Carlos Tenreiro

Apontamentos de Teoria das Probabilidades

Coimbra, 2002

Versão de Dezembro de 2004

Nota prévia

Os presentes apontamentos têm por base as notas do curso de Teoria das Probabilidades que leccionámos no segundo semestre dos anos lectivos de 2000/01 e 2001/02, a alunos do Ramo Científico, especialização em Matemática Pura, do terceiro ano da licenciatura em Matemática da Universidade de Coimbra. Uma versão preliminar destes apontamentos foi utilizada como texto de apoio ao curso no último dos anos lectivos referidos.

Ao longo dos dez capítulos que constituem este texto, desenvolvemos temas habituais num primeiro curso de Teoria das Probabilidades, cujo principal objectivo é o estabelecimento dos teoremas limite clássicos: leis dos grandes números de Kolmogorov e teorema do limite central de Lindeberg.

Estando os alunos já familiarizados com tópicos como o do prolongamento de medidas, da integração relativamente a uma medida, dos espaços L^p de Lebesgue, das medidas produto, da transformação de medidas, ou dos teoremas de Radon-Nikodym e da decomposição de Lebesgue, a abordagem às probabilidades feita nesta disciplina, é fortemente influenciada por tal facto.

Ao fazermos referência a um dos resultados anteriores, ou a outro qualquer resultado de Medida e Integração que sabemos ser do conhecimento do aluno, remetemos o leitor para os nossos Apontamentos de Medida e Integração (Coimbra, 2000) que neste texto designaremos pelas iniciais AMI.

Carlos Tenreiro

Índice

Parte I

Distribuições de probabilidade

Capítulo 1

Espaços de probabilidade

Modelo matemático para uma experiência aleatória. Propriedades duma probabilidade. Modelação de algumas experiências aleatórias. Algumas construções de espaços de probabilidade. Produto infinito de espaços de probabilidade. Probabilidade condicionada. Teorema de Bayes. Produto generalizado de probabilidades. Breve referência à simulação de experiências aleatórias.

1.1 Modelo matemático para uma experiência aleatória

Em 1933 A.N. Kolmogorov ([1]) 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 no desenvolvimento da teoria das probabilidades.

Ao conjunto das realizações possíveis duma **experiência aleatória** Kolmogorov começou por associar um conjunto Ω, a que chamamos **espaço dos resultados** ou **espaço fundamental**, em que cada elemento $\omega \in \Omega$ caracteriza completamente uma realização possível da experiência aleatória. Identificou os **acontecimentos aleatórios** associados à experiência 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 pelos conjuntos $\emptyset$ e Ω, respectivamente. Os subconjuntos singulares de Ω dizem-se **acontecimentos elementares**.

As operações usuais entre conjuntos, reunião, intersecção, diferença, etc, permitem exprimir ou construir acontecimentos em função ou a partir de outros acontecimentos:

[1] Kolmogorov, A.N., *Grundbegriffe der Wahrscheinlichkeitrechnung*, 1933.

$A \cup B \equiv$ 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 que se realiza quando A não se realiza; $A - B \equiv$ acontecimento que se realiza quando A se realiza e B não se realiza; $\bigcup_{n=1}^{\infty} A_n \equiv$ 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; $\liminf A_n \equiv$ acontecimento que se realiza quando se realizam todos os acontecimentos A_n com excepção dum número finito deles; $\limsup A_n \equiv$ acontecimento que se realiza quando se realiza um infinidade de acontecimentos A_n.

Finalmente, com a axiomatização do conceito de probabilidade, Kolmogorov estabelece regras gerais a que deve satisfazer a atribuição de probabilidade aos acontecimentos duma experiência aleatória.

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 "saída de número par", "saída de número inferior a 3", etc., podem ser identificados com os subconjuntos do espaço dos resultados $\{2, 4, 6\}$, $\{1, 2\}$, etc., respectivamente. 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 $\mathrm{P}(A)$ denotarmos a probabilidade do acontecimento A, $\mathrm{P}(A)$ deverá ser um número real do intervalo $[0, 1]$, com $\mathrm{P}(\Omega) = 1$ e $\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B)$, 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. O domínio natural de definição duma tal aplicação é assim uma **σ-álgebra**. Recordemos que uma classe $\mathcal{A}$ de partes de Ω é uma σ-álgebra se contém o conjunto vazio, e é estável para a complementação e para a reunião numerável. Uma σ-álgebra contém claramente Ω, e é estável para a intersecção numerável bem como para a intersecção e reunião finitas.

Definição 1.1.1 *Uma* ***probabilidade*** P *sobre uma σ-álgebra $\mathcal{A}$ de partes de Ω é uma aplicação de $\mathcal{A}$ em $[0,1]$ tal que:*

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

b) Para todo o $A_n \in \mathcal{A}$, $n = 1, 2, \ldots$ disjuntos dois a dois

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

Ao terno $(\Omega, \mathcal{A}, \mathrm{P})$ chamamos **espaço de probabilidade**. Quando a uma experiência aleatória associamos o espaço de probabilidade $(\Omega, \mathcal{A}, \mathrm{P})$ dizemos também que este espaço é um **modelo probabilístico** para a experiência aleatória em causa. Os elementos de $\mathcal{A}$ dizem-se **acontecimentos aleatórios**. Fazendo em *b*), $A_1 = \Omega$ e $A_n = \emptyset$, para $n \geq 2$, obtemos $\mathrm{P}(\Omega) = \mathrm{P}(\Omega) + \sum_{n=2}^{\infty} \mathrm{P}(\emptyset)$, o que implica $\mathrm{P}(\emptyset) = 0$. Por outras palavras, uma probabilidade é uma **medida** definida num espaço mensurável $(\Omega, \mathcal{A})$ em que a medida de todo o espaço é igual à unidade (ver AMI, §2.1).

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 1.1.2 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 $\mathcal{A} = \mathcal{P}(\Omega)$ por $\mathrm{P}(\{x\}) = 1/6$, para $x \in \Omega$. Duma forma geral, se o espaço Ω dos resultados duma experiência aleatória é finito e todos os seus elementos têm a mesma possibilidade de ocorrer, será natural tomar

$$\mathrm{P}(A) = \frac{\sharp A}{\sharp \Omega}, \text{ para } A \subset \Omega,$$

isto é,

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

que não é mais do que a **definição clássica de probabilidade**.

Exemplo 1.1.3 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, $\mathrm{P}(]c, d]) = (d - c)/(b - a)$, para $a \leq c < d \leq b$. Mais geralmente, se Q é uma região mensurável de $\mathbb{R}^d$ com volume $0 < \lambda(Q) < +\infty$, onde λ é a medida de Lebesgue em $\mathbb{R}^d$, a extracção ao acaso dum ponto de Q pode ser modelada pela probabilidade

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

dita **probabilidade geométrica**.

Exercícios

1. **(Paradoxo dos dados ([2]))** 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, respectivamente. 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?

2. (**Paradoxo do dia de aniversário**) Se não mais que 365 pessoas estão a assistir a um espectáculo, é possível que todas elas tenham um dia de aniversário diferente. Com 366 pessoas é certo que pelo menos duas delas têm o mesmo dia de aniversário. Admitindo que os nascimentos se distribuem uniformemente pelos 365 dias do ano, e que há n (≤ 365) pessoas a assistir ao espectáculo, calcule a probabilidade p_n de pelo menos duas delas terem o mesmo dia de aniversário. Verifique que $p_{23} > 0.5$ e que $p_{56} > 0.99$.

   Suponha agora que também está a assistir ao espectáculo. Qual é a probabilidade q_n de alguém com seu dia de aniversário estar também a assistir ao espectáculo? Verifique que $q_{23} < 0.059$ e que $q_{56} < 0.141$.

3. Num segmento de recta de comprimento L dois pontos são escolhidos ao acaso. Qual é a probabilidade da distância entre eles não exceder x, com $0 \leq x \leq L$?

4. Qual é a probabilidade das raízes da equação quadrática $x^2 + 2Ax + B = 0$ serem reais, se (A, B) é um ponto escolhido ao acaso sobre o rectângulo $[-R, R] \times [-S, S]$?

5. Suponhamos que extraímos ao acaso um ponto x do intervalo $[0, 1]$, e que não estamos interessados em x mas no seu quadrado y. Se pretendemos calcular a probabilidade de y pertencer ao subintervalo $]c, d]$ de $[0, 1]$, conclua que deverá tomar $\Omega = [0, 1]$ e P tal que

$$\mathrm{P}(]c, d]) = \sqrt{d} - \sqrt{c}, \text{ para } 0 \leq c \leq d \leq 1.$$

[2]Este problema foi colocado a Galileu Galilei (1564–1642), o que o levou a escrever *Sopra le scoperte dei dadi* (Sobre uma descoberta acerca de dados) entre 1613 e 1623.

1.2 Propriedades duma probabilidade

As propriedades seguintes são consequência do facto duma probabilidade ser uma medida definida num espaço mensurável $(\Omega, \mathcal{A})$ em que a medida de todo o espaço é igual à unidade. A sua demonstração é deixada ao cuidado do aluno.

Proposição 1.2.1 (Aditividade finita) *Se $A_1, \ldots, A_n$ são acontecimentos aleatórios disjuntos dois a dois, então $\mathrm{P}(\bigcup_{k=1}^n A_k) = \sum_{k=1}^n \mathrm{P}(A_k)$.*

Proposição 1.2.2 *Para $A, B \in \mathcal{A}$, temos:*
a) $\mathrm{P}(A^c) = 1 - \mathrm{P}(A)$*;*
b) Se $A \subset B$, então $\mathrm{P}(B - A) = \mathrm{P}(B) - \mathrm{P}(A)$*;*
c) Se $A \subset B$, então $\mathrm{P}(A) \leq \mathrm{P}(B)$ *(***monotonia***);*
d) $\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cap B)$*.*

Proposição 1.2.3 (Subaditividade completa) *Se $A_n \in \mathcal{A}$, para $n = 1, 2, \ldots$, então $\mathrm{P}(\bigcup_{n=1}^\infty A_n) \leq \sum_{n=1}^\infty \mathrm{P}(A_n)$.*

Proposição 1.2.4 (Continuidade) *Se $A_n \in \mathcal{A}$, para $n = 1, 2, \ldots$, e $A_n \to A$ então $\mathrm{P}(A_n) \to \mathrm{P}(A)$.*

Dizemos que uma função de conjunto P definida numa classe $\mathcal{B}$ de partes de Ω, é **ascendentemente contínua** (resp. **descendentemente contínua**) em $A \in \mathcal{B}$, se para toda a sucessão (A_n) em $\mathcal{B}$ com $A_n \uparrow A$ (resp. $A_n \downarrow A$), se tem $\mathrm{P}(A_n) \to \mathrm{P}(A)$. P diz-se ascendentemente contínua (resp. descendentemente contínua) se for ascendentemente contínua (resp. descendentemente contínua) em todo o $A \in \mathcal{B}$.

Do resultado seguinte fica claro que quando exigimos que uma probabilidade seja não só aditiva mas também σ-aditiva, o que estamos a exigir a P é uma propriedade de continuidade. Recordemos que uma **semi-álgebra** $\mathcal{C}$ de partes dum conjunto Ω é um **semi-anel** de partes de Ω que contém Ω, isto é, é uma classe não-vazia de subconjuntos de Ω que contém Ω, que é estável para a intersecção finita, e o complementar de qualquer elemento de $\mathcal{C}$ é reunião finita disjunta de elementos de $\mathcal{C}$ (ver AMI, §1.2).

Teorema 1.2.5 *Seja P uma função de conjunto não-negativa e aditiva numa semi--álgebra $\mathcal{B}$ de partes de Ω com $\mathrm{P}(\Omega) = 1$. As afirmações seguintes são equivalentes:*
i) P é σ-aditiva em $\mathcal{B}$;
ii) P é ascendentemente contínua;
iii) P é ascendentemente contínua em Ω;
iv) P é descendentemente contínua;
v) P é descendentemente contínua em $\emptyset$.

Exercícios

1. (**Fórmula de Daniel da Silva** ([3]) ou da **Inclusão-Exclusão**:) Se $A_1, \ldots, A_n$, para $n \geq 2$, são acontecimentos, mostre que

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

2. (**Paradoxo das coincidências** ([4])) Numa festa de natal os n funcionários de uma empresa decidem dar entre si presentes. Cada um trás um presente que é misturado com os outros e distribuído ao acaso pelos funcionários. Este procedimento é utilizado acreditando-se que a probabilidade p_n de alguém receber o seu próprio presente é pequena se o número de funcionários é grande. Calcule p_n e mostre que $p_n \to 1 - e^{-1}$. Verifique que $p_n \approx 0.6321$, para $n \geq 7$.

   (Sugestão: Utilize a fórmula de Daniel da Silva aplicada aos acontecimentos $A_i =$"o i-ésimo funcionário recebe o seu presente".)

3. (**Desigualdades de Bonferroni**) Se $A_1, \ldots, A_n$ são acontecimentos, mostre que:

   (a) $\mathrm{P}\Big(\bigcup_{i=1}^{n} A_i\Big) \geq \sum_{i=1}^{n} \mathrm{P}(A_i) - \sum_{1\leq i<j\leq n} \mathrm{P}(A_i \cap A_j)$.

   (b) $\mathrm{P}\Big(\bigcup_{i=1}^{n} A_i\Big) \leq \sum_{i=1}^{n} \mathrm{P}(A_i) - \sum_{1\leq i<j\leq n} \mathrm{P}(A_i \cap A_j) + \sum_{1\leq i<j<k\leq n} \mathrm{P}(A_i \cap A_j \cap A_k)$.

4. Se (A_n) é uma sucessão de acontecimentos mostre que $\mathrm{P}(\bigcap_{n=1}^{\infty} A_n) = 1$ sse $\mathrm{P}(A_n) = 1$, para todo o $n \in \mathbb{N}$.

1.3 Modelação de algumas experiências aleatórias

Dando continuidade ao parágrafo 1.1, apresentamos agora mais alguns exemplos de modelações de experiências aleatórias.

Exemplo 1.3.1 Consideremos n lançamentos sucessivos duma moeda equilibrada. Se representarmos por 1 a saída de "cara" e por 0 a saída de "coroa", o espaço dos resultados é $\Omega = \{0,1\}^n = \{(x_1, \ldots, x_n) : x_i = 0 \text{ ou } 1\}$. Tal como no Exemplo 1.1.2, sendo a moeda equilibrada, todos os elementos de Ω têm a mesma possibilidade de ocorrer. Poderemos assim tomar P definida em $\mathcal{A} = \mathcal{P}(\Omega)$ por

$$\mathrm{P}(\{(x_1, \ldots, x_n)\}) = 1/2^n, \text{ para } (x_1, \ldots, x_n) \in \{0,1\}^n.$$

[3] Daniel Augusto da Silva (1814-1878).

[4] Este problema é pela primeira vez considerado por Pierre Rémond de Montmort (1678–1719) em *Essay d'Analyse sur les Jeux de Hazard*, 1708.

Exemplo 1.3.2 Consideremos agora n lançamentos sucessivos duma moeda não necessariamente equilibrada, isto é, em cada lançamento a probabilidade de obtermos 1 (cara) é p e a probabilidade de obtermos 0 (coroa) é $1-p$. Qual é o espaço de probabilidade que devemos associar a esta experiência aleatória? O espaço dos resultados é, tal como no exemplo anterior, $\Omega = \{0,1\}^n$. No entanto, os elementos de Ω não têm agora, para $p \neq 1/2$, a mesma possibilidade de ocorrer. Para determinarmos a probabilidade que devemos associar a esta experiência, tentemos reduzir-nos ao exemplo anterior considerando uma experiência auxiliar que consiste em n extracções sucessivas de uma bola dum saco com ℓ bolas idênticas em que ℓp estão numeradas com 1 e $\ell(1-p)$ são numeradas com 0 (se p é racional é sempre possível determinar ℓ; por exemplo, se $p = 0.1$ basta tomar $\ell = 10$ e passamos a ter uma experiência que consiste na repetição n vezes duma outra, esta com 10 resultados igualmente prováveis, em que um deles é do tipo 1 e os restantes são de tipo 0). A ocorrência do acontecimento $\{(x_1,\ldots,x_n)\}$ com $\sum_{i=1}^n x_i = k$, corresponde na experiência auxiliar à ocorrência de um conjunto de resultados elementares em número de $(\ell p)^k(\ell(1-p))^{n-k}$. Sendo ℓ^n o número total de acontecimentos elementares, e sendo estes igualmente prováveis, então P deverá ser dada por

$$\mathrm{P}\{(x_1,\ldots,x_n)\}) = (\ell p)^k(\ell(1-p))^{n-k}/\ell^k = p^k(1-p)^{n-k},$$

isto é,

$$\mathrm{P}\{(x_1,\ldots,x_n)\}) = p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i},$$

para $(x_1,\ldots,x_n) \in \{0,1\}^n$.

Exemplo 1.3.3 Consideremos n repetições, sempre nas mesmas condições, duma experiência aleatória com k resultados possíveis $1,\ldots,k$, sendo $p_1,\ldots,p_k$ as respectivas probabilidades de ocorrência, onde $\sum_{i=1}^k p_i = 1$. Seguindo o raciocínio anterior o espaço dos resultados é $\Omega = \{1,\ldots,k\}^n$ e P deverá ser dada por

$$\mathrm{P}(\{(x_1,\ldots,x_n)\}) = p_1^{\sum_{i=1}^n \mathbb{1}_{\{1\}}(x_i)} \ldots p_k^{\sum_{i=1}^n \mathbb{1}_{\{k\}}(x_i)},$$

para $(x_1,\ldots,x_n) \in \{1,\ldots,k\}^n$, onde $\mathbb{1}_A$ representa a função indicatriz do conjunto A.

Nos exemplos que a seguir apresentamos não é simples, sem mais, associar ou mesmo garantir a existência dum modelo probabilístico para a experiência aleatória em causa. Os dois primeiros casos são clássicos tendo sido considerados por Carl Friedrich Gauss (1777-1855) ([5]) e por Francis Galton (1822-1911) ([6]), respectivamente. Em ambos,

[5] Gauss, C.F., *Theoria motus corporum celestium in sectionibus conicis solem ambientium*, 1809.
[6] Galton, F., *Typical laws of heredity in man*, 1877.

a probabilidade P é definida pela exibição da sua **densidade** f, dita **de probabilidade**, relativamente à medida de Lebesgue, isto é, $\mathrm{P} = f\lambda$ (ver AMI, §7.1). O último exemplo é ilustrativo duma classe de modelos probabilísticos conhecidos por **processos estocásticos**. A teoria dos processos estocásticos não será desenvolvida neste curso.

Exemplo 1.3.4 (Distribuição dos erros de medida) Consideremos o erro $x = y - \mu$ cometido ao tomarmos o valor observado y como medida do verdadeiro valor μ, desconhecido. Por razões que detalharemos no Capítulo 9, a experiência aleatória que consiste na observação de y, pode ser descrita pela probabilidade definida, para $a \leq b$, por

$$\mathrm{P}(]a, b]) = \int_{]a,b]} \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-(x-\mu)^2/(2\sigma^2)} d\lambda(x),$$

onde o parâmetro $\sigma > 0$ pode ser interpretado como uma medida da precisão das observações. Na Figura 1.1 apresentam-se os gráficos da função integranda anterior para vários valores de σ, a que chamamos **densidade normal de parâmetros** μ **e** σ^2.

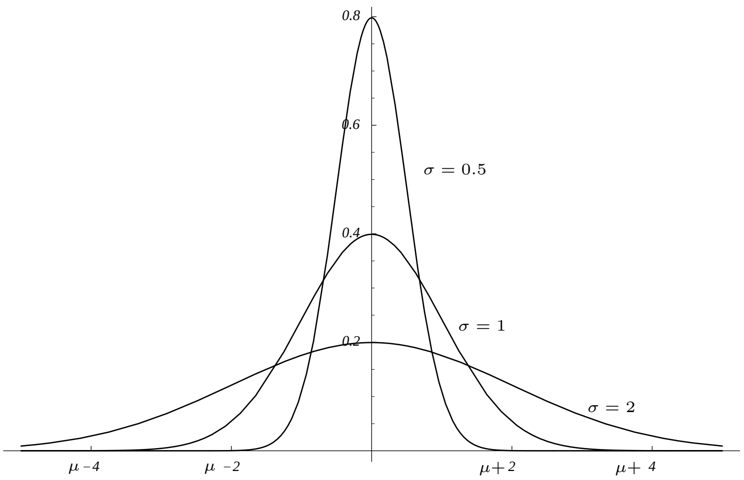
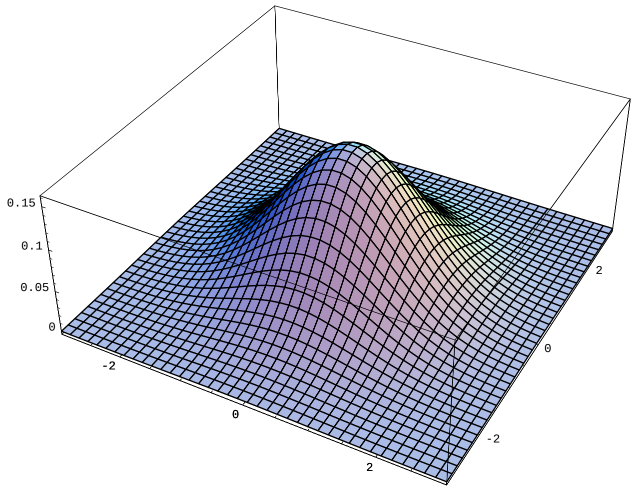


Figura 1.1: Densidade normal univariada

Exemplo 1.3.5 (Densidade normal bivariada) Quando se estuda a relação entre as alturas dos filhos (y) e dos pais (x) convenientemente normalizadas, é habitual descrever as observações realizadas (x, y), através da probabilidade definida, para $a \leq b$ e $c \leq d$, por

$$\mathrm{P}(]a, b]\times]c, d]) = \int_{]a,b]\times]c,d]} \frac{1}{2\pi\sqrt{1-\rho^2}}\, e^{-(x^2-2\rho xy+y^2)/(2(1-\rho^2))} d\lambda(x)d\lambda(y),$$

onde o parâmetro $\rho \in]-1, 1[$ quantifica a associação ou dependência existente entre as quantidades numéricas em estudo. Nas Figuras 1.2 e 1.3, e para os valores $\rho = 0$ e



$\rho = 0$

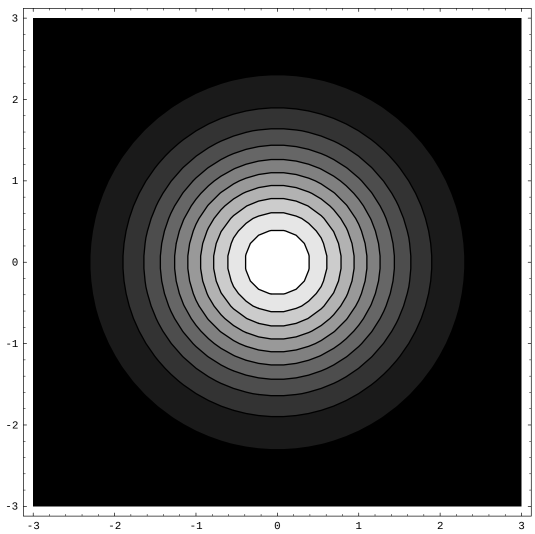
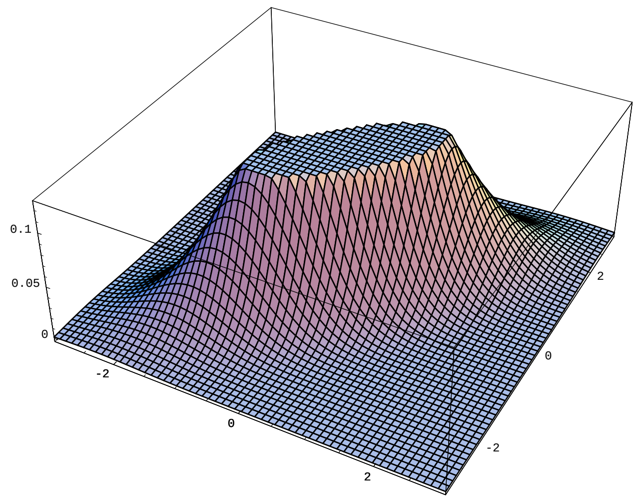


Figura 1.2: Densidade normal bivariada

e $\rho = 0.75$, respectivamente, apresentam-se o gráfico e as curvas de nível relativos à função integranda anterior.

Exemplo 1.3.6 Suponhamos que lançamos uma moeda equilibrada até ocorrer "cara". Nesta situação, será natural tomarmos para conjunto dos resultados $\Omega = \{0,1\}^\infty = \{(x_1, x_2, \ldots) : x_i = 0 \text{ ou } 1\}$, isto é, o conjunto de todas as sucessões de zeros e uns. Para podermos responder a qualquer pergunta sobre esta experiência, por exemplo, a de sabermos qual é a probabilidade de não ocorrer "cara" em nenhum dos lançamentos (ou melhor, para que esta pergunta faça sentido), temos, tal como nos exemplos anteriores, de garantir que lhe está associado um espaço de probabilidade que a descreve. Admi-



$\rho = 0.75$

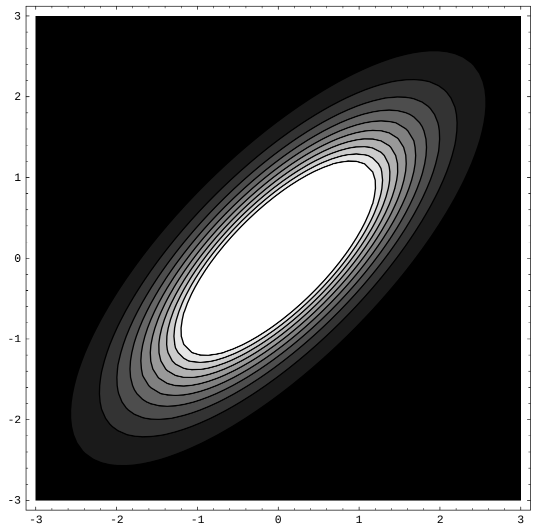


Figura 1.2 (cont.): Densidade normal bivariada

tindo que P é uma tal probabilidade definida numa apropriada σ-álgebra de partes de Ω, e que F_n é um acontecimento que depende apenas dos n primeiros lançamentos, será natural que P satisfaça $\mathrm{P}(F_n) = \mathrm{P}_n(F_n)$, onde P_n é a probabilidade em $\Omega_n = \{0,1\}^n$ definida no Exemplo 1.3.1. A existência duma tal probabilidade será estabelecida no §1.5. Se F_n é o acontecimento $F_n =$"ocorre pela primeira vez cara no n-ésimo lançamento", a probabilidade de não ocorrer "cara" em nenhum dos lançamentos será então dada pela probabilidade do acontecimento $(\bigcup_{n=1}^\infty F_n)^c$, isto é, por $1 - \sum_{n=1}^\infty \mathrm{P}(F_n) = 1 - \sum_{n=1}^\infty \mathrm{P}_n(F_n)$, sendo assim igual a zero a probabilidade não ocorrer "cara" em nenhum dos lançamentos.

