z um vetor normal standard em $\mathbb{R}^d$, A uma matriz real quadrada de dimensão d e b um vetor em $\mathbb{R}^d$. Para qualquer $\alpha \in \mathbb{R}^d$ temos

$$\alpha^T X = \alpha^T (AZ + b) = (A^T \alpha)^T Z + \alpha^T b = \sum_{k=1}^{d} \beta_k Z_k + \beta,$$

onde $Z = (Z_1, \ldots, Z_d)$, $A^T \alpha = (\beta_1, \ldots, \beta_d)$ e $\alpha^T b = \beta$. Sendo $\alpha^T X$ não degenerada por hipótese, da independência das margens de Z temos $\mathrm{Var}(\alpha^T X) = \sum_{k=1}^{d} \beta_k^2 > 0$, o que, pelo Corolário 6.5.5, permite concluir que $\alpha^T X \sim N(\beta, \sum_{k=1}^{d} \beta_k^2)$. ■

A expressão da função caraterística dum vetor aleatório normal é apresentada no resultado seguinte. Como podemos confirmar, uma tal expressão depende exclusivamente de $E(X)$ e C_X o que permite concluir que a distribuição dum vetor normal fica caraterizada pela respetiva média e matriz de covariância. Por essa razão, e tal como fazemos no caso univariado, é habitual escrever $X \sim N(\mu, \Sigma)$ quando o vetor X possui uma distribuição normal com média μ e matriz de covariância Σ.

Proposição 6.7.2. *Se X é um ve.a. normal em $\mathbb{R}^d$ a sua função caraterística é dada por*

$$\phi_X(t) = \exp(\,\mathrm{i}t^T E(X)) \exp(-t^T C_X t/2),\ t \in \mathbb{R}^d.$$

Dem: Para $t \in \mathbb{R}^d$ temos

$$\phi_X(t) = E\big(e^{\,\mathrm{i}t^T X}\big) = \phi_{t^T X}(1),$$

onde $t^T X$ é uma v.a.r. com $E(t^T X) = t^T E(X)$ e $\mathrm{Var}(t^T X) = t^T C_X t$. Se $t^T X$ é degenerada, então $t^T C_X t = 0$ e $t^T X$ é uma variável de Dirac no ponto $t^T E(X)$. Se $t^T X$ não é degenerada, sabemos da Proposição 6.7.1 que $t^T X$ é uma variável aleatória normal com média $t^T E(X)$ e variância $t^T C_X t$. Usando agora as expressões das funções caraterísticas das leis de Dirac e normal obtidas na Secção 6.3, concluímos que, num e noutro caso,

$$\phi_{t^T X}(1) = e^{\,\mathrm{i}t^T E(X)} e^{-t^T C_X t/2},$$

o que prova o pretendido. ■

Vimos na Secção 5.11 que no caso dos vetores normais bivariados as respetivas margens são independentes quando e só quando são não correlacionadas. Estamos agora em condições de mostrar que esta propriedade é válida para qualquer vetor aleatório normal em $\mathbb{R}^d$.

Proposição 6.7.3. *Se $X = (X_1, \ldots, X_d)$ é um vetor aleatório normal em $\mathbb{R}^d$, então $X_1, \ldots, X_d$ são independentes sse* $\mathrm{Cov}(X_i, X_j) = 0$ *para $i \neq j$.*

Dem: Atendendo ao Corolário 5.7.7, basta provar que a não correlação das variáveis $X_1, \ldots, X_d$ implica a sua independência. Uma vez que

$$t^T E(X) = \sum_{j=1}^{d} t_j E(X_j)$$

e

$$t^T C_X t = \sum_{j,k=1}^{d} t_j t_k \mathrm{Cov}(X_j, X_k) = \sum_{j=1}^{d} t_j^2 \mathrm{Var}(X_j),$$

então, pelo Teorema 6.7.2, temos

$$\begin{aligned}\phi_X(t) &= \exp\Big(\mathrm{i}\sum_{j=1}^{d} t_j E(X_j)\Big)\exp\Big(-\sum_{j=1}^{d} t_j^2 \mathrm{Var}(X_j)/2\Big)\\ &= \prod_{j=1}^{d}\exp(\mathrm{i}t_j E(X_j))\exp(-t_j^2\mathrm{Var}(X_j)/2) = \prod_{j=1}^{d}\phi_{X_j}(t_j),\end{aligned}$$

o que, pela Proposição 6.6.4, permite concluir que $X_1, \ldots, X_d$ são independentes. ■

6.8 Bibliografia

Gonçalves, E., Mendes Lopes, N. (2013). *Probabilidades: princípios teóricos*. Lisboa: Escolar Editora.

Jacod, J., Protter, P. (2000). *Probability essentials*. Berlin: Springer.

James, B.R. (1981). *Probabilidade: um curso em nível intermediário*. Rio de Janeiro: IMPA.

Métivier, M. (1972). *Notions fondamentales de la théorie des probabilités*, Paris: Dunod.

Monfort, A. (1980). *Cours de probabilités*. Paris: Economica.

Resnick, S.I. (2005). *A probability path*. Boston: Birkhäuser.

Tenreiro, C. (2002). *Apontamentos de Teoria das Probabilidades*. Coimbra: DMUC. https://www.mat.uc.pt/∼tenreiro/apontamentos/ATP0204.pdf

7

Leis dos Grandes Números

Convergências quase certa, em probabilidade e em média quadrática. Teorema da convergência dominada de Lebesgue. Primeiras leis dos grandes números: leis fracas de Chebyshev, Bernoulli e Poisson, e lei forte de Borel. Leis dos grandes números de Khintchine e de Kolmogorov. Aplicações à Estatística.

7.1 Convergências funcionais

Sendo $X, X_1, X_2, \ldots$ variáveis aleatórias reais definidas sobre um mesmo espaço de probabilidade $(\Omega, \mathscr{A}, P)$, no presente capítulo obtemos condições suficientes para que a sucessão de variáveis aleatórias reais $\frac{1}{n}S_n$, onde $S_n = X_1 + \cdots + X_n$, convirja, num sentido a precisar, para uma constante real μ quando n tende para infinito. Como veremos, no caso em que as variáveis X_n são identicamente distribuídas com $X_n \sim X$, essa constante não é mais do que a esperança matemática de X. Resultados deste tipo são conhecidos por leis dos grandes números.

Antes de estabelecermos tais leis, começamos por estudar modos de convergência que nos permitirão dar um sentido preciso à convergência anterior.

Definição 7.1.1. *Dizemos que* (X_n) *converge para* X *quase certamente, e escrevemos* $X_n \xrightarrow{qc} X$, *se*

$$P\big(\{\omega \in \Omega : \lim X_n(\omega) = X(\omega)\}\big) = 1.$$

Dizer que a sucessão (X_n) converge para X quase certamente é assim dizer que a menos dum conjunto com probabilidade nula, a sucessão (X_n) converge pontualmente para X. Por outras palavras, existe $\Omega_1 \in \mathscr{A}$, com $P(\Omega_1) = 1$, tal que $\lim X_n(\omega) = X(\omega)$, para todo o $\omega \in \Omega_1$.

Das propriedades dos conjuntos quase certos, isto é, com probabilidade 1, verificamos que as propriedades da convergência quase certa duma sucessão de variáveis aleatórias são essencialmente iguais às da convergência pontual. Uma das exceções é o da não unicidade do limite quase certo. No entanto, mesmo esta propriedade pode ser

recuperada através da identificação de variáveis aleatórias que coincidem a menos dum conjunto de probabilidade nula, isto é, identificando variáveis quase certamente iguais.

Definição 7.1.2. *Dizemos que* (X_n) converge para X em probabilidade, *e escrevemos* $X_n \xrightarrow{p} X$, *se*

$$\forall\, \epsilon > 0 \quad P\big(\{\omega \in \Omega : |X_n(\omega) - X(\omega)| \geq \epsilon\}\big) \to 0,$$

isto é,

$$\forall\, \epsilon > 0 \quad \lim_{n\to\infty} P\big(\{\omega \in \Omega : |X_n(\omega) - X(\omega)| \geq \epsilon\}\big) = 0.$$

Tal como acontece para a convergência quase certa, provemos que se X e Y são limite em probabilidade duma sucessão de variáveis aleatórias então $P(X = Y) = 1$.

Proposição 7.1.3. *Se* $X_n \xrightarrow{p} X$ *e* $X_n \xrightarrow{p} Y$, *então* $X = Y$ *q.c.*

Dem: Uma vez que

$$\{X = Y\} = \{|X - Y| = 0\} = \bigcap_{k=1}^{\infty} \{|X - Y| < 1/k\},$$

basta provar que $P(|X - Y| < \epsilon) = 1$, ou ainda que $P(|X - Y| \geq \epsilon) = 0$, para todo o $\epsilon > 0$. Dado então $\epsilon > 0$, temos

$$\begin{aligned} \{|X - Y| \geq \epsilon\} &= \{|X - X_n + X_n - Y| \geq \epsilon\} \\ &\subset \{|X_n - X| + |X_n - Y| \geq \epsilon\} \\ &\subset \{|X_n - X| \geq \epsilon/2\} \cup \{|X_n - Y| \geq \epsilon/2\}, \end{aligned}$$

e assim

$$P(|X - Y| \geq \epsilon) \leq P(|X_n - X| \geq \epsilon/2) + P(|X_n - Y| \geq \epsilon/2),$$

onde, por hipótese, ambas as parcelas do segundo membro convergem para zero quando n tende para infinito, o que conclui a demonstração. ■

Proposição 7.1.4. *Se* $X_n \xrightarrow{qc} X$, *então* $X_n \xrightarrow{p} X$.

Dem: Comecemos por notar que

$$\{\omega : \lim X_n(\omega) = X(\omega)\} = \bigcap_{\epsilon>0} \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} \{\omega : |X_k(\omega) - X(\omega)| < \epsilon\}, \tag{7.1.5}$$

de onde concluímos que para todo o $\epsilon > 0$ se tem

$$\{\omega : \lim X_n(\omega) = X(\omega)\} \subset \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} \{\omega : |X_k(\omega) - X(\omega)| < \epsilon\}.$$

Por hipótese temos então (simplificando a notação)

$$P\Big(\bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty}\{|X_k - X| < \epsilon\}\Big) = 1,$$

ou ainda,

$$\lim_{n\to\infty} P\Big(\bigcup_{k=n}^{\infty}\{|X_k - X| \geq \epsilon\}\Big) = 0.$$

Por maioria de razão temos também

$$\lim_{n\to\infty} P\big(|X_n - X| \geq \epsilon\big) = 0,$$

o que conclui a demonstração. ■

Não sendo nosso objetivo fazer aqui um estudo detalhado sobre estes e outros modos de convergência estocástica, limitamo-nos a apresentar no resultado seguinte uma condição suficiente para a convergência quase certa que nos será útil a seguir.

Proposição 7.1.6. *Se a sucessão (X_n) de v.a.r. é tal que*

$$\sum_{n=1}^{\infty} P\big(|X_n - X| \geq \epsilon\big) < +\infty, \text{ para todo o } \epsilon > 0,$$

então $X_n \xrightarrow{qc} X$.

Dem: Comecemos por reescrever a igualdade (7.1.5) na forma

$$\{\omega : \lim X_n(\omega) = X(\omega)\} = \bigcap_{m=1}^{\infty}\bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty}\{\omega : |X_k(\omega) - X(\omega)| < 1/m\}.$$

Para provar que $X_n \xrightarrow{qc} X$ basta assim provar que (simplificando a notação)

$$P\Big(\bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty}\{|X_k - X| < \epsilon\}\Big) = 1,$$

para todo o $\epsilon > 0$ (ver Exercício 8), ou ainda que,

$$P\Big(\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty}\{|X_k - X| \geq \epsilon\}\Big) = 0,$$

para todo o $\epsilon > 0$. Atendendo a que

$$\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty}\{|X_k - X| \geq \epsilon\} = \limsup\{|X_n - X| \geq \epsilon\},$$

o resultado é agora consequência do teorema de Borel-Cantelli aplicado aos acontecimentos $A_n = \{|X_n - X| \geq \epsilon\}$ (ver Teorema 2.3.11). ■

As convergências quase certa e em probabilidade são preservadas por transformações contínuas.

Proposição 7.1.7. *Se* $X_n \xrightarrow{p} X$ *(resp.* $X_n \xrightarrow{qc} X$*) então* $g(X_n) \xrightarrow{p} g(X)$ *(resp.* $g(X_n) \xrightarrow{qc} g(X)$*), para toda a função contínua* g *de* $\mathbb{R}$ *em* $\mathbb{R}$*.*

Dem: Demonstramos o resultado apenas para a convergência quase certa. Por hipótese sabemos que existe Ω_1, com $P(\Omega_1) = 1$, tal que $X_n(\omega) \to X(\omega)$, para todo o $\omega \in \Omega_1$. Sendo g contínua temos também $g(X_n(\omega)) \to g(X(\omega))$, para todo o $\omega \in \Omega_1$, o que prova o pretendido. ■

Definição 7.1.8. *Dizemos que a sucessão* (X_n) *de v.a.r. em* $\mathscr{L}^2$ *converge para* X *em média quadrática, e escrevemos* $X_n \xrightarrow{mq} X$*, se*

$$E(X_n - X)^2 \to 0.$$

Notemos que a variável limite X pertence necessariamente a $\mathscr{L}^2$ uma vez que $X = X_n - (X_n - X)$ onde $X_n \in \mathscr{L}^2$ e $X_n - X \in \mathscr{L}^2$. Não existe nenhuma relação geral entre as convergências quase certa e em média quadrática. No entanto, a convergência em média quadrática implica a convergência em probabilidade.

Proposição 7.1.9. *Se* $X_n \xrightarrow{mq} X$ *então* $X_n \xrightarrow{p} X$*.*

Dem: Usando a desigualdade Chebyshev-Markov com $p = 2$ (Proposição 4.3.2), para $\epsilon > 0$ temos

$$P\big(|X_n - X| \geq \epsilon\big) \leq \frac{1}{\epsilon^2} E(X_n - X)^2 \to 0. \qquad ■$$

Quando a sucessão (X_n) converge em média quadrática para a variável X, a sucessão das respetivas médias, $(E(X_n))$, converge para a média da variável limite, $E(X)$. Com efeito, usando a desigualdade de Cauchy-Schwarz (Proposição 5.7.1) com $X_n - X$ no lugar de X e 1 no lugar de Y, temos

$$|E(X_n) - E(X)| = |E(X_n - X)| \leq \sqrt{E(X_n - X)^2} \to 0.$$

Sob certas condições adicionais, a convergência em probabilidade ou a convergência quase certa da sucessão (X_n) para a variável X também implicam a convergência da sucessão $(E(X_n))$ para $E(X)$ (a este propósito, ver o Teorema 6.2.3). Por nos ser útil mais à frente, enunciamos a seguir este importante resultado, cuja demonstração ultrapassa, em muito, os objetivos do curso.

Teorema 7.1.10 (da convergência dominada de Lebesgue)**.** *Se a sucessão* (X_n) *de v.a.r. é tal que*

a) $X_n \xrightarrow{p} X$ *ou* $X_n \xrightarrow{qc} X$*, para alguma v.a.r.* X*;*

b) $|X_n| \leq Y$ *q.c. para todo o* n*, com* $Y \in \mathscr{L}^1$*;*

então $X \in \mathscr{L}^1$ *e* $E(X_n) \to E(X)$*.*

7.2 Primeiras leis dos grandes números

Estudados, mesmo que brevemente, os modos de convergência estocástica quase certa e em probabilidade, dirigimos agora a nossa atenção para o principal objetivo deste capítulo que é, como referimos, o estabelecimento de condições suficientes (e se possível necessárias) para que a sucessão de variáveis aleatórias reais

$$\frac{1}{n}S_n,$$

onde

$$S_n = X_1 + \cdots + X_n,$$

convirja, em probabilidade ou quase certamente, para uma constante real μ. Quando a convergência envolvida é a convergência em probabilidade, dizemos que temos uma lei fraca dos grandes números. Quando a convergência envolvida é a convergência quase certa, dizemos que temos uma lei forte dos grandes números.

No que se segue assumiremos que a sucessão $X_1, \ldots, X_n, \ldots$ é constituída por v.a.r. independentes, ou mais geralmente, que tais variáveis são duas a duas não correlacionadas. Como estabelecemos no resultado seguinte, em ambos os casos a variância da soma S_n exprime-se como soma das variâncias das suas parcelas.

Proposição 7.2.1. *Se $X_1, \ldots, X_n \in \mathscr{L}^2$ então*

$$\mathrm{Var}(S_n) = \sum_{i=1}^{n} \mathrm{Var}(X_i) + 2 \sum_{1 \leq i < j \leq n} \mathrm{Cov}(X_i, X_j).$$

Além disso, se $\mathrm{Cov}(X_i, X_j) = 0$ para $i \neq j$, temos

$$\mathrm{Var}(S_n) = \sum_{i=1}^{n} \mathrm{Var}(X_i).$$

Dem: Da alínea c) da Proposição 5.7.3 concluímos facilmente que a covariância é uma função bilinear, isto é, é linear em cada um dos seus argumentos. Assim,

$$\begin{aligned}
\mathrm{Var}(S_n) &= \mathrm{Cov}\Big(\sum_{i=1}^{n} X_i, \sum_{j=1}^{n} X_j \Big) = \sum_{i,j=1}^{n} \mathrm{Cov}(X_i, X_j) \\
&= \sum_{i=1}^{n} \mathrm{Cov}(X_i, X_i) + \sum_{i,j=1; i \neq j}^{n} \mathrm{Cov}(X_i, X_j) \\
&= \sum_{i=1}^{n} \mathrm{Var}(X_i) + 2 \sum_{1 \leq i < j \leq n} \mathrm{Cov}(X_i, X_j).
\end{aligned}$$

■

Os resultados seguintes estabelecem leis dos grandes números usando técnicas de demonstração simples baseadas em desigualdades de tipo Markov (estudadas na secção 4.3) e no cálculo de momentos de ordem superior ou igual à segunda. Começamos pela lei fraca de P. Tchebychev (1821-1894), a partir da qual deduziremos as leis fracas de J. Bernoulli (1655-1705) e de S.D. Poisson (1781-1840).

Teorema 7.2.2 (Lei fraca de Tchebychev ([1])). *Se (X_n) é uma sucessão de v.a.r. em $\mathscr{L}^2$, duas a duas não correlacionadas, com $\sup_{n\in\mathbb{N}} E(X_n^2) < \infty$, então*

$$\frac{1}{n}\big(S_n - E(S_n)\big) \xrightarrow{p} 0.$$

Além disso, se $E(X_n) = \mu \in \mathbb{R}$, então

$$\frac{1}{n}S_n \xrightarrow{p} \mu.$$

Dem: Usando a desigualdade de Tchebychev-Markov (Proposição 4.3.2) com $p = 2$, ou a desigualdade de Bienaymé-Tchebychev (Proposição 4.3.3), para todo o $\epsilon > 0$ temos

$$P\big(|(S_n - E(S_n))/n| \geq \epsilon\big) \leq \frac{\mathrm{Var}(S_n)}{n^2\epsilon^2},$$

onde, pela Proposição 7.2.1,

$$\mathrm{Var}(S_n) = \sum_{i=1}^{n} \mathrm{Var}(X_i) \leq n \sup_{i\in\mathbb{N}} E(X_i^2).$$

Assim,

$$P\big(|(S_n - E(S_n))/n| \geq \epsilon\big) \leq \frac{1}{n\epsilon^2} \sup_{i\in\mathbb{N}} E(X_i^2) \to 0,\ n \to \infty. \qquad \blacksquare$$

Corolário 7.2.3. *Se (X_n) é uma sucessão de v.a.r. independentes e identicamente distribuídas em $\mathscr{L}^2$ com $X_n \sim X$, então*

$$\frac{1}{n}S_n \xrightarrow{p} \mu = E(X).$$

Corolário 7.2.4 (Lei fraca de Bernoulli ([2])). *Se (X_n) é uma sucessão de v.a.r. independentes com distribuições de Bernoulli de parâmetro $p \in]0,1[$, então*

$$\frac{1}{n}S_n \xrightarrow{p} p.$$

[1] Tchébychef, P.L. (1867). Des valeurs moyennes. *J. Math. Pures Appl. (2^e série)* 12, 177–184.
[2] Bernoulli, J. (1713). *Ars conjectandi*, Basel.

Se A é um acontecimento aleatório com $P(A) \in]0,1[$, sabemos que a variável aleatória X_i que toma o valor 1 se A ocorre e 0 se A não ocorre na i-ésima repetição, sempre nas mesmas condições, duma experiência aleatória, possui uma distribuição de Bernoulli de parâmetro $P(A)$. Neste caso, $S_n = X_1 + \cdots + X_n$ e S_n/n são, respetivamante, o número de vezes e a proporção de vezes, que o acontecimento A ocorre em n repetições da experiência. A S_n/n chamamos também frequência relativa do acontecimento A. A lei fraca de Bernoulli estabelece assim que

$$\frac{1}{n}S_n \xrightarrow{p} P(A),$$

isto é, a frequência relativa do acontecimento A converge (em probabilidade) para a probabilidade do acontecimento A, resultado este que é, por vezes, referido como definição frequencista de probabilidade.

Exemplo 7.2.5. É a lei fraca de Bernoulli (e também a lei forte de Borel de que falaremos a seguir) que justifica as aproximações que apresentámos em §2.4.1 para a sensibilidade e para a especificidade de um teste de diagnóstico. Pretendendo obter a partir de uma amostra de indivíduos doentes uma aproximação para a sensibilidade do teste de diagnóstico para a doença em causa, isto é, para a probabilidade do teste dar positivo quando a doença está presente, de acordo com a lei fraca de Bernoulli uma tal probabilidade pode ser aproximada pela frequência relativa dos indivíduos em que o teste deu positivo. De forma análoga, para obter uma aproximação para a especificidade do teste de diagnóstico para a doença em causa, isto é, para a probabilidade do teste dar negativo quando a doença está ausente, devemos considerar a frequência relativa dos indivíduos em que o teste deu negativo numa amostra de indivíduos saudáveis.

Corolário 7.2.6 (Lei fraca de Poisson ([3]))**.** *Se (X_n) é uma sucessão de v.a.r. independentes com distribuições de Bernoulli de parâmetro $p_n \in]0,1[$, então*

$$\frac{1}{n}S_n - p_n \xrightarrow{p} 0.$$

Todos os resultados anteriores envolveram a convergência em probabilidade, sendo, por isso, leis fracas dos grandes números. No resultado seguinte estabelecemos uma primeira lei forte dos grandes números. Para tal, condições mais restritivas, do que as que até agora considerámos, sobre os momentos das variáveis intervenientes são necessárias.

[3] Poisson, S.D. (1837). *Recherches sur la probabilité des judgements en matière criminale et matière civile*, Paris.

Teorema 7.2.7. *Seja (X_n) uma sucessão de v.a.r. independentes em $\mathscr{L}^4$ com $E(X_n) = \mu \in \mathbb{R}$ para todo o n. Se $\sup_{n\in\mathbb{N}} E(X_n^4) < \infty$, então*

$$\frac{1}{n}S_n \xrightarrow{qc} \mu.$$

Dem: Para simplificar a notação, comecemos por verificar que basta demonstrar o resultado no caso da sucessão (X_n) ser centrada, uma vez que

$$\frac{1}{n}S_n - \mu = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu).$$

Seja então (X_n) uma sucessão de v.a.r. com $E(X_n) = 0$ para todo o n, nas condições do enunciado. Pretendemos provar que $S_n/n \xrightarrow{qc} 0$. Usando a desigualdade de Chebyshev-Markov (Proposição 4.3.2), para todo o $\epsilon > 0$ temos

$$P\big(|S_n/n| \geq \epsilon\big) = P\big(|S_n| \geq n\epsilon\big) \leq \frac{E(S_n^4)}{n^4\epsilon^4},$$

onde

$$E(S_n^4) = E\bigg(\Big(\sum_{i=1}^{n} X_i\Big)^4\bigg) = \sum_{i,j,k,l=1}^{n} E(X_iX_jX_kX_l).$$

Uma vez que as variáveis X_i são independentes com média zero, são nulas todas as parcelas da soma anterior, com exceção daquelas em que os quatro índices são iguais, ou aquelas em que os índices são iguais dois a dois. Assim,

$$\begin{aligned} E(S_n^4) &= \sum_{i=1}^{n} E(X_i^4) + 3\sum_{i,j=1,i\neq j}^{n} E(X_i^2)E(X_j^2) \\ &\leq (n + 3n(n-1)) \sup_{n\in\mathbb{N}} E(X_n^4) \leq 3n^2 \sup_{n\in\mathbb{N}} E(X_n^4), \end{aligned}$$

o que permite concluir que

$$P\big(|S_n/n| \geq \epsilon\big) \leq \frac{3\sup_{n\in\mathbb{N}} E(X_n^4)}{n^2\epsilon^4}.$$

Usando agora a Proposição 7.1.6, concluímos finalmente que $\dfrac{S_n}{n} \xrightarrow{qc} 0$. ■

Corolário 7.2.8 (Lei forte de Borel ([4]))**.** *Se (X_n) é uma sucessão de v.a.r. independentes com distribuições de Bernoulli de parâmetro $p \in]0,1[$, então*

$$\frac{1}{n}S_n \xrightarrow{qc} p.$$

[4]Borel, E. (1909). Les probabilités dénombrables et leurs applications arithmétiques. *Rend. Circ. Mat. Palermo* 27, 247–271.

7.3 Leis de Khintchine e de Kolmogorov

Refinando a técnica de demonstração é possível obter leis fracas e fortes dos grandes números sobre condições menos restritivas que as consideradas anteriormente. Tal acontece, por exemplo, no caso das sucessões de variáveis aleatórias reais independentes e identicamente distribuídas. Vamos limitar-nos aqui a apresentar, sem demonstração, a lei fraca de A. Khintchine (1894-1959) e a lei forte de A. Kolmogorov (1903-1987).

Teorema 7.3.1 (Lei fraca de Khintchine ([5])). *Se (X_n) é uma sucessão de v.a.r. independentes e identicamente distribuídas com $X_n \sim X \in \mathscr{L}^1$, então*

$$\frac{1}{n} S_n \xrightarrow{p} \mu = E(X).$$

A lei fraca de Khintchine tem um interesse meramente histórico. Pouco depois de ter sido estabelecida, sob as mesmas condições teóricas Kolmogorov estabelece uma lei forte dos grandes números. Além disso, mostra que a existência da esperança matemática das variáveis da sucessão envolvida é condição não só suficiente, mas também necessária para que S_n/n convirja quase certamente para um valor real μ. É este resultado fundamental que a literatura matemática guarda em lugar de destaque sob a designação de lei forte dos grandes números de Kolmogorov.

Teorema 7.3.2 (Lei forte de Kolmogorov ([6])). *Seja (X_n) uma sucessão de v.a.r. independentes e identicamente distribuídas com $X_n \sim X$. Então existe $\mu \in \mathbb{R}$ tal que*

$$\frac{1}{n} S_n \xrightarrow{qc} \mu$$

sse $X \in \mathscr{L}^1$. Nesse caso $\mu = E(X)$.