Exemplo 1.3.7 (Processo de Poisson) Consideremos o número de ocorrências de um determinado fenómeno aleatório no intervalo de tempo $]0, t]$ para todo o $t > 0$. Pensemos, por exemplo, na chegada de chamadas a uma central telefónica, na chegada de clientes a uma caixa de supermercado, na emissão de partículas por uma substância radioactiva, etc. Se o fenómeno ocorre nos instantes $t_1, t_2, t_3, \ldots$ com $t_1 < t_2 < t_3 < \cdots$, o resultado da experiência não é mais do que uma função de $]0, +\infty[$ em $\mathbb{N}_0$, cujo gráfico é apresentado na Figura 1.3. O conjunto Ω dos resultados possíveis da experiência pode ser assim identificado com o conjunto das funções escalonadas de $]0, +\infty[$ em $\mathbb{N}_0$, não-decrescentes e contínuas à direita.

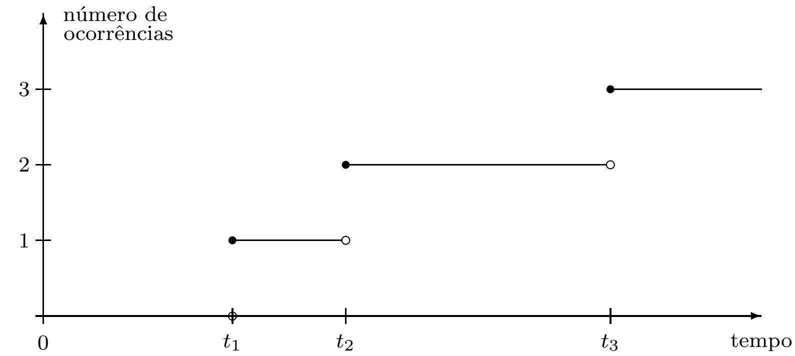


Figura 1.3: Acontecimento elementar dum processo de Poisson

Admitamos que: H1) a probabilidade de se verificarem k ocorrências num determinado intervalo de tempo finito depende apenas da sua amplitude; H2) dados dois intervalos de tempo finitos e disjuntos, a probabilidade de se verificarem k ocorrências num deles não nos dá qualquer informação sobre a probabilidade de se verificarem j ocorrências no outro; H3) não há ocorrências simultâneas. Poderíamos demonstrar que as hipóteses anteriores determinam, numa apropriada σ-álgebra de partes de Ω, uma famílias de probabilidades indexada por um parâmetro real $\lambda > 0$ que pode ser interpretado como o número médio de chegadas num intervalo de tempo unitário.

Exercícios

1. Vou lançar dois dados equilibrados n vezes consecutivas e aposto com outro jogador que pelo menos um par de 6 irá sair. Para que o jogo me seja favorável deverei lançar o dado 24 ou 25 vezes?

2. (**Problema da divisão das apostas** ([7])) Dois jogadores jogam uma série de partidas justas até que um deles obtenha 6 vitórias. Por motivos exteriores ao jogo, este é interrompido quando um dos jogadores somava 5 vitórias e o outro 3 vitórias. Como devemos dividir o montante apostado por ambos os jogadores?

[7] Este problema e o anterior foram colocados por Antoine Gombaud (1607-1684), chevalier de Méré,

3. Eu e outro jogador aceitamos lançar sucessivamente dois dados nas condições seguintes: eu ganho se tirar 7 pontos, ele ganha se tirar 6 pontos e é ele que lança em primeiro lugar. Que probabilidade tenho eu de ganhar?

4. (**Problema da ruína do jogador** ([8])) A e B têm cada um 12 moedas e jogam com três dados. Se saem 11 pontos, A dá uma moeda a B, e se saem 14 pontos, B dá uma moeda a A. Ganha aquele que primeiro ficar com todas as moedas. Qual é a probabilidade de A ganhar?

   (Sugestão: Para $m \in \{-12, \ldots, 12\}$, denote por p_m a probabilidade de A ganhar quando possui $12 + m$ moedas, e verifique que p_m satisfaz uma relação de recorrência linear.)

5. Uma caixa contém b bolas brancas e p bolas pretas. Uma bola é extraída ao acaso da caixa, e sem ser nela reposta, uma segunda bola é extraída ao acaso. Qual o espaço de probabilidade que associa à experiência descrita? Qual é a probabilidade: De ambas as bolas serem brancas? Da primeira bola ser branca e da segunda ser preta? Da segunda ser preta? Da segunda ser preta, sabendo que a primeira bola é branca?

1.4 Algumas construções de espaços de probabilidade

Recordamos neste parágrafo construções de espaços de probabilidade já nossas conhecidas da disciplina de Medida e Integração. Alguns dos exemplos apresentados nos parágrafos anteriores são casos particulares das construções seguintes.

Exemplo 1.4.1 Se $\Omega = \{\omega_i : i \in I\}$, com I finito ou numerável, e p_i, $i \in I$, são números reais não-negativos com $\sum_{i \in I} p_i = 1$, então

$$\mathrm{P}(A) = \sum_{i:\omega_i \in A} p_i, \text{ para } A \in \mathcal{P}(\Omega),$$

é uma probabilidade em $(\Omega, \mathcal{P}(\Omega))$. As probabilidades consideradas nos Exemplos 1.1.2, 1.3.1 e 1.3.2, são casos particulares desta. No caso em que $I = \{1, 2, \ldots, n\}$ e $p_i = 1/n$, para todo o $i \in I$, obtemos a definição clássica de probabilidade.

Exemplo 1.4.2 Se $F : \mathbb{R} \to \mathbb{R}$ é uma função não-decrescente, contínua à direita com $F(x) \to 0$ ou 1, se $x \to -\infty$ ou $x \to +\infty$, respectivamente, então existe uma e uma só probabilidade P sobre $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ tal que

$$\mathrm{P}(]-\infty, x]) = F(x), \text{ para todo o } x \in \mathbb{R}.$$

a Blaise Pascal (1623-1662). O problema da divisão das apostas é resolvido por este e por Pierre de Fermat (1601–1665) numa célebre troca de correspondência no verão de 1654. A resolução do problema por Pascal é publicada em *Traité du Triangle Arithmétique*, 1665. Este problema era já na altura clássico, sendo referido por Luca Paccioli (1445–1517) em *Summa de arithmetica, geometria, proportioni et proportionalita*, 1494.

[8] Este problema e o anterior são dois dos problemas resolvidos por Christian Huygens (1629-1695) em *De ratiociniis in aleae ludo* (Sobre a lógica do jogo de dados), 1657. O problema da ruína do jogador foi colocado por Pascal a Fermat, tendo chegado posteriormente ao conhecimento de Huygens.

F diz-se **função de distribuição de** P (ver AMI, §2.9). A probabilidade definida no Exemplo 1.1.3 é um caso particular desta, em que $F(x) = (x-a)/(b-a)$, se $a \le x \le b$, $F(x) = 0$, se $x < 0$, e $F(x) = 1$, se $x > b$.

Exemplo 1.4.3 O exemplo anterior pode ser generalizado ao caso multidimensional. Para $x = (x_1, \ldots, x_d)$ e $y = (y_1, \ldots, y_d)$ em $\mathbb{R}^d$, escrevemos $x \le y$ (resp. $x < y$) se $x_i \le y_i$ (resp. $x_i < y_i$) para todo o $i = 1, \ldots, d$. Tal com em $\mathbb{R}$, os conjuntos dos pontos x tais que $a < x \le b$ ou dos pontos x tais que $x \le b$, serão denotados por $]a, b]$ ou $]-\infty, b]$, respectivamente. Dado um rectângulo semi-aberto à esquerda $]a, b]$, denotamos por V o conjunto dos seus vértices, isto é, o conjunto dos pontos da forma $(x_1, \ldots, x_d)$ com $x_i = a_i$ ou $x_i = b_i$, para $i = 1, \ldots, d$. Se $x \in V$, designamos por $\mathrm{sgn}(x)$ o sinal de x, que é definido por $\mathrm{sgn}(x) = (-1)^{\sharp\{i: x_i = a_i\}}$. Dada uma função $F : \mathbb{R}^d \to \mathbb{R}$, tal que: *i)* F é não-decrescente, isto é, $F]a, b] = \sum_{x\in V} \mathrm{sgn}(x)F(x) \ge 0$, se $a < b$; *ii)* F é contínua à direita, isto é, $\lim_{x \to y,\, y \le x} F(x) = F(y)$, para todo o $y \in \mathbb{R}^d$; *iii)* $F(x) \to 0$ ou 1, se $\min_{i=1,\ldots,d} x_i \to -\infty$ ou $+\infty$, respectivamente; então existe uma e uma só probabilidade P sobre $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ tal que

$$\mathrm{P}(]-\infty, x]) = F(x), \text{ para todo o } x \in \mathbb{R}^d.$$

F diz-se **função de distribuição de** P. A demonstração da existência de P pode ser encontrada em Billingsley, 1986, pg. 177–180. A unicidade é consequência imediata do lema da igualdade de medidas (cf. AMI, §2.6).

Exemplo 1.4.4 Se μ é uma medida em $(\Omega, \mathcal{A})$ e f é uma aplicação $\mathcal{B}(\overline{\mathbb{R}})$-mensurável definida em $(\Omega, \mathcal{A})$, não-negativa com $\int f d\mu = 1$, então

$$\mathrm{P}(A) = \int_A f d\mu, \text{ para } A \in \mathcal{A},$$

é uma probabilidade. P diz-se probabilidade com densidade f relativamente a μ, e f diz-se **densidade de probabilidade de** P **relativamente a** μ (ver AMI, §7.1).

Note que a construção descrita no Exemplo 1.4.1 é um caso particular desta se tomarmos $f = \sum_{i\in I} p_i \mathbb{I}_{\{\omega_i\}}$ e μ a medida contagem em Ω. Verifique que o mesmo acontece com as construções consideradas nos Exemplos 1.1.3, 1.3.4 e 1.3.5. No caso da extracção ao acaso dum ponto do intervalo $[a, b]$, P tem densidade f relativamente à medida de Lebesgue em $\mathbb{R}$, onde

$$f(x) = \begin{cases} \frac{1}{b-a}, & \text{se } a \le x \le b \\ 0, & \text{senão} \end{cases} \tag{1.4.5}$$

A densidade assim definida diz-se **densidade uniforme** sobre o intervalo $[a, b]$.

Exemplo 1.4.6 Se Q é uma probabilidade num espaço mensurável $(E, \mathcal{B})$, e f é uma aplicação mensurável de $(E, \mathcal{B})$ em $(\Omega, \mathcal{A})$, então P definida por

$$\mathrm{P}(A) = \mathrm{Q}(f^{-1}(A)), \text{ para } A \in \mathcal{A},$$

é uma probabilidade, dita **probabilidade imagem de** Q **por** f (ver AMI, §7.1). Este é, em particular, o caso da probabilidade definida no Exercício 1.1.5 (porquê?).

Exemplo 1.4.7 Se P_i é uma probabilidade sobre $(\Omega_i, \mathcal{A}_i)$, para $i = 1, \ldots, d$, podemos definir sobre o espaço produto $(\Omega, \mathcal{A}) = (\prod_{i=1}^d \Omega_i, \bigotimes_{i=1}^d \mathcal{A}_i)$ a probabilidade $\mathrm{P} = \bigotimes_{i=1}^d \mathrm{P}_i$, dita **probabilidade produto das probabilidade** $\mathrm{P}_1, \ldots, \mathrm{P}_d$ (ver AMI, §§6.1, 6.2). Sabemos que P é a única probabilidade sobre $(\Omega, \mathcal{A})$ que satisfaz

$$\mathrm{P}(A_1 \times \ldots \times A_d) = \prod_{i=1}^d \mathrm{P}_i(A_i),$$

para todo o $A_i \in \mathcal{A}_i$, $i = 1, \ldots, d$. A probabilidade construída no Exemplo 1.3.2 é um caso particular desta bastando tomar, para $i = 1, \ldots, n$, $(\Omega_i, \mathcal{A}_i) = (\{0,1\}, \mathcal{P}(\{0,1\}))$ e $\mathrm{P}_i(\{1\}) = p = 1 - \mathrm{P}_i(\{0\})$. O mesmo acontece com a probabilidade definida no Exemplo 1.3.5 quando $\rho = 0$.

1.5 Produto de espaços de probabilidade

No Exemplo 1.3.6, deixámos em aberto a questão da existência de uma probabilidade definida num produto infinito de espaços de probabilidade verificando propriedades semelhantes às da probabilidade produto definida num produto finito de espaços de probabilidade (cf. Exemplo 1.4.7). Respondemos neste parágrafo a essa questão.

No que se segue, $(\Omega_t, \mathcal{A}_t, \mathrm{P}_t)$, $t \in T$, é uma qualquer família de espaços de probabilidade, e vamos denotar por $\prod_{t\in T} \Omega_t$, o produto cartesiano dos espaços anteriores, isto é, o conjunto de todos os elementos da forma $(\omega_t, t \in T)$, onde $\omega_t \in \Omega_t$, para $t \in T$. Quando $T = \{1, \ldots, n\}$ ou $T = \mathbb{N}$ escrevemos habitualmente $\Omega_1 \times \ldots \times \Omega_n$ ou $\Omega_1 \times \Omega_2 \times \ldots$, respectivamente. Se $\Omega_t = \Omega$, para todo o $t \in T$, usamos a notação Ω^T, Ω^n ou Ω^∞, respectivamente.

Sendo $S \subset T$, e π_S a aplicação projecção de $\prod_{t\in T} \Omega_t$ em $\prod_{t\in S} \Omega_t$ definida por $\pi_S(\omega_t, t \in T) = (\omega_t, t \in S)$, todo o subconjunto de $\prod_{t\in T} \Omega_t$ da forma $\pi_S^{-1}(A)$, com $A \subset \prod_{t\in S} \Omega_t$, diz-se **cilindro de base A**. Um tal cilindro diz-se de **dimensão finita** se S é finito.

Definição 1.5.1 *Chamamos* σ***-álgebra produto*** *das* σ*-álgebras* $\mathcal{A}_t$, $t \in T$, *à* σ*-álgebra* $\bigotimes_{t\in T} \mathcal{A}_t$, *gerada pelos cilindros de dimensão finita cujas bases são rectângulos men-*

suráveis. Por outras palavras, se

$$
\begin{aligned}
\mathcal{S} &= \Big\{\pi_S^{-1}(A) : S \subset T, \sharp S < \infty, A = \prod_{t\in S} A_t \text{ com } A_t \in \mathcal{A}_t, \text{ para } t \in S\Big\} \\
&= \Big\{\prod_{t\in S} A_t : A_t \in \mathcal{A}_t \text{ e } A_t = \Omega_t \text{ excepto para um número finito de índices}\Big\} \\
&= \bigcup_{S\subset T, \sharp S<\infty} \pi_S^{-1}\Big(\prod_{t\in S} \mathcal{A}_t\Big),
\end{aligned}
$$

então

$$\bigotimes_{t\in T} \mathcal{A}_t = \sigma(\mathcal{S}).$$

O espaço mensurável $(\prod_{t\in T}\Omega_t, \bigotimes_{t\in T}\mathcal{A}_t)$ diz-se **produto dos espaços mensuráveis** $(\Omega_t, \mathcal{A}_t)$, $t \in T$. Como anteriormente, denotamos a σ-álgebra anterior por $\mathcal{A}_1 \bigotimes \ldots \bigotimes \mathcal{A}_n$ ou $\mathcal{A}_1 \bigotimes \mathcal{A}_2 \bigotimes \ldots$, quando $T = \{1, \ldots, n\}$ ou $T = \mathbb{N}$. Se $\mathcal{A}_t = \mathcal{A}$, para todo o $t \in T$, usaremos as notações $\mathcal{A}^T$, $\mathcal{A}^n$ ou $\mathcal{A}^\infty$.

Proposição 1.5.2 *A σ-álgebra produto $\bigotimes_{t\in T}\mathcal{A}_t$ é a σ-álgebra gerada pelas aplicações projecção $\pi_S : \prod_{t\in T}\Omega_t \to (\prod_{t\in S}\Omega_t, \bigotimes_{t\in S}\mathcal{A}_t)$, com $S \subset T$ finito.*

Dem: Como $\sigma(\pi_S; S \subset T, \sharp S < \infty) = \sigma(\cup_{S\subset T,\sharp S<\infty}\pi_S^{-1}(\bigotimes_{t\in S}\mathcal{A}_t))$, obtemos $\mathcal{S} \subset \sigma(\pi_S; S \subset T, \sharp S < \infty)$, ou ainda, $\bigotimes_{t\in S}\mathcal{A}_t \subset \sigma(\pi_S; S \subset T, \sharp S < \infty)$. Para estabelecer a inclusão contrária vamos mostrar que $\pi_S^{-1}(\bigotimes_{t\in S}\mathcal{A}_t) \subset \bigotimes_{t\in S}\mathcal{A}_t$. Como $\bigotimes_{t\in S}\mathcal{A}_t \subset \sigma(\prod_{t\in T}\mathcal{A}_t)$ e $\pi_S^{-1}(\prod_{t\in S}\mathcal{A}_t) \subset \mathcal{S}$, obtemos $\pi_S^{-1}(\bigotimes_{t\in S}\mathcal{A}_t) = \pi_S^{-1}(\sigma(\prod_{t\in T}\mathcal{A}_t)) = \sigma(\pi_S^{-1}(\prod_{t\in T}\mathcal{A}_t)) \subset \sigma(\mathcal{S}) = \bigotimes_{t\in S}\mathcal{A}_t$. $\square$

Proposição 1.5.3 *$\bigotimes_{t\in T}\mathcal{A}_t$ é também gerada pelas aplicações $\pi_t : \prod_{t\in T}\Omega_t \to (\Omega_t, \mathcal{A}_t)$, com $t \in T$.*

Dem: Para $S \subset T$ finito e $A_t \in \mathcal{A}_t$, para $t \in S$, temos $\pi_S^{-1}(\prod_{t\in S}A_t) = \cap_{t\in S}\pi_t^{-1}(A_t) \in \sigma(\pi_t; t \in T)$. Assim, $\mathcal{S} \subset \sigma(\pi_t; t \in T)$, e também $\bigotimes_{t\in S}\mathcal{A}_t \subset \sigma(\pi_t; t \in T)$. A inclusão contrária é imediata pela proposição anterior. $\square$

Proposição 1.5.4 *Uma aplicação $f = (f_t, t \in T) : (E, \mathcal{F}) \to (\prod_{t\in S}\Omega_t, \bigotimes_{t\in S}\mathcal{A}_t)$ é mensurável sse $f_t : (E, \mathcal{F}) \to (\Omega_t, \mathcal{A}_t)$ é mensurável para todo o $t \in T$.*

Dem: Sendo f mensurável, a mensurabilidade de f_t, para $t \in T$, é consequência da proposição anterior, uma vez que $f_t = \pi_t \circ f$. Reciprocamente, para $A = \prod_{t\in T}A_t$, com $A_t \in \mathcal{A}_t$ e $A_t = \Omega_t$, excepto para um conjunto finito S de índices, temos $f^{-1}(A) = \{x \in E : f_t(x) \in A_t, t \in S\} = \cap_{t\in S}f_t^{-1}(A_t) \in \mathcal{F}$, pela mensurabilidade de cada uma das aplicações f_t. $\square$

A proposição anterior permite-nos concluir, em particular, que a σ-álgebra gerada por f, $\sigma(f)$, não é mais do que a σ-álgebra gerada pela família de aplicações $f_t, t \in T$, isto é, $\sigma(f) = \sigma(f_t, t \in T)$.

O resultado seguinte estabelece a existência duma probabilidade sobre $\bigotimes_{t\in T} \mathcal{A}_t$ que sobre os cilindros de dimensão finita cujas bases sejam rectângulos mensuráveis $\prod_{t\in S} A_t$, coincide com a probabilidade produto $\bigotimes_{t\in S} \mathrm{P}_t$.

Teorema 1.5.5 *Existe uma única probabilidade* P *sobre* $(\prod_{t\in T} \Omega_t, \bigotimes_{t\in T} \mathcal{A}_t)$ *tal que para todo o* $S \subset T$ *finito, e* $A = \prod_{t\in S} A_t$*, com* $A_t \in \mathcal{A}_t$ *para* $t \in S$*,*

$$\mathrm{P}(\pi_S^{-1}(A)) = \prod_{t\in S} \mathrm{P}_t(A_t) = \Big(\bigotimes_{t\in S} \mathrm{P}_t\Big)(A).$$

A probabilidade P *denota-se por* $\bigotimes_{t\in T} \mathrm{P}_t$ *e denomina-se* ***probabilidade produto das probabilidades*** P_t*,* $t \in T$*. O espaço* $(\prod_{t\in T} \Omega_t, \bigotimes_{t\in T} \mathcal{A}_t, \bigotimes_{t\in T} \mathrm{P}_t)$ *diz-se* ***produto cartesiano dos espaços de probabilidade*** $(\Omega_t, \mathcal{A}_t, \mathrm{P}_t)$*,* $t \in T$*.*

DEM: Seguindo a demonstração apresentada em Monfort, 1980, pg. 105–108, limitamo--nos a dar conta das suas principais etapas. O primeiro passo da demonstração consiste em mostrar que $\mathcal{S}$ é uma semi-álgebra de partes de $\Omega = \prod_{t\in T} \Omega_t$ e que P definida pela fórmula anterior é aí aditiva e satisfaz $\mathrm{P}(\Omega) = 1$. Usando o Teorema 1.2.5, estabelece-se a seguir a σ-aditividade de P em $\mathcal{S}$. Finalmente, utilizando o teorema do prolongamento (ver AMI, §2.5), concluímos que existe um único prolongamento σ-aditivo de P a $\bigotimes_{t\in T} \mathcal{A}_t$, o que conclui a demonstração. □

Exercícios

1. Suponha que lança uma moeda um número infinito de vezes sempre nas mesmas condições e que em cada lançamento a probabilidade de obter "cara" é igual a $p \in]0, 1[$. Calcule a probabilidade:

   (a) de não ocorrer "cara" em nenhum dos lançamentos;

   (b) de ocorrer "cara" um número infinito de vezes;

   (c) de obter uma infinidade de vezes uma sequência particular e finita de "caras" e "coroas".

2. Uma moeda equilibrada é lançada até ocorrer "cara" pela primeira vez, e suponhamos que estamos interessados no número de lançamentos efectuados.

   (a) Que espaço de probabilidade associaria a esta experiência?

   (b) Sendo E o acontecimento "ocorrência de "cara" pela primeira vez depois dum número par de "coroas"" e F o acontecimento "ocorrência de "cara" pela primeira vez depois dum número ímpar de "coroas"", calcule a probabilidade de E e de F.

1.6 Probabilidade condicionada

Retomemos o Exemplo 1.1.2 e suponhamos agora que lançamos o dado e que, apesar de não sabermos qual foi a face que ocorreu, sabemos 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. Será natural substituir a probabilidade P pela probabilidade P_B definida por $\mathrm{P}_B(A) = \sharp A \cap B/\sharp B$.

Duma forma geral, se $(\Omega, \mathcal{A}, \mathrm{P})$ é o espaço de probabilidade associado a uma experiência aleatória, e se sabemos que $B \in \mathcal{A}$, com $\mathrm{P}(B) > 0$, se realiza ou vai realizar, a probabilidade dum acontecimento $A \in \mathcal{A}$ 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 assim natural medir a probabilidade de A se realizar por um número proporcional a $\mathrm{P}(A \cap B)$, isto é, devemos associar a esta experiência o novo espaço de probabilidade $(\Omega, \mathcal{A}, \mathrm{P}_B)$ onde

$$\mathrm{P}_B(A) = \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)}, \quad \text{para } A \in \mathcal{A}.$$

Notemos que P_B é efectivamente uma probabilidade sobre $\mathcal{A}$.

Definição 1.6.1 *Para $B \in \mathcal{A}$, com $\mathrm{P}(B) > 0$, e $A \in \mathcal{A}$, $\mathrm{P}_B(A)$ diz-se* ***probabilidade condicionada de*** *A* ***sabendo*** *B ou* ***probabilidade condicionada de*** *A* ***dado*** *B. $\mathrm{P}_B(A)$ denota-se também por $\mathrm{P}(A|B)$.*

O conhecimento de $\mathrm{P}(B)$ e de $\mathrm{P}_B(A)$ permitem calcular a probabilidade da intersecção $A \cap B$. O resultado seguinte generaliza tal facto à intersecção dum número finito de acontecimentos.

Teorema 1.6.2 (Fórmula da probabilidade composta) *Se $A_1, \ldots, A_n$, com $n \geq 2$, são acontecimentos aleatórios com $\mathrm{P}(A_1 \cap \ldots \cap A_{n-1}) > 0$, então*

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

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 $\mathrm{P}(A_1 \cap \ldots \cap A_{n-1}) > 0$, basta ter em conta que $\mathrm{P}(A_1 \cap \ldots \cap A_n) = \mathrm{P}(A_1 \cap \ldots \cap A_{n-1})\mathrm{P}(A_n|A_1 \cap \ldots \cap A_{n-1})$. $\square$

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

Teorema 1.6.3 (Fórmula da probabilidade total) *Sejam $A_1, \ldots, A_n$ acontecimentos aleatórios dois a dois disjuntos de probabilidade positiva e $B \in \mathcal{A}$ tal que $B \subset A_1 \cup \ldots \cup A_n$. Então*

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

A modelação duma experiência aleatória consiste, como vimos até agora, na fixação dum espaço de probabilidade que descreve completamente (ou acreditamos que descreve) a experiência em causa. A realização dum acontecimento aleatório particular, não traz qualquer informação suplementar sobre futuras realizações da experiência uma vez que acreditamos que esta é completamente descrita pelo espaço de probabilidade considerado. Outra perspectiva é no entanto possível. Se admitirmos que o espaço de probabilidade considerado não descreve completamente a experiência em causa, mas que a descreve apenas de uma forma aproximada, a realização dum acontecimento aleatório particular pode melhorar o conhecimento que temos sobre a experiência aleatória. Nesse caso será de todo o interesse saber como devemos calcular a probabilidade dum acontecimento à luz desta nova informação.