7.4 Aplicações à Estatística

Na área da estatística matemática os resultados anteriores são de grande importância permitindo justificar, de um ponto de vista assintótico (isto é, quando n tende para infinito), a prática comum de, quando recolhemos uma amostra de uma população tomarmos a média amostral ou média empírica ($\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$), para estimar a respetiva média populacional ($\mu = E(X)$). Com efeito, a partir da lei dos grandes números de Kolmogorov sabemos que se (X_n) é uma sucessão de v.a.r. independentes com $X_n \sim X$ e $X \in \mathscr{L}^1$, então

$$\bar{X}_n \xrightarrow{qc} E(X).$$

[5] Khintchine, A. (1929). Sur la loi des grands nombres. *C. R. Acad. Sci. Paris* 188, 477–479.

[6] Kolmogorov, A.N. (1933). *Grundbegriffe der Wahrscheinlichkeitrechnung*, Berlin (o resultado é aqui enunciado sem demonstração).

Usando novamente a lei dos grandes números de Kolmogorov é possível estabelecer resultados semelhantes quando pretendemos estimar a variância $\mathrm{Var}(X)$, ou mesmo a covariância $\mathrm{Cov}(X,Y)$ a partir de observações independentes do vetor (X,Y).

Proposição 7.4.1. *Se (X_n) é uma sucessão de v.a.r. independentes com $X_n \sim X$ e $X \in \mathscr{L}^2$, então*

$$S_X^2 \xrightarrow{qc} \mathrm{Var}(X),$$

onde $S_X^2 = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$ é a variância empírica ou variância amostral.

Dem: Tendo em conta que

$$S_X^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}_n^2,$$

do Teorema 7.3.2 sabemos que

$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 \xrightarrow{qc} E(X^2)$$

e que

$$\bar{X}_n^2 \xrightarrow{qc} (E(X))^2.$$

Assim,

$$S_X^2 \xrightarrow{qc} E(X^2) - (E(X))^2 = \mathrm{Var}(X). \qquad \blacksquare$$

Proposição 7.4.2. *Se (X_n, Y_n) uma sucessão de ve.a.r. independentes com $(X_n, Y_n) \sim (X,Y)$ e $X, Y \in \mathscr{L}^2$, então*

$$S_{XY} \xrightarrow{qc} \mathrm{Cov}(X,Y),$$

onde $S_{XY} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(X_i - \bar{X}_n)(Y_i - \bar{Y}_n)$ é a covariância empírica ou covariância amostral.

Dem: Tendo em conta que

$$S_{XY} = \frac{1}{n}\sum_{i=1}^{n} X_i Y_i - \bar{X}_n\,\bar{Y}_n,$$

do Teorema 7.3.2 sabemos que

$$\frac{1}{n}\sum_{i=1}^{n} X_i Y_i \xrightarrow{qc} E(XY)$$

e que

$$\bar{X}_n\,\bar{Y}_n \xrightarrow{qc} E(X)E(Y).$$

Assim,

$$S_{XY} \xrightarrow{qc} E(XY) - E(X)E(Y) = \mathrm{Cov}(X,Y). \qquad \blacksquare$$

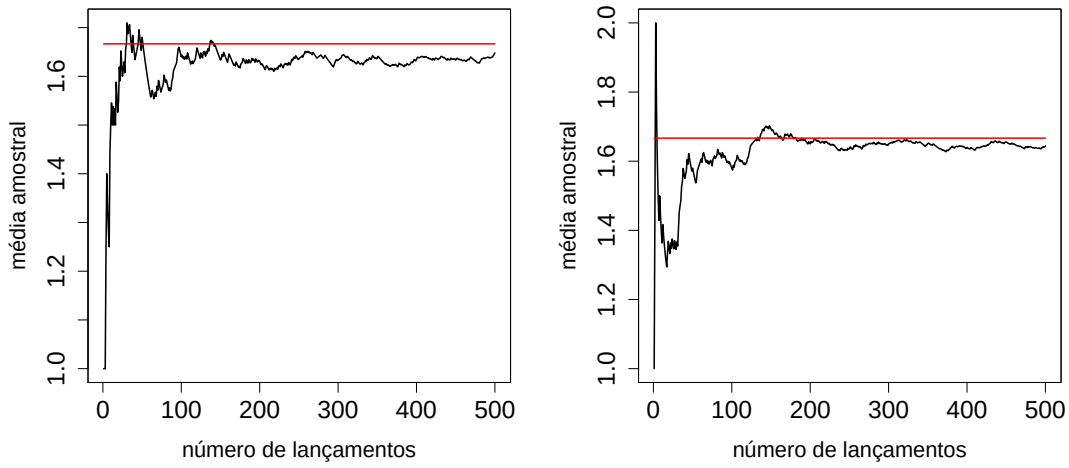


Figura 7.1: Evoluções do número médio de pontos por lançamento ao longo de 500 lançamentos.

7.5 Duas ilustrações

Nos dois exemplos seguintes ilustramos as leis dos grandes números estabelecidas anteriormente.

Exemplo 7.5.1. Suponhamos que um dado equilibrado tem marcados os números 1, em três das faces, 2, em duas das faces, e o número 3 na face restante. Se X representar o número de pontos obtidos num lançamento do dado, a distribuição de probabilidade de X é dada por

valores de X	1	2	3
probabilidade	1/2	1/3	1/6

sendo a média de X igual a $E(X) = 5/3$. A lei dos grandes números é ilustrada na Figura 7.1 onde se mostra a evolução da média amostral S_n/n (número médio de pontos por lançamento) com o aumento das observações, para dois conjuntos de observações $X_1, \ldots, X_n$ da variável X, onde X_i representa o número de pontos obtidos no i-ésimo lançamento do dado.

Exemplo 7.5.2. No jogo da roleta, a roda da roleta está dividida em 37 partes iguais numeradas de 0 a 36, e um jogador, que à partida aposta num dos números de 1 a 36, recebe em caso de vitória 36 vezes mais do que aquilo que apostou. Admitindo que a aposta do jogador é sempre de 10 euros, ele recebe os 10 euros que apostou mais 350 euros pagos pelo casino se sair o número em que apostou. Caso contrário, perde o que apostou. Representando por X o ganho líquido do jogador em cada partida, X tem como distribuição de probabilidade

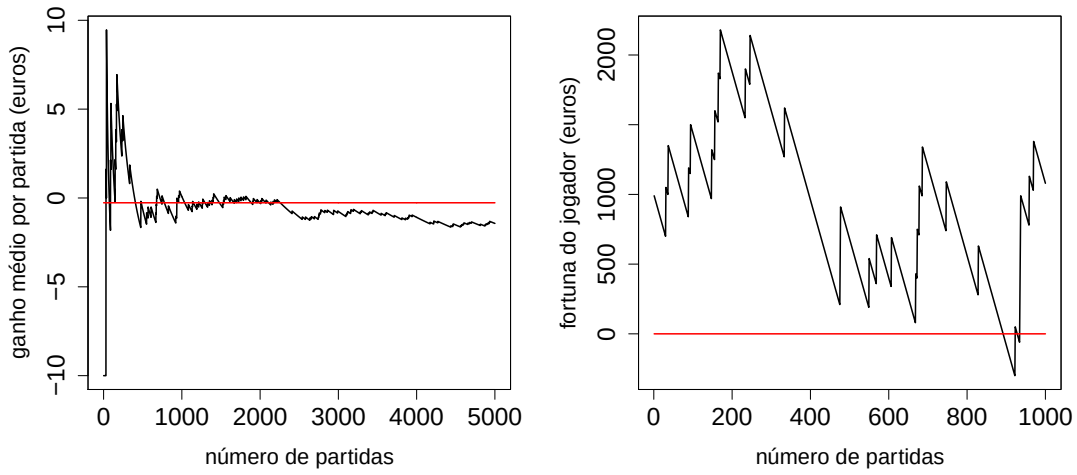


Figura 7.2: Evolução do ganho médio por partida e correspondente evolução da fortuna de um jogador.

valores de X	-10	350
probabilidade	$36/37$	$1/37$

O ganho médio por partida é dado por

$$E(X) = \frac{36}{37} \times (-10) + \frac{1}{37} \times 350 = -\frac{10}{37} = -0{,}27,$$

isto é, em cada partida o jogador perde em média 27 cêntimos. Atendendo à lei dos grandes números, quer isto dizer que, independentemente do dinheiro que o jogador leva para o casino, ao fim dum grande número de partidas ficará sem dinheiro nenhum.

Para ilustrar os factos referidos, apresentamos na Figura 7.2 a evolução da média amostral, ou seja, do ganho médio por partida de um jogador com uma grande fortuna inicial, e também a correspondente evolução da fortuna, até pouco depois de começar a ficar em dívida, de um jogador que entra para o casino com 1000 euros para jogar na roleta.

7.6 Bibliografia

Gonçalves, E., Mendes Lopes, N. (2013). *Probabilidades: princípios teóricos*. Lisboa: Escolar Editora.

Jacod, J., Protter, P. (2000). *Probability essentials*. Berlin: Springer.

James, B.R. (1981). *Probabilidade: um curso em nível intermediário*. Rio de Janeiro: IMPA.

Monfort, A. (1980). *Cours de probabilités*. Paris: Economica.

Resnick, S.I. (2005). *A probability path*. Boston: Birkhäuser.

Tenreiro, C. (2002). *Apontamentos de Teoria das Probabilidades*. Coimbra: DMUC. https://www.mat.uc.pt/∼tenreiro/apontamentos/ATP0204.pdf

8

Teorema do Limite Central

Introdução. Convergência em distribuição. Caraterizações e propriedades da convergência em distribuição. O Teorema de Lévy-Bochner. O Teorema do Limite Central de Laplace-Liapounov. Algumas consequências e aplicações.

8.1 Introdução

Terminámos o último capítulo analisando no Exemplo 7.5.2 a evolução da fortuna de um jogador de roleta cujo ganho líquido por partida, que representámos por X, possuía uma distribuição de probabilidade dada por

valores de X	-10	350
probabilidade	$36/37$	$1/37$

Suponhamos agora que o jogador decide jogar n partidas numa das suas idas ao casino. Será possível obter uma aproximação para a probabilidade dele ganhar mais do que aquilo que perde nessas n partidas? Denotando por X_i o que o jogador ganha na i-ésima partida, o ganho líquido obtido pelo jogador no final das n partidas é dado por

$$S_n = X_1 + X_2 + \cdots + X_n.$$

Assim, pretendemos saber se é possível obter uma aproximação para a probabilidade

$$P(S_n > 0).$$

A lei dos grandes números (Teorema 7.3.2), que estabelece, como vimos, que $S_n/n \xrightarrow{qc} -0{,}27$, apenas nos permite concluir que

$$P(S_n > 0) = P(\frac{1}{n}S_n > 0) \approx 0,$$

quando n é "grande". Para obtermos uma aproximação mais informativa para a probabilidade anterior necessitamos de informação adicional sobre a distribuição de probabilidade de S_n quando n é grande, informação essa que nos será dada pelo teorema

do limite central que estudaremos neste capítulo. Sob certas condições, veremos que, independentemente da distribuição das variáveis $X_1, \ldots, X_n$, a distribuição de probabilidade de S_n pode ser aproximada, quando n tende para $+\infty$, por uma distribuição normal, o que nos permitirá obter uma aproximação para a probabilidade $P(S_n > 0)$ a partir da distribuição normal.

Antes de estabelecermos o teorema do limite central, começamos por estudar um novo modo de convergência, a que chamaremos convergência em distribuição, que nos permitirá formalizar a referida aproximação da distribuição de probabilidade de S_n por uma distribuição normal.

8.2 Convergência em distribuição

A noção de convergência duma sucessão (X_n) de v.a.r. para uma v.a.r. X que estudamos neste parágrafo é de natureza distinta das convergências funcionais consideradas em §7.1. Para tais modos de convergência interessam os valores particulares que tomam as variáveis X_n e X em pontos do espaço onde estão definidas. Para a noção de convergência que a seguir consideramos interessam apenas as probabilidades com que essas variáveis tomam tais valores. Os diferentes modos de convergência são habitualmente designados por modos de convergência estocástica.

Se X é uma v.a.r. com função de distribuição F_X, denotaremos por $C(F_X)$ o conjunto dos pontos de continuidade de F_X. Sabemos já que o conjunto dos pontos de descontinuidade de F_X, que não é mais do que o conjunto dos ponto $x \in \mathbb{R}$, para os quais $P(X = x) = 0$ (ver Proposição 3.3.4.e), é finito ou infinito numerável. Assim $C(F_X)$ denso em $\mathbb{R}$, isto é, qualquer intervalo aberto $]a, b[$ contém pontos de $C(F_X)$.

Definição 8.2.1. *Dizemos que uma sucessão (X_n) de v.a.r., não necessariamente definidas num mesmo espaço de probabilidade,* converge em distribuição (ou em lei) *para uma v.a.r. X, e escrevemos $X_n \xrightarrow{d} X$, se*

$$\lim F_{X_n}(x) = F_X(x), \ \forall\, x \in C(F_X).$$

Notemos que seria desapropriado impor que a condição anterior fosse verificada para todo o ponto de $\mathbb{R}$ como ilustra o exemplo da sucessão $X_n = 1/n$, que deverá convergir para $X = 0$ segundo um qualquer modo de convergência aceitável. Reparemos que $F_{X_n}(x)$ converge para $F_X(x)$, para todo o $x \in \mathbb{R}$, com exceção do ponto $x = 0$, único ponto de descontinuidade de F_X. No caso da sucessão $X_n = -1/n$, $F_{X_n}(x)$ converge para $F_X(x)$, para todo o $x \in \mathbb{R}$.

O exemplo da sucessão $X_n = (-1)^n X$, onde $X \sim N(0, 1)$, é ilustrativo da diferença entre a noção de convergência agora introduzida e as anteriormente estudadas, uma

vez que $X_n \sim X$, o que implica que $X_n \xrightarrow{d} X$, e no entanto X_n não converge em probabilidade para X (ver Exercício 79).

Provemos agora a unicidade do limite em distribuição no sentido seguinte:

Proposição 8.2.2. *Se $X_n \xrightarrow{d} X$ e $X_n \xrightarrow{d} Y$, então $X \sim Y$.*

Dem: Por hipótese $F_X(x) = F_Y(x)$, para todo o $x \in C(F_X) \cap C(F_Y)$. Atendendo agora a que $C(F_X) \cap C(F_Y)$ é denso em $\mathbb{R}$ (porquê?), se $x \notin C(F_X) \cap C(F_Y)$ existe uma sucessão (x_n) com $x_n \in C(F_X) \cap C(F_Y)$ e $x_n > x$ tal que $\lim x_n = x$. Assim, uma vez que F_X e F_Y são contínuas à direita e coincidem em $C(F_X) \cap C(F_Y)$, temos

$$F_X(x) = \lim F_X(x_n) = \lim F_Y(x_n) = F_Y(x).$$

Concluímos assim que $F_X(x) = F_Y(x)$, para todo o $x \in \mathbb{R}$, ou seja, $X \sim Y$. ■

Proposição 8.2.3. *As seguintes proposições são equivalentes:*

i) $X_n \xrightarrow{d} X$;

ii) $P(X_n \in A) \to P(X \in A)$, *para todo o* $A \in \mathscr{B}(\mathbb{R})$ *com* $P(X \in fr(A)) = 0$;

iii) $E(f(X_n)) \to E(f(X))$, *para toda a função* f *contínua e limitada de* $\mathbb{R}$ *em* $\mathbb{R}$.

Dem: A demonstração completa deste resultado ultrapassa os objetivos do curso. A implicação *ii)* $\Rightarrow$ *i)* é a única que é simples de obter. Com efeito, para $x \in C(F_X)$, uma vez que $A =]-\infty, x]$, é tal que $P(X \in fr(A)) = P(X = x) = 0$, então por hipótese temos $P(X_n \in]-\infty, x]) \to P(X \in]-\infty, x])$, ou seja, $F_{X_n}(x) \to F_X(x)$. Das restantes implicações, limitamo-nos a estabelecer *i)* $\Rightarrow$ *ii)* quando $A =]a, b]$ com $P(X \in fr(A)) = 0$. Uma vez que $fr(A) = \{a, b\}$ então $P(X = a) = 0$ e $P(X = b) = 0$, ou seja, $a, b \in C(F_X)$. Usando agora a hipótese, sabemos que $F_{X_n}(a) \to F_X(a)$ e $F_{X_n}(b) \to F_X(b)$, de onde tiramos que

$$\begin{aligned} P(X_n \in A) = P\big(X_n \in]a, b]\big) &= F_{X_n}(b) - F_{X_n}(a) \\ &\to F_X(b) - F_X(a) = P\big(X \in]a, b]\big) = P(X \in A). \end{aligned}$$ ■

Tal como para os outros modos de convergência estudados, a convergência em distribuição é preservada por transformações contínuas.

Proposição 8.2.4. *Se $X_n \xrightarrow{d} X$ então $g(X_n) \xrightarrow{d} g(X)$, para toda a função contínua g de $\mathbb{R}$ em $\mathbb{R}$.*

Dem: Usando a alínea *iii)* da Proposição 8.2.3, seja f uma função contínua e limitada de $\mathbb{R}$ em $\mathbb{R}$. Uma vez que a composição $f \circ g$ é uma função contínua e limitada, por hipótese e pela alínea c) da Proposição 8.2.3 temos que $E((f \circ g)(X_n)) \to E((f \circ g)(X))$, ou ainda $E(f(g(X_n))) \to E(f(g(X))$, o que permite concluir. ■

O resultado seguinte dá-nos condições suficientes para a convergência em distribuição.

Teorema 8.2.5 (de Scheffé ([1])). *a) Se (X_n) e X são v.a.r. discretas que tomam valores num conjunto finito ou numerável S com $f_{X_n}(x) \to f_X(x)$, para todo o $x \in S$, então $X_n \xrightarrow{d} X$.*

b) Se (X_n) e X são v.a.r. absolutamente contínuas com $f_{X_n}(x) \to f_X(x)$, em quase todo o ponto $x \in \mathbb{R}$, então $X_n \xrightarrow{d} X$;

Dem: Vamos demonstrar apenas a) no caso em que $S = \mathbb{N}_0$. Para $x < 0$, temos $F_{X_n}(x) = 0$ e $F_X(x) = 0$, e portanto $F_{X_n}(x) \to F_X(x)$. Para $x \geq 0$, temos

$$\begin{aligned} F_{X_n}(x) = P(X_n \leq x) &= \sum_{k \in \mathbb{N}_0, k \leq x} P(X_n = k) = \sum_{k \in \mathbb{N}_0, k \leq x} f_{X_n}(k) \\ &\to \sum_{k \in \mathbb{N}_0, k \leq x} f_X(k) = \sum_{k \in \mathbb{N}_0, k \leq x} P(X = k) = F_X(x). \end{aligned}$$

■

Quando no Capítulo 3 falámos da lei de Poisson (ver §3.5.1), mostrámos que se $X_n \sim B(n, p_n)$, onde $p_n \in]0,1[$ é tal que $\lim np_n = \lambda \in]0, +\infty[$, então

$$P(X_n = k) \to P(X = k),$$

para todo o $k \in \mathbb{N}_0$, onde $X \sim \mathscr{P}(\lambda)$. Tendo em conta a alínea a) do resultado anterior, podemos então dizer que

$$X_n \xrightarrow{d} X,$$

isto é, a distribuição binomial $B(n, p_n)$ converge para a distribuição de Poisson $\mathscr{P}(\lambda)$, quando $\lim np_n = \lambda$. De forma simbólica, esta convergência é por vezes traduzida por

$$B(n, p_n) \approx \mathscr{P}(np_n),$$

o que põe em relevo o facto da probabilidade $P(X_n = k)$, para $k \in \{0, 1, \ldots, n\}$, poder ser aproximada, quando p_n é próximo de zero e n é grande, pela probabilidade $P(X = k)$, onde $X \sim \mathscr{P}(np_n)$.

Estabelecemos a seguir que a convergência em distribuição é implicada pela convergência em probabilidade (e também, por maioria de razão, pela convergência quase certa).

Proposição 8.2.6. *Se $X_n \xrightarrow{p} X$ então $X_n \xrightarrow{d} X$.*

[1] Scheffé, H. (1947). A useful convergence theorem for probability distributions. *Ann. Math. Statist.* 28, 434–458.

Dem: Usando a alínea *iii)* da Proposição 8.2.3, seja f uma função contínua e limitada de $\mathbb{R}$ em $\mathbb{R}$. Uma vez que $X_n \xrightarrow{p} X$, concluímos que $f(X_n) \xrightarrow{p} f(X)$ por f ser contínua (Proposição 7.1.7). Por outro lado, sendo f limitada, estamos nas condições do teorema da convergência dominada de Lebesgue (Teorema 7.1.10), o que permite concluir que $E(f(X_n)) \xrightarrow{p} E(f(X))$. ■

Proposição 8.2.7. *Se $X_n \xrightarrow{d} a$, com $a \in \mathbb{R}$, então $X_n \xrightarrow{p} a$.*

Dem: Se $X = a$ então $F_X = \mathbb{1}_{[a,+\infty[}$ e $C(F_X) = \mathbb{R} \setminus \{a\}$. Assim, por hipótese, temos $\lim F_{X_n}(x) = 0$, se $x < a$, e $\lim F_{X_n}(x) = 1$, se $x > a$. Dado $\epsilon > 0$, temos então

$$\begin{aligned} P(|X_n - a| \geq \epsilon) &= 1 - P(|X_n - a| < \epsilon) \\ &= 1 - P(a - \epsilon < X_n < a + \epsilon) \\ &\leq 1 - P(a - \epsilon/2 < X_n \leq a + \epsilon/2) \\ &= 1 - F_{X_n}(a + \epsilon/2) + F_{X_n}(a - \epsilon/2) \to 0. \end{aligned}$$ ■

O resultado seguinte, que nos será muito útil nas secções seguintes, dá-nos uma caraterização da convergência em distribuição de uma sucessão de v.a.r. em termos da convergência das respetivas funções caraterísticas.

Teorema 8.2.8 (da continuidade de Lévy ([2]))**.** *Se $\phi_{X_n}(t) \to \phi_\infty(t)$, para todo o $t \in \mathbb{R}$, onde ϕ_∞ é contínua na origem, então $X_n \xrightarrow{d} X$ para alguma v.a.r. X e $\phi_X = \phi_\infty$.*

Corolário 8.2.9. *$X_n \xrightarrow{d} X$ sse $\phi_{X_n}(t) \to \phi_X(t)$, para todo o $t \in \mathbb{R}$.*

Dem: Para $t \in \mathbb{R}$ fixo, as funções $x \mapsto \cos(tx)$ e $x \mapsto \sin(tx)$ são contínuas e limitadas em $\mathbb{R}$. Usando a Proposição 8.2.3 concluímos que se $X_n \xrightarrow{d} X$ então $E(\cos(tX_n)) \to E(\cos(tX))$ e $E(\sin(tX_n)) \to E(\sin(tX))$, e portanto

$$\begin{aligned} \phi_{X_n}(t) &= E(\mathrm{e}^{\,\mathrm{i}tX_n}) \\ &= E(\cos(tX_n)) + \mathrm{i}E(\sin(tX_n)) \\ &\to E(\cos(tX)) + \mathrm{i}E(\sin(tX)) \\ &= E(\mathrm{e}^{\,\mathrm{i}tX}) = \phi_X(t). \end{aligned}$$

A implicação recíproca é consequência do Teorema 8.2.8 e da alínea e) da Proposição 6.2.2, uma vez que ϕ_X é contínua na origem (efetivamente, é contínua em $\mathbb{R}$). ■

[2] Lévy, P. (1922). Sur la détermination des lois de probabilité par leurs fonctions caractéristiques. *C. R. Acad. Sci. Paris* 175, 854–856; Bochner, S. (1933). Monotone funktionen, Stieltjessche integrale und harmonische analyse. *Mathematische Annalen* 108, 378–410.

8.3 O Teorema do Limite Central: dois exemplos

Se $X_1, \ldots, X_n, \ldots$ são variáveis aleatórias independentes e identicamente distribuídas com distribuições normais de média μ e desvio padrão σ, sabemos pela lei fraca dos grandes números que

$$\frac{1}{n} S_n \xrightarrow{p} \mu,$$

onde

$$S_n = X_1 + \cdots + X_n.$$

Sendo a convergência em distribuição implicada pela convergência em probabilidade, a distribuição assintótica de S_n/n é assim degenerada (isto é, de Dirac). Uma tal distribuição assintótica não é útil na avaliação da probabilidade de acontecimentos do tipo

$$\left\{ S_n/n \in]a, b] \right\}$$

para $a, b \in \mathbb{R}$. Com efeito, atendendo à Proposição 8.2.3, se $a, b \neq \mu$ (μ é o único ponto de descontinuidade da distribuição assintótica), apenas podemos dizer que

$$P\big(S_n/n \in]a, b]\big) \to P\big(\mu \in]a, b]\big) = \begin{cases} 0, & \text{se } \mu \notin]a, b] \\ 1, & \text{se } \mu \in]a, b]. \end{cases}$$

No entanto, para todo o $n \in \mathbb{N}$, sabemos que

$$\frac{1}{n} S_n \sim N\left(\mu, \frac{\sigma^2}{n}\right),$$

ou ainda,

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

o que implica que

$$\frac{\sqrt{n}}{\sigma}\left(\frac{1}{n} S_n - \mu\right) \xrightarrow{d} N(0, 1). \tag{8.3.1}$$

Concluímos assim que apesar de S_n/n possuir uma distribuição assintótica (isto é, quando $n \to \infty$) degenerada, quando convenientemente normalizada (centragem e redução) a variável S_n/n possui uma distribuição assintótica não degenerada.

O facto de uma tal distribuição assintótica ser normal, não é, como veremos neste capítulo, uma propriedade exclusiva das variáveis normais.

Exemplo 8.3.2. Consideremos o caso em que $X_1, \ldots, X_n$ são v.a.r. exponenciais de parâmetro $\lambda = 1$. Sabemos que $\phi_{X_k}(t) = 1/(1 - \mathrm{i}t)$ e assim

$$\phi_{S_n}(t) = \frac{1}{(1 - \mathrm{i}t)^n},\ t \in \mathbb{R},$$

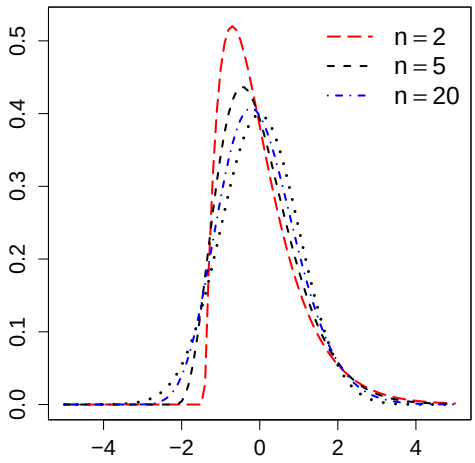


Figura 8.1: Densidade de probabilidade de $\frac{S_n/n-\mu}{\sigma/\sqrt{n}}$ quando $X_1, \ldots, X_n \sim \text{Exp}(1)$, e densidade normal centrada e reduzida.

o que permite concluir que S_n possui uma distribuição gama de parâmetros $\alpha = n$ e $\lambda = 1$ (ver §6.3). Tendo em conta que $\mu = E(X_k) = 1$, $\sigma^2 = \text{Var}(X_k) = 1$, na Figura 8.1 comparamos a densidade de probabilidade da variável

$$\frac{\sqrt{n}}{\sigma}\left(\frac{1}{n}S_n - \mu\right) = \frac{1}{\sqrt{n}}S_n - \sqrt{n}$$

com a da densidade normal centrada e reduzida para vários valores de n. Atendendo ao Teorema 8.2.5.b), os gráficos sugerem que a convergência em distribuição (8.3.1) também ocorre neste caso.