Retomando os comentários que precederam o resultado anterior, significa isto que se conhecermos as probabilidades $\mathrm{P}(\cdot|A_i)$ para $i = 1, \ldots, n$, e as probabilidades $\alpha_1, \ldots, \alpha_n$ de cada um dos acontecimentos $A_1, \ldots, A_n$, respectivamente, será natural considerar numa primeira abordagem à modelação da experiência aleatória o espaço de probabilidade $(\Omega, \mathcal{A}, \mathrm{P}_\alpha)$ onde, para $C \in \mathcal{A}$, P_α é definida por $\mathrm{P}_\alpha(C) = \sum_{i=1}^{n} \alpha_i \mathrm{P}(C|A_i)$ (verifique que, para todo o i, $\mathrm{P}_\alpha(\cdot|A_i) = \mathrm{P}(\cdot|A_i)$ e $\mathrm{P}_\alpha(A_i) = \alpha_i$). Se admitirmos que a realização dum acontecimento B nos vai permitir conhecer melhor o fenómeno aleatório em estudo, e que as probabilidades $\mathrm{P}(\cdot|A_i)$ não são alteradas com a observação de B, devemos então, numa segunda etapa, substituir $\alpha_1, \ldots, \alpha_n$ por $\beta_1, \ldots, \beta_n$, onde $\beta_i = \mathrm{P}_\alpha(A_i|B)$, e considerar o novo espaço de probabilidade $(\Omega, \mathcal{A}, \mathrm{P}_\beta)$ onde $\mathrm{P}_\beta(C) = \sum_{i=1}^{n} \beta_i \mathrm{P}(C|A_i)$, para $C \in \mathcal{A}$. Os α_i e os β_i dizem-se probabilidades **a priori** e **a posteriori** dos A_i, respectivamente.

O resultado seguinte permite concluir que cada β_i, pode ser calculado a partir das probabilidades *a priori* $\alpha_1, \ldots, \alpha_n$ e das probabilidades condicionais $\mathrm{P}(\cdot|A_1), \ldots, \mathrm{P}(\cdot|A_n)$. Mais precisamente, $\beta_i = \mathrm{P}(B|A_i)\alpha_i / \sum_{j=1}^{n} \alpha_j \mathrm{P}(B|A_j)$, para $i = 1, \ldots, n$.

Teorema 1.6.4 (Teorema de Bayes) *Nas condições do teorema anterior, se* $\mathrm{P}(B) > 0$*, então, para* $i = 1, \ldots, n$*,*

$$\mathrm{P}(A_i|B) = \frac{\mathrm{P}(B|A_i)\mathrm{P}(A_i)}{\sum_{j=1}^{n} \mathrm{P}(A_j)\mathrm{P}(B|A_j)}.$$

Os dois resultados anteriores são válidos para uma infinidade numerável de acontecimentos $A_1, A_2, \ldots$ com probabilidades positivas. Em particular, se (A_i) é uma partição de Ω, a condição $B \subset A_1 \cup A_2 \cup \ldots$, é sempre verificada.

Exercícios

1. Demonstre os Teoremas 1.6.3 e 1.6.4.

2. Uma urna contém r bolas brancas e s bolas pretas. Uma bola é extraída ao acaso da urna, e é de seguida reposta na urna com mais t bolas da sua cor. Este processo é repetido novamente. Qual é a probabilidade: Da segunda bola extraída ser preta? Da primeira bola ser branca sabendo que a segunda é branca?

3. Numa determinada espécie animal, os espécimes com genótipos PP e PC são pretos e os espécimes com genótipos CC são castanhos. Um animal de cor preta, que sabemos resultar dum cruzamento $PC \times PC$, é cruzado com um animal castanho, sendo os três descendentes deste cruzamento todos pretos.

   (a) Quais as probabilidades do progenitor preto ter genótipos PP e PC, respectivamente?

   (b) Calcule as probabilidades anteriores, no caso do progenitor de cor preta resultar dum cruzamento $PP \times PC$.

   (c) Poderão as probabilidades anteriores ser calculadas no caso de apenas sabermos que o progenitor de cor preta resultou dum cruzamento $PC \times PC$ ou $PP \times PC$?

4. (**Paradoxo do teste para despiste duma doença rara**) Um teste ao sangue é utilizado para despiste duma doença rara: em 98.5% dos casos o teste dá um resultado positivo quando a doença está presente (sensibilidade do teste); em 97.5% dos casos o teste dá um resultado negativo quando a doença não está presente (especificidade do teste); 0.41% da população sofre dessa doença.

   (a) Qual a probabilidade do teste indicar que uma pessoa sofre da doença, sem sabermos nada acerca dessa pessoa?

   (b) Qual a probabilidade de efectivamente estar doente uma pessoa cujo teste indica que sofre dessa doença?

   (c) Calcule a probabilidade do teste fornecer um diagnóstico correcto.

5. Um homem acusado num caso de paternidade possui uma característica genética presente em 2% dos adultos do sexo masculino. Esta característica só pode ser transmitida de pai para filho e quando presente no progenitor é sempre transmitida para cada um dos seus descendentes. Admitindo que a probabilidade p do homem ser o pai da criança em

causa é de 0.5, determine a probabilidade do homem ser pai da criança sabendo que esta possui a referida característica genética. Calcule esta última probabilidade para $p = 0.01$ e $p = 0.001$.

6. Um saco contém duas moedas: uma normal com cara de um lado e coroa do outro, e outra com cara dos dois lados. Uma moeda é tirada ao acaso do saco.
   (a) Se pretendesse calcular a probabilidade de obter cara em dois lançamentos da moeda, qual era o espaço de probabilidade que consideraria?
   (b) A moeda tirada do saco é lançada n vezes, e os resultados obtidos são todos cara. Qual é a probabilidade da moeda que lançámos ser a que tem cara nos dois lados?
   (c) Se pretendesse calcular a probabilidade de obter cara nos próximos dois lançamentos da moeda, qual era o espaço de probabilidade que consideraria?
   (Sugestão: Utilize o Teorema de Bayes.)
7. (**Probabilidade das causas**) Sobre uma mesa estão sete urnas em tudo idênticas que denotamos por $U_0, \ldots, U_7$, contendo a urna U_i, i bolas pretas e $6 - i$ bolas brancas. De uma das urnas escolhida ao acaso, são feitas duas tiragens com reposição, tendo-se observado duas bolas brancas. Qual é a composição mais provável da urna escolhida?

1.7 Produto generalizado de probabilidades

Dados dois espaços de probabilidade $(\Omega_1, \mathcal{A}_1, \mathrm{P}_1)$ e $(\Omega_2, \mathcal{A}_2, \mathrm{P}_2)$, sabemos já que é possível definir no produto cartesiano $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2)$ uma única probabilidade $\mathrm{P}_1 \otimes \mathrm{P}_2$ que satisfaz $(\mathrm{P}_1 \otimes \mathrm{P}_2)(A_1 \times A_2) = \mathrm{P}_1(A_1)\mathrm{P}_2(A_2)$, para todo o $A_1 \in \mathcal{A}_1$ e $A_2 \in \mathcal{A}_2$. Grosso modo, e tendo em mente os Exemplos 1.3.1, 1.3.2 e 1.3.5 (com $\rho = 0$), podemos dizer que um resultado particular (x, y) da experiência aleatória descrita pela probabilidade $\mathrm{P}_1 \otimes \mathrm{P}_2$ resulta da realização de duas experiências aleatórias descritas pelas probabilidades P_1 e P_2, respectivamente, em que a probabilidade de ocorrência de y como resultado da segunda experiência não depende da ocorrência do resultado x na primeira experiência.

Tal situação não se verifica no Exemplo 1.3.5 quando $\rho \neq 0$. Na modelação da experiência aleatória aí descrita, em vez de optarmos por definir uma probabilidade P no produto cartesiano dos espaços associados às alturas normalizadas dos pais e dos filhos, poderiamos optar por decompor o problema em dois problemas mais simples, começando por modelar a experiência aleatória associada à observação das alturas normalizadas dos pais através duma probabilidade P_1 com densidade normal de parâmetros 0 e 1 (por exemplo), isto é,

$$\mathrm{P}_1(A_1) = \int_{A_1} \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2} d\lambda(x),$$

para $A_1 \in \mathcal{B}(\mathbb{R})$, modelando a seguir a experiência aleatória associada à observação das alturas dos filhos correspondentes a um progenitor cuja altura normalizada é igual a x,

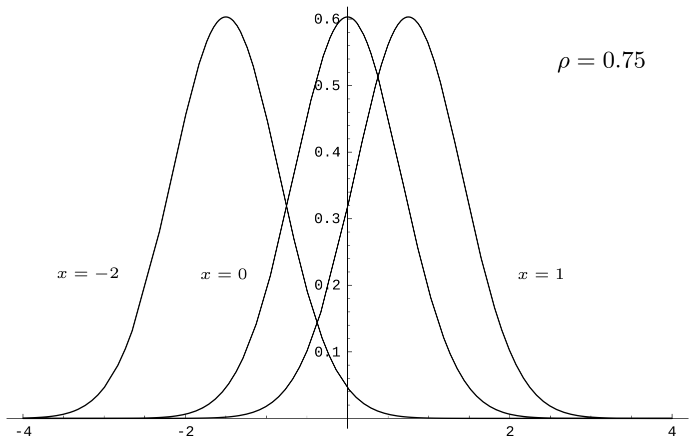


Figura 1.4: Densidade da probabilidade de transição no caso normal bivariado

por uma probabilidade $\mathrm{P}_2^1(x,\cdot)$ com densidade normal cujos parâmetros dependem de x. Tomando a densidade normal de parâmetros ρx e $1-\rho^2$, obteríamos

$$\mathrm{P}_2^1(x, A_2) = \int_{A_2} \frac{1}{\sqrt{2\pi(1-\rho^2)}}\, e^{-(y-\rho x)^2/(2(1-\rho^2))} d\lambda(y),$$

para $A_2 \in \mathcal{B}(\mathbb{R})$.

Colocado num contexto geral, o problema que naturalmente se levanta é saber se é possível a partir duma probabilidade P_1 definida sobre $(\Omega_1, \mathcal{A}_1)$, e duma família de probabilidades $\mathrm{P}_2^1(x,\cdot)$ sobre $(\Omega_2, \mathcal{A}_2)$ indexada por $x \in \Omega_1$, definir uma probabilidade P sobre $\mathcal{A}_1 \otimes \mathcal{A}_2$ que preserve as interpretações anteriores atribuídas a P_1 e a P_2^1, isto é, $\mathrm{P}_1(A_1)$ deverá ser a probabilidade P de $A_1 \times \Omega_2$, e $\mathrm{P}_2^1(x, A_2)$ deverá ser a probabilidade condicional de $\Omega_1 \times A_2$ dado $\{x\} \times \Omega_2$, sempre que este último acontecimento tenha probabilidade não-nula.

O resultado seguinte estabelece a possibilidade de definir uma tal probabilidade sobre $\mathcal{A}_1 \otimes \mathcal{A}_2$. Note que quando a família de probabilidades $\mathrm{P}_2^1(x,\cdot)$, $x \in \Omega_1$, se reduz a um único elemento P_2, a probabilidade P não é mais do que $\mathrm{P}_1 \otimes \mathrm{P}_2$.

Definição 1.7.1 *Chamamos* ***probabilidade de transição*** *sobre* $\Omega_1 \times \mathcal{A}_2$*, a uma aplicação* P_2^1 *de* $\Omega_1 \times \mathcal{A}_2$ *em* $[0,1]$ *tal que para todo o* $x \in \Omega_1$*,* $\mathrm{P}_2^1(x,\cdot)$ *é uma probabilidade sobre* $(\Omega_2, \mathcal{A}_2)$*, e para todo o* $A_2 \in \mathcal{A}_2$*,* $\mathrm{P}_2^1(\cdot, A_2)$ *é* $\mathcal{A}_1$*-mensurável.*

Teorema 1.7.2 *Sejam* $(\Omega_1, \mathcal{A}_1, \mathrm{P}_1)$ *um espaço de probabilidade,* $(\Omega_2, \mathcal{A}_2)$ *um espaço mensurável e* P_2^1 *uma probabilidade de transição sobre* $\Omega_1 \times \mathcal{A}_2$*. Então, existe uma única probabilidade* P *sobre* $\mathcal{A}_1 \otimes \mathcal{A}_2$ *tal que*

$$\mathrm{P}(A_1 \times A_2) = \int_{A_1} \mathrm{P}_2^1(x, A_2) d\mathrm{P}_1(x),$$

para todo o $A_1 \in \mathcal{A}_1$ *e* $A_2 \in \mathcal{A}_2$*.*

DEM: A fórmula anterior define P sobre a semi-álgebra $\mathcal{A}_1 \times \mathcal{A}_2$ de partes de $\Omega_1 \times \Omega_2$. Além disso, $\mathrm{P}(\Omega_1 \times \Omega_2) = 1$ e P é σ-aditiva em $\mathcal{A}_1 \times \mathcal{A}_2$ (para estabelecer a σ-aditividade de P adapte a demonstração do Teorema 6.2.1 de AMI, sobre a existência da medida produto). Para concluir basta agora usar o teorema do prolongamento (cf. AMI, §2.6) que garante a existência de um e um só prolongamento σ-aditivo de P a $\mathcal{A}_1 \otimes \mathcal{A}_2$. □

Exercícios

1. (**Problema do concurso das portas**) É-lhe proposto o seguinte jogo. Tem à sua frente três portas das quais uma contém um prémio, estando as outras duas vazias. Começa por escolher um das portas. Sem lhe ser dada nenhuma informação sobre o que contém a porta que escolheu, uma das outras duas, a que não tem o prémio, é aberta. É-lhe agora pedido para escolher entre as duas portas fechadas restantes. Qual o espaço de probabilidade que devemos associar a esta experiência? Calcule a probabilidade de ganhar o prémio considerando cada uma das seguintes estratégias:

   (a) na segunda escolha mantém a porta inicialmente escolhida;
   (b) na segunda escolha muda de porta;
   (c) na segunda escolha escolhe ao acaso uma nova porta (entre as duas que ainda estão fechadas).

2. Retome os Exercícios 1.6.2 e 1.6.6. Identifique os modelos de probabilidade associados às experiências aleatórias aí descritas.

3. Mostre que a probabilidade definida no Exemplo 1.3.5 é um caso particular do produto generalizado de probabilidades, podendo ser definida a partir duma probabilidade P_1 sobre $\mathcal{B}(\mathbb{R})$ com densidade normal de parâmetros 0 e 1, e duma probabilidade de transição P_2^1 sobre $\mathbb{R} \times \mathcal{B}(\mathbb{R})$, onde $\mathrm{P}_2^1(x, \cdot)$ tem densidade normal de parâmetros ρx e $1 - \rho^2$ (ver Figura 1.4).

1.8 Breve referência à simulação de experiências aleatórias

Algumas das experiências aleatórias descritas no §1.1 podem ser facilmente simuladas com a ajuda dum computador. Na base de todo o processo está a simulação da extracção ao acaso de pontos do intervalo $]0, 1[$ (ver Exemplo 1.1.3). É por ela que começamos.

Os algoritmos utilizados para esse fim passam pela obtenção duma sucessão $x_0, x_1, x_2, \ldots, x_n, \ldots$ de inteiros entre 1 e $m-1$, com m "grande", que pareça comportar-se como se da extracção ao acaso de pontos do conjunto $\{1, \ldots, m-1\}$ se tratasse. O método mais usado para gerar uma tal sucessão, é o **método de congruência linear**. Começando com uma "semente" x_0, x_{n+1} é obtido de x_n através da fórmula

$$x_{n+1} = ax_n + b \ (\mathrm{mod}\ m),$$

onde a e b são constantes convenientemente escolhidas, de modo que se obtenha, por um lado, uma sucessão com um período grande, e, por outro lado, que a sucessão imite a extracção ao acaso de pontos de $\{1, \ldots, m-1\}$. Para obter uma sucessão de números em $]0, 1[$ basta dividir cada x_n por m. Os números assim obtidos dizem-se **números pseudo-aleatórios**, ou mais simplesmente, **números aleatórios**.

Sendo a sucessão anterior completamente determinada pela semente x_0, para obter diferentes sucessões, diferentes valores de x_0 têm de ser escolhidos, ou pelo utilizador, ou, de forma automática, com base no relógio do computador. A partir da função "random" do compilador de Pascal dum computador Compaq (Workstation Alpha Unix) obtivemos os seguintes 50 números aleatórios (primeiras seis casas decimais):

$$\begin{gathered}0.750923,\ 0.514810,\ 0.989085,\ 0.676017,\ 0.582768,\ 0.992278,\ 0.900570,\ 0.276358,\\ 0.154543,\ 0.896320,\ 0.631060,\ 0.799246,\ 0.093678,\ 0.344508,\ 0.520097,\ 0.426544,\\ 0.189514,\ 0.070280,\ 0.458262,\ 0.145676,\ 0.270472,\ 0.428466,\ 0.193471,\ 0.095973,\\ 0.438925,\ 0.171107,\ 0.073370,\ 0.986646,\ 0.940340,\ 0.777523,\ 0.356934,\ 0.691263,\\ 0.292333,\ 0.346020,\ 0.367280,\ 0.875102,\ 0.338298,\ 0.267851,\ 0.151460,\ 0.492841,\\ 0.164171,\ 0.782520,\ 0.292087,\ 0.257849,\ 0.127028,\ 0.812184,\ 0.684393,\ 0.316542,\\ 0.882464,\ 0.142655.\end{gathered}$$

Quando nada é dito em contrário o compilador atrás referido usa o número 7774755 como semente. Para uma semente (número natural) escolhida pelo utilizador deverá utilizar a instrução "seed(semente)", e para uma semente baseada no relógio da máquina use "seed(wallclock)".

Utilizando o gerador de números aleatórios podemos também simular a experiência descrita no Exemplo 1.1.2. Se for r o número aleatório gerado, basta associar-lhe a face do dado com o número $\lfloor 6r \rfloor + 1$, onde $\lfloor x \rfloor$ denota a parte inteira de x. Por outras palavras, ocorre a face i do dado se r pertence ao subintervalo $[(i-1)/6, i/6[$ de $[0, 1[$. A partir dos números aleatórios anteriores obtemos os resultados seguintes resultados para o lançamento simulado dum dado equilibrado:

$$\begin{gathered}5,\ 4,\ 6,\ 5,\ 4,\ 6,\ 6,\ 2,\ 1,\ 6,\ 4,\ 5,\ 1,\ 3,\ 4,\ 3,\ 2,\ 1,\ 3,\ 1,\ 2,\ 3,\ 2,\ 1,\ 3,\\ 2,\ 1,\ 6,\ 6,\ 5,\ 3,\ 5,\ 2,\ 3,\ 3,\ 6,\ 3,\ 2,\ 1,\ 3,\ 1,\ 5,\ 2,\ 2,\ 1,\ 5,\ 5,\ 2,\ 6,\ 1\end{gathered}$$

De forma análoga, ainda a partir dos números aleatórios anteriores, obtemos os resultados seguintes para o lançamento simulado duma moeda equilibrada:

$$\begin{gathered}1,\ 1,\ 1,\ 1,\ 1,\ 1,\ 1,\ 0,\ 0,\ 1,\ 1,\ 1,\ 0,\ 0,\ 1,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\\ 0,\ 0,\ 1,\ 1,\ 1,\ 0,\ 1,\ 0,\ 0,\ 0,\ 1,\ 0,\ 0,\ 0,\ 0,\ 0,\ 1,\ 0,\ 0,\ 0,\ 1,\ 1,\ 0,\ 1,\ 0\end{gathered}$$

As técnicas que conjugam os métodos de simulação anterior e a interpretação frequencista de probabilidade para efectuar cálculos são conhecidos na literatura como

métodos de Monte Carlo. Alguns exemplos são apresentados nos exercícios seguintes.

Exercícios

1. Como poderia simular num computador a extracção ao acaso dum ponto do quadrado $[0,1]\times[0,1]$? Utilizando a interpretação frequencista de probabilidade (que justificaremos mais à frente), como poderia calcular de forma aproximada a área do círculo inscrito nesse quadrado?

2. Simule as experiências descritas nos Exercícios 1.1.1 e 1.3.1 num computador, e ensaie uma resposta às pergunta feitas nesses exercícios apenas com base nessa simulação.

3. Escreva um algoritmo para simular a extracção ao acaso dum ponto do intervalo $[a,b[$, para a e b quaisquer.

4. No casino de Monte Carlo a roda da roleta é dividida em 37 casas iguais, 18 vermelhas, 18 pretas e uma verde. Se um jogador aposta 1 euro na cor vermelha tem probabilidade 18/37 de ganhar e 19/37 de perder. Por simulação, e para $n = 200, 1000$ e 2000, obtenha aproximações para a probabilidade do ganho líquido do jogador ao fim de n partidas ser não-negativo.

1.9 Bibliografia

Billingsley, P. (1986). *Probability and Measure*, Wiley.

James, B.R. (1981). *Probabilidades: um curso de nível intermediário*, IMPA.

Kallenberg, O. (1997). *Foundations of Modern Probability*, Springer.

Kolmogorov, A.N. (1950). *Foundations of the Theory of Probability*, Chelsea Publishing Company (tradução do original *Grundbegriffe der Wahrscheinlichkeitrechnung*, datado de 1933).

Monfort, A. (1980). *Cours de Probabilités*, Economica.

Resnick, S.I. (1999). *A Probability Path*, Birkhäuser.

Sobre a história das Probabilidades (e não só)

Borel, E. (1950). *Éléments de la Théorie des Probabilités*, Éditions Albin Michel.

Hald, A. (1990). *A History of Probability and Statistics and their applications before 1750*, Wiley.

Hald, A. (1998). *A History of Mathematical Statistics from 1759 to 1930*, Wiley.

Sobre números aleatórios e simulação de experiências aleatórias

Grycko, E., Pohl, C., Steinert, F. (1998). *Experimental Stochastics*, Springer.

Knuth, D.E. (1981). *The Art of Computer Programming*, vol. II, Addison-Wesley.

Tompson, J.R. (2000). *Simulation: a Modeler's Approach*, Wiley.

Capítulo 2

Variáveis aleatórias e distribuições de probabilidade

Variáveis aleatórias e suas distribuições de probabilidade. Classificação das distribuições de probabilidade sobre $\mathbb{R}^d$. Função de distribuição duma variável aleatória real e dum vector aleatório. Transformação de vectores aleatórios absolutamente contínuos. Distribuições condicionais.

2.1 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. Pense no que acontece quando joga ao Monopólio e lança os dados: interessa-lhe a soma dos pontos obtidos e não os pontos ocorridos em cada um dos dados. Por outras palavras, sendo $(\Omega, \mathcal{A}, \mathrm{P})$ um modelo probabilístico para a experiência aleatória em causa, e observado um ponto $\omega \in \Omega$, interessamo-nos por uma função de ω. Surge assim de forma natural a noção de variável aleatória.

Definição 2.1.1 *Chamamos* ***variável aleatória*** *em $(E, \mathcal{B})$, onde E é um conjunto não-vazio munido duma σ-álgebra $\mathcal{B}$ de partes de E, a toda a aplicação mensurável X com valores em $(E, \mathcal{B})$ definida num espaço de probabilidade $(\Omega, \mathcal{A}, \mathrm{P})$.*

Uma variável aleatória (v.a.) X diz-se **variável aleatória real** (v.a.r.) se $E = \mathbb{R}$, **vector aleatório** (ve.a.) se $E = \mathbb{R}^d$ para algum número natural d, **sucessão aleatória** se $E = \mathbb{R}^\infty$, e **processo estocástico** ou **função aleatória** se $E = \mathbb{R}^T$ com T um conjunto infinito de índice. De acordo com a Proposição 1.5.4, se X_t, $t \in T$, é uma família qualquer de variáveis aleatórias reais definidas num mesmo espaço de probabilidade, então $X = (X_t, t \in T)$ é uma variável aleatória em $(\mathbb{R}^T, \mathcal{B}(\mathbb{R})^T)$.

Definição 2.1.2 *Se X é uma variável aleatória definida em $(\Omega, \mathcal{A}, \mathrm{P})$ com valores em $(E, \mathcal{B})$, chamamos* ***lei de probabilidade*** *ou* ***distribuição de probabilidade*** *de X, à medida imagem de P por X. Denotando por P_X uma tal medida, temos $\mathrm{P}_X = \mathrm{P}X^{-1}$, isto é,*

$$\mathrm{P}_X(B) = \mathrm{P}(X^{-1}(B)) = \mathrm{P}(\{\omega \in \Omega : X(\omega) \in B\}), \quad \textit{para } B \in \mathcal{B}.$$

Por simplicidade de escrita, escreveremos $\mathrm{P}(X \in B)$ em vez de $\mathrm{P}(\{\omega \in \Omega : X(\omega) \in B\})$. P_X é claramente uma probabilidade sobre $(E, \mathcal{B})$. Se X e Y são variáveis aleatórias com valores num mesmo espaço mensurável (mas não necessariamente definidos num mesmo espaço de probabilidade), escrevemos $X \sim Y$ sempre que X e Y tenham a mesma distribuição, isto é, sempre que $\mathrm{P}_X = \mathrm{P}_Y$. Se X e Y estão definidas num mesmo espaço de probabilidade e $X = Y$ P-quase certamente (q.c.), isto é, $\mathrm{P}(X = Y) = 1$, então $X \sim Y$. O recíproco não é verdadeiro (ver Exercício 2.1.7).

Notemos que a σ-álgebra $X^{-1}(\mathcal{B})$, que não é mais do que a σ-álgebra gerada por X, $\sigma(X)$ (cf. AMI §3.6), contém toda a "informação" sobre X necessária ao cálculo da sua distribuição de probabilidade. Quando afirmamos que uma variável aleatória tem distribuição μ sobre $(E, \mathcal{B})$, estamos a dizer que existe um espaço de probabilidade de base $(\Omega, \mathcal{A}, \mathrm{P})$ e uma variável aleatória X nele definida tal que $\mathrm{P}_X = \mu$. Normalmente apenas $(E, \mathcal{B}, \mu)$ tem interesse e nenhum relevo é assumido pelo espaço de base (ver Exercício 2.1.1).

Exemplo 2.1.3 Consideremos um espaço de probabilidade $(\Omega, \mathcal{A}, \mathrm{P})$ e seja $A \in \mathcal{A}$, com $\mathrm{P}(A) = p$. A função $X = \mathbb{1}_A$, é uma v.a. com valores em $(\{0,1\}, \mathcal{P}(\{0,1\}))$. Claramente $\sigma(X) = \sigma(A) = \{\emptyset, A, A^c, \Omega\}$ e a lei de probabilidade P_X de X é dada por $\mathrm{P}_X(B) = 0$ se $B = \emptyset$, $\mathrm{P}_X(B) = p$ se $B = \{1\}$, $\mathrm{P}_X(B) = 1 - p$ se $B = \{0\}$ e $\mathrm{P}_X(B) = 1$ se $B = \{0, 1\}$. Qualquer variável aleatória com esta distribuição será representada por $B(p)$. Assim, indicamos $X \sim B(p)$ e dizemos que X é uma variável de **Bernoulli de parâmetro** p. Dizemos também que X tem (ou segue) uma lei (ou distribuição) de Bernoulli de parâmetro p.