Exemplo 8.3.3. Consideremos agora o caso em que $X_1, X_2, \ldots$ são v.a.r. independentes com $P(X_k = \pm 1) = 1/2$. Procedendo como no caso da distribuição binomial, podemos determinar a distribuição de probabilidade da variável S_n que tem como suporte o conjunto $\{-n, -n+2, \ldots, n-2, n\}$. Analisando separadamente os casos em que n é par ou é ímpar, concluímos que

$$P(S_{2n} = \pm 2k) = C_{n+k}^{2n}\left(\frac{1}{2}\right)^{2n} \quad \text{e} \quad P(S_{2n+1} = \pm(2k+1)) = C_{n+k}^{2n+1}\left(\frac{1}{2}\right)^{2n+1},$$

para $k = 0, \ldots, n$.

Por outro lado, sabemos que

$$\phi_{X_k}(t) = E(\cos(tX_k)) = \frac{1}{2}\cos(-t) + \frac{1}{2}\cos(t) = \cos(t), \ t \in \mathbb{R}.$$

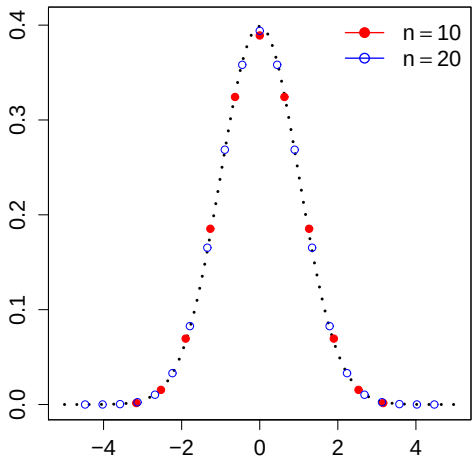


Figura 8.2: Gráficos da função $x \mapsto \sqrt{n}\, P(S_n/\sqrt{n} = x)/2$, para $x \in \mathscr{S}_n$, e da densidade normal centrada e reduzida.

Denotando por ϕ_n a função caraterística de $\frac{S_n/n-\mu}{\sigma/\sqrt{n}} = \frac{S_n}{\sqrt{n}}$, onde $\mu = E(X_k) = 0$ e $\sigma^2 = \mathrm{Var}(X_k) = 1$, para $t \in \mathbb{R}$ temos

$$\phi_n(t) = \phi_{S_n}(t/\sqrt{n}) = \big(\cos(t/\sqrt{n})\big)^n = \left(1 + \frac{x_n(t)}{n}\right)^n,$$

onde

$$x_n(t) = -\frac{t^2}{2} + \frac{t^4}{4!n} - \cdots \to -\frac{t^2}{2}.$$

Concluímos assim que

$$\phi_n(t) \to e^{-t^2/2} = \phi_{N(0,1)}(t),$$

para todo o $t \in \mathbb{R}$ (note que se $x_n \to x$ então $(1+x_n/n)^n \to \mathrm{e}^x$), o que, pelo teorema de Lévy-Bochner, permite concluir que tem lugar a convergência em distribuição (8.3.1).

Sendo $\mathscr{S}_n = \{-\sqrt{n}, -\sqrt{n}+2/\sqrt{n}, \ldots, \sqrt{n}-2/\sqrt{n}, \sqrt{n}\}$ o suporte da variável $S_n/\sqrt{n}$, tendo em conta a convergência anterior, podemos esperar que, para n grande e para $x \in \mathscr{S}_n$, se tenha

$$\begin{aligned} P\left(\frac{S_n}{\sqrt{n}} = x\right) &= P\left(\frac{S_n}{\sqrt{n}} \in \left]x - \frac{1}{\sqrt{n}}, x + \frac{1}{\sqrt{n}}\right]\right) \\ &\approx P\left(Z \in \left]x - \frac{1}{\sqrt{n}}, x + \frac{1}{\sqrt{n}}\right]\right) \\ &= \frac{2}{\sqrt{n}} \frac{F_{N(0,1)}\big(x + 1/\sqrt{n}\big) - F_{N(0,1)}\big(x - 1/\sqrt{n}\big)}{2/\sqrt{n}} \\ &\approx \frac{2}{\sqrt{n}}\, f_{N(0,1)}(x) \end{aligned}$$

A aproximação anterior é ilustrada na Figura 8.2, onde se apresentam, para alguns valores de n e para $x \in \mathscr{S}_n$, o gráfico da função $x \mapsto \sqrt{n}P(S_n/\sqrt{n} = x)/2$, bem como o da densidade normal centrada e reduzida.

8.4 O Teorema do Limite Central clássico

Mostramos a seguir que a convergência em distribuição (8.3.1) ocorre para uma vasta família de variáveis aleatórias. Um resultado deste tipo é conhecido como teorema do limite central ou teorema central do limite. Para que possamos generalizar os argumentos utilizados no Exemplo 8.3.3 a outras distribuições de probabilidade, é essencial o resultado seguinte que não é mais do que um desenvolvimento de Taylor para a função caraterística onde o respetivo resto é apresentado numa forma que nos será útil (ver Teorema 6.4.1).

Lema 8.4.1. *Se $X \in \mathscr{L}^2$, então para todo o $t \in \mathbb{R}$,*

$$\phi_X(t) = 1 + \mathrm{i}tE(X) - \frac{t^2}{2}E(X^2) + r_2(t),$$

onde

$$|r_2(t)| \leq E\bigg(\min\Big(\frac{|tX|^3}{6}, |tX|^2\Big)\bigg).$$

Dem: Sendo g uma função real de variável real com derivada de terceira ordem contínua em $\mathbb{R}$, sabemos da fórmula de Taylor com resto integral que, para $x \in \mathbb{R}$, vale o desenvolvimento

$$g(x) = g(0) + g'(0)x + g''(0)\frac{x^2}{2} + \frac{1}{2}x^3\int_0^1 (1-s)^2 g'''(sx)ds.$$

Um desenvolvimento semelhante é válido para a função $x \mapsto \mathrm{e}^{\mathrm{i}x} = \cos x + \mathrm{i}\sin x$, bastando aplicar o desenvolvimento anterior à parte real e à parte imaginária de $\mathrm{e}^{\mathrm{i}x}$. Assim, $x \in \mathbb{R}$, temos

$$\mathrm{e}^{\mathrm{i}x} = 1 + \mathrm{i}x - \frac{x^2}{2} - \frac{\mathrm{i}x^3}{2}\int_0^1 (1-s)^2\mathrm{e}^{\mathrm{i}sx}ds = 1 + \mathrm{i}x - \frac{x^2}{2} + r(x),$$

onde

$$r(x) = -\frac{\mathrm{i}x^3}{2}\int_0^1 (1-s)^2\mathrm{e}^{\mathrm{i}sx}ds.$$

Por um lado, temos

$$|r(x)| \leq \frac{|x|^3}{2}\int_0^1 (1-s)^2 ds = \frac{|x|^3}{6}.$$

Por outro lado, integrando por partes temos

$$\begin{aligned} r(x) &= -\frac{x^2}{2}\int_0^1 (1-s)^2\, \mathrm{i}x\mathrm{e}^{\mathrm{i}sx}ds \\ &= -\frac{x^2}{2}\left((1-s)^2\mathrm{e}^{\mathrm{i}sx}\Big]_0^1 + 2\int_0^1 (1-s)\mathrm{e}^{\mathrm{i}sx}ds\right) \\ &= \frac{x^2}{2}\left(1 - 2\int_0^1 (1-s)\mathrm{e}^{\mathrm{i}sx}ds\right), \end{aligned}$$

de onde tiramos que

$$|r(x)| \leq \frac{x^2}{2}\left(1 + 2\int_0^1 (1-s)ds\right) = x^2.$$

Em resumo, para $x \in \mathbb{R}$ temos que

$$\mathrm{e}^{\mathrm{i}x} = 1 + \mathrm{i}x - \frac{x^2}{2} + r(x),$$

onde

$$|r(x)| \leq \min\left(\frac{|x|^3}{6}, x^2\right).$$

Fazendo agora $x = tX$ e tomando esperança matemática em ambos os membros da igualdade anterior obtemos

$$\phi_X(t) = E\big(\mathrm{e}^{\mathrm{i}tX}\big) = 1 + \mathrm{i}tE(X) - \frac{t^2}{2}E(X^2) + r_2(t),$$

onde $r_2(t) = E(r(tX))$ é tal que

$$|r_2(t)| = |E(r(tX))| \leq E|r(tX)| = E\left(\min\left(\frac{|tX|^3}{6}, |tX|^2\right)\right),$$

o que conclui a demonstração. ■

Teorema 8.4.2 (do limite central de Laplace-Liapounov ([3])). *Seja (X_n) uma sucessão de v.a.r. independentes e identicamente distribuídas em $\mathscr{L}^2$, com $X_n \sim X$, onde $E(X) = \mu$ e $\mathrm{Var}(X) = \sigma^2 > 0$. Então*

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

[3]Laplace, P.S. (1810). Mémoire sur les approximations des formules qui sont fonctions de très grands nombres d'observations, et sur leur applications aux probabilités. *Mém. Acad. Sci. Paris* 10, 353–415, 559–565; Liapounov, A. (1900). Sur une proposition de théorie des probabilités. *Bull. Acad. Sci. St. Petersbourg* 13, 359–386.

Dem: Atendendo a que

$$\frac{\sqrt{n}}{\sigma}\left(\frac{1}{n}S_n - \mu\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{X_i - \mu}{\sigma},$$

onde $\frac{X_i-\mu}{\sigma}$ são variáveis independentes e identicamente distribuídas em $\mathscr{L}^2$ com média zero e variância unitária, basta considerar o caso $\mu = 0$ e $\sigma = 1$. Denotando por ϕ_n a função caraterística de $S_n/\sqrt{n}$, para $t \in \mathbb{R}$, temos

$$\phi_n(t) = \phi_{S_n}(t/\sqrt{n}) = \big(\phi_X(t/\sqrt{n})\big)^n,$$

onde, pelo Lema 8.4.1 e pelo teorema da convergência dominada,

$$\begin{aligned}\phi_X(t/\sqrt{n}) &= 1 + \mathrm{i}tE(X)/\sqrt{n} - t^2E(X)^2/(2n) + r_2(t/\sqrt{n})\\ &= 1 - t^2/(2n) + r_2(t/\sqrt{n})\\ &= 1 + \frac{-t^2/2 + nr_2(t/\sqrt{n})}{n},\end{aligned}$$

e

$$\big|nr_2(t/\sqrt{n})\big| \leq nE\bigg(\min\Big(\frac{|tX|^3}{6n^{3/2}}, \frac{|tX|^2}{n}\Big)\bigg) = E\bigg(\min\Big(\frac{|tX|^3}{6n^{1/2}}, |tX|^2\Big)\bigg) \to 0.$$

Finalmente, para todo o $t \in \mathbb{R}$,

$$\phi_n(t) = \left(1 + \frac{-t^2/2 + nr_2(t/\sqrt{n})}{n}\right)^n \to e^{-t^2/2} = \phi_{N(0,1)}(t),$$

o que, pelo teorema de Lévy-Bochner, conclui a demonstração. ■

8.5 Algumas consequências

Os resultados seguintes são consequências imediatas do teorema do limite central de Laplace-Liapounov (Teorema 8.4.2) e dos Corolários 6.5.2, 6.5.3 e 6.5.4. Os dois primeiros resultados são ilustrados pelos gráficos da Figura 3.4 e pelo gráfico da esquerda da Figura 3.5.

Proposição 8.5.1. *Se (X_n) é uma sucessão de v.a.r. com $X_n \sim B(n,p)$ e $p \in]0,1[$, então*

$$(np(1-p))^{-1/2}\big(X_n - np\big) \xrightarrow{d} N(0,1),$$

ou seja

$$B(n,p) \approx N\big(np, np(1-p)\big).$$

Proposição 8.5.2. *Se (X_n) é uma sucessão de v.a.r. com $X_n \sim \mathscr{P}(n\lambda)$ e $\lambda > 0$, então*

$$(n\lambda)^{-1/2}\big(X_n - n\lambda\big) \xrightarrow{d} N(0,1),$$

ou seja,

$$\mathscr{P}(n\lambda) \approx N(n\lambda, n\lambda).$$

Proposição 8.5.3. *Se (X_n) é uma sucessão de v.a.r. com $X_n \sim \gamma(n\alpha, \lambda)$ e $\alpha, \lambda > 0$, então*

$$\lambda(n\alpha)^{-1/2}\Big(X_n - \frac{n\alpha}{\lambda}\Big) \xrightarrow{d} N(0,1),$$

ou seja,

$$\gamma(n\alpha, \lambda) \approx N\left(\frac{n\alpha}{\lambda}, \frac{n\alpha}{\lambda^2}\right).$$

8.6 Duas ilustrações

Retomamos aqui os dois exemplos considerados em §7.5. Em ambos os casos o cálculo das probabilidades envolvidas é feito à custa do teorema do limite central que estabelece que para n grande é válida a aproximação em distribuição

$$\frac{1}{n}S_n \approx N\left(\mu, \frac{\sigma^2}{n}\right),$$

ou ainda

$$S_n \approx N\big(n\mu, n\sigma^2\big),$$

onde $S_n = X_1 + \cdots + X_n$ e $X_1, \ldots, X_n$ são v.a.r. independentes e identicamente distribuídas em $\mathscr{L}^2$, com $E(X_1) = \mu$ e $\mathrm{Var}(X_1) = \sigma^2 > 0$.

Exemplo 8.6.1. Suponhamos que decidimos lançar o dado descrito no Exemplo 7.5.1 (pág. 153) 100 vezes consecutivas, e que apostamos com um amigo que vamos obter pelo menos 165 pontos na soma dos pontos obtidos nos vários lançamentos, e com outro amigo que vamos obter mais de 170 pontos. Qual é a probabilidade de ganharmos a aposta com cada um dos amigos? Se representarmos por $X_1, X_2, \ldots, X_{100}$ os pontos obtidos em cada um dos 100 lançamentos e por S_{100} a sua soma, isto é, $S_{100} = X_1 + X_2 + \cdots + X_{100}$, as probabilidades pedidas são dadas por $P(S_{100} \geq 165)$ e $P(S_{100} > 170)$, respetivamente. As variáveis são independentes e identicamente distribuídas com $E(X_1) = 5/3$ e $\mathrm{Var}(X_1) = 5/9$. Atendendo ao teorema do limite central, sabemos que

$$S_{100} \approx N\left(100\frac{5}{3}, 100\,\frac{5}{9}\right).$$

Assim, denotando por Z a variável normal standard, temos (a primeira igualdade, chamada correção de continuidade, tem em conta o facto de S_n tomar valores de 1 em 1 pontos)

$$\begin{aligned} P(S_{100} \geq 165) &= P(S_{100} \geq 164{,}5) \\ &\approx P\left(Z \geq \frac{164{,}5 - 100(5/3)}{10\sqrt{5/9}}\right) \\ &= P(Z \geq -0{,}291) = 0{,}614, \end{aligned}$$

e

$$\begin{aligned} P(S_{100} > 170) &= P(S_{100} > 170{,}5) \\ &\approx P\left(Z > \frac{170{,}5 - 100(5/3)}{10\sqrt{5/9}}\right) \\ &\approx P(Z > 0{,}514) = 0{,}304. \end{aligned}$$

Em vez de utilizarmos o teorema do limite central, as probabilidades anteriores podiam ser aproximadas se tivessemos a possibilidade de repetir muitas vezes a experiência anterior. Com efeito, se 100 lançamentos do dado fossem repetidos M vezes ($M = 5000$, por exemplo), bastaria calcular as frequências relativas dos acontecimentos anteriores nas M repetições da experiência para obter aproximações para as suas probabilidades. Um tal procedimento é justificado pelas lei dos grandes números de Bernoulli (Corolário 7.2.4) e de Borel (Corolário 7.2.8). Para repetir M vezes 100 lançamentos do dado, vamos usar um computador e o método de simulação mencionado em §3.7, também chamado método de Monte Carlo.

Para gerar n lançamentos do dado podemos usar a função seguinte escrita em R:

```
rdado = function(n)
{
 u <- runif(n)
 1*(u<3/6)+2*(3/6<=u)*(u<5/6)+3*(u>=5/6)
}
```

As frequências relativas dos acontecimentos $\{S_{100} \geq 165\}$ e $\{S_{100} > 170\}$ podem ser obtidas usando o código R:

```
n <- 100 # numero de lancamentos
M <- 5000 # numero de repeticoes da experiencia
somapontos <- array(dim=M)
for (i in 1:M)
```

```
  somapontos[i] <- sum(rdado(n))
sum(somapontos>=165)/M
sum(somapontos>170)/M
```

Usando a instrução set.seed(346762) para fixar a semente do gerador de números aleatórios, obtemos as aproximações

$$P(S_{100} \geq 165) \approx 0{,}6104 \quad \text{e} \quad P(S_{100} > 170) \approx 0{,}3074,$$

que são muito próximas das obtidas anteriormente a partir do teorema do limite central.

Exemplo 8.6.2. Suponhamos que no jogo da roleta descrito no Exemplo 7.5.2 (pág. 153), o jogador decide jogar 150 partidas numa das suas idas ao casino. Calculemos uma aproximação para a probabilidade dele ganhar mais do que aquilo que perde. Representando por X_i o ganho (ou perda) líquido do jogador na i-ésima partida, o ganho líquido do jogador no fim das n partidas é dado por $S_n = X_1 + X_2 + \cdots + X_n$. As variáveis X_i são independentes, identicamente distribuídas com média $-10/37$ euros e desvio padrão $2160/37$ euros. Usando o teorema do limite central, sabemos que

$$S_{150} \approx N\big(-150\,(10/37); 150\,(2160/37)^2\big).$$

Assim, denotando por Z a variável normal standard, temos

$$\begin{aligned} P(S_{150} > 0) &\approx P\left(Z > \frac{0 - (-150\,(10/37))}{(2160/37)\sqrt{150}}\right) \\ &= P(Z > 0{,}0567) = 0{,}4773. \end{aligned}$$

Vejamos agora o que acontece à probabilidade anterior, se o jogador decide jogar 1500 e 15000 partidas em vez de 150. Quando joga 1500 partidas temos

$$S_{1500} \approx N\big(-1500\,(10/37); 1500\,(2160/37)^2\big),$$

e assim

$$\begin{aligned} P(S_{1500} > 0) &\approx P\left(Z > \frac{0 - (-1500\,(10/37))}{(2160/37)\sqrt{1500}}\right) \\ &= P(Z > 0{,}1793) = 0{,}4289. \end{aligned}$$

Quando joga 15000 partidas, temos

$$S_{15000} \approx N\big(-15000\,(10/37); 15000\,(2160/37)^2\big),$$

e

$$\begin{aligned} P(S_{15000} > 0) &\approx P\left(Z > \frac{0 - (-15000\,(10/37))}{(2160/37)\sqrt{15000}}\right) \\ &= P(Z > 0{,}5670) = 0{,}2854. \end{aligned}$$

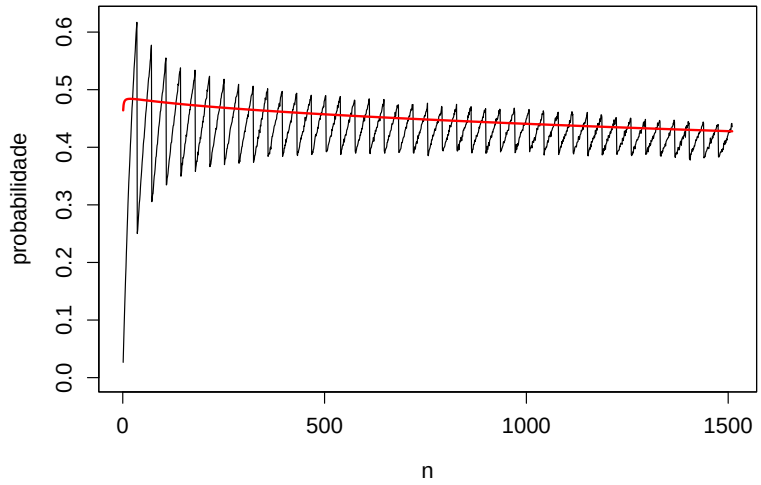


Figura 8.3: Aproximações da função $n \mapsto P(S_n > 0)$, para $n \in \{1, \ldots, 1500\}$, obtidas pelo método de Monte Carlo com $M = 10000$ (a preto) e pelo teorema do limite central (a vermelho).

Também neste caso é simples usar o método de Monte Carlo para obter aproximações para as probabilidades anteriores. Para simularmos o ganho do jogador em n partidas da roleta podemos usar a função seguinte escrita em R:

```
rganho = function(n)
{
 u <- runif(n)
 350*(u<1/37)-10*(u>=1/37)
}
```

As frequências relativas dos acontecimentos $\{S_{150} > 0\}$, $\{S_{1500} > 0\}$ e $\{S_{15000} > 0\}$ podem ser obtidas usando o código R:

```
n <- 150 # numero de partidas
M<- 5000 # numero de repeticoes da experiencia
set.seed(346762)
ganhototal <- array(dim=M)
for (i in 1:M)
 ganhototal[i] <- sum(rganho(n))
sum(ganhototal>0)/M
```

Usando a instrução set.seed(346762) para fixar a semente do gerador de números

aleatórios, obtemos as aproximações

$$P(S_{150} > 0) \approx 0{,}3712,\ P(S_{1500} > 0) \approx 0{,}4218 \text{ e } P(S_{15000} > 0) \approx 0{,}2800.$$

que, nos últimos casos, são muito próximas das obtidas anteriormente a partir do teorema do limite central. No caso $n = 150$, a aproximação obtida pelo teorema de limite central é fraca, refletindo o caráter assintótico de teorema do limite central, mas também o comportamento da função $n \mapsto P(S_n > 0)$, cujas aproximações, obtidas pelo método de Monte Carlo e pelo teorema do limite central, são apresentadas na Figura 8.3.

8.7 Bibliografia

Gonçalves, E., Mendes Lopes, N. (2013). *Probabilidades: princípios teóricos*. Lisboa: Escolar Editora.

Jacod, J., Protter, P. (2000). *Probability essentials*. Berlin: Springer.

James, B.R. (1981). *Probabilidade: um curso em nível intermediário*. Rio de Janeiro: IMPA.

Kallenberg, O. (1997). *Foundations of modern probability*. New York: Springer.

Monfort, A. (1980). *Cours de probabilités*. Paris: Economica.

Resnick, S.I. (2005). *A probability path*. Boston: Birkhäuser.

Tenreiro, C. (2002). *Apontamentos de Teoria das Probabilidades*. Coimbra: DMUC. https://www.mat.uc.pt/∼tenreiro/apontamentos/ATP0204.pdf

Tabelas das distribuições Binomial, Poisson e Normal Standard

$$P(X \leq k),\ X \sim B(n,p)$$

n	$k \backslash p$	0,05	0,10	0,15	0,20	0,25	0,30	0,35	0,40	0,45	0,50
1	0	0,9500	0,9000	0,8500	0,8000	0,7500	0,7000	0,6500	0,6000	0,5500	0,5000
2	0	0,9025	0,8100	0,7225	0,6400	0,5625	0,4900	0,4225	0,3600	0,3025	0,2500
	1	0,9975	0,9900	0,9775	0,9600	0,9375	0,9100	0,8775	0,8400	0,7975	0,7500
3	0	0,8574	0,7290	0,6141	0,5120	0,4219	0,3430	0,2746	0,2160	0,1664	0,1250
	1	0,9928	0,9720	0,9393	0,8960	0,8438	0,7840	0,7183	0,6480	0,5748	0,5000
	2	0,9999	0,9990	0,9966	0,9920	0,9844	0,9730	0,9571	0,9360	0,9089	0,8750
4	0	0,8145	0,6561	0,5220	0,4096	0,3164	0,2401	0,1785	0,1296	0,0915	0,0625
	1	0,9860	0,9477	0,8905	0,8192	0,7383	0,6517	0,5630	0,4752	0,3910	0,3125
	2	0,9995	0,9963	0,9880	0,9728	0,9492	0,9163	0,8735	0,8208	0,7585	0,6875
	3	1,0000	0,9999	0,9995	0,9984	0,9961	0,9919	0,9850	0,9744	0,9590	0,9375
5	0	0,7738	0,5905	0,4437	0,3277	0,2373	0,1681	0,1160	0,0778	0,0503	0,0313
	1	0,9774	0,9185	0,8352	0,7373	0,6328	0,5282	0,4284	0,3370	0,2562	0,1875
	2	0,9988	0,9914	0,9734	0,9421	0,8965	0,8369	0,7648	0,6826	0,5931	0,5000
	3	1,0000	0,9995	0,9978	0,9933	0,9844	0,9692	0,9460	0,9130	0,8688	0,8125
	4	1,0000	1,0000	0,9999	0,9997	0,9990	0,9976	0,9947	0,9898	0,9815	0,9688
6	0	0,7351	0,5314	0,3771	0,2621	0,1780	0,1176	0,0754	0,0467	0,0277	0,0156
	1	0,9672	0,8857	0,7765	0,6554	0,5339	0,4202	0,3191	0,2333	0,1636	0,1094
	2	0,9978	0,9842	0,9527	0,9011	0,8306	0,7443	0,6471	0,5443	0,4415	0,3438
	3	0,9999	0,9987	0,9941	0,9830	0,9624	0,9295	0,8826	0,8208	0,7447	0,6563
	4	1,0000	0,9999	0,9996	0,9984	0,9954	0,9891	0,9777	0,9590	0,9308	0,8906
	5	1,0000	1,0000	1,0000	0,9999	0,9998	0,9993	0,9982	0,9959	0,9917	0,9844
7	0	0,6983	0,4783	0,3206	0,2097	0,1335	0,0824	0,0490	0,0280	0,0152	0,0078
	1	0,9556	0,8503	0,7166	0,5767	0,4449	0,3294	0,2338	0,1586	0,1024	0,0625
	2	0,9962	0,9743	0,9262	0,8520	0,7564	0,6471	0,5323	0,4199	0,3164	0,2266
	3	0,9998	0,9973	0,9879	0,9667	0,9294	0,8740	0,8002	0,7102	0,6083	0,5000
	4	1,0000	0,9998	0,9988	0,9953	0,9871	0,9712	0,9444	0,9037	0,8471	0,7734
	5	1,0000	1,0000	0,9999	0,9996	0,9987	0,9962	0,9910	0,9812	0,9643	0,9375
	6	1,0000	1,0000	1,0000	1,0000	0,9999	0,9998	0,9994	0,9984	0,9963	0,9922
8	0	0,6634	0,4305	0,2725	0,1678	0,1001	0,0576	0,0319	0,0168	0,0084	0,0039
	1	0,9428	0,8131	0,6572	0,5033	0,3671	0,2553	0,1691	0,1064	0,0632	0,0352
	2	0,9942	0,9619	0,8948	0,7969	0,6785	0,5518	0,4278	0,3154	0,2201	0,1445
	3	0,9996	0,9950	0,9786	0,9437	0,8862	0,8059	0,7064	0,5941	0,4770	0,3633
	4	1,0000	0,9996	0,9971	0,9896	0,9727	0,9420	0,8939	0,8263	0,7396	0,6367
	5	1,0000	1,0000	0,9998	0,9988	0,9958	0,9887	0,9747	0,9502	0,9115	0,8555
	6	1,0000	1,0000	1,0000	0,9999	0,9996	0,9987	0,9964	0,9915	0,9819	0,9648
	7	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9998	0,9993	0,9983	0,9961