Proposição 2.1.4 *Se X é uma variável aleatória em $(E, \mathcal{B})$ e $g : (E, \mathcal{B}) \to (F, \mathcal{C})$ é uma aplicação mensurável, a distribuição P_X de X e g determinam a distribuição de $g(X)$. Mais precisamente, $\mathrm{P}_{g(X)}$ é a medida imagem de P_X por g:*

$$\mathrm{P}_{g(X)} = \mathrm{P}_X\, g^{-1}.$$

Dem: Para $C \in \mathcal{C}$, $\mathrm{P}_{g(X)}(C) = \mathrm{P}(X^{-1}(g^{-1}(C))) = \mathrm{P}_X(g^{-1}(C)) = (\mathrm{P}_X\, g^{-1})(C)$. $\square$

Se $X_1, \ldots, X_n$ são variáveis aleatórias definidas num mesmo espaço de probabilidade com valores em $(E_1, \mathcal{B}_1), \ldots (E_n, \mathcal{B}_n)$, respectivamente, sabemos que $X = (X_1, \ldots, X_n)$

é uma variável aleatória com valores em $(\prod_{i=1}^n E_i, \bigotimes_{i=1}^n \mathcal{B}_i)$. O resultado anterior permite concluir que conhecendo a distribuição P_X de X conhecemos também as distribuições P_{X_j} ditas **distribuições marginais de** X, uma vez que $X_j = \pi_j \circ X$ com $\pi_j : (\prod_{i=1}^n E_i, \bigotimes_{i=1}^n \mathcal{B}_i) \to (E_j, \mathcal{B}_j)$ a projecção $\pi_j(x_1, \ldots, x_n) = x_j$. As variáveis aleatórias X_j dizem-se **margens de** X.

O conhecimento das distribuições marginais de X não permite, duma forma geral, caracterizar a distribuição de X. Com efeito, os vectores (X_1, X_2) e (Y_1, Y_2) com valores em $(\{0,1\}^2, \mathcal{P}(\{0,1\}^2))$ e distribuições distintas definidas, para $(i,j) \in \{0,1\}^2$, por $\mathrm{P}_{(X_1,X_2)}(\{(i,j)\}) = 1/8$, se $i = j$, $\mathrm{P}_{(X_1,X_2)}(\{(i,j)\}) = 3/8$, se $i \neq j$, e $\mathrm{P}_{(Y_1,Y_2)}(\{(i,j)\}) = 1/4$, para todo o (i,j), têm por distribuições marginais variáveis de Bernoulli de parâmetro $1/2$.

A seguir apresentamos alguns exemplos importantes de variáveis aleatórias que estão relacionadas com os espaços de probabilidade considerados no Capítulo 1.

Exemplo 2.1.5 Considere um modelo probabilístico $(\Omega, \mathcal{A}, \mathrm{P})$ que descreva a repetição n vezes duma experiência sempre nas mesmas condições. Cada experiência tem dois resultados possíveis que vamos designar por "sucesso" e "insucesso", sendo $p \in [0,1]$ a probabilidade de sucesso em cada experiência. Se X é a v.a. que nos dá o número de sucessos obtidos nas n repetições da experiência, então P_X é uma probabilidade sobre $(\{0,1,\ldots,n\}, \mathcal{P}(\{0,1,\ldots,n\}))$, com

$$P_X(\{k\}) = \binom{n}{k} p^k (1-p)^{n-k}, \text{ para } k = 0, 1, \ldots, n.$$

Dizemos que X segue uma **distribuição binomial de parâmetros** n **e** p, e indicamos $X \sim B(n,p)$.

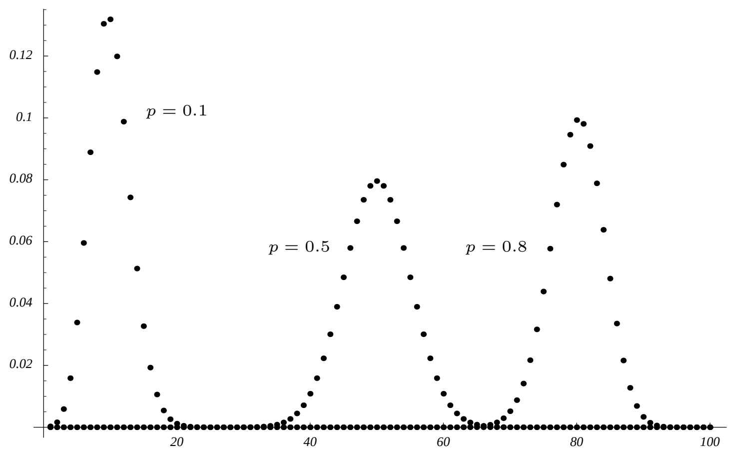


Figura 2.1: Distribuição binomial ($n = 100$)

(Obs: A distribuição binomial é um modelo para problemas de amostragem com reposição, como no caso dum problema controlo de qualidade em que um lote de peças é aceite se uma amostra escolhida ao acaso do lote não contiver “muitas” peças defeituosas.)

Exemplo 2.1.6 Seja $(\Omega, \mathcal{A}, \mathrm{P})$ o modelo probabilístico que descreve n repetições, sempre nas mesmas condições, duma experiência aleatória com k resultados possíveis $1, \ldots, k$, sendo $p_1, \ldots, p_k$ as respectivas probabilidades, onde $\sum_{i=1}^k p_i = 1$ (ver Exemplo 1.3.3). Para $i = 1, \ldots, k$, denotemos por X_i o número de ocorrências do resultado i nas n repetições da experiência. $X = (X_1, \ldots, X_k)$ é um vector aleatório em $\{0, 1, \ldots, n\}^k$, e, para $(x_1, \ldots, x_k) \in \{0, 1, \ldots, n\}^k$, temos

$$P_X(\{(x_1, \ldots, x_k)\}) = \frac{n!}{x_1! \ldots x_k!}\, p_1^{x_1} \ldots p_k^{x_k}.$$

Dizemos neste caso que X é um ve.a. **multinomial de parâmetros** $n \in \mathbb{N}$ **e** $(p_1, \ldots, p_k)$, e indicamos $X \sim M(n, p_1, \ldots, p_k)$.

Exemplo 2.1.7 Se X é uma v.a. com valores no intervalo $[a, b]$ $(a < b)$, cuja distribuição de probabilidade tem densidade, relativamente à medida de Lebesgue em $\mathbb{R}$, dada por (1.4.5), dizemos que X é uma v.a. **uniforme sobre o intervalo** $[a, b]$ e escrevemos $X \sim U([a, b])$ (ver Exemplo 1.1.3).

Exemplo 2.1.8 Se X é uma v.a. real cuja densidade de probabilidade é normal de parâmetros μ e σ^2 (cf. Exemplo 1.3.4), dizemos que X é uma v.a. **normal de parâmetros** μ **e** σ^2 e escrevemos $X \sim N(\mu, \sigma^2)$. Se $\mu = 0$ e $\sigma^2 = 1$, X diz-se **normal standard**, ou, por razões que veremos mais à frente, **normal centrada e reduzida**. (Obs: 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 denominado teorema do limite central que estudaremos no Capítulo 9.)

Exemplo 2.1.9 Se (X, Y) é um ve.a. em $\mathbb{R}^2$ com densidade de probabilidade dada por

$$f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\, \exp\left(-\frac{1}{2(1-\rho^2)}\left(\frac{(x-m_1)^2}{\sigma_1^2} - \frac{2\rho(x-m_1)(y-m_2)}{\sigma_1\sigma_2} + \frac{(y-m_2)^2}{\sigma_2^2}\right)\right),$$

para $(x, y) \in \mathbb{R}^2$, dizemos que (X, Y) é um ve.a. **normal de parâmetros** $m_1, m_2 \in \mathbb{R}$, $\sigma_1, \sigma_2 > 0$ e $-1 < \rho < 1$ (ver Exemplo 1.3.5).

Exercícios

1. Se X é uma v.a. com valores em $(E,\mathcal{B})$, sabemos que a sua lei de probabilidade é uma probabilidade sobre $(E,\mathcal{B})$. Mostre agora que se Q é uma probabilidade sobre $(E,\mathcal{B})$, existe uma v.a. X com valores em $(E,\mathcal{B})$ definida num apropriado espaço de probabilidade $(\Omega,\mathcal{A},\mathrm{P})$ tal que $\mathrm{P}_X = Q$.

2. Sejam P_n, $n \in \mathbb{N}$, medidas de probabilidade sobre $(E,\mathcal{B})$ e P definida em $(\Omega,\mathcal{A}) = (E^\infty,\mathcal{B}^\infty)$ por $\mathrm{P} = \otimes_{n=1}^\infty \mathrm{P}_n$. Considere a sucessão (X_n) definida, para $\omega = (\omega_1,\omega_2,\ldots) \in \Omega$, por $X_n(\omega) = \omega_n$ (projecção), e mostre que $\mathrm{P}_{X_n} = \mathrm{P}_n$, para todo o $n \in \mathbb{N}$.

3. Sejam T um qualquer conjunto de índices e $X = (X_t, t \in T)$ e $Y = (Y_t, t \in T)$ variáveis aleatórias com valores em $(\otimes_{t\in T} E_t, \otimes_{t\in T}\mathcal{B}_t)$. Mostre que $X \sim Y$ sse $(X_{t_1},\ldots,X_{t_n}) \sim (Y_{t_1},\ldots,Y_{t_n})$, para todo o $n \in \mathbb{N}$ e $t_1,\ldots,t_n \in T$.

4. Determine a lei de probabilidade da variável aleatória que nos dá a soma dos pontos obtidos no lançamento de dois dados equilibrados.

5. Se X é uma v.a. binomial de parâmetros n e p, mostre que $n - X$ é uma v.a. binomial de parâmetros n e $1-p$.

6. Retome o Exercício 1.8.4 e denote por S_n o ganho líquido do jogador ao fim de n partidas. Apresente uma fórmula para o cálculo de $\mathrm{P}(S_n \geq 0)$. Utilize-a quando $n = 200, 1000$ e 2000. Compare os resultados com os obtidos por simulação.

7. Sejam X e Y variáveis aleatórias definidas em $(\Omega,\mathcal{A},\mathrm{P}) = ([0,1],\mathcal{B}([0,1]),\lambda)$ por
$$X(\omega) = \omega \quad \text{e} \quad Y(\omega) = 1-\omega.$$
Mostre que $X \sim Y$ e no entanto $\mathrm{P}(X = Y) = 0$.

8. Considere um modelo probabilístico $(\Omega,\mathcal{A},\mathrm{P})$ que descreva a repetição duma experiência sempre nas mesmas condições. Cada experiência tem dois resultados possíveis que vamos designar por “sucesso” e “insucesso”, sendo $p \in [0,1]$ a probabilidade de sucesso em cada experiência. Seja X a v.a. que nos dá o número de lançamentos efectuados para obtermos o primeiro sucesso. Mostre que X tem uma **distribuição geométrica de parâmetro** $p \in [0,1]$, isto é,
$$\mathrm{P}_X(\{k\}) = (1-p)^{k-1}p, \text{ para } k \in \mathbb{N}.$$

9. No contexto do exercício anterior seja X a v.a. que nos dá o número de insucessos observados antes de obtermos o r-ésimo sucesso. Mostre que X tem uma **distribuição binomial negativa**, dita também **distribuição de Pascal**, e escrevemos $X \sim BN(r,p)$, isto é,
$$\mathrm{P}_X(\{k\}) = \tbinom{k+r-1}{r-1}p^r(1-p)^k, \text{ para } k \in \mathbb{N}_0.$$

10. Para cada $n \in \mathbb{N}$, seja X_n uma v.a. binomial de parâmetros $n \in \mathbb{N}$ e $p_n \in]0,1[$, onde $np_n \to \lambda > 0$, e X uma v.a. de **Poisson de parâmetro** λ, isto é, P_X é uma probabilidade sobre $(\mathbb{N}_0,\mathcal{P}(\mathbb{N}_0))$ definida por
$$\mathrm{P}_X(\{n\}) = e^{-\lambda}\,\frac{\lambda^n}{n!}, \text{ para } n \in \mathbb{N}_0.$$

(a) Para todo o $k \in \mathbb{N}$, mostre que

$$\frac{\mathrm{P}_{X_n}(\{k\})}{\mathrm{P}_{X_n}(\{k-1\})} \to \frac{\lambda}{k}.$$

(b) (**Convergência da binomial para a Poisson**) Para todo o $k \in \mathbb{N}_0$, conclua que

$$\mathrm{P}_{X_n}(\{k\}) \to \mathrm{P}_X(\{k\}),$$

o que justifica a designação de lei dos acontecimentos raros que é atribuída à distribuição de Poisson.

(Obs: A distribuição de Poisson é usada em problemas de filas de espera para descrever o número de chegadas de clientes a um posto de atendimento num determinado intervalo de tempo, ou, mais geralmente, para representar a realização de *acontecimentos independentes* que ocorrem com frequência constante. É também usada para descrever o número de defeitos em peças semelhantes de um dado material.)

2.2 Classificação das leis de probabilidade sobre $\mathbb{R}^d$

No parágrafo anterior vimos exemplos de leis de probabilidade discretas, como as dos Exemplos 2.1.3, 2.1.5 e 2.1.6, e de leis de probabilidade absolutamente contínuas, como as dos Exemplos 2.1.7, 2.1.8 e 2.1.9. Recordemos que uma medida ν sobre $\mathcal{B}(\mathbb{R}^d)$ se diz: **absolutamente contínua** relativamente à medida de Lebesgue, e escrevemos $\nu \ll \lambda$, se para todo o $A \in \mathcal{B}(\mathbb{R}^d)$ com $\lambda(A) = 0$, então $\nu(A) = 0$; **discreta**, se existe S quando muito numerável tal que $\nu(S^c) = 0$; **difusa**, se $\nu(\{x\}) = 0$, para todo o $x \in \mathbb{R}^d$; **alheia** relativamente à medida de Lebesgue, e escrevemos $\nu \perp \lambda$, se existe $A \in \mathcal{B}(\mathbb{R}^d)$ tal que $\nu(A) = \lambda(A^c) = 0$; **singular**, se é difusa e alheia relativamente à medida de Lebesgue.

O teorema da decomposição de Lebesgue já nosso conhecido da disciplina de Medida e Integração, e que enunciamos de seguida para medidas finitas, permitir-nos-á classificar de forma simples as leis de probabilidade sobre $\mathbb{R}^d$ (ver AMI, §8.6).

Teorema da decomposição de Lebesgue: *Se ν é uma medida finita em $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, então $\nu = \nu_0 + \nu_1$ onde ν_0 e ν_1 são medidas em $\mathbb{R}^d$ tais que $\nu_0 \perp \lambda$ e $\nu_1 \ll \lambda$. A decomposição anterior de ν, a que chamamos* ***decomposição de Lebesgue*** *de ν em relação a λ, é única.*

Teorema 2.2.1 *Seja X um vector aleatório em $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Então existem medidas ν_{ac}, ν_d e ν_s sobre $\mathcal{B}(\mathbb{R}^d)$ tais que*

$$\mathrm{P}_X = \nu_{ac} + \nu_d + \nu_s,$$

onde $\nu_{ac} \ll \lambda$, ν_d é discreta e ν_s é singular. A decomposição anterior é única. A ν_{ac}, ν_d e ν_s, chamamos ***parte absolutamente contínua, discreta*** *e* ***singular*** *de P_X, respectivamente.*

Dem: Pelo teorema da decomposição de Lebesgue, $\mathrm{P}_X = \nu_0 + \nu_1$, onde $\nu_0 \perp \lambda$ e $\nu_1 \ll \lambda$. Denotando por S, o conjunto dos pontos x para os quais $\nu_0(\{x\}) \neq 0$, um tal conjunto é quando muito numerável (porquê?). Tomando agora, para $A \in \mathcal{B}(\mathbb{R}^d)$, $\nu_2(A) = \nu_0(A \cap S)$ e $\nu_3(A) = \nu_0(A \cap S^c)$, obtemos $\nu_0 = \nu_2 + \nu_3$, com ν_2 discreta e ν_3 singular. Atendendo à unicidade da decomposição $\mathrm{P}_X = \nu_0 + \nu_1$, basta, para concluir, mostrar a unicidade da decomposição $\nu_0 = \nu_2 + \nu_3$. Suponhamos então que $\nu_0 = \nu_2' + \nu_3'$, com ν_2' discreta e ν_3' singular. Sendo S' quando muito numerável tal que $\nu_2'((S')^c) = 0$, e ν_3 e ν_3' difusas, temos $\nu_2(A) = \nu_2(A \cap (S \cup S')) = \sum_{x \in A \cap (S \cup S')} \nu_2(\{x\}) = \sum_{x \in A \cap (S \cup S')} \nu_2'(\{x\}) = \nu_2'(A \cap (S \cup S')) = \nu_2'(A)$, para $A \in \mathcal{B}(\mathbb{R}^d)$. Finalmente, sendo ν_2 finita, $\nu_3 = \nu_0 - \nu_2 = \nu_0 - \nu_2' = \nu_3'$. $\square$

Definição 2.2.2 *Se X é uma variável aleatória em $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ e ν_{ac}, ν_d e ν_s as partes absolutamente contínua, discreta e singular de P_X, respectivamente, dizemos que X (ou a sua lei de probabilidade) é* ***absolutamente contínua*** *se $\nu_d = \nu_s = 0$,* ***discreta*** *se $\nu_{ac} = \nu_s = 0$, e* ***singular*** *se $\nu_{ac} = \nu_d = 0$.*

Atendendo ao teorema de Radon-Nikodym (ver AMI, §8.4), sabemos que ν_{ac} admite a representação $\nu_{ac}(A) = \int_A f d\lambda$, $A \in \mathcal{B}(\mathbb{R}^d)$, para alguma função f mensurável de $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ em $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, não-negativa e integrável. À função f, que é única a menos dum conjunto de medida de Lebesgue nula, chamamos **derivada de Radon-Nikodym** de ν_{ac} relativamente a λ. Assim, X é absolutamente contínua sse $\mathrm{P}_X(A) = \int_A f d\lambda$, para todo o $A \in \mathcal{B}(\mathbb{R}^d)$, para alguma função f mensurável, não-negativa com $\int f d\lambda = 1$. Neste caso f diz-se **densidade de probabilidade de** X (ou de P_X).

Tendo em conta a definição de medida discreta, podemos dizer que X é discreta sse existe um subconjunto S de $\mathbb{R}^d$, quando muito numerável, tal que $\mathrm{P}_X(S) = 1$. Ao mais pequeno conjunto S (no sentido da inclusão) com estas propriedades chamamos **suporte de** X (ou de P_X) e denotamo-lo por S_X. Claramente, $S_X = \{x \in \mathbb{R}^d : \mathrm{P}_X(\{x\}) > 0\}$. A função $g : \mathbb{R}^d \to \mathbb{R}$ definida por $g(x) = \mathrm{P}_X(\{x\}) \mathbb{1}_{S_X}(x)$, diz-se **função de probabilidade de** X. Notemos que g é a derivada de Radon-Nikodym de P_X relativamente à medida contagem definida em $\mathbb{R}^d$.

Como veremos de seguida, subvectores de vectores absolutamente contínuos são absolutamente contínuos e subvectores de vectores discretos são ainda discretos.

Teorema 2.2.3 *Se $(X_1, \ldots, X_d)$ é um vector aleatório absolutamente contínuo de densidade f, então, para todo o $\{i_1, \ldots, i_m\} \subset \{1, \ldots, d\}$, $(X_{i_1}, \ldots, X_{i_m})$ é absolutamente contínuo de densidade*

$$g(x_{i_1}, \ldots, x_{i_m}) = \int_{\mathbb{R}^{d-m}} f(x_1, \ldots, x_d) d\lambda_{d-m},$$

onde λ_{d-m} representa a medida de Lebesgue em $\mathbb{R}^{d-m}$.

Dem: Para $B \in \mathcal{B}(\mathbb{R}^m)$, temos $\mathrm{P}_{(X_{i_1},\ldots,X_{i_m})}(B) = \mathrm{P}((X_1,\ldots,X_d) \in \pi^{-1}_{\{i_1,\ldots,i_m\}}(B)) = \int_{\pi^{-1}_{\{i_1,\ldots,i_m\}}(B)} f(x_1,\ldots,x_d)d\lambda_d = \int \mathbb{I}_B(x_{i_1},\ldots,x_{i_m})f(x_1,\ldots,x_d)d\lambda_d = \int_{\mathbb{R}^m} \mathbb{I}_B(x_{i_1},\ldots,x_{i_m}) \int_{\mathbb{R}^{d-m}} f(x_1,\ldots,x_d)d\lambda_{d-m}d\lambda_m = \int_B g(x_{i_1},\ldots,x_{i_m})d\lambda_m$. □

Teorema 2.2.4 *Se $(X_1,\ldots,X_d)$ é um vector aleatório discreto com suporte S e função de probabilidade g, então, para todo o $\{i_1,\ldots,i_m\} \subset \{1,\ldots,d\}$, $(X_{i_1},\ldots,X_{i_m})$ é discreto com suporte $\pi_{i_1,\ldots,i_m}(S)$ e função de probabilidade*

$$\begin{aligned} h(x_{i_1},\ldots,x_{i_m}) &= \int_{\mathbb{R}^{d-m}} g(x_1,\ldots,x_d)d\mu_{d-m} \\ &= \sum_{(x_1,\ldots,x_d)\in\pi^{-1}_{i_1,\ldots,i_m}(\{(x_{i_1},\ldots,x_{i_m})\})} g(x_1,\ldots,x_d), \end{aligned}$$

onde μ_{d-m} representa a medida contagem em $\mathbb{R}^{d-m}$.

Exercícios

1. Seja (X,Y) o ve.a. definido no Exemplo 2.1.9. Mostre que $X \sim N(m_1,\sigma_1^2)$.
2. Se $X \sim M(n,p_1,\ldots,p_k)$, mostre que $X_i \sim B(n,p_i)$, para $i=1,\ldots,k$.
3. Considere os vectores aleatórios (X,Y) de densidade
$$f(x,y) = \frac{1}{2\pi}\, e^{-(x^2+y^2)/2},$$
e (U,V) de densidade
$$g(x,y) = \frac{1}{\pi}\, e^{-(x^2+y^2)/2}\mathbb{I}_{(]-\infty,0]\times]-\infty,0])\cup([0,+\infty[\times[0,+\infty[)}(x,y),$$
para $(x,y)\in\mathbb{R}^2$. Mostre que $X \sim U$ e $Y \sim V$, e, no entanto, $(X,Y) \not\sim (U,V)$.

2.3 Função de distribuição duma variável aleatória real

Apresentamos neste parágrafo um instrumento importante no estudo da distribuição de probabilidade duma variável aleatória real X definida num espaço de probabilidade $(\Omega,\mathcal{A},\mathrm{P})$.

Definição 2.3.1 *Chamamos **função de distribuição de** X, e denotamo-la por F_X, à função de distribuição de P_X, isto é,*

$$F_X(x) = \mathrm{P}_X(]-\infty,x]) = \mathrm{P}(X \le x),\ x\in\mathbb{R}.$$

Proposição 2.3.2 *F_X satisfaz as seguintes propriedades:*

a) F_X é não-decrescente e contínua à direita.

b) $F_X(x) \to 0$ ou 1, se $x \to -\infty$ ou $x \to +\infty$, respectivamente.

c) $\mathrm{P}_X(\{a\}) = F_X(a) - F_X(a^-)$, $\mathrm{P}_X(]a,b]) = F_X(b) - F_X(a)$, $\mathrm{P}_X([a,b]) = F_X(b) - F_X(a^-)$, $\mathrm{P}_X(]a,b[) = F_X(b^-) - F_X(a)$ e $\mathrm{P}_X([a,b[) = F_X(b^-) - F_X(a^-)$, para todo o $-\infty < a < b < +\infty$.

d) F_X é contínua em $x \in \mathbb{R}$ sse $\mathrm{P}_X(\{x\}) = 0$.

e) O conjunto dos pontos de descontinuidade de F_X é quando muito numerável.

f) F_X caracteriza P_X (isto é, $F_X = F_Y$ sse $X \sim Y$)

Dem: Demonstraremos apenas a alínea f). A demonstração das restantes alíneas fica ao cuidado do aluno. Se $X \sim Y$ então $\mathrm{P}_X = \mathrm{P}_Y$ e consequentemente $F_X = F_Y$. Reciprocamente, se $F_X = F_Y$ para $a, b \in \mathbb{R}$, temos $\mathrm{P}_X(]a,b]) = F_X(b) - F_X(a) = F_Y(b) - F_Y(a) = \mathrm{P}_Y(]a,b])$, ou ainda, $\mathrm{P}_X = \mathrm{P}_Y$ pelo lema da igualdade de medidas (ver AMI, §2.6). □

Notemos que, atendendo à alínea d), X é difusa sse F_X é contínua em $\mathbb{R}$. Além disso, das alíneas d) e e), e da decomposição de Lebesgue, concluímos que a parte discreta de P_X tem por suporte o conjunto dos pontos de descontinuidade de F_X.

O resultado seguinte dá-nos duas caracterizações da continuidade absoluta duma variável aleatória real em termos da sua função de distribuição. A sua demonstração fica como exercício.

Teorema 2.3.3 *Se X é uma variável aleatória real, são equivalentes as seguintes proposições:*

i) X é absolutamente contínua.

ii) $F_X(x) = \int_{]-\infty,x]} f d\lambda$, para alguma função não-negativa e mensurável f, com $\int f d\lambda = 1$.

O resultado anterior e o teorema da diferenciação de Lebesgue que a seguir enunciamos (ver Rudin, 1974, pg. 176, e AMI, §9.3), permitem-nos, no caso de X ser absolutamente contínua, garantir a diferenciabilidade quase em todo o ponto de F_X, bem como relacionar F'_X com a densidade de probabilidade de X.

Teorema da diferenciação de Lebesgue: *Se $F(x) = \int_{]-\infty,x]} f\, d\lambda$, para $x \in \mathbb{R}$, onde $f : \mathbb{R} \to \mathbb{R}$ é $\mathcal{B}(\mathbb{R})$-mensurável e integrável, então F possui derivada em quase todo o ponto de $\mathbb{R}$ e $F' = f$, λ-q.t.p.*

Teorema 2.3.4 *Se X é uma variável aleatória real absolutamente contínua de densidade f, então F_X possui derivada em λ-quase todo o ponto de $\mathbb{R}$ e $F'_X = f$, λ-q.t.p.*

Mesmo no caso em que X não é necessariamente uma v.a. absolutamente contínua, é possível obter o resultado seguinte (ver Rudin, 1974, pg. 177).

Teorema 2.3.5 *Se X é uma variável aleatória real então F_X possui derivada em λ-quase todo o ponto de $\mathbb{R}$ e $F'_X = f_{ac}$, λ-q.t.p., onde f_{ac} é a derivada de Radon-Nikodym da parte absolutamente contínua de P_X.*

Terminamos este parágrafo estabelecendo duas condições suficientes para a continuidade absoluta duma variável aleatória em termos da sua função de distribuição.

Teorema 2.3.6 *Se X é uma variável aleatória real e F_X satisfaz pelo menos uma das condições a) $\int F'_X d\lambda = 1$ ou b) F_X é continuamente diferenciável em $\mathbb{R}$, então X é absolutamente contínua.*

DEM: a) Atendendo aos Teoremas 2.2.1 e 2.3.5, podemos escrever $\mathrm{P}_X = F'_X\lambda + \nu_d + \nu_s$. Se F'_X é tal que $\int F'_X d\lambda = 1$, obtemos então $\mathrm{P}_X(\mathbb{R}) = 1 + \nu_d(\mathbb{R}) + \nu_s(\mathbb{R})$, ou ainda, $\nu_d = \nu_s = 0$, isto é, X é absolutamente contínua. b) Pelo teorema fundamental do cálculo, $\int_{]a,b]} F'_X d\lambda = \int_{]a,b]} F'_X(t)dt$ (integral de Riemann) $= F_X(b) - F_X(a) = \mathrm{P}_X(]a,b])$, para todo o $a < b$ em $\mathbb{R}$. Como F'_X é não-negativa concluímos que F'_X é λ-integrável e que $\int F'_X d\lambda = 1$. $\square$

Exercícios

1. Sejam $a \in \mathbb{R}$ e X uma v.a. constantemente igual a a (dizemos que X é **degenerada**). Mostre que $\mathrm{P}_X = \delta_a$, isto é, a lei de probabilidade de X é a medida de Dirac no ponto a, e determine a função de distribuição F_X de X.

2. Seja X uma v.a. **uniforme discreta** sobre o conjunto $\{1, 2, \ldots, n\}$, isto é, X toma valores no conjunto $\{1, 2, \ldots, n\}$ e
$$\mathrm{P}_X(\{j\}) = 1/n, \text{ para } j = 1, \ldots, n.$$
Determine a função de distribuição de X.

3. Sejam U uma v.a.r. centrada e reduzida, isto é, $U \sim N(0,1)$, e X definida por $X = \sigma U + \mu$, com $\mu \in \mathbb{R}$ e $\sigma > 0$ fixos. Mostre que $X \sim N(\mu, \sigma^2)$.

4. Sejam X uma v.a. uniforme sobre o intervalo $[a,b]$, e Y a v.a.r. definida em $([0,1], \mathcal{B}([0,1]), \lambda)$ por $Y(\omega) = (1-\omega)a + \omega b$.
   (a) Determine a função de distribuição de X.
   (b) Mostre que $Y \sim X$.

5. Denotemos por X a v.a. que descreve a "extracção ao acaso dum ponto do intervalo $[0,1]$". Determine a função de distribuição de X^2 e conclua que X^2 é absolutamente contínua. Descreverá X^2 a extracção ao acaso dum ponto do intervalo $[0,1]$?

6. Sendo X uma v.a. normal de parâmetros 0 e 1, mostre que X^2 admite por densidade de probabilidade
$$f(x) = \begin{cases} \frac{1}{\sqrt{2\pi}} x^{-1/2} e^{-x/2}, & \text{se } x \geq 0 \\ 0, & \text{se } x < 0. \end{cases}$$

7. Considere a v.a. X de $([0,1], \mathcal{B}([0,1), \lambda)$ em $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, definida por $X(\omega) = \omega$, se $0 \leq \omega < 1/2$, $X(\omega) = 1/2$, se $1/2 \leq \omega \leq 3/4$, e $X(\omega) = 2\omega$, se $3/4 < \omega \leq 1$. Determine a função de distribuição de X e identifique as partes absolutamente contínua, discreta e singular de P_X.

8. Dizemos que uma v.a.r. X tem uma distribuição **exponencial de parâmetro** $\lambda > 0$, e escrevemos $X \sim E(\lambda)$, se admite uma densidade de probabilidade da forma
$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{se } x > 0 \\ 0, & \text{se } x \leq 0. \end{cases}$$

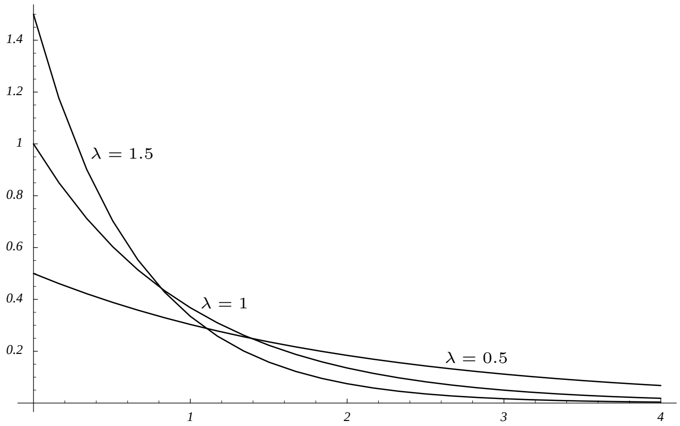


Figura 2.2: Distribuição exponencial

(a) Determine a função de distribuição F_X.

(b) Mostre que se $U \sim U([0,1[)$, então, para $\lambda > 0$, $X \sim -\frac{1}{\lambda}\ln(1-U)$.

(Obs: A distribuição exponencial é usada como modelo para o tempo de funcionamento duma componente ou sistema, quando assumimos que o número de falhas por unidade de tempo é constante, ou para descrever o tempo que medeia entre chegadas consecutivas de clientes a um posto de atendimento, quando assumimos que o número de chegadas por unidade de tempo é constante.)

9. (**Representação de Skorokhod duma v.a.r.**) Sejam X uma v.a.r. com função de distribuição F e
$$F^{\leftarrow}(x) = \inf\{s \in \mathbb{R} : F(s) \geq x\},$$
para $x \in]0,1[$ ($F^{\leftarrow}$ diz-se **inversa generalizada de** F ou **função quantil de** F).

(a) Mostre que:
i. $F^{\leftarrow}(x) \leq u$ sse $x \leq F(u)$, para $u \in \mathbb{R}$; ii. Se $U \sim U(]0,1[)$, então $F^{\leftarrow}(U) \sim X$.

(b) Se X está definida num espaço de probabilidade $(E, \mathcal{F}, Q)$, mostre que existe uma v.a. real Y definida em $(]0,1[, \mathcal{B}(]0,1[), \lambda)$ tal que $X \sim Y$.

(c) Sendo F contínua, mostre que:
i. $F(F^{\leftarrow}(x)) = x$; ii. $F(X) \sim U([0,1])$.

10. Se X é uma v.a.r. com função de distribuição F contínua em $\mathbb{R}$ e estritamente crescente quando $0 < F(x) < 1$, sabemos do exercício anterior que $F^{-1}(U) \sim X$, quando $U \sim U(]0,1[)$. Atendendo a que pode simular uma v.a. uniforme sobre o intervalo $]0,1[$ (ver §1.8), implemente a simulação das variáveis aleatórias reais seguintes cuja densidade de probabilidade se indica (ver Figuras 2.3-2.6):

(a) **Cauchy** de parâmetros α e β:

$$f(x) = (\beta\pi(1 + (x-\alpha)^2/\beta^2))^{-1},\ x \in \mathbb{R} \quad (\alpha \in \mathbb{R},\ \beta > 0);$$

(b) **Laplace** de parâmetros α e β:

$$f(x) = \beta e^{-\beta|x-\alpha|}/2,\ x \in \mathbb{R} \quad (\alpha \in \mathbb{R},\ \beta > 0);$$

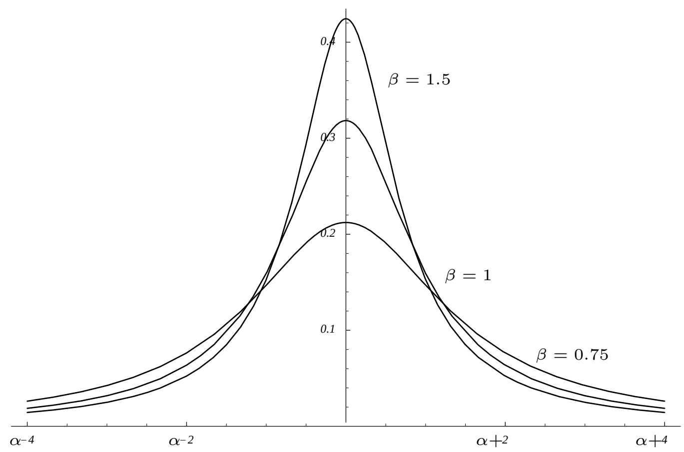


Figura 2.3: Distribuição de Cauchy

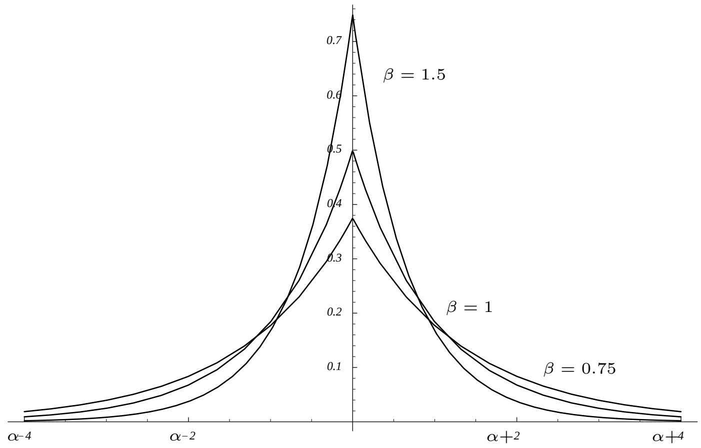


Figura 2.4: Distribuição de Laplace

(c) **Logística** de parâmetros α e β:

$$f(x) = e^{-(x-\alpha)/\beta}(1 + e^{-(x-\alpha)/\beta})^{-2}/\beta,\ x \in \mathbb{R} \quad (\alpha \in \mathbb{R},\ \beta > 0);$$

(d) **Weibull** de parâmetros α e β:

$$f(x) = \alpha^\beta \beta x^{\beta-1} e^{-(\alpha x)^\beta},\ x \geq 0 \quad (\alpha,\ \beta > 0).$$

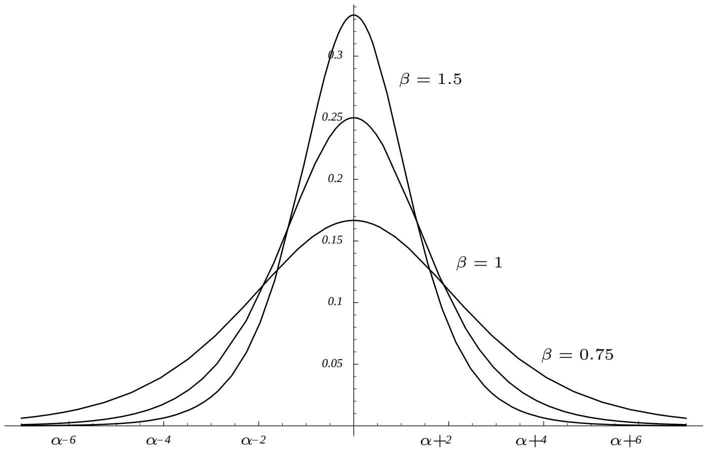


Figura 2.5: Distribuição logística

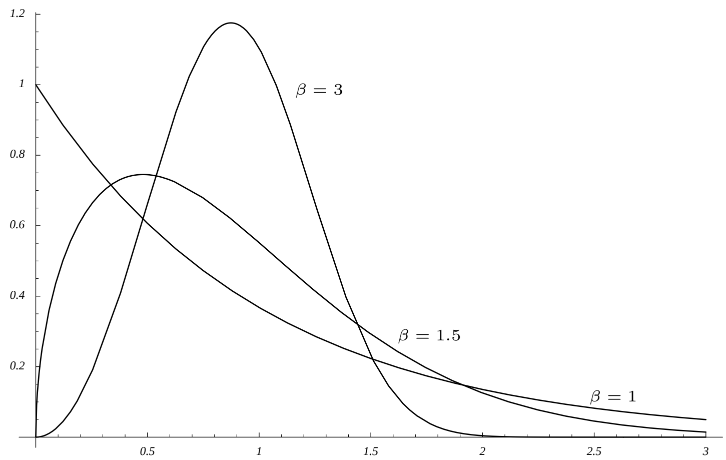


Figura 2.6: Distribuição de Weibull ($\alpha = 1$)

2.4 Função de distribuição dum vector aleatório

Neste parágrafo generalizamos a noção de função de distribuição ao caso multivariado. A notação que a seguir utilizamos foi introduzida no Exemplo 1.4.3.

Definição 2.4.1 *Chamamos* ***função de distribuição*** *do vector aleatório* $X = (X_1, \ldots, X_d)$*, e denotamo-la por* F_X*, à função de distribuição de* P_X*, isto é,*

$$F_X(x) = \mathrm{P}_X(]-\infty, x]) = \mathrm{P}(X \leq x),\ x \in \mathbb{R}^d.$$

Proposição 2.4.2 *F_X goza das seguintes propriedades:*

a) F_X é contínua à direita e não-decrescente coordenada a coordenada;

b) $F_X(x) \to 0$ ou 1, se $\min_{i=1,\ldots,d} x_i \to -\infty$ ou $+\infty$, respectivamente;

c) Para $a \leq b$, $\mathrm{P}_X(]a,b]) = \sum_{x \in V} \mathrm{sgn}(x) F_X(x)$, onde V é o conjunto dos vértices de $]a,b]$;

d) F_X caracteriza P_X.

Dem: As alíneas a) e b) obtêm-se como no caso real. A alínea c) é consequência da decomposição $]a,b] =]-\infty, b] - \bigcup_{i=1}^{d}]-\infty, (b_1, \ldots, b_{i-1}, a_i, b_{i+1}, \ldots, b_d)]$ e da Fórmula de Daniel da Silva. A alínea d) obtém-se de c) e do lema da igualdade de medidas. □

Sendo F_X contínua à direita e não-decrescente coordenada a coordenada, a continuidade de F_X num ponto é equivalente à continuidade à esquerda nesse ponto. No resultado seguinte estabelecemos uma condição necessária e suficiente para que um ponto de $\mathbb{R}^d$ seja ponto de continuidade de F_X.

Teorema 2.4.3 *Sejam X um vector aleatório em $\mathbb{R}^n$, $x \in \mathbb{R}^d$, fixo, e $fr(]-\infty, x])$ a fronteira de $]-\infty, x]$. Então F_X é contínua em x sse $\mathrm{P}_X(fr(]-\infty, x])) = 0$.*

Dem: Sendo (ϵ_n) uma sucessão em $\mathbb{R}^d$ com $0 \leq \epsilon_n \downarrow 0$, temos, para $x \in \mathbb{R}^d$, $]-\infty, x] -]-\infty, x - \epsilon_n] \downarrow fr(]-\infty, x])$, e assim $\mathrm{P}_X(fr(]-\infty, x])) = F_X(x) - \lim F_X(x - \epsilon_n)$, o que permite concluir. □

No caso real, a continuidade de F_X em $\mathbb{R}$ é condição necessária e suficiente para que X seja difusa. Como podemos concluir do resultado anterior, no caso multidimensional a continuidade de F_X em $\mathbb{R}^d$ apesar de suficiente não é condição necessária para que X seja difuso.

Aplicações sucessivas do teorema da diferenciação de Lebesgue, permitem generalizar o Teorema 2.3.4 ao caso multidimensional.

Teorema 2.4.4 *Se X é um vector aleatório em $\mathbb{R}^d$ absolutamente contínuo de função de distribuição F_X, então $\dfrac{\partial^d F_X}{\partial x_1 \ldots \partial x_d}$ existe em λ-quase todo o ponto de $\mathbb{R}^d$ e é uma versão da densidade de probabilidade de X.*

Terminamos este parágrafo, notando que conhecida a função de distribuição dum vector X, podemos facilmente obter a função de distribuição dum seu subvector.

Teorema 2.4.5 *Se F_X é a função de distribuição de $(X_1, \ldots, X_d)$, então para $\{i_1, \ldots, i_m\} \subset \{1, \ldots, d\}$, a função de distribuição de $(X_{i_1}, \ldots, X_{i_m})$ é dada por*

$$F_{(X_{i_1}, \ldots, X_{i_m})}(x_{i_1}, \ldots, x_{i_m}) = \lim F_X(x_1, \ldots, x_d),$$

onde o limite anterior é tomado quando $x_j \to +\infty$*, para todo o* $j \in \{1, \ldots, d\} \setminus \{i_1, \ldots, i_m\}$.

Exercícios

1. Se $U \sim N(0,1)$, mostre que o ve.a. $(U,0)$ em $\mathbb{R}^2$ é difuso e estude a sua função de distribuição quanto à continuidade.
2. Se (X,Y) é um ve.a. em $\mathbb{R}^2$ com função de distribuição F, mostre que $\frac{\partial^2 F}{\partial x \partial y}$ está definida em quase todo o ponto de $\mathbb{R}^2$ e é não-negativa. Além disso, mostre que se F é de classe C^2 então (X,Y) é absolutamente contínuo.

2.5 Transformação de vectores absolutamente contínuos

Suponhamos que X e Y são vectores aleatórios em $\mathbb{R}^d$ tais que $Y = g(X)$ com $g : U \to V$, bijectiva entre os abertos U e V, e g e g^{-1} de classe C^1. Mostramos neste parágrafo que Y é absolutamente contínuo se X o for, e determinamos a densidade de probabilidade de Y em função da de X. Um tal resultado é uma consequência imediata do teorema da mudança de variável no integral de Lebesgue que recordamos de seguida (ver AMI, §§7.3, 7.4).

Teorema da mudança de variável: *Nas condições anteriores, seja* $f : U \to \overline{\mathbb{R}}$ $\mathcal{B}(U)$*-mensurável. Se* f *é não-negativa, então*

$$\int_U f d\lambda = \int_V (f \circ g^{-1})(x) |det(J_{g^{-1}}(x))| d\lambda(x),$$

onde $J_{g^{-1}}(x)$ *representa a matriz jacobiana de* g^{-1} *no ponto* x*. Além disso, para* f *qualquer, a* λ*-integrabilidade de* f *é equivalente à* λ*-integrabilidade de* $(f \circ g^{-1})(\cdot)|det(J_{g^{-1}}(\cdot))|$*, e nesse caso vale a igualdade anterior.*

Teorema 2.5.1 *Nas condições anteriores, se* X *é absolutamente contínuo com densidade* f*, então* Y *é absolutamente contínuo e uma versão da sua densidade de probabilidade é dada por*

$$h(x) = \begin{cases} (f \circ g^{-1})(x) |\det(J_{g^{-1}}(x))|, & se\ x \in V \\ 0 & se\ x \notin V. \end{cases}$$

Dem: Para $B \in \mathcal{B}(V)$, temos $\mathrm{P}_Y(B) = \mathrm{P}(g(X) \in B) = \mathrm{P}(X \in g^{-1}(B)) = \int_{g^{-1}(B)} f d\lambda = \int_U f \mathbb{1}_{g^{-1}(B)} d\lambda = \int_V (f \mathbb{1}_{g^{-1}(B)} \circ g^{-1})(x) |\det(J_{g^{-1}}(x))| d\lambda(x) = \int_V (f \circ g^{-1})(x) \mathbb{1}_B(x) |\det(J_{g^{-1}}(x))| d\lambda(x) = \int_B (f \circ g^{-1})(x) |\det(J_{g^{-1}}(x))| d\lambda(x)$. $\square$

Uma aplicação interessante do resultado anterior surge na determinação da densidade de probabilidade da soma de duas variáveis aleatórias X e Y com valores em $\mathbb{R}^d$, quando o vector (X, Y) tem por densidade $(x, y) \to f(x)g(y)$, com f e g densidades de probabilidade em $\mathbb{R}^d$. Pelo teorema anterior, o vector $(X + Y, Y)$ tem por densidade $(u, v) \to f(u - v)g(v)$, e pelo Teorema 2.2.3 a densidade h de $X + Y$ é dada por $h(u) = \int f(u - v)g(v)d\lambda(v)$, a que chamamos **convolução das densidades** f e g, e que denotamos por $f \star g$. Voltaremos a este assunto no Capítulo 4.

Exercícios

1. Retome o Exercício 2.3.5. Use o Teorema da transformação de variáveis aleatórias absolutamente contínuas para determinar a densidade de probabilidade de X^2.

2. Sejam (X, Y) o ve.a. definido no Exercício 2.2.3, e $Z = X + Y$. Mostre que $Z \sim N(0, 2)$.

3. Seja (X, Y) um ponto escolhido ao acaso no quadrado $[0, 1] \times [0, 1]$. Determine a distribuição de $Z = X + Y$, dita **distribuição triangular** sobre o intervalo $[0, 2]$.

4. Se (X, Y) é um ve.a. com valores em $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$ e densidade f, mostre que as v.a. $Z_1 = XY$ e $Z_2 = X/Y$ são absolutamente contínuas com densidades
$$g_1(z) = \int f(u, z/u)/|u|\, d\lambda(u), \text{ para } z \in \mathbb{R},$$
e
$$g_2(z) = \int f(zv, v)|v|\, d\lambda(v), \text{ para } z \in \mathbb{R},$$
respectivamente. Se (X, Y) é o ve.a. definido no Exercício 2.2.3, conclua que Z_2 possui uma distribuição de Cauchy de parâmetros 0 e 1.

5. Sejam (X, Y) o ve.a. definido no Exercício 2.2.3 e $Z = X^2 + Y^2$.

   (a) Mostre que, para $A \in \mathcal{B}(\mathbb{R})$,
   $$\mathrm{P}(Z \in A) = \int\int \mathbb{I}_A(x^2 + y^2)\frac{1}{2\pi}\, e^{-(x^2+y^2)/2}d\lambda(x)d\lambda(y).$$

   (b) Conclua que Z segue uma lei exponencial de parâmetro 1/2.

6. (**Método de Box-Muller para simulação de variáveis normais** ([1])) Seja (U, V) um ve.a. com distribuição uniforme sobre o rectângulo $[0, 1[\times[0, 1[$.

   (a) Determine a densidade de probabilidade do vector $(R, \Theta) = (\sqrt{-2\ln(1 - U)}, 2\pi V)$ e conclua que Θ possui uma distribuição uniforme sobre o intervalo $[0, 2\pi[$ e que R possui uma **distribuição de Rayleigh**, isto é, R tem por densidade
   $$f_R(r) = re^{-r^2/2}\,\mathbb{I}_{[0,+\infty[}(r).$$

   (b) Mostre que $X = R\cos\Theta$ possui uma distribuição normal standard.

[1] Box, G.E.P., Muller, M.E., *Ann. Math. Stat.* 29, 610–611, 1958.

2.6 Distribuições condicionais

Dada uma probabilidade P_1 sobre $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ e uma probabilidade de transição Q sobre $\mathbb{R}^n \times \mathcal{B}(\mathbb{R}^m)$, sabemos do §1.7 que existe um vector aleatório (X, Y) definido num espaço de probabilidade $(\Omega, \mathcal{A}, \mathrm{P})$ tal que $\mathrm{P}_X = \mathrm{P}_1$ e

$$\mathrm{P}_{(X,Y)}(A \times B) = \int_A Q(x, B) d\mathrm{P}_X(x), \tag{2.6.1}$$

para todo o $A \times B \in \mathcal{B}(\mathbb{R}^n) \times \mathcal{B}(\mathbb{R}^m)$.

O problema que agora consideramos pode ser visto como o inverso do anterior. Dado um vector aleatório (X, Y) definido num espaço de probabilidade $(\Omega, \mathcal{A}, \mathrm{P})$ e com valores em $(\mathbb{R}^n \times \mathbb{R}^m, \mathcal{B}(\mathbb{R}^n) \otimes \mathcal{B}(\mathbb{R}^m))$, será possível escrever a sua distribuição de probabilidade na forma (2.6.1) para alguma probabilidade de transição Q sobre $\mathbb{R}^n \times \mathcal{B}(\mathbb{R}^m)$? A resposta a esta questão é afirmativa mas a sua justificação completa ultrapassa largamente os objectivos deste curso ([2]). Vamos contentar-nos com algumas respostas parciais.

Admitamos em primeiro lugar que X é discreto. Tomando, para $B \in \mathcal{B}(\mathbb{R}^m)$,

$$Q(x, B) = \begin{cases} \mathrm{P}(Y \in B | X = x), & \text{se } \mathrm{P}(X = x) > 0 \\ \nu(B), & \text{se } \mathrm{P}(X = x) = 0, \end{cases}$$

onde ν é uma probabilidade fixa sobre $\mathcal{B}(\mathbb{R}^m)$, concluímos que Q é uma probabilidade de transição sobre $\mathbb{R}^n \times \mathcal{B}(\mathbb{R}^m)$ e, para $A \times B \in \mathcal{B}(\mathbb{R}^n) \times \mathcal{B}(\mathbb{R}^m)$,

$$\begin{aligned} &\int_A Q(x, B) d\mathrm{P}_X(x) \\ &\quad = \sum_{x \in A : \mathrm{P}(X=x)>0} \mathrm{P}(Y \in B | X = x) \mathrm{P}(X = x) \\ &\quad = \sum_{x \in A : \mathrm{P}(X=x)>0} \mathrm{P}(X = x, Y \in B) \\ &\quad = \mathrm{P}_{(X,Y)}(A \times B). \end{aligned}$$

O mesmo acontece quando (X, Y) é um vector absolutamente contínuo com densidade f, bastando definir

$$Q(x, B) = \begin{cases} \displaystyle\int_B \frac{f(x, y)}{f_X(x)} d\lambda(y), & \text{se } f_X(x) > 0 \\ \nu(B), & \text{se } f_X(x) = 0, \end{cases}$$

[2] No caso das variáveis X e Y tomarem valores em espaços gerais, o resultado pode não ser verdadeiro (ver Hennequin e Tortrat, 1965, pg. 236–238).

onde $f_X(x) = \int f(x,y)d\lambda(y)$ e ν é uma probabilidade fixa sobre $\mathcal{B}(\mathbb{R}^m)$. Com efeito,

$$\begin{aligned}
&\int_A Q(x,B)d\mathrm{P}_X(x) \\
&\qquad = \int_A \int_B \frac{f(x,y)}{f_X(x)}\, d\lambda(y) f_X(x)\, d\lambda(x) \\
&\qquad = \int_{A\times B} f(x,y)\, d\lambda(y) d\lambda(x) \\
&\qquad = \mathrm{P}_{(X,Y)}(A\times B),
\end{aligned}$$

para $A \times B \in \mathcal{B}(\mathbb{R}^n) \times \mathcal{B}(\mathbb{R}^m)$. A aplicação $y \to f_Y(y|X = x) = \frac{f(x,y)}{f_X(x)}$, que não é mais do que uma versão de derivada de Radon-Nikodym de $Q(x,\cdot)$ relativamente a λ, diz-se **densidade condicional de Y dado $X = x$**. A densidade de (X,Y) pode ser assim obtida a partir de f_X e de $f_Y(\cdot|X = \cdot)$ pela fórmula $f(x,y) = f_X(x)f_Y(y|X = x)$.

Definição 2.6.2 *Sejam X e Y são vectores aleatórios definidos num espaço de probabilidade $(\Omega, \mathcal{A}, \mathrm{P})$ com valores em $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ e $(\mathbb{R}^m, \mathcal{B}(\mathbb{R}^n))$, respectivamente. Toda a probabilidade de transição Q sobre $\mathbb{R}^n \times \mathcal{B}(\mathbb{R}^m)$ satisfazendo*

$$\int_A Q(x,B)d\mathrm{P}_X(x) = \mathrm{P}_{(X,Y)}(A\times B),$$

para todo o $A \times B \in \mathcal{B}(\mathbb{R}^n) \times \mathcal{B}(\mathbb{R}^m)$, é dita ***lei ou distribuição condicional de Y dado X****, e é denotada por $\mathrm{P}_Y(\cdot|X = \cdot)$. A $\mathrm{P}_Y(\cdot|X = x)$ chamamos* ***lei ou distribuição condicional de Y dado $X = x$****.*

Observemos que no caso particular em que X é discreto, e tal como a notação sugere, $\mathrm{P}_Y(\cdot|X = x)$, para $x \in \mathbb{R}^n$ com $\mathrm{P}(X = x) > 0$, é efectivamente a distribuição de probabilidade de Y quando Y é considerada definida no espaço de probabilidade $(\Omega, \mathcal{A}, \mathrm{P}(\cdot|X = x))$.

Notemos também que se $\mathrm{P}_{Y,1}(\cdot|X = \cdot)$ e $\mathrm{P}_{Y,2}(\cdot|X = \cdot)$ são distribuições condicionais de Y dado X, então $\mathrm{P}_{Y,1}(\cdot|X = x) = \mathrm{P}_{Y,2}(\cdot|X = x)$, para P_X-quase todo o ponto x de $\mathbb{R}^n$.

Exercícios

1. Sejam X uma v.a. com valores em $\mathbb{R}^n$ e $Y = g(X)$ com $g : \mathbb{R}^n \to \mathbb{R}^m$ uma aplicação mensurável. Determine $\mathrm{P}_Y(\cdot|X = \cdot)$.

2. Seja (X,Y) um ve.a. em $\mathbb{R}^2$ com $X \sim N(0,1)$ e cuja distribuição condicional de Y dado $X = x$ tem uma distribuição $N(x,1)$. Prove que $Y \sim N(0,2)$.

3. Um ponto X é escolhido ao acaso do intervalo $[a, b]$ e a seguir um ponto Y é escolhido ao acaso do intervalo $[X, b]$. Mostre que a densidade de probabilidade de Y é dada, para $y \in \mathbb{R}$, por
$$f_Y(y) = \frac{1}{b-a} \ln\Big(\frac{b-a}{b-y}\Big) 1\!\!1_{[a,b[}(y).$$

4. Um animal põe um certo número X de ovos segundo uma distribuição de Poisson de parâmetro λ. Cada um desses ovos, independentemente dos outros, dá origem a um novo animal com probabilidade p. Denotando por Y o número de crias de cada ninhada, determine a distribuição de Y.

   (Sugestão: Comece por determinar a distribuição condicional de Y dado $X = n$.)

2.7 Bibliografia

Hennequin, P.L., Tortrat, A. (1965). *Théorie des Probabilités et Quelques Applications*, Masson.

Jacod, J., Protter, P. (2000). *Probability Essentials*, Springer.

Kallenberg, O. (1997). *Foundations of Modern Probability*, Springer.

Rudin, W. (1974). *Real and Complex Analysis*, McGraw-Hill.

Capítulo 3

Independência

Independência de acontecimentos aleatórios, de classes e de variáveis aleatórias. Caracterizações da independência duma família de variáveis aleatórias. Distribuição da soma de variáveis aleatórias independentes. Leis zero-um de Borel e de Kolmogorov.

3.1 Independência de classes de acontecimentos aleatórios

Introduzimos neste capítulo uma das mais importantes noções que abordamos neste curso. Trata-se da noção de independência cujas implicações serão exploradas neste e nos próximos capítulos.

Se A e B são acontecimentos aleatórios dum espaço de probabilidade $(\Omega, \mathcal{A}, \mathrm{P})$, com $\mathrm{P}(B) > 0$, a probabilidade condicionada $\mathrm{P}(A|B)$ pode ser interpretada como a probabilidade do acontecimento A quando sabemos que o acontecimento B se realizou. O facto de sabermos que B se realizou, pode, ou não, alterar a probabilidade $\mathrm{P}(A)$ do acontecimento A, isto é, pode, ou não, verificar-se a igualdade $\mathrm{P}(A|B) = \mathrm{P}(A)$, ou ainda, $\mathrm{P}(A \cap B) = \mathrm{P}(A)\mathrm{P}(B)$. Tal facto motiva a definição seguinte.

Definição 3.1.1 *Os acontecimentos aleatórios $A_t, t \in T$, onde T denota um qualquer conjunto de índices, dizem-se* ***independentes****, se para qualquer conjunto finito de índices distintos $t_1, \ldots, t_n \in T$, $\mathrm{P}\left(\bigcap_{k=1}^n A_{t_k}\right) = \prod_{k=1}^n \mathrm{P}(A_{t_k})$.*

Notemos que os acontecimentos duma família podem ser dois a dois independentes sem serem (colectivamente) independentes. Para ilustrar tal situação, considere, por exemplo, $\Omega = \{0,1\}^2$, $\mathcal{A} = \mathcal{P}(\Omega)$ e P tal que $\mathrm{P}(\{(i,j)\}) = 1/4$, para $(i,j) \in \Omega$, e os acontecimentos $A = \{(0,0),(0,1)\}$, $B = \{(0,0),(1,0)\}$ e $C = \{(0,0),(1,1)\}$.

A noção de independência de acontecimentos aleatórios pode, de forma natural, ser extendida a uma família arbitrária de classes.

Definição 3.1.2 *Dizemos que as subclasse não-vazias $\mathcal{C}_t, t \in T$, de $\mathcal{A}$ são **independentes**, quando, para qualquer conjunto finito de índices distintos $t_1, \ldots, t_n \in T$, e de acontecimentos $A_{t_1} \in \mathcal{C}_{t_1}, \ldots, A_{t_n} \in \mathcal{C}_{t_n}$, os acontecimentos $A_{t_k}, k = 1, \ldots, n$, forem independentes.*

No resultado seguinte lançamos mão das noções de **π-sistema** e de **d-sistema**. Recordemos que um π-sistema é uma classe de partes de Ω que é estável para a intersecção finita, enquanto que um d-sistema, ou sistema de Dynkin, contém Ω e é estável para a complementação e para a reunião numerável disjunta (ver AMI, §1.2).

Teorema 3.1.3 *Sejam $\mathcal{C}_t, t \in T$, subclasses não-vazias de $\mathcal{A}$, tais que:*
a) $\mathcal{C}_t$ é um π-sistema, para todo o $t \in T$;
b) $\mathcal{C}_t$, $t \in T$, são independentes.
Então as σ-álgebras $\sigma(\mathcal{C}_t)$, $t \in T$, são independentes.

Dem: Para $t_1, \ldots, t_n \in T$, distintos, e $A_{t_1}, \ldots, A_{t_n}$ fixos em $\mathcal{C}_{t_1}, \ldots, \mathcal{C}_{t_n}$, respectivamente, consideremos a classe $\mathcal{L} = \{A \in \mathcal{A} : \mathrm{P}(AA_{t_2} \ldots A_{t_n}) = \mathrm{P}(A)\mathrm{P}(A_{t_1}) \ldots \mathrm{P}(A_{t_n})\}$. $\mathcal{L}$ é um d-sistema e, sendo $\mathcal{C}_{t_1}, \ldots, \mathcal{C}_{t_n}$ independentes, $\mathcal{C}_{t_1} \subset \mathcal{L}$. Consequentemente, $d(\mathcal{C}_{t_1}) \subset \mathcal{L}$. Sendo $\mathcal{C}_{t_1}$ um π-sistema, $d(\mathcal{C}_{t_1}) = \sigma(\mathcal{C}_{t_1})$ (cf. AMI, Teorema 1.3.3), o que prova que $\sigma(\mathcal{C}_{t_1}), \mathcal{C}_{t_2}, \ldots, \mathcal{C}_{t_n}$ são independentes. Repetindo o raciocínio para as classes $\mathcal{C}_{t_2}, \ldots, \mathcal{C}_{t_n}, \sigma(\mathcal{C}_{t_1})$ concluímos que $\sigma(\mathcal{C}_{t_2}), \mathcal{C}_{t_3}, \ldots, \mathcal{C}_{t_n}, \sigma(\mathcal{C}_{t_1})$ são independentes, e finalmente que $\sigma(\mathcal{C}_{t_1}), \sigma(\mathcal{C}_{t_2}), \ldots, \sigma(\mathcal{C}_{t_n})$ são independentes. $\square$

Corolário 3.1.4 *Os acontecimentos $A_t, t \in T$, são independentes sse as σ-álgebras $\sigma(A_t), t \in T$, o forem.*

Teorema 3.1.5 *Sejam $\mathcal{B}_t$, $t \in T$, sub-σ-álgebras de $\mathcal{A}$ independentes e $\mathcal{P}$ uma partição de T. Então as σ-álgebras $\mathcal{B}_S = \sigma(\mathcal{B}_t, t \in S)$, $S \in \mathcal{P}$, são ainda independentes.*

Dem: Para $S \in \mathcal{P}$, seja $\mathcal{C}_S = \{\bigcap_{\alpha \in K} : B_\alpha \in \mathcal{B}_\alpha, K \subset S, K \text{ finito}\}$. Vamos provar que $\mathcal{C}_S, S \in \mathcal{P}$, é uma família de π-sistemas independentes com $\sigma(\mathcal{C}_S) = \mathcal{B}_S$, o que permite concluir pelo teorema anterior. 1) $\mathcal{C}_S$ é claramente um π-sistema. 2) Sejam agora $S_1, \ldots, S_k \in \mathcal{P}$ distintos (logo disjuntos) e $A_i \in \mathcal{C}_{S_i}$, $i = 1, \ldots, k$. Então $A_i = \bigcap_{\alpha_i \in K_i} B^i_{\alpha_i}$, com $B^i_{\alpha_i} \in \mathcal{B}_{\alpha_i}$ e $K_i \subset S_i$ finito. Uma vez que $\mathrm{P}(\bigcap_{i=1}^k A_i) = \mathrm{P}(\bigcap_{i=1}^k \bigcap_{\alpha_i \in K_i} B^i_{\alpha_i}) = \prod_{i=1}^k \prod_{\alpha_i \in K_i} \mathrm{P}(B^i_{\alpha_i}) = \prod_{i=1}^k \mathrm{P}(A_i)$, concluímos que $\mathcal{C}_S, S \in \mathcal{P}$, é uma família de π-sistemas independentes. 3) Claramente $\mathcal{C}_S \subset \mathcal{B}_S$, e também $\sigma(\mathcal{C}_S) \subset \mathcal{B}_S$. Por outro lado, $\mathcal{B}_\alpha \subset \mathcal{C}_S$, para $\alpha \in S$, e também $\bigcup_{\alpha \in S} \mathcal{B}_\alpha \subset \mathcal{C}_S$. Assim, $\mathcal{B}_S = \sigma(\mathcal{B}_\alpha, \alpha \in S) = \sigma(\bigcup_{\alpha \in S} \mathcal{B}_\alpha) \subset \sigma(\mathcal{C}_S)$. $\square$

Exercícios

1. Utilizando a definição, mostre que se A e B são acontecimentos aleatórios independentes, também o são os pares de acontecimentos A e B^c, A^c e B, e A^c e B^c.

2. Mostre que $A_1, \ldots, A_n$ são acontecimentos independentes sse para todo o $j \in \{1, \ldots, n\}$ e $I \subset \{1, \ldots, n\} - \{j\}$ com $\mathrm{P}(\bigcap_{i \in I} A_i) > 0$, então $\mathrm{P}(A_j | \bigcap_{i \in I} A_i) = \mathrm{P}(A_j)$.

3. Se $A_n, n \geq 1$, são acontecimentos independentes, mostre que $\mathrm{P}(\bigcap_{n=1}^{\infty} A_n) = \prod_{n=1}^{\infty} \mathrm{P}(A_n)$.

4. Sejam $(\Omega, \mathcal{A})$ o produto dos espaços mensuráveis $(\Omega_n, \mathcal{A}_n)$, $n \in \mathbb{N}$, e P uma probabilidade sobre $\mathcal{A}$. Para $n \in \mathbb{N}$ e $A_n \in \mathcal{A}_n$, considere os acontecimentos
$$B_n = \Omega_1 \times \ldots \times \Omega_{n-1} \times A_n \times \Omega_{n+1} \times \ldots$$
e as probabilidades P_n definidas em $(\Omega_n, \mathcal{A}_n)$ por $\mathrm{P}_n(A_n) = \mathrm{P}(B_n)$. Mostre que os acontecimentos $B_n, n \geq 1$, são independentes sse $\mathrm{P} = \otimes_{n=1}^{\infty} \mathrm{P}_n$.

5. Se $A_n, n \geq 1$, são acontecimentos independentes, mostre que $\cup_{i=1}^{n} A_i$ e $\cup_{i=n+1}^{\infty} A_i$ são independentes, com $n \in \mathbb{N}$ fixo.

6. Para $s > 1$, fixo, sejam $\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}$, e X uma variável aleatória com valores em $\mathbb{N}$ tal que $\mathrm{P}(X = n) = \frac{1}{\zeta(s)} \frac{1}{n^s}$, para $n \in \mathbb{N}$.
   (a) Para $p \in \mathbb{N}$, considere o conjunto $E_p = \{X$ é divisível por $p\}$, e mostre que $\mathrm{P}(E_p) = 1/p^s$.
   (b) Mostre que os conjuntos E_p, com p primo, são independentes.
   (c) Estabeleça a fórmula de Euler: $\frac{1}{\zeta(s)} = \prod_{p \text{ primo}} \left(1 - \frac{1}{p^s}\right)$.

3.2 Independência de variáveis aleatórias

As variáveis aleatórias que consideramos neste parágrafo estão definidos sobre um mesmo espaço de probabilidade $(\Omega, \mathcal{A}, \mathrm{P})$, podendo, no entanto, tomar valores em espaços mensuráveis diversos. No que se segue, T é um qualquer conjunto de índices.

Definição 3.2.1 *Dizemos que X_t, $t \in T$, é uma família de **variáveis aleatórias independentes** se $\sigma(X_t)$, $t \in T$, forem σ-álgebras independentes.*

Uma caracterização da independência duma qualquer família X_t, $t \in T$, de variáveis aleatórias em termos da distribuição da variável aleatória $(X_t, t \in T)$, é apresentada no resultado seguinte. Fica assim clara a relação estreita entre as noções de independência da família X_t, $t \in T$, de variáveis aleatórias e a forma produto para a distribuição de probabilidade da variável aleatória $(X_t, t \in T)$.

Teorema 3.2.2 *As variáveis aleatórias X_t, $t \in T$, onde cada X_t toma valores em $(E_t, \mathcal{B}_t)$, são independentes sse $\mathrm{P}_{(X_t, t \in T)} = \bigotimes_{t \in T} \mathrm{P}_{X_t}$.*

Dem: Comecemos por notar que como a σ-álgebra $\bigotimes_{t \in T} \mathcal{B}_t$ é gerada pelos conjuntos do tipo $\pi_S^{-1}(\prod_{t \in S} B_t)$, com $B_t \in \mathcal{B}_t$, $t \in T$, e $S \subset T$ finito, a igualdade de medidas expressa no enunciado é equivalente à igualdade $\mathrm{P}_{(X_t, t \in S)} = \bigotimes_{t \in S} \mathrm{P}_{X_t}$, para todo o subconjunto finito S de T. Suponhamos então que X_t, $t \in T$, são variáveis aleatórias independentes, e para $S \subset T$ finito, consideremos $B_t \in \mathcal{B}_t$, para $t \in S$. Como $\mathrm{P}_{(X_t, t\in S)}(\prod_{t\in T} B_t) = \mathrm{P}(\bigcap_{t\in S}\{X_t \in B_t\}) = \prod_{t\in S}\mathrm{P}(\{X_t \in B_t\}) = \bigotimes_{t\in S}\mathrm{P}_{X_t}(\prod_{t\in T} B_t)$, concluímos que $\mathrm{P}_{(X_t,t\in T)} = \bigotimes_{t\in T}\mathrm{P}_{X_t}$. Reciprocamente, sejam $S \subset T$ finito, e $A_t \in \sigma(X_t)$, para $t \in S$. Por definição de σ-álgebra gerada por X_t, $A_t = X_t^{-1}(B_t)$, com $B_t \in \mathcal{B}_t$. Assim, $\mathrm{P}(\bigcap_{t\in S} A_t) = \mathrm{P}((X_t, t \in S) \in \prod_{t\in S} B_t) = \mathrm{P}_{(X_t,t\in S)}(\prod_{t\in S} B_t) = \bigotimes_{t\in S}\mathrm{P}_{X_t}(\prod_{t\in S} B_t) = \prod_{t\in S}\mathrm{P}_{X_t}(B_t) = \prod_{t\in S}\mathrm{P}(A_t)$, ou seja, X_t, $t \in T$, são independentes. $\square$

Nos dois resultados seguintes apresentamos caracterizações da independência das margens dum vector aleatório em termos da sua função de distribuição e, no caso deste ser absolutamente contínuo, da sua densidade de probabilidade. Um resultado do mesmo tipo, mas em termos da sua função de probabilidade, vale para vectores aleatórios discretos.

Teorema 3.2.3 *Seja $(X_1,\ldots,X_n)$ um vector aleatório em $\mathbb{R}^n$ com função de distribuição $F_{(X_1,\ldots,X_n)}$. As variáveis aleatórias reais $X_1,\ldots,X_n$ são independentes sse*

$$F_{(X_1,\ldots,X_n)} = \prod_{i=1}^n F_{X_i},$$

onde F_{X_i} denota a função de distribuição da variável aleatória X_i. Além disso, se $F_{(X_1,\ldots,X_n)} = \prod_{i=1}^n G_i$, onde cada G_i é uma distribuição de probabilidade em $\mathbb{R}$, então $G_i = F_{X_i}$, para $i = 1,\ldots,n$, e as variáveis aleatórias $X_1,\ldots,X_n$ são independentes.

Dem: 1) Se $X_1,\ldots,X_n$ são independentes, $P_{(X_1,\ldots,X_n)} = \bigotimes_{i=1}^n \mathrm{P}_{X_i}$, o que implica que $F_{(X_1,\ldots,X_n)}(x_1,\ldots,x_n) = P_{(X_1,\ldots,X_n)}(\prod_{i=1}^n]-\infty,x_i]) = \bigotimes_{i=1}^n \mathrm{P}_{X_i}(\prod_{i=1}^n]-\infty,x_i]) = \prod_{i=1}^n \mathrm{P}_{X_i}(]-\infty,x_i]) = \prod_{i=1}^n F_{X_i}(x_i)$, para $(x_1,\ldots,x_n) \in \mathbb{R}^n$. Reciprocamente, se $F_{(X_1,\ldots,X_n)} = \prod_{i=1}^n F_{X_i}$, então $P_{(X_1,\ldots,X_n)}$ e $\bigotimes_{i=1}^n \mathrm{P}_{X_i}$ coincidem sobre o π-sistema dos borelianos da forma $\prod_{i=1}^n]-\infty,x_i]$, que gera $\mathcal{B}(\mathbb{R}^n)$. Pelo lema da igualdade de medida, $P_{(X_1,\ldots,X_n)}$ e $\bigotimes_{i=1}^n \mathrm{P}_{X_i}$ coincidem sobre $\mathcal{B}(\mathbb{R}^n)$ (cf. AMI, §2.6), o que atendendo ao teorema anterior é equivalente à independência das variáveis $X_1,\ldots,X_n$. 2) Suponhamos agora que $F_{(X_1,\ldots,X_n)} = \prod_{i=1}^n G_i$, onde cada G_i é uma distribuição de probabilidade em $\mathbb{R}$. Assim, para $i = 1,\ldots,n$, e $x_i \in \mathbb{R}$, $F_{X_i}(x_i) = \lim_{\substack{x_j\to+\infty\\ j\neq i}} F_{(X_1,\ldots,X_n)}(x_1,\ldots,x_n) = \lim_{\substack{x_j\to+\infty\\ j\neq i}} \prod_{k=1}^n G_k(x_k) = G_i(x_i)$. Além disso, $F_{(X_1,\ldots,X_n)} = \prod_{i=1}^n F_i$, o que pela primeira parte da demonstração é equivalente à independência de $X_1,\ldots,X_n$. $\square$

Teorema 3.2.4 *Seja $(X_1,\ldots,X_n)$ um vector aleatório em $\mathbb{R}^n$ com densidade de probabilidade $f_{(X_1,\ldots,X_n)}$. As variáveis aleatórias reais $X_1,\ldots,X_n$ são independentes sse*

$$f_{(X_1,\ldots,X_n)} = \prod_{i=1}^n f_{X_i},$$

onde f_{X_i} denota a densidade de probabilidade da variável aleatória X_i. Além disso, se $f_{(X_1,\ldots,X_n)} = \prod_{i=1}^n g_i$, onde cada g_i é uma densidade de probabilidade em $\mathbb{R}$, então $g_i = f_{X_i}$, para $i = 1,\ldots,n$, e as variáveis aleatórias $X_1,\ldots,X_n$ são independentes.

DEM: 1) Se $X_1,\ldots,X_n$ são independentes, o teorema de Fubini (ver AMI, §6.4) permite concluir que $P_{(X_1,\ldots,X_n)} = (\prod_{i=1}^n f_{X_i})\,\lambda_n$, ou ainda, $f_{(X_1,\ldots,X_n)} = \prod_{i=1}^n f_{X_i}$. Reciprocamente, e ainda pelo teorema de Fubini, se $f_{(X_1,\ldots,X_n)} = \prod_{i=1}^n f_{X_i}$, concluímos que $P_{(X_1,\ldots,X_n)} = \bigotimes_{i=1}^n \mathrm{P}_{X_i}$, isto é, $X_1,\ldots,X_n$ são independentes. 2) Se $f_{(X_1,\ldots,X_n)} = \prod_{i=1}^n g_i$, onde cada g_i é uma densidade de probabilidade em $\mathbb{R}$, então, para $i = 1,\ldots,n$ e $x_i \in \mathbb{R}$, $f_{X_i}(x_i) = \int_{\mathbb{R}^{n-1}} \prod_{j=1}^n g_j(x_j) d\lambda_{n-1} = g(x_i) \prod_{\substack{j=1 \\ j\neq i}}^n \int g(x_j) d\lambda = g(x_i)$. Assim, $f_{(X_1,\ldots,X_n)} = \prod_{i=1}^n f_{X_i}$, o que pela primeira parte da demonstração é equivalente à independência das variáveis $X_1,\ldots,X_n$. □

Terminamos este parágrafo com uma caracterização da independência de dois vectores aleatórios em termos de distribuições condicionais.

Teorema 3.2.5 *Sejam X e Y são vectores aleatórios com valores em $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ e $(\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m))$, respectivamente. X e Y são independentes sse $\mathrm{P}_Y(\cdot|X = x)$ é independente de x, para P_X-quase todo o ponto x. Neste caso $\mathrm{P}_Y(\cdot|X = x) = P_Y$, para P_X-quase todo o ponto x.*

DEM: Basta ter em conta que, para $A \in \mathcal{B}(\mathbb{R}^n)$ e $B \in \mathcal{B}(\mathbb{R}^m)$, $\mathrm{P}_{(X,Y)}(A \times B) = \int_A \mathrm{P}_Y(B|X = x)\, d\mathrm{P}_X(x)$ e $\mathrm{P}_X(A)\mathrm{P}_Y(B) = \int_A P_Y(B)\, d\mathrm{P}_X(x)$. □

Exercícios

1. Dada uma família de acontecimentos aleatórios $A_t, t \in T$, mostre que $\mathbb{I}_{A_t}, t \in T$, são independentes sse os acontecimentos $A_t, t \in T$, o forem.

2. Se $X_t : (\Omega, \mathcal{A}, \mathrm{P}) \to (E_t, \mathcal{B}_t)$, com $t \in T$, são variáveis aleatórias independentes, e $f_t : (E_t, \mathcal{B}_t) \to (F_t, \mathcal{C}_t)$, são aplicações mensuráveis, mostre que $f_t \circ X_t$, $t \in T$, são também variáveis aleatórias independentes.

3. Sejam $X_1,\ldots,X_n$ são v.a. reais independentes, e $m < n$ natural. Mostre que:

   (a) Os vectores aleatórios $(X_1,\ldots,X_m)$ e $(X_{m+1},\ldots,X_n)$, são independentes;

   (b) $\sum_{i=1}^m X_i$ e $\sum_{i=m+1}^n X_i$ são v.a. independentes.

4. (**Construção de v.a. independentes**) Mostre que as variáveis aleatórias (X_n) definidas no Exercício 2.1.2 são independentes.

5. Dadas variáveis aleatórias $X_i : (\Omega_i, \mathcal{A}_i, \mathrm{P}_i) \to (E_i, \mathcal{B}_i)$, para $i = 1,\ldots,n$, mostre que existe um espaço de probabilidade $(\Omega, \mathcal{A}, \mathrm{P})$ e variáveis aleatórias independentes $Y_i : (\Omega, \mathcal{A}, \mathrm{P}) \to (E_i, \mathcal{B}_i)$, $i = 1,\ldots,n$, tais que $Y_i \sim X_i$ para todo o i.

6. Dada uma sucessão (X_n) de v.a.r. identicamente distribuídas, mostre que existem v.a.r. $Y_1, Y_2, \ldots, Z_1, Z_2, \ldots$, definidas num mesmo espaço de probabilidade que satisfazem: a) $X_n \sim Y_n \sim Z_n$, para todo o $n \in \mathbb{N}$; b) $Y_1, Y_2, \ldots, Z_1, Z_2, \ldots$ são independentes.

7. Mostre que o resultado expresso no exercício anterior continua válido para uma qualquer sucessão (X_n) de v.a.r. não necessariamente identicamente distribuídas.

8. (**Método de Box-Muller para simulação de variáveis normais, II**) Sejam R e Θ as variáveis aleatórias definidas no Exercício 2.5.6. Mostre que $X = R\cos\Theta$ e $Y = R\sin\Theta$, são variáveis independentes com distribuições normal standard.

3.3 Soma de variáveis aleatórias independentes

Se X e Y são variáveis aleatórias com valores em $\mathbb{R}^d$, absolutamente contínuas e independentes, isto é, se (X,Y) é um vector com densidade $(x,y)\to f_X(x)f_Y(y)$, vimos no §2.5 que a soma $X+Y$ é uma variável absolutamente contínua cuja densidade é a **convolução das densidades** f_X e f_Y, isto é,

$$f_{X+Y}(x) = (f_X \star f_Y)(x) = \int f_X(x-y)f_Y(y)d\lambda(y).$$

No caso discreto é também possível obter uma fórmula do tipo anterior. Com efeito, se X e Y são variáveis discretas e independentes com funções de probabilidade g_X e g_Y, temos, para $x \in \mathbb{R}^d$,

$$\begin{aligned} g_{X+Y}(x) &= \sum_{y\in\mathbb{R}^d} \mathrm{P}(X+Y=x, Y=y) \\ &= \sum_{y\in\mathbb{R}^d} \mathrm{P}(X=x-y, Y=y) \\ &= \sum_{y\in\mathbb{R}^d} g_X(x-y)g_Y(y) \\ &=: (g_X \star g_Y)(x), \end{aligned}$$

a que chamamos **convolução das funções de probabilidade** g_X e g_Y.

Se denotarmos agora por $S_n = X_1 + \ldots + X_n$, a soma de n variáveis aleatórias independentes e identicamente distribuídas, com densidade ou função de probabilidade comum f, a densidade ou função de distribuição f_{S_n} de S_n pode ser obtida por indução a partir de $f_{S_{n-1}}$ e de f, pois $S_n = S_{n-1} + X_n$, e S_{n-1} e X_n são independentes.

Nos casos seguintes é simples obter a distribuição de S_n pelo método anterior.

Exemplo 3.3.1 Se $X_1, \ldots, X_n$ são v.a. independentes com $X_i \sim N(0,1)$, então

$$f_{S_n}(x) = \frac{1}{\sqrt{2n\pi}}\, e^{-x^2/(2n)}, \text{ para } x \in \mathbb{R}.$$

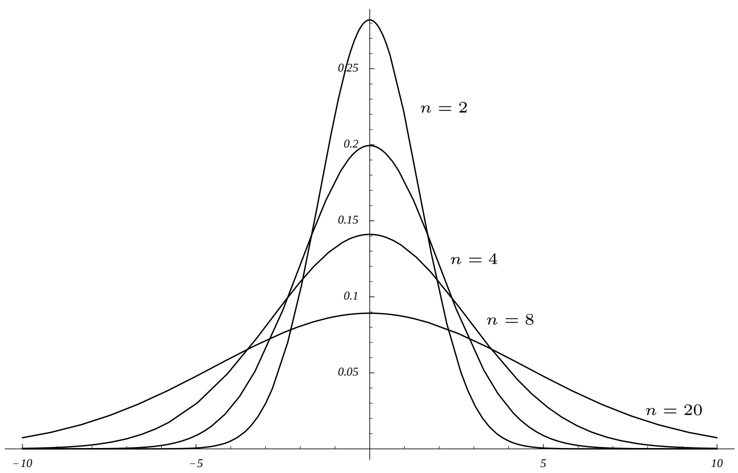


Figura 4.1: Distribuição da soma de n v.a. i.i.d. $N(0,1)$

Exemplo 3.3.2 Se $X_1, \ldots, X_n$ são v.a. independentes com $X_i \sim B(p)$, então

$$f_{S_n}(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x}, & \text{se } x \in \{0, 1, \ldots, n\} \\ 0, & \text{caso contrário.} \end{cases}$$

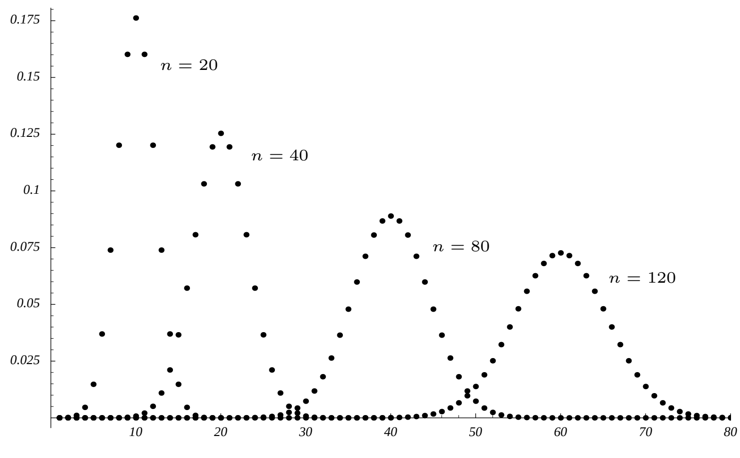


Figura 4.2: Distribuição da soma de n v.a. i.i.d. $B(1/3)$

Exemplo 3.3.3 Se $X_1, \ldots, X_n$ são v.a. independentes com $X_i \sim E(\lambda)$, temos

$$f_{S_n}(x) = \begin{cases} \lambda e^{-\lambda x} \frac{(\lambda x)^{n-1}}{(n-1)!}, & \text{se } x \geq 0 \\ 0, & \text{se } x < 0. \end{cases}$$

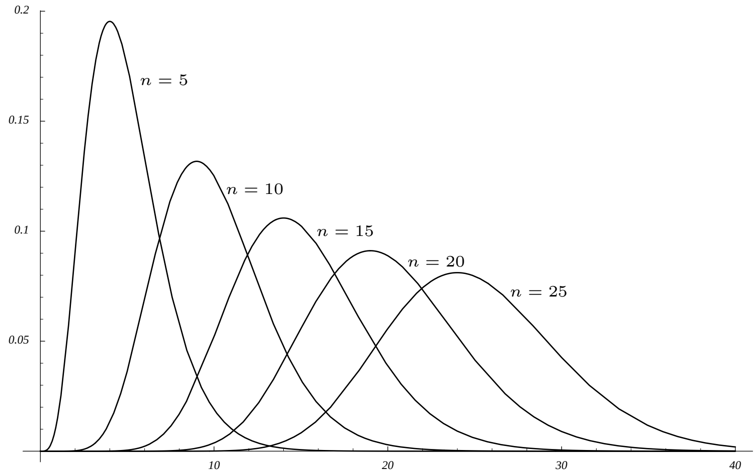


Figura 4.3: Distribuição da soma de n v.a. i.i.d. $E(1)$

No primeiro exemplo $S_n \sim N(0,n)$, enquanto que no segundo $S_n \sim B(n,p)$, o que seria de esperar atendendo à definição de distribuição binomial. No último exemplo, dizemos que S_n possui uma **distribuição de Erlang** de parâmetros n e λ.

Exercícios

1. Estabeleça os resultados enunciados nos exemplos anteriores.

2. Se $X_1, \ldots, X_n$ são v.a. independentes com distribuições geométricas de parâmetro p, mostre que $S_n = X_1 + \ldots + X_n$ verifica $S_n \sim Y + n$, onde $Y \sim BN(n,p)$ (ver Exercício 2.1.9).

3. Sejam X e Y independentes, e $Z = X + Y$. Determine a densidade de Z quando:

   (a) $X \sim E(\mu)$ e $Y \sim E(\lambda)$;

   (b) $X \sim N(m_1, \sigma_1^2)$ e $Y \sim N(m_2, \sigma_2^2)$.

4. Dizemos que uma v.a.r. X tem uma distribuição do **qui-quadrado com n graus de liberdade** ($n \in \mathbb{N}$), e escrevemos $X \sim \chi_n^2$, se admite uma densidade de probabilidade da forma
$$f(x) = \begin{cases} \frac{1}{\Gamma(n/2)2^{n/2}}\, x^{n/2-1}e^{-x/2}, & \text{se } x \geq 0 \\ 0, & \text{se } x < 0, \end{cases}$$
onde $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}dx$, para $\alpha > 0$, é a função Gamma. Mostre que se $X_1, X_2, \ldots, X_n$ são v.a. normais standard independentes, então $X_1^2 + X_2^2 + \ldots + X_n^2 \sim \chi_n^2$.

   (Sugestão: Tenha em conta o Exercício 2.3.6 e a igualdade $\int_0^1 x^{p-1}(1-x)^{q-1}dx = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}$, válida para $p, q > 0$.)

5. Sejam $X_1, \ldots, X_n$ v.a.r. independentes e $Y_1, \ldots, Y_n$ v.a.r. independentes, com $X_i \sim Y_i$ para $i = 1, \ldots, n$. Mostre que $\sum_{j=1}^n X_j \sim \sum_{j=1}^n Y_j$. Verifique que a hipótese de independência é essencial para a validade do resultado.

3.4 Leis zero-um de Borel e de Kolmogorov

Atendendo ao teorema de Borel-Cantelli já nosso conhecido da disciplina de Medida e Integração, sabemos que, sob certas condições sobre a sucessão de acontecimentos (A_n), o acontecimento $A_n\ i.o. = \limsup A_n$, isto é, o acontecimento que se realiza quando se realiza uma infinidade de acontecimentos A_n, tem probabilidade zero. Mais precisamente:

Teorema 3.4.1 (de Borel–Cantelli ([1])) *Se os acontecimentos aleatórios A_n, $n \geq 1$, satisfazem $\sum_{n=1}^{\infty} \mathrm{P}(A_n) < +\infty$, então $\mathrm{P}(A_n\ i.o.) = 0$.*

No caso dos acontecimentos (A_n) serem independentes este resultado pode ser precisado. Mostramos de seguida que a probabilidade do acontecimento $A_n\ i.o.$ só pode tomar dois valores possíveis: zero ou um.

Teorema 3.4.2 (Lei zero-um de Borel ([2])) *Se os acontecimentos aleatórios A_n, $n \geq 1$, são independentes então*

$$\mathrm{P}(A_n\ i.o.) = \begin{cases} 0 & sse \quad \sum_{n=1}^{\infty} \mathrm{P}(A_n) < +\infty \\ 1 & sse \quad \sum_{n=1}^{\infty} \mathrm{P}(A_n) = +\infty. \end{cases}$$

DEM: Pelo teorema de Borel-Cantelli, basta mostrar que $\sum_{n=1}^{\infty} \mathrm{P}(A_n) = +\infty$ implica $\mathrm{P}(A_n \text{ i.o.}) = 1$. Tal é equivalente a provar que $\mathrm{P}(\bigcup_{k=n}^{\infty} A_k) = 1$, para todo o $n \in \mathbb{N}$. Atendendo à independência dos acontecimentos A_k^c, $k \in \mathbb{N}$, e à desigualdade $1 - x \leq \exp(-x)$, válida para todo o $x \in [0,1]$, obtemos $\mathrm{P}(\bigcap_{k=n}^{\infty} A_k^c) = \lim_m \mathrm{P}(\bigcap_{k=n}^{m} A_k^c) = \lim_m \prod_{k=n}^{m} \mathrm{P}(A_k^c) = \lim_m \prod_{k=n}^{m} (1 - \mathrm{P}(A_k)) \leq \lim_m \exp(-\sum_{k=n}^{m} \mathrm{P}(A_k)) = 0$. $\square$

Como veremos de seguida, a propriedade exibida pelo acontecimento $A_n\ i.o.$ da sua probabilidade só poder tomar dois valores, zero ou um, é partilhada por uma classe mais vasta de acontecimentos aleatórios. Um tal resultado é conhecido como lei zero-um de Kolmogorov.

Definição 3.4.3 *Uma σ-álgebra $\mathcal{B} \subset \mathcal{A}$, diz-se* P***-trivial*** *se $\mathrm{P}(A) = 0$ ou $\mathrm{P}(A) = 1$, para todo o $A \in \mathcal{B}$.*

Claramente $\{\emptyset, \Omega\}$ é P-trivial para toda a probabilidade P.

Lema 3.4.4 *Uma sub-σ-álgebra $\mathcal{B}$ de $\mathcal{A}$ é P-trivial sse é independente de si própria.*

[1]Cantelli, F.P., *Rend. Accad. Naz. Lincei.* 26, 295–302, 1917.
[2]Borel, E, *Rend. Circ. Mat. Palermo* 27, 247–271, 1909.

DEM: Se $\mathcal{B}$ é independente de si própria, então para todo o $A \in \mathcal{B}$, $\mathrm{P}(A) = \mathrm{P}(A \cap A) = \mathrm{P}(A)\mathrm{P}(A)$, ou seja, $\mathrm{P}(A) = 0$ ou $\mathrm{P}(A) = 1$. Reciprocamente, se $\mathcal{B}$ é P-trivial e A e B são elementos de $\mathcal{B}$ com $\mathrm{P}(A) = 0$ ou $\mathrm{P}(B) = 0$, então $\mathrm{P}(A \cap B) \leq \min(\mathrm{P}(A), \mathrm{P}(B)) = 0$. Se $\mathrm{P}(A) = \mathrm{P}(B) = 1$, sabemos que $\mathrm{P}(A \cap B) = 1$, para toda a probabilidade P. Em ambos os casos, $\mathrm{P}(A \cap B) = \mathrm{P}(A)\mathrm{P}(B)$. $\square$

Teorema 3.4.5 (Lei zero-um de Kolmogorov ([3])) *Sejam $\mathcal{B}_1, \mathcal{B}_2, \ldots$ sub-σ-álgebras independentes de $\mathcal{A}$, e $\mathcal{B}_\infty$ a* **σ-álgebra assintótica** *associada à sucessão $(\mathcal{B}_n)$, isto é,*

$$\mathcal{B}_\infty = \bigcap_{n=1}^{\infty} \sigma(\mathcal{B}_k, k \geq n).$$

Então $\mathcal{B}_\infty$ é P*-trivial.*

DEM: Consideremos $n \geq 2$, e denotemos por $\overline{\mathcal{B}}_n$ a σ-álgebra $\sigma(\mathcal{B}_k, k \geq n)$. Pelo Teorema 3.1.5, as σ-álgebras $\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_{n-1}, \overline{\mathcal{B}}_n$ são independentes, e por maioria de razão, são ainda independentes as σ-álgebras $\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_{n-1}, \mathcal{B}_\infty$ pois $\mathcal{B}_\infty \subset \overline{\mathcal{B}}_n$. Sendo n qualquer, isto significa que $\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_\infty$ são independentes, sendo, pelo Teorema 3.1.5, também independentes as σ-álgebras $\sigma(\mathcal{B}_k, k \geq 1)$ e $\mathcal{B}_\infty$. Finalmente, como $\mathcal{B}_\infty \subset \sigma(\mathcal{B}_k, k \geq 1)$, concluímos que $\mathcal{B}_\infty$ é independente de si própria, ou seja, que $\mathcal{B}_\infty$ é P-trivial. $\square$

Teorema 3.4.6 *Seja $\mathcal{B}$ uma sub-σ-álgebra* P*-trivial de $\mathcal{A}$. Uma variável aleatória X $\mathcal{B}$-mensurável com valores em $(\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$ é degenerada, isto é, X é* P*-q.c. constante.*

DEM: Seja X $\mathcal{B}$-mensurável com valores em $\overline{\mathbb{R}}$. Como $X^{-1}(]-\infty, x]) = \{X \leq x\} \in \mathcal{B}$, então $\mathrm{P}(X \leq x) = 0$ ou 1, para todo o $x \in \overline{\mathbb{R}}$. Seja $c = \sup\{x \in \overline{\mathbb{R}} : \mathrm{P}(X \leq x) = 0\}$. Se $c = -\infty$ então $\mathrm{P}(X \leq x) = 1$, para todo o $x \in \mathbb{R}$, e assim $\mathrm{P}(X = -\infty) = \lim \mathrm{P}(X \leq -n) = 1$. Se $c = +\infty$, então $\mathrm{P}(X \leq x) = 0$, para todo o $x \in \mathbb{R}$, e assim $\mathrm{P}(X = +\infty) = 1 - \lim \mathrm{P}(X \leq n) = 1$. Se $c \in \mathbb{R}$, concluímos que $\mathrm{P}(X \leq x) = 0$, para todo o $x < c$ e $\mathrm{P}(X \leq x) = 1$, para todo o $x > c$. Consequentemente, $\mathrm{P}(X = c) = \mathrm{P}(X \leq c) - \mathrm{P}(X < c) = \lim \mathrm{P}(X \leq c + 1/n) - \lim \mathrm{P}(X \leq c - 1/n) = 1 - 0 = 1$. $\square$

Se $X_1, X_2, \ldots$ é uma sucessão de variáveis aleatórias reais independentes, e

$$S_n = X_1 + \ldots + X_n,$$

estudaremos mais à frente o comportamento assintótico das sucessões

$$S_n \quad \text{e} \quad S_n/n.$$

[3]Kolmogorov, A.N., *Grundbegriffe der Wahrscheinlichkeitrechnung*, Berlin, 1933.

De acordo com o resultado seguinte, estas sucessões ou convergem ou divergem quase certamente, isto é, o conjunto dos pontos $w \in \Omega$ onde convergem ou tem probabilidade zero ou ou tem probabilidade um. Além disso, sendo S_n/n convergente, a variável aleatória limite é quase certamente degenerada. Mais precisamente:

Corolário 3.4.7 *Nas condições anteriores, se (a_n) é uma sucessão de números reais com $a_n \to +\infty$, então:*
a) S_n e S_n/a_n convergem ou divergem quase certamente;
b) $\limsup S_n/a_n$ e $\liminf S_n/a_n$, são quase certamente constantes.

Exercícios

1. Se $A_n, n \geq 1$, são acontecimentos independentes e $A_n \to A$, mostre que $\mathrm{P}(A) = 0$ ou $\mathrm{P}(A) = 1$.

2. Sejam $X_n, n \geq 1$, variáveis de Bernoulli, com

$$\mathrm{P}(X_n = 1) = p_n = 1 - \mathrm{P}(X_n = 0), \quad \text{para} \quad n \in \mathbb{N}.$$

   (a) Mostre que $\{\lim X_n = 0\} = (\limsup A_n)^c$, onde $A_n = X_n^{-1}(1)$ para $n \in \mathbb{N}$.
   (b) Conclua que $\mathrm{P}(\lim X_n = 0) = 1$ se $\sum_{n=1}^{\infty} p_n < +\infty$.
   (c) Se $X_n, n \geq 1$ são independentes, mostre que $\mathrm{P}(\lim X_n = 0) = 1$ sse $\sum_{n=1}^{\infty} p_n < +\infty$.

3.5 Bibliografia

Jacod, J., Protter, P. (2000). *Probability Essentials*, Springer.

Kallenberg, O. (1997). *Foundations of Modern Probability*, Springer.

Resnick, S.I. (1999). *A Probability Path*, Birkhäuser.

Williams, D. (1991). *Probability with Martingales*, Cambridge University Press.

Capítulo 4

Integração de variáveis aleatórias

Esperança matemática duma variável aleatória real e suas principais propriedades. Momentos duma variável aleatória real. Parâmetros de dispersão e de forma. Covariância e correlação. Integração de vectores aleatórios.

4.1 Esperança matemática

Introduzimos neste parágrafo o primeiro dos parâmetros de resumo da distribuição de probabilidade duma variável aleatória real X de que falaremos neste capítulo. Para motivar a definição que dele apresentaremos, suponhamos, em primeiro lugar, que X é uma variável discreta que toma os valores $x_1, \ldots, x_n$ com probabilidades $p_1, \ldots, p_n$, onde $p_1 + \ldots + p_n = 1$. Pretendendo resumir a distribuição de probabilidade de X através dum parâmetro que descreva o centro duma tal distribuição, é natural recorrer à analogia deste problema com o da definição do centro de massa dum sistema discreto de pontos materiais com massas p_i em x_i. Somos assim levados a definir um tal parâmetro por $\sum_{i=1}^{n} x_i p_i$. No caso de X ser absolutamente contínua com densidade de probabilidade f, vale o mesmo tipo de analogia, sendo natural definir um tal parâmetro de resumo por $\int x f(x) dx$, isto é, como o centro de massa dum sistema contínuo de pontos materiais com densidade de massa $f(x)$ em x.

Lançando mão da noção de integral duma função real relativamente a uma medida (ver AMI, §§4.1–4.3), as duas fórmulas anteriores podem ser escritas de forma unificada como o integral da função identidade relativamente a P_X,

$$\int x \, d\mathrm{P}_X(x),$$

onde $(\Omega, \mathcal{A}, \mathrm{P})$ é o espaço de probabilidade onde admitimos que X está definida, ou ainda, pelo teorema da mudança de variável (ver AMI, §7.2), como o integral de X

relativamente à medida de probabilidade P,

$$\int X d\mathrm{P}.$$

No contexto das probabilidades o integral anterior é denominado e denotado duma forma especial.

Definição 4.1.1 *Chamamos* ***esperança matemática*** *(também dita* **valor médio***,* **valor esperado** *ou* **média***) da variável aleatória real X, que denotamos por $\mathrm{E}(X)$, ao integral*

$$\mathrm{E}(X) = \int X d\mathrm{P},$$

sempre que este integral exista.

Pelas razões já avançadas, dizemos que a esperança matemática, como parâmetro de resumo da distribuição de probabilidade duma variável aleatória, é um parâmetro de **localização**.

Recordemos, que se X é uma variável aleatória com valores em $([0,+\infty], \mathcal{B}([0,+\infty]))$, sabemos que o integral de X relativamente à medida de probabilidade P é um elemento de $[0,+\infty]$. Se X toma valores em $(\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$, X admite a decomposição $X = X^+ - X^-$, onde $X^+ = X \vee 0$ e $X^- = X \wedge 0$, são ditas **parte positiva** e **parte negativa** de X, respectivamente. Tal decomposição permite generalizar a noção de integral a X através da fórmula

$$\int X d\mathrm{P} = \int X^+ d\mathrm{P} - \int X^- d\mathrm{P},$$

sempre que $\int X^+ d\mathrm{P} < +\infty$ ou $\int X^- d\mathrm{P} < +\infty$. Se além disso $\int X d\mathrm{P} < \infty$, dizemos que X é P**-integrável**, ou simplesmente que X é integrável.

Claramente, a esperança matemática existe quando e só quando uma das variáveis X^+ ou X^- for integrável, e existe e é finita quando e só quando X for integrável.

Mostramos a seguir que a esperança matemática duma função mensurável de X depende unicamente dessa função e da distribuição de probabilidade de X. Em particular, a esperança matemática duma variável aleatória real depende apenas da sua distribuição de probabilidade.

Teorema 4.1.2 *Se X é uma variável aleatória com valores em $(E, \mathcal{B})$ e g é uma aplicação mensurável de $(E, \mathcal{B})$ em $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, então $\mathrm{E}(g(X))$ existe sse $\int g\, d\mathrm{P}_X$ existe e nesse caso*

$$\mathrm{E}(g(X)) = \int g(x)\, d\mathrm{P}_X(x).$$

Dem: Se g é não-negativa, pelo teorema da mudança de variável (ver AMI, §7.2) obtemos $\mathrm{E}(g(X)) = \int g(X)d\mathrm{P} = \int g \circ X d\mathrm{P} = \int g d(\mathrm{P}X^{-1}) = \int g d\mathrm{P}_X$. Sendo g qualquer, basta considerar a decomposição $g = g^+ - g^-$ e ter em conta que $(g \circ X)^+ = g^+(X)$ e $(g \circ X)^- = g^-(X)$. (Apresente uma demonstração alternativa usando a Proposição 2.1.4.) $\square$

No caso de X ser uma variável aleatória em $\mathbb{R}^d$ discreta ou absolutamente contínua (mais precisamente se P_X não tem parte singular), o resultado anterior permite obter fórmulas para o cálculo de $\mathrm{E}(g(X))$. Assim, se X é discreta com $\mathrm{P}_X = \sum_{i=1}^{\infty} p_i \delta_{x_i}$, onde $p_i = \mathrm{P}(X = x_i)$, então

$$\mathrm{E}(g(X)) = \sum_{i=1}^{\infty} p_i \int g(x) d\delta_{x_i}(x) = \sum_{i=1}^{\infty} p_i g(x_i).$$

Se X é absolutamente contínua com densidade f, então

$$\mathrm{E}(g(X)) = \int g(x) d\mathrm{P}_X(x) = \int g(x) f(x) d\lambda(x).$$

As propriedades que a seguir enunciamos são consequência imediata das propriedades do integral.

Teorema 4.1.3 *Sejam X e Y variáveis aleatórias reais definidas num mesmo espaço de probabilidade.*

a) X é integrável sse $|X|$ é integrável, e nesse caso $|\mathrm{E}(X)| \leq \mathrm{E}(|X|)$.

b) Se X e Y são integráveis, e $\alpha, \beta \in \mathbb{R}$, então $\alpha X + \beta Y$ é integrável e $\mathrm{E}(\alpha X + \beta Y) = \alpha \mathrm{E}(X) + \beta \mathrm{E}(Y)$.

c) Se $|X| \leq Y$, com Y integrável então X é integrável.

d) Se $|X| \leq M$, q.c., com $M > 0$, então X é integrável. Além disso, se $X = a$, q.c., com $a \in \mathbb{R}$, então $\mathrm{E}(X) = a$.

O resultado seguinte permite simplificar o cálculo da esperança matemática, no caso das variáveis aleatórias integráveis e simétricas.

Teorema 4.1.4 *Se X é integrável e simétrica relativamente a $a \in \mathbb{R}$, isto é, se $X - a \sim -(X - a)$, então $\mathrm{E}(X) = a$.*

Dem: Atendendo a que a esperança matemática duma variável aleatória real depende apenas da sua distribuição de probabilidade, concluímos que $\mathrm{E}(X - a) = \mathrm{E}(-(X - a))$, ou ainda, $\mathrm{E}(X) = a$. $\square$

Se X é discreta com função de probabilidade simétrica relativamente a a, ou absolutamente contínua com densidade de probabilidade simétrica relativamente a a, então X é claramente simétrica relativamente a a.

Notemos que a hipótese de integrabilidade é essencial para a validade do resultado anterior. Por exemplo, se X é uma variável aleatória de Cauchy com densidade $f(x) = (\pi(1+x^2))^{-1}$, para $x \in \mathbb{R}$, X é simétrica relativamente à origem e no entanto X não possui esperança matemática. Com efeito, $\int (x \wedge 0) d\mathrm{P}_X(x) = \int (x \vee 0) d\mathrm{P}_X(x) = \int_{[0,+\infty[} \frac{x}{\pi(1+x^2)} \, d\lambda(x) = \frac{2}{\pi} \lim \int_0^n \frac{2x}{1+x^2} \, d\lambda(x) = \frac{2}{\pi} \lim \int_0^n \frac{2x}{1+x^2} \, dx$ (integral de Riemann) $= \frac{2}{\pi} \lim \ln(1+n^2) = +\infty$.

Apresentamos agora alguns exemplos de cálculo da esperança matemática.

Exemplos 4.1.5 1. Se X é uma variável de Bernoulli de parâmetro p, então $\mathrm{E}(X) = 0 \times (1-p) + 1 \times p = p$.

2. Se X é uma variável de Poisson de parâmetro λ, temos $\mathrm{E}(X) = \sum_{n=0}^{\infty} n e^{-\lambda} \lambda^n/n! = e^{-\lambda}\lambda \sum_{n=0}^{\infty} \lambda^n/n! = \lambda$.

3. Se X é uma v.a. normal de parâmetros m e σ^2, então $\mathrm{E}(X) = m$. Para justificarmos esta afirmação, e tendo em conta que $X \sim \sigma U + m$, com $U \sim N(0,1)$, basta mostrar que $\mathrm{E}(U) = 0$, ou ainda, atendendo à simetria de U relativamente à origem, que U é integrável. Tal é verdade, pois tomando $M > 0$ tal que $x \leq e^x$, para $x \geq M$, obtemos $\mathrm{E}(|U|) = \int_{\mathbb{R}} |u| f_U(u) d\lambda(u) = \frac{2}{\sqrt{2\pi}} \int_{[0,+\infty[} u e^{-u^2/2} d\lambda(u) \leq M + \frac{2}{\sqrt{2\pi}} \int_{[M,+\infty[} e^{-u^2/2+u} d\lambda(u) = M + \frac{2e^{1/2}}{\sqrt{2\pi}} \int_{[M,+\infty[} e^{-(u-1)^2/2} d\lambda(u) \leq M + e^{1/2} < +\infty$.

Exercícios

1. Suponhamos que lançamos sucessivamente uma moeda equilibrada e seja X o número de lançamentos efectuados até ocorrer a primeira cara. Determine a distribuição de X, bem como o número médio de lançamentos necessários para obter a primeira cara.

2. Para cada uma das seguintes v.a. calcule a respectiva esperança matemática:

   (a) Binomial de parâmetro n e p.

   (b) Geométrica de parâmetro p.

   (c) Exponencial de parâmetro λ.

   (d) Uniforme sobre o intervalo $[a, b]$.

3. Deduza uma fórmula que lhe permita calcular a esperança matemática duma variável aleatória Y, a partir das densidades $f_Y(\cdot|X = \cdot)$ e f_X, e aplique-a ao cálculo da esperança matemática da v.a. Y definida no Exercício 2.6.3.

4. No casino de Monte Carlo a roda da roleta possui 37 divisões iguais, numeradas de 0 a 36, podendo um jogador apostar um euro num dos números com excepção do 0. Ele recebe 36 euros se a bola pára nesse número, obtendo assim ganho líquido de 35 euros, e perde o que apostou caso contrário. Qual é o seu ganho (líquido) médio? Um jogo que decorre em várias partidas idênticas diz-se **justo** (no sentido clássico), se o nosso ganho líquido médio for nulo, ou de forma equivalente, se o valor que pagamos para jogar cada uma das partidas (aposta), for igual ao nosso de ganho ilíquido médio. Caso contrário,

dizemos que o jogo nos é **favorável** ou **desfavorável**, consoante o nosso ganho líquido médio for positivo ou negativo, respectivamente. Para que valor da aposta é o jogo da roleta justo?

5. (**Paradoxo de São Petersburgo** ([1])) Pedro joga contra Paulo, e pagará a este uma quantia que depende do resultado duma série de lançamentos duma moeda equilibrada: se ocorre "coroa" nos $n-1$ primeiros lançamentos e "cara" no n-ésimo lançamento, Paulo recebe 2^n euros. Por sua vez, Paulo pagará inicialmente uma quantia Q a Pedro. Deverá o Paulo aceitar pagar 15 euros por partida para jogar? Verifique que independentemente do valor Q pago pelo Paulo, o seu ganho médio líquido por partida é superior a Q. Será possível determinar Q de modo que o jogo seja justo? Simule este jogo num computador e ensaie uma resposta à pergunta anterior com base unicamente nessa simulação.

4.2 Momentos

Da disciplina de Medida e Integração conhecemos os espaços vectoriais $\mathcal{L}^p(\Omega, \mathcal{A}, \mathrm{P})$, com $0 < p < +\infty$, das variáveis aleatórias reais X de potência p integrável, isto é, tais que $\mathrm{E}|X|^p < +\infty$ (cf. AMI, §5.2). Identificando variáveis aleatórias que coincidem a menos dum conjunto de probabilidade P nula, obtemos os espaços $L^p(\Omega, \mathcal{A}, \mathrm{P})$, que são espaços de Banach para a norma $\|X\|_p = \mathrm{E}^{1/p}|X|^p$ se $p \geq 1$, e são espaços métricos com distância $d(X, Y) = \|X - Y\|_p^p$, para $0 < p < 1$. Para $0 < p < q < +\infty$ sabemos também que $\mathcal{L}^q \subset \mathcal{L}^p$.

A par da esperança matemática que estudámos no parágrafo anterior e que definimos para toda a variável aleatória de $\mathcal{L}^1$, 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 $X \in \mathcal{L}^p$. Chamamos* ***momento de ordem** p **de** X* *a $\mathrm{E}(X^p)$, e* ***momento centrado de ordem** p **de** X* *a $\mu_p = \mathrm{E}(X - \mathrm{E}(X))^p$.*

Atendendo à desigualdade de Hölder (cf. AMI, §5.3), para $p \leq q$, é válida a desigualdade $\mu_p^{1/p} \leq \mu_q^{1/q}$.

Como parâmetros de resumo da distribuição de probabilidade duma variável aleatória, particular interesse têm para nós o momento de primeira ordem, já estudado no parágrafo anterior, e o momento centrado de segunda ordem. Este último, por razões que decorrem da sua definição é um parâmetro de **dispersão** (em torno da média) da distribuição de probabilidade duma variável aleatória.

[1] Este jogo conceptual foi pela primeira vez estudado por Nicolaus Bernoulli (1687–1759), que o discute com Montmort numa troca de correspondência entre 1713 e 1716. O jogo torna-se conhecido através dum artigo de Daniel Bernoulli (1700–1782), primo de Nicolaus, publicado na revista da Academia Imperial de Ciências de São Petersburgo em 1738.

Definição 4.2.2 *Se $X \in \mathcal{L}^2$, chamamos* ***variância de*** *X, que denotamos por* $\mathrm{Var}(X)$, *ao seu momento centrado de segunda ordem,* $\mathrm{Var}(X) = \mathrm{E}(X - \mathrm{E}(X))^2$. *A $\sigma(X) = \sqrt{\mathrm{Var}(X)}$, chamamos* ***desvio-padrão de*** *X.*

As demonstrações das propriedades da variância expressas nas proposições seguintes são deixadas ao cuidado do aluno.

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

Proposição 4.2.4 *Se $X \in \mathcal{L}^2$ e $a, b \in \mathbb{R}$, então:*
a) $\mathrm{Var}(X) = \mathrm{E}(X^2) - \mathrm{E}^2(X)$;
b) $\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$.

As fórmulas anteriores são de grande utilidade no cálculo da variância. Para as variáveis aleatórias consideradas nos Exemplos 4.1.5, efectuamos agora o cálculo da sua variância.

Exemplos 4.2.5 1. Se X é uma variável de Bernoulli de parâmetro p, então $\mathrm{E}(X^2) = 0 \times (1-p) + 1 \times p = p$, e portanto $\mathrm{Var}(X) = p - p^2 = p(1-p)$.

2. Se X é uma variável de Poisson de parâmetro λ, comecemos por efectuar o cálculo de $\mathrm{E}(X(X-1)) = \sum_{n=0}^{\infty} n(n-1)e^{-\lambda}\,\lambda^n/n! = \lambda^2 e^{-\lambda}\sum_{n=2}^{\infty}\,\lambda^{n-2}/(n-2)! = \lambda^2$. Assim, $\mathrm{Var}(X) = \lambda$.

3. Se $X \sim N(m, \sigma^2)$, sabemos que $X \sim \sigma U + m$, com $U \sim N(0,1)$, e portanto $\mathrm{Var}(X) = \mathrm{Var}(\sigma U + m) = \sigma^2\mathrm{Var}(U) = \sigma^2\mathrm{E}(U^2)$, pois $\mathrm{E}(U) = 0$. Finalmente, integrando por partes, obtemos $\mathrm{E}(U^2) = \int x^2 \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2} d\lambda(x) = \int \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2} d\lambda(x) = 1$, donde $\mathrm{Var}(X) = \sigma^2$ (ver Figura 1.1). Em particular concluímos que a variável normal de parâmetros 0 e 1 tem média zero e variância unitária. Toda a variável aleatória com esta propriedade diz-se **centrada e reduzida**.

Terminamos este parágrafo 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 4.2.6 *Se $X \in \mathcal{L}^3$ chamamos* ***coeficiente de assimetria*** *de X a $\beta_1 = \mu_3/\mu_2^{3/2}$. Se $X \in \mathcal{L}^4$ chamamos* ***coeficiente de achatamento*** *de X a $\beta_2 = \mu_4/\mu_2^2$.*

Notemos que se $X \in \mathcal{L}^3$ 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 que traduz "o peso nas caudas" da distribuição de X é habitualmente comparado com o da distribuição normal para a qual $\beta_2 = 3$.

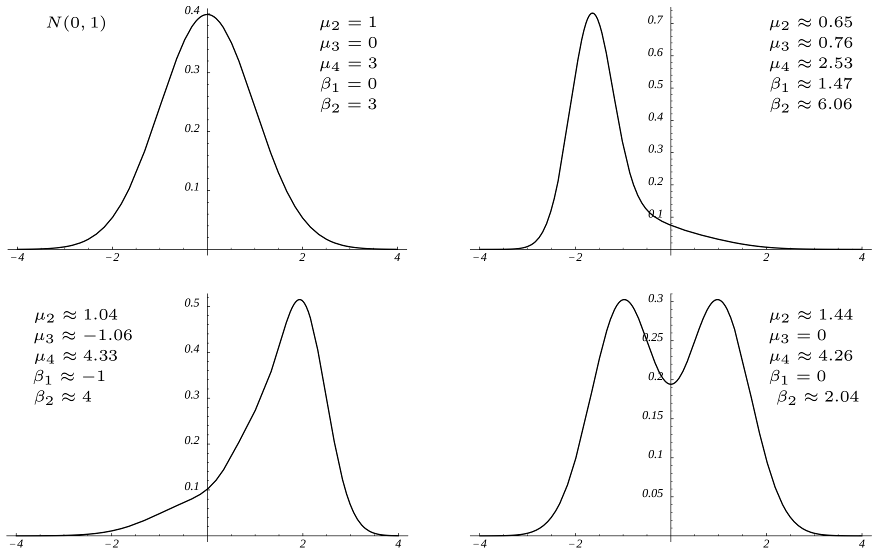


Figura 3.1

Exercícios

1. Se X é uma variável de quadrado integrável com média m e variância $\sigma^2 > 0$, mostre que $U = (X - m)/\sigma$ é uma v.a. centrada e reduzida.

2. Para cada uma das seguintes v.a. calcule a variância respectiva:

   (a) Geométrica de parâmetro p.
   (b) Uniforme sobre o intervalo $[a, b]$.
   (c) Exponencial de parâmetro λ.

3. Seja Y a v.a. definida no Exercício 2.6.2. Sem explicitar a distribuição de Y, calcule $\mathrm{E}(Y)$ e $\mathrm{Var}(Y)$.

4. Se X é uma v.a.r. de quadrado integrável, mostre que $\mathrm{E}(X)$ é a v.a. constante que melhor aproxima X no sentido de L^2, isto é,
$$\forall\, a \in \mathbb{R}, \quad \mathrm{E}(X - \mathrm{E}(X))^2 \leq \mathrm{E}(X - a)^2.$$

5. Se $X \sim N(m, \sigma^2)$, mostre que $X \in \mathcal{L}^p$ para todo o $p \geq 1$.

6. Seja X uma v.a.r. absolutamente contínua com densidade de probabilidade
$$f(x) = \begin{cases} \dfrac{1}{\sqrt{2\pi}\,\sigma\, x} \exp\!\left(-\dfrac{(\ln x - m)^2}{2\sigma^2}\right) & , \text{se } x > 0 \\ 0 & , \text{se } x \leq 0, \end{cases}$$
onde $m \in \mathbb{R}$ e $\sigma > 0$. Dizemos neste caso que X segue uma distribuição **log-normal de parâmetros m e σ**, e escrevemos $X \sim LN(m, \sigma)$.

(a) Para $c > 0$ e $\alpha > 0$, mostre que $cX^\alpha \sim LN(\ln c + \alpha m, \alpha\sigma)$.

(b) Prove que $\mathrm{E}(X) = \exp(m + \sigma^2/2)$.

(c) Utilizando as alíneas anteriores, calcule os momentos de ordem k, $k \in \mathbb{N}$, e a variância de X.

7. (a) **(Desigualdade de Bienaymé** ([2])**-Tchebychev** ([3])**)** Mostre que se X é uma variável aleatória real integrável, então para todo o $\alpha > 0$,

$$\mathrm{P}(|X - \mathrm{E}(X)| \geq \alpha) \leq \frac{\mathrm{Var}(X)}{\alpha^2}.$$

(Sugestão: Comece por verificar que $\mathbb{1}_{\{|X-\mathrm{E}(X)|\geq\alpha\}} \leq (X - \mathrm{E}(X))^2/\alpha^2$.)

(b) Mostre que a desigualdade anterior é óptima no sentido em que para qualquer $\alpha > 0$, existe uma variável aleatória X que verifica a igualdade.

(c) Conclua que para qualquer variável aleatória de quadrado integrável, a probabilidade do seu desvio relativamente à média ser superior ou igual a k vezes o seu desvio-padrão, não é superior a $1/k^2$ (se $k = 3$ obtemos $1/k^2 = 0.111\ldots$, e para $k = 5$ obtemos $1/k^2 = 0.04$).

4.3 Covariância e correlação

Se (X, Y) é um vector aleatório em $\mathbb{R}^2$, os parâmetros de resumo das distribuições de X e de Y que estudámos no parágrafo anterior, são também parâmetros de resumo da distribuição de (X, Y). Contrariamente a tais parâmetros que incidem unicamente sobre as distribuições marginais do vector, vamos neste parágrafo estudar um parâmetro de resumo da distribuição de (X, Y) que, como veremos, nos dá uma medida da dependência linear (afim) entre as variáveis X e Y.

Para tal vamos lançar mão das propriedades particulares do espaço de Banach $L^2(\Omega, \mathcal{A}, \mathrm{P})$. Este espaço vectorial, é um espaço com produto interno definido por $\langle X, Y\rangle = \mathrm{E}(XY)$. Como $\|X\|_2 = \sqrt{\langle X, X\rangle}$, dizemos que L^2 é um **espaço de Hilbert**. Sabemos também que em $\mathcal{L}^2$ é válida a propriedade seguinte conhecida como **desigualdade de Cauchy-Schwarz**:

Teorema 4.3.1 *Se $X, Y \in \mathcal{L}^2$ então $|\mathrm{E}(XY)| \leq \sqrt{\mathrm{E}(X^2)}\,\sqrt{\mathrm{E}(Y^2)}$. Além disso, tem-se a igualdade sse X e Y são linearmente dependentes.*

Sempre que X e Y não sejam quase certamente nulas, o quociente $\frac{\mathrm{E}(XY)}{\sqrt{\mathrm{E}(X^2)}\sqrt{\mathrm{E}(Y^2)}}$ surge assim como uma medida natural da dependência linear entre X e Y. Se pretendemos avaliar não só a dependência linear mas também a dependência afim, o quociente anterior deixa de ser indicado para o efeito.

[2]Bienaymé, I.-J., *C. R. Acad. Sci. Paris* 37, 309–324, 1853.

[3]Tchebychev, P.L., *J. Math. Pures et Appl.* (Sér. 2) 12, 177–184, 1867.

Definição 4.3.2 *Se $X, Y \in \mathcal{L}^2$, chamamos* ***covariância de*** (X,Y) *ao número real*

$$\mathrm{Cov}(X,Y) = \mathrm{E}((X - \mathrm{E}(X))(Y - \mathrm{E}(Y))).$$

Se além disso X e Y são de variância não-nula, chamamos ***coeficiente de correlação de*** (X,Y) *ao número do intervalo* $[-1,1]$ *dado por*

$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma(X)\sigma(Y)}.$$

Notemos que se $X, Y \in \mathcal{L}^2$, então $\mathrm{Cov}(X,Y) = \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y)$ e $\mathrm{Var}(X) = \mathrm{Cov}(X,X)$. Além disso, $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X,Y)$. O cálculo anterior da variância da soma simplifica-se se $X - \mathrm{E}(X)$ e $Y - \mathrm{E}(Y)$ são ortogonais (no sentido do produto interno de L^2), uma vez que neste caso $\mathrm{Cov}(X,Y) = 0$. Dizemos então que X e Y são **não-correlacionadas**. Neste caso $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$. A generalização das duas igualdades anteriores à soma dum número finito de variáveis $X_1, \ldots, X_n \in \mathcal{L}^2$, é simples, obtendo-se $\mathrm{Var}(\sum_{i=1}^n X_i) = \sum_{i=1}^n \mathrm{Var}(X_i) + 2\sum_{1\leq i<j\leq n} \mathrm{Cov}(X_i, X_j)$, e também, $\mathrm{Var}(\sum_{i=1}^n X_i) = \sum_{i=1}^n \mathrm{Var}(X_i)$, se as variáveis são duas a duas não-correlacionadas.

Do resultado seguinte concluímos que duas variáveis reais independentes são, em particular, não-correlacionadas. Reparemos ainda que a integrabilidade do produto de duas variáveis independentes é consequência da integrabilidade de cada um dos factores.

Teorema 4.3.3 *Se X e Y são variáveis aleatórias reais integráveis e independentes, então XY é integrável e* $\mathrm{E}(XY) = \mathrm{E}(X)\mathrm{E}(Y)$.

DEM: Sejam então X e Y variáveis aleatórias reais integráveis e comecemos por mostrar que XY é ainda integrável. Com efeito, pelo teorema de Fubini, $\mathrm{E}(|XY|) = \int |xy|\, d\mathrm{P}_{(X,Y)} = \int |xy|\, d\mathrm{P}_X \otimes \mathrm{P}_Y = \int |x||y|\, d\mathrm{P}_X d\mathrm{P}_Y = \int |x|\, d\mathrm{P}_X \int |y|\, d\mathrm{P}_Y < +\infty$. Utilizando os mesmos argumentos obtemos $\mathrm{E}(XY) = \mathrm{E}(X)\mathrm{E}(Y)$. □

Terminamos este parágrafo estabelecendo um resultado que reforça a interpretação do coeficiente de correlação entre duas variáveis aleatórias, como uma medida da dependência afim entre essas variáveis.

Teorema 4.3.4 *Se $X, Y \in \mathcal{L}^2$ são de variância não-nula, então:*

a) $\rho(aX + c, bY + d) = \rho(X,Y)$, *para* $a, b > 0$ *e* $c, d \in \mathbb{R}$;

b) $\rho(X, aX + b) = a/|a|$, *para* $a \neq 0$ *e* $b \in \mathbb{R}$;

c) $\rho(X,Y) = \pm 1$ *sse existem* $a, b, c \in \mathbb{R}$, *com* $ab \neq 0$, *tais que*

$$aX + bY + c = 0,\ \mathrm{P}\text{-}q.c.$$

DEM: As duas primeiras alíneas obtêm-se directamente da definição de ρ. Para estabelecer c), consideremos a variável aleatória $Z = Y/\sigma(Y) - X\rho(X,Y)/\sigma(X)$ que satisfaz $\sigma^2(Z) = 1 - \rho^2(X,Y)$. Basta agora usar a alínea b) e a Proposição 4.2.3. □

Exercícios

1. Mostre que a covariância é uma função bilinear, isto é, se $X_1,\dots,X_n,Y_1,\dots,Y_m$ são variáveis de quadrado integrável e $a_1,\dots,a_n,b_1,\dots,b_m$ números reais, então
$$\mathrm{Cov}\Big(\sum_{i=1}^n a_iX_i, \sum_{j=1}^m b_jY_j\Big) = \sum_{i=1}^n\sum_{j=1}^m a_ib_j\mathrm{Cov}(X_i,Y_j).$$
2. Mostre que se $X_1,\dots,X_n$ são variáveis aleatórias reais integráveis e independentes, então $\prod_{i=1}^n X_i$ é integrável e $\mathrm{E}\big(\prod_{i=1}^n X_i\big) = \prod_{i=1}^n \mathrm{E}(X_i)$.
3. Verifique que o coeficiente de correlação pode ser igual a 0 para variáveis não necessariamente independentes. Para tal considere X em $\mathcal{L}^3$ simétrica relativamente à origem e $Y = X^2$.

4.4 Integração de vectores aleatórios

As noções de integração de variáveis aleatórias que até agora estudámos, podem ser extendidas de forma natural ao caso dos vectores aleatórios. No que se segue, denotaremos por $\|\cdot\|$ a norma euclideana de $\mathbb{R}^d$.

Definição 4.4.1 *Um vector aleatório $X = (X_1,\dots,X_d)$ com valores em $(\mathbb{R}^d,\mathcal{B}(\mathbb{R}^d))$ diz-se **integrável** se $\mathrm{E}\|X\| < +\infty$. Nesse caso, chamamos **esperança matemática** de X ao vector de $\mathbb{R}^d$ dado por*
$$\mathrm{E}(X) = (\mathrm{E}(X_1),\dots,\mathrm{E}(X_d)).$$

Claramente, a noção de integrabilidade não depende da norma considerada ser a euclideana. Além disso, X é integrável sse $\|X\|$ é integrável, ou ainda, sse cada uma das variáveis aleatórias X_i, $i = 1,\dots,d$, é integrável.

Para $0 < p < +\infty$, podemos definir o espaço vectorial real dos vectores aleatórios X com valores em $\mathbb{R}^d$ de potência p integrável, isto é, tais que $\mathrm{E}\|X\|^p < +\infty$. Um tal conjunto é denotado por $\mathcal{L}^p(\Omega,\mathcal{A},\mathrm{P},\mathbb{R}^d)$, ou simplesmente por $\mathcal{L}^p$. Claramente, a aplicação $X \to \mathrm{E}(X)$, de $\mathcal{L}^1$ em $\mathbb{R}^d$, é uma aplicação linear.

A par da esperança matemática, a noção 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.

Definição 4.4.2 *Se $X \in \mathcal{L}^2$, chamamos* ***matriz de covariância de*** $X = (X_1, \ldots, X_d)$ *(dita também matriz de dispersão ou de variância-covariância) à matriz*

$$\mathrm{C}_X = [\mathrm{Cov}(X_i, X_j)]_{1 \leq i,j \leq d}.$$

A matriz de covariância é simétrica e semi-definida positiva, pois $\mathrm{Var}(\sum_{i=1}^d \lambda_i X_i) = \lambda^T C_X \lambda$, para todo o $\lambda \in \mathbb{R}^d$.

Da alínea c) do Teorema 4.3.4 sabemos que a matriz de covariância $\mathrm{C}_{(X,Y)}$ dum vector aleatório em $\mathbb{R}^2$ nos dá informação sobre o tipo de distribuição de (X, Y). Mais precisamente, sabemos que se $\mathrm{C}_{(X,Y)}$ possui característica 1 então a distribuição de (X, Y) está concentrada numa recta, não sendo, por isso, absolutamente contínua. Generalizamos a seguir este resultado ao caso dum vector aleatório em $\mathbb{R}^d$:

Teorema 4.4.3 *Sejam X um ve.a. em $\mathbb{R}^d$ de quadrado integrável e C_X a sua matriz de covariância. Se $\mathrm{car}(\mathrm{C}_X) = r$, então a distribuição de X está concentrada num subespaço afim de $\mathbb{R}^d$ de dimensão r.*

Exercícios

1. Seja $U = (X, Y)$ o ve.a. definido no Exemplo 2.1.9. Calcule $\mathrm{E}(U)$ e C_U.
2. Sejam A uma matriz real de tipo $n \times m$ e b um vector em $\mathbb{R}^n$. Se X é um ve.a. em $\mathbb{R}^m$ de quadrado integrável, mostre que a esperança matemática e a matriz de covariância de X e $AX + b$ se encontram relacionadas da seguinte forma:
$$\mathrm{E}(AX + b) = A\mathrm{E}(X) + b \quad \text{e} \quad \mathrm{C}_{AX+b} = A\mathrm{C}_X A^T.$$
3. Demonstre o Teorema 4.4.3. Conclua que no caso em que $\mathrm{car}(\mathrm{C}_X) = d$, X pode ser ou não absolutamente contínuo.

4.5 Bibliografia

Hennequin, P.L., Tortrat, A. (1965). *Théorie des Probabilités et Quelques Applications*, Masson.

Jacod, J., Protter, P. (2000). *Probability Essentials*, Springer.

Monfort, A. (1980). *Cours de Probabilités*, Economica.

Parte II

Leis dos grandes números

Capítulo 5

Convergências funcionais de variáveis aleatórias

Convergência quase certa, em probabilidade e em média de ordem p duma sucessão de variáveis aleatórias. Relações entre os diversos modos de convergência. Principais propriedades e caracterizações. Teorema da convergência dominada em $\mathcal{L}^p$. Convergências funcionais de vectores aleatórios.

5.1 Convergência quase certa

Neste capítulo $X, X_1, X_2, \ldots$ representam variáveis aleatórias reais definidas sobre um mesmo espaço de probabilidade $(\Omega, \mathcal{A}, \mathrm{P})$.

Definição 5.1.1 *Dizemos que (X_n)* ***converge para X quase certamente****, e escrevemos $X_n \xrightarrow{qc} X$, se*

$$\mathrm{P}(\{\omega \in \Omega : \lim X_n(\omega) = X(\omega)\}) = 1.$$

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 $N \in \mathcal{A}$, com $\mathrm{P}(N) = 0$, tal que $\lim X_n(\omega) = X(\omega)$, para todo o $\omega \in N^c$.

Das propriedades dos conjuntos de probabilidade nula, verificamos assim que as propriedades da convergência quase certa duma sucessão de variáveis aleatórias são essencialmente iguais às da convergência pontual. Uma das excepções 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.

Proposição 5.1.2 *Se* $X_n \xrightarrow{qc} X$ *e* $X_n \xrightarrow{qc} Y$, *então* $X = Y$ *q.c..*

No resultado seguinte apresentamos uma caracterização da convergência quase certa bastante útil quando pretendemos estabelecer a existência do limite quase certo.

Teorema 5.1.3 *Seja* (X_n) *uma sucessão de variáveis aleatórias reais. As condições seguintes são equivalentes:*

(i) $X_n \xrightarrow{qc} X$, para alguma variável aleatória real X;

(ii) (X_n) é de Cauchy quase certamente, isto é,

$$\sup_{n,m\geq k} |X_n - X_m| \xrightarrow{qc} 0,\ k \to +\infty.$$

Dem: A implicação (i) $\Rightarrow$ (ii) é óbvia. Estabeleçamos a implicação recíproca. Sendo (X_n) de Cauchy quase certamente, concluímos que existe $N \in \mathcal{A}$ com $\mathrm{P}(N) = 0$ tal que para todo o $w \in N^c$ a sucessão $(X_n(\omega))$ é de Cauchy em $\mathbb{R}$. Definindo $X(\omega) = \lim X_n(\omega)$, para $\omega \in N^c$ e $X(\omega) = 0$, para $\omega \in N$, temos claramente $X_n \xrightarrow{qc} X$. $\square$

Exercícios

1. Sendo f uma função contínua real de variável real, prove que se $X_n \xrightarrow{qc} X$, então $f(X_n) \xrightarrow{qc} f(X)$.

2. Mostre que as seguintes condições são equivalentes:

   (i) $X_n \xrightarrow{qc} X$;

   (ii) $\forall\, \epsilon > 0 \quad \mathrm{P}\Big(\bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} \{|X_n - X| \geq \epsilon\}\Big) = 0$;

   (iii) $\forall\, \epsilon > 0 \quad \mathrm{P}\Big(\bigcup_{n=k}^{\infty} \{|X_n - X| \geq \epsilon\}\Big) \to 0,\ k \to +\infty$.

3. Diz-se que uma sucessão (X_n) de v.a.r. converge