$$P(X \leq k),\ X \sim B(n,p)$$

n	$k\backslash p$	0,05	0,10	0,15	0,20	0,25	0,30	0,35	0,40	0,45	0,50
9	0	0,6302	0,3874	0,2316	0,1342	0,0751	0,0404	0,0207	0,0101	0,0046	0,0020
	1	0,9288	0,7748	0,5995	0,4362	0,3003	0,1960	0,1211	0,0705	0,0385	0,0195
	2	0,9916	0,9470	0,8591	0,7382	0,6007	0,4628	0,3373	0,2318	0,1495	0,0898
	3	0,9994	0,9917	0,9661	0,9144	0,8343	0,7297	0,6089	0,4826	0,3614	0,2539
	4	1,0000	0,9991	0,9944	0,9804	0,9511	0,9012	0,8283	0,7334	0,6214	0,5000
	5	1,0000	0,9999	0,9994	0,9969	0,9900	0,9747	0,9464	0,9006	0,8342	0,7461
	6	1,0000	1,0000	1,0000	0,9997	0,9987	0,9957	0,9888	0,9750	0,9502	0,9102
	7	1,0000	1,0000	1,0000	1,0000	0,9999	0,9996	0,9986	0,9962	0,9909	0,9805
	8	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9997	0,9992	0,9980
10	0	0,5987	0,3487	0,1969	0,1074	0,0563	0,0282	0,0135	0,0060	0,0025	0,0010
	1	0,9139	0,7361	0,5443	0,3758	0,2440	0,1493	0,0860	0,0464	0,0233	0,0107
	2	0,9885	0,9298	0,8202	0,6778	0,5256	0,3828	0,2616	0,1673	0,0996	0,0547
	3	0,9990	0,9872	0,9500	0,8791	0,7759	0,6496	0,5138	0,3823	0,2660	0,1719
	4	0,9999	0,9984	0,9901	0,9672	0,9219	0,8497	0,7515	0,6331	0,5044	0,3770
	5	1,0000	0,9999	0,9986	0,9936	0,9803	0,9527	0,9051	0,8338	0,7384	0,6230
	6	1,0000	1,0000	0,9999	0,9991	0,9965	0,9894	0,9740	0,9452	0,8980	0,8281
	7	1,0000	1,0000	1,0000	0,9999	0,9996	0,9984	0,9952	0,9877	0,9726	0,9453
	8	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9995	0,9983	0,9955	0,9893
	9	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9997	0,9990
11	0	0,5688	0,3138	0,1673	0,0859	0,0422	0,0198	0,0088	0,0036	0,0014	0,0005
	1	0,8981	0,6974	0,4922	0,3221	0,1971	0,1130	0,0606	0,0302	0,0139	0,0059
	2	0,9848	0,9104	0,7788	0,6174	0,4552	0,3127	0,2001	0,1189	0,0652	0,0327
	3	0,9984	0,9815	0,9306	0,8389	0,7133	0,5696	0,4256	0,2963	0,1911	0,1133
	4	0,9999	0,9972	0,9841	0,9496	0,8854	0,7897	0,6683	0,5328	0,3971	0,2744
	5	1,0000	0,9997	0,9973	0,9883	0,9657	0,9218	0,8513	0,7535	0,6331	0,5000
	6	1,0000	1,0000	0,9997	0,9980	0,9924	0,9784	0,9499	0,9006	0,8262	0,7256
	7	1,0000	1,0000	1,0000	0,9998	0,9988	0,9957	0,9878	0,9707	0,9390	0,8867
	8	1,0000	1,0000	1,0000	1,0000	0,9999	0,9994	0,9980	0,9941	0,9852	0,9673
	9	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9998	0,9993	0,9978	0,9941
	10	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9998	0,9995
12	0	0,5404	0,2824	0,1422	0,0687	0,0317	0,0138	0,0057	0,0022	0,0008	0,0002
	1	0,8816	0,6590	0,4435	0,2749	0,1584	0,0850	0,0424	0,0196	0,0083	0,0032
	2	0,9804	0,8891	0,7358	0,5583	0,3907	0,2528	0,1513	0,0834	0,0421	0,0193
	3	0,9978	0,9744	0,9078	0,7946	0,6488	0,4925	0,3467	0,2253	0,1345	0,0730
	4	0,9998	0,9957	0,9761	0,9274	0,8424	0,7237	0,5833	0,4382	0,3044	0,1938
	5	1,0000	0,9995	0,9954	0,9806	0,9456	0,8822	0,7873	0,6652	0,5269	0,3872
	6	1,0000	0,9999	0,9993	0,9961	0,9857	0,9614	0,9154	0,8418	0,7393	0,6128
	7	1,0000	1,0000	0,9999	0,9994	0,9972	0,9905	0,9745	0,9427	0,8883	0,8062
	8	1,0000	1,0000	1,0000	0,9999	0,9996	0,9983	0,9944	0,9847	0,9644	0,9270
	9	1,0000	1,0000	1,0000	1,0000	1,0000	0,9998	0,9992	0,9972	0,9921	0,9807
	10	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9997	0,9989	0,9968
	11	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9998

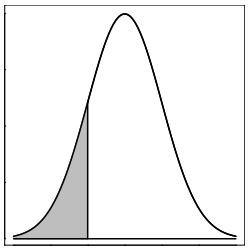
$$P(X \leq k),\ X \sim \mathscr{P}(\lambda)$$

$k\backslash\lambda$	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
0	0,9048	0,8187	0,7408	0,6703	0,6065	0,5488	0,4966	0,4493	0,4066
1	0,9953	0,9825	0,9631	0,9384	0,9098	0,8781	0,8442	0,8088	0,7725
2	0,9998	0,9989	0,9964	0,9921	0,9856	0,9769	0,9659	0,9526	0,9371
3	1,0000	0,9999	0,9997	0,9992	0,9982	0,9966	0,9942	0,9909	0,9865
4	1,0000	1,0000	1,0000	0,9999	0,9998	0,9996	0,9992	0,9986	0,9977
5	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9998	0,9997
6	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
$k\backslash\lambda$	1,0	1,5	2,0	2,5	3,0	3,5	4,0	4,5	5,0
0	0,3679	0,2231	0,1353	0,0821	0,0498	0,0302	0,0183	0,0111	0,0067
1	0,7358	0,5578	0,4060	0,2873	0,1991	0,1359	0,0916	0,0611	0,0404
2	0,9197	0,8088	0,6767	0,5438	0,4232	0,3208	0,2381	0,1736	0,1247
3	0,9810	0,9344	0,8571	0,7576	0,6472	0,5366	0,4335	0,3423	0,2650
4	0,9963	0,9814	0,9473	0,8912	0,8153	0,7254	0,6288	0,5321	0,4405
5	0,9994	0,9955	0,9834	0,9580	0,9161	0,8576	0,7851	0,7029	0,6160
6	0,9999	0,9991	0,9955	0,9858	0,9665	0,9347	0,8893	0,8311	0,7622
7	1,0000	0,9998	0,9989	0,9958	0,9881	0,9733	0,9489	0,9134	0,8666
8	1,0000	1,0000	0,9998	0,9989	0,9962	0,9901	0,9786	0,9597	0,9319
9	1,0000	1,0000	1,0000	0,9997	0,9989	0,9967	0,9919	0,9829	0,9682
10	1,0000	1,0000	1,0000	0,9999	0,9997	0,9990	0,9972	0,9933	0,9863
11	1,0000	1,0000	1,0000	1,0000	0,9999	0,9997	0,9991	0,9976	0,9945
12	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9997	0,9992	0,9980
13	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9997	0,9993
14	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9998
15	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999
16	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
$k\backslash\lambda$	5,5	6,0	6,5	7,0	7,5	8,0	8,5	9,0	9,5
0	0,0041	0,0025	0,0015	0,0009	0,0006	0,0003	0,0002	0,0001	0,0001
1	0,0266	0,0174	0,0113	0,0073	0,0047	0,0030	0,0019	0,0012	0,0008
2	0,0884	0,0620	0,0430	0,0296	0,0203	0,0138	0,0093	0,0062	0,0042
3	0,2017	0,1512	0,1118	0,0818	0,0591	0,0424	0,0301	0,0212	0,0149
4	0,3575	0,2851	0,2237	0,1730	0,1321	0,0996	0,0744	0,0550	0,0403
5	0,5289	0,4457	0,3690	0,3007	0,2414	0,1912	0,1496	0,1157	0,0885
6	0,6860	0,6063	0,5265	0,4497	0,3782	0,3134	0,2562	0,2068	0,1649
7	0,8095	0,7440	0,6728	0,5987	0,5246	0,4530	0,3856	0,3239	0,2687
8	0,8944	0,8472	0,7916	0,7291	0,6620	0,5925	0,5231	0,4557	0,3918
9	0,9462	0,9161	0,8774	0,8305	0,7764	0,7166	0,6530	0,5874	0,5218
10	0,9747	0,9574	0,9332	0,9015	0,8622	0,8159	0,7634	0,7060	0,6453
11	0,9890	0,9799	0,9661	0,9467	0,9208	0,8881	0,8487	0,8030	0,7520
12	0,9955	0,9912	0,9840	0,9730	0,9573	0,9362	0,9091	0,8758	0,8364
13	0,9983	0,9964	0,9929	0,9872	0,9784	0,9658	0,9486	0,9261	0,8981
14	0,9994	0,9986	0,9970	0,9943	0,9897	0,9827	0,9726	0,9585	0,9400
15	0,9998	0,9995	0,9988	0,9976	0,9954	0,9918	0,9862	0,9780	0,9665
16	0,9999	0,9998	0,9996	0,9990	0,9980	0,9963	0,9934	0,9889	0,9823
17	1,0000	0,9999	0,9998	0,9996	0,9992	0,9984	0,9970	0,9947	0,9911
18	1,0000	1,0000	0,9999	0,9999	0,9997	0,9993	0,9987	0,9976	0,9957
19	1,0000	1,0000	1,0000	1,0000	0,9999	0,9997	0,9995	0,9989	0,9980
20	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9998	0,9996	0,9991
21	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9998	0,9996
22	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9999
23	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999
24	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000

$$P(X \leq k),\ X \sim \mathscr{P}(\lambda)$$

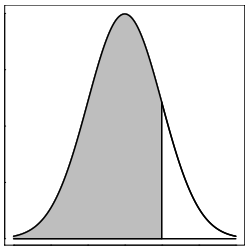
$k\backslash\lambda$	10,0	11,0	12,0	13,0	14,0	15,0	16,0	17,0	18,0	19,0
0	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
1	0,0005	0,0002	0,0001	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
2	0,0028	0,0012	0,0005	0,0002	0,0001	0,0000	0,0000	0,0000	0,0000	0,0000
3	0,0103	0,0049	0,0023	0,0011	0,0005	0,0002	0,0001	0,0000	0,0000	0,0000
4	0,0293	0,0151	0,0076	0,0037	0,0018	0,0009	0,0004	0,0002	0,0001	0,0000
5	0,0671	0,0375	0,0203	0,0107	0,0055	0,0028	0,0014	0,0007	0,0003	0,0002
6	0,1301	0,0786	0,0458	0,0259	0,0142	0,0076	0,0040	0,0021	0,0010	0,0005
7	0,2202	0,1432	0,0895	0,0540	0,0316	0,0180	0,0100	0,0054	0,0029	0,0015
8	0,3328	0,2320	0,1550	0,0998	0,0621	0,0374	0,0220	0,0126	0,0071	0,0039
9	0,4579	0,3405	0,2424	0,1658	0,1094	0,0699	0,0433	0,0261	0,0154	0,0089
10	0,5830	0,4599	0,3472	0,2517	0,1757	0,1185	0,0774	0,0491	0,0304	0,0183
11	0,6968	0,5793	0,4616	0,3532	0,2600	0,1848	0,1270	0,0847	0,0549	0,0347
12	0,7916	0,6887	0,5760	0,4631	0,3585	0,2676	0,1931	0,1350	0,0917	0,0606
13	0,8645	0,7813	0,6815	0,5730	0,4644	0,3632	0,2745	0,2009	0,1426	0,0984
14	0,9165	0,8540	0,7720	0,6751	0,5704	0,4657	0,3675	0,2808	0,2081	0,1497
15	0,9513	0,9074	0,8444	0,7636	0,6694	0,5681	0,4667	0,3715	0,2867	0,2148
16	0,9730	0,9441	0,8987	0,8355	0,7559	0,6641	0,5660	0,4677	0,3751	0,2920
17	0,9857	0,9678	0,9370	0,8905	0,8272	0,7489	0,6593	0,5640	0,4686	0,3784
18	0,9928	0,9823	0,9626	0,9302	0,8826	0,8195	0,7423	0,6550	0,5622	0,4695
19	0,9965	0,9907	0,9787	0,9573	0,9235	0,8752	0,8122	0,7363	0,6509	0,5606
20	0,9984	0,9953	0,9884	0,9750	0,9521	0,9170	0,8682	0,8055	0,7307	0,6472
21	0,9993	0,9977	0,9939	0,9859	0,9712	0,9469	0,9108	0,8615	0,7991	0,7255
22	0,9997	0,9990	0,9970	0,9924	0,9833	0,9673	0,9418	0,9047	0,8551	0,7931
23	0,9999	0,9995	0,9985	0,9960	0,9907	0,9805	0,9633	0,9367	0,8989	0,8490
24	1,0000	0,9998	0,9993	0,9980	0,9950	0,9888	0,9777	0,9594	0,9317	0,8933
25	1,0000	0,9999	0,9997	0,9990	0,9974	0,9938	0,9869	0,9748	0,9554	0,9269
26	1,0000	1,0000	0,9999	0,9995	0,9987	0,9967	0,9925	0,9848	0,9718	0,9514
27	1,0000	1,0000	0,9999	0,9998	0,9994	0,9983	0,9959	0,9912	0,9827	0,9687
28	1,0000	1,0000	1,0000	0,9999	0,9997	0,9991	0,9978	0,9950	0,9897	0,9805
29	1,0000	1,0000	1,0000	1,0000	0,9999	0,9996	0,9989	0,9973	0,9941	0,9882
30	1,0000	1,0000	1,0000	1,0000	0,9999	0,9998	0,9994	0,9986	0,9967	0,9930
31	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9997	0,9993	0,9982	0,9960
32	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9996	0,9990	0,9978
33	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9998	0,9995	0,9988
34	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9998	0,9994
35	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9997
36	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9998
37	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999
38	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000

$$P(Z \leq z),\ Z \sim N(0,1)$$



z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
-3,5	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002
-3,4	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0002
-3,3	0,0005	0,0005	0,0005	0,0004	0,0004	0,0004	0,0004	0,0004	0,0004	0,0003
-3,2	0,0007	0,0007	0,0006	0,0006	0,0006	0,0006	0,0006	0,0005	0,0005	0,0005
-3,1	0,0010	0,0009	0,0009	0,0009	0,0008	0,0008	0,0008	0,0008	0,0007	0,0007
-3,0	0,0013	0,0013	0,0013	0,0012	0,0012	0,0011	0,0011	0,0011	0,0010	0,0010
-2,9	0,0019	0,0018	0,0018	0,0017	0,0016	0,0016	0,0015	0,0015	0,0014	0,0014
-2,8	0,0026	0,0025	0,0024	0,0023	0,0023	0,0022	0,0021	0,0021	0,0020	0,0019
-2,7	0,0035	0,0034	0,0033	0,0032	0,0031	0,0030	0,0029	0,0028	0,0027	0,0026
-2,6	0,0047	0,0045	0,0044	0,0043	0,0041	0,0040	0,0039	0,0038	0,0037	0,0036
-2,5	0,0062	0,0060	0,0059	0,0057	0,0055	0,0054	0,0052	0,0051	0,0049	0,0048
-2,4	0,0082	0,0080	0,0078	0,0075	0,0073	0,0071	0,0069	0,0068	0,0066	0,0064
-2,3	0,0107	0,0104	0,0102	0,0099	0,0096	0,0094	0,0091	0,0089	0,0087	0,0084
-2,2	0,0139	0,0136	0,0132	0,0129	0,0125	0,0122	0,0119	0,0116	0,0113	0,0110
-2,1	0,0179	0,0174	0,0170	0,0166	0,0162	0,0158	0,0154	0,0150	0,0146	0,0143
-2,0	0,0228	0,0222	0,0217	0,0212	0,0207	0,0202	0,0197	0,0192	0,0188	0,0183
-1,9	0,0287	0,0281	0,0274	0,0268	0,0262	0,0256	0,0250	0,0244	0,0239	0,0233
-1,8	0,0359	0,0351	0,0344	0,0336	0,0329	0,0322	0,0314	0,0307	0,0301	0,0294
-1,7	0,0446	0,0436	0,0427	0,0418	0,0409	0,0401	0,0392	0,0384	0,0375	0,0367
-1,6	0,0548	0,0537	0,0526	0,0516	0,0505	0,0495	0,0485	0,0475	0,0465	0,0455
-1,5	0,0668	0,0655	0,0643	0,0630	0,0618	0,0606	0,0594	0,0582	0,0571	0,0559
-1,4	0,0808	0,0793	0,0778	0,0764	0,0749	0,0735	0,0721	0,0708	0,0694	0,0681
-1,3	0,0968	0,0951	0,0934	0,0918	0,0901	0,0885	0,0869	0,0853	0,0838	0,0823
-1,2	0,1151	0,1131	0,1112	0,1093	0,1075	0,1056	0,1038	0,1020	0,1003	0,0985
-1,1	0,1357	0,1335	0,1314	0,1292	0,1271	0,1251	0,1230	0,1210	0,1190	0,1170
-1,0	0,1587	0,1562	0,1539	0,1515	0,1492	0,1469	0,1446	0,1423	0,1401	0,1379
-0,9	0,1841	0,1814	0,1788	0,1762	0,1736	0,1711	0,1685	0,1660	0,1635	0,1611
-0,8	0,2119	0,2090	0,2061	0,2033	0,2005	0,1977	0,1949	0,1922	0,1894	0,1867
-0,7	0,2420	0,2389	0,2358	0,2327	0,2296	0,2266	0,2236	0,2206	0,2177	0,2148
-0,6	0,2743	0,2709	0,2676	0,2643	0,2611	0,2578	0,2546	0,2514	0,2483	0,2451
-0,5	0,3085	0,3050	0,3015	0,2981	0,2946	0,2912	0,2877	0,2843	0,2810	0,2776
-0,4	0,3446	0,3409	0,3372	0,3336	0,3300	0,3264	0,3228	0,3192	0,3156	0,3121
-0,3	0,3821	0,3783	0,3745	0,3707	0,3669	0,3632	0,3594	0,3557	0,3520	0,3483
-0,2	0,4207	0,4168	0,4129	0,4090	0,4052	0,4013	0,3974	0,3936	0,3897	0,3859
-0,1	0,4602	0,4562	0,4522	0,4483	0,4443	0,4404	0,4364	0,4325	0,4286	0,4247
-0,0	0,5000	0,4960	0,4920	0,4880	0,4840	0,4801	0,4761	0,4721	0,4681	0,4641

$$P(Z \leq z),\ Z \sim N(0,1)$$



z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,0	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990
3,1	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,2	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995
3,3	0,9995	0,9995	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997
3,4	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998
3,5	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998

Bibliografia

Adams, W.J. (2009). *The life and times of the central limit theorem*. American Mathematical Society, London Mathematical Society.

Alexandrov, A.D., Kolmogorov, A.N., Lavrentiev, M.A. (2020). *Mathématiques, leur contenu, leurs méthodes, leur signification*. Saint-Cyr-sur-Mer: Les Éditions du Bec de l'Aigle.

Chae, S.B. (1980). *Lebesgue integration*. New York: Marcel Dekker.

Chow, Y.S., Teicher, H. (1997). *Probability theory: independence, interchangeability, martingales*. New York: Springer.

Deheuvels, P. (1990). *La probabilité de hasard et la certitude*. Paris: Presses Universitaires de France.

Feller, W. (1945). The fundamental limit theorems in probability. *Bull. Amer. Math. Soc.* 51, 800–832.

Gonçalves, E., Mendes Lopes, N. (2013). *Probabilidades: princípios teóricos*. Lisboa: Escolar Editora.

Haigh, J. (2002). *Probability models*. London: Springer.

Hald, A. (1990). *A history of probability and statistics and their applications before 1750*. New York: Wiley.

Hald, A. (1998). *A history of mathematical statistics from 1750 to 1930*. New York: Wiley.

Jacod, J., Protter, P. (2000). *Probability essentials*. Berlin: Springer.

James, B.R. (1981). *Probabilidade: um curso em nível intermediário*. Rio de Janeiro: IMPA.

Kallenberg, O. (1997). *Foundations of modern probability*. New York: Springer.

Kolmogorov, A.N. (1950). *Foundations of the theory of probability*. New York: Chelsea.

Le Cam, L. (1986). The central limit theorem around 1935. *Statist. Sci.* 1, 78–91.

Lima, E.L. (1995). *Curso de análise, Vol. 1*. Rio de Janeiro: IMPA.

Métivier, M. (1972). *Notions fondamentales de la théorie des probabilités*, Paris: Dunod.

Monfort, A. (1980). *Cours de probabilités*. Paris: Economica.

Natanson, I.P. (1961). *Theory of functions of a real variable*. New York: Frederick Ungent.

Resnick, S.I. (2005). *A probability path*. Boston: Birkhäuser.

Tenreiro, C. (2000). *Apontamentos de Medida e Integração*. Coimbra: DMUC. https://www.mat.uc.pt/∼tenreiro/apontamentos/AMI0004.pdf

Tenreiro, C. (2002). *Apontamentos de Teoria das Probabilidades*. Coimbra: DMUC. https://www.mat.uc.pt/∼tenreiro/apontamentos/ATP0204.pdf

Índice Remissivo

σ-álgebra, 15
    de Borel, 16, 17
    gerada, 17

acontecimento(s)
    aleatório(s), 13, 16
    certo, 13
    elementar, 14
    incompatíveis, 15
    independentes, 31
    limite inferior, 22
    limite superior, 22
amplitude interquartil, 87

Bayes, Thomas, 27
Bernoulli, Jakod, 5, 148
Bienaymé, Iréneé-Jules, 85
Bochner, Salomon, 161
Borel, Émile, 16, 24, 33
boreliano, 17

Cantelli, Francesco, 24
Cardano, Girolamo, 1
coeficiente
    de achatamento, 84
    de assimetria, 84
    de correlação
        definição, 116
        propriedades, 116
    de forma, 84
conjuntos mensuráveis, 17
convergência
    em distribuição, 158
    em média quadrática, 146
    em probabilidade, 144
    quase certa, 143
convolução de densidades, 122
covariância
    definição, 116
    empírica, 152
    propriedades, 116
curtose, 84

densidade de probabilidade
    condicional, 110
    dum vetor aleatório, 98
    duma variável aleatória, 40
    normal standard, 8, 123
Desigualdade
    de Bienaymé-Tchebychev, 85
    de Cauchy-Schwarz, 115
    de Hölder, 118
    de Markov, 84
    de Minkowski, 119
    de Tchebychev-Markov, 85
desvio padrão, 80
distribuição

assimétrica negativa, 84
assimétrica positiva, 84
binomial, 52
condicional, 110
de Bernoulli, 52
de Dirac, 51
de Poisson, 54
de probabilidade, 36, 92
exponencial, 57
gama, 58, 59
geométrica, 56
hipergeométrica, 56
leptocúrtica, 84
marginal, 92
mesocúrtica, 84
normal ou gaussiana, 59
platicúrtica, 84
simétrica, 70
uniforme, 56
uniforme discreta, 52
distribuição, binomial, 5

espaço
$\mathscr{L}^p$, 80
de probabilidade, 16
esperança condicional, 111
esperança matemática
cálculo da, 76
dum ve.a. em $\mathbb{R}^d$, 119
duma v.a. complexa, 130
duma v.a. real, 67
propriedades da, 71
experiência aleatória, 13, 17
simulação, 63

Fórmula
da probabilidade composta, 26
da probabilidade total, 26
de Daniel da Silva, 21
de Poincaré, 21
Feller, William, 11
Fermat, Pierre, 1
função caraterística
cálculo, 132
definição, 130
derivadas e momentos, 134, 140
dum ve.a. em $\mathbb{R}^d$, 139
propriedades, 131, 139
função de distribuição
definição, 43
dum vetor aleatório, 100
propriedades, 45, 100
função de probabilidade
condicional, 110
dum vetor aleatório, 96
duma variável aleatório, 39
função gama, 59

Galilei, Galileu, 4
Gombaud, Antoine, 1

Huygens, Christiaan, 2

Khintchine, Alexandre, 151
Kolmogorov, Andrei, 13, 15, 151

Lévy, Paul, 11, 161
Laplace, Pierre-Simon de, 3, 9, 166
Lebesgue, Henri, 16
Lei dos grandes números
de Bernoulli, 6, 18, 148
de Borel, 150
de Khintchine, 151
de Kolmogorov, 151
de Poisson, 149
de Tchebychev, 148
Lei zero-um de Borel, 33

Lema de Borel-Cantelli, 24
Liapounov, Alexandre, 10, 166
Lindeberg, Jarl, 11

Markov, Andrei, 10, 84, 85
matriz de covariância, 120
medida de Lebesgue, 16, 17
modelo probabilístico, 16
Moivre, Abraham de, 7
momento
    centrado de ordem p, 79
    centrado de segunda ordem, 80
    de ordem p, 79

Pólya, George, 11
Pacioli, Luca, 1
Pascal, Blaise, 1
Poisson, Siméon Denis, 6, 148
probabilidade
    condicionada, 25
    definição, 16
    definição de Laplace, 17
    espaço de, 16
    geométrica, 19
    propriedades, 19
problema
    da divisão das apostas, 2
    da ruína do jogador, 2
produto de convolução, 122

quantil de ordem p, 86
quartis da distribuição, 87

Scheffé, Henri, 160
Stirling, James, 7

Tartaglia, Niccolò, 1
Tchebychev, Pafnuty, 10, 85, 148
Teorema
    da continuidade de Lévy, 161
    da convergência dominada, 131, 146
    de Bayes, 27
    de Lévy-Bochner, 161
    de Scheffé, 160
Teorema do limite central
    de Laplace, 10
    de de Moivre, 7
    de Lévy–Feller, 11
    de Laplace-Liapounov, 10, 166
    de Lindeberg, 11
testes de diagnóstico
    especificidade, 27, 149
    falso negativo, 27
    falso positivo, 27
    sensibilidade, 27, 149
    valor preditivo negativo, 29
    valor preditivo positivo, 29

variáveis aleatórias
    iguais em distribuição, 37
    independentes, 104
    quase certamente iguais, 38
variável aleatória
    absolutamente contínua, 40
    binomial, 52
    complexa, 129
    de Bernoulli, 52
    de Dirac, 51
    de Poisson, 54
    definição, 36
    discreta, 39
    exponencial, 57
    gama, 59
    geométrica, 56
    hipergeométrica, 56
    integrável, 80
    normal ou gaussiana, 59
    simétrica, 70

uniforme, 56
uniforme discreta, 52
variância, 80
cálculo, 81
empírica, 152
propriedades da, 80
vetor aleatório, 91
absolutamente contínuo, 98
discreto, 96
normal, 123
normal standard, 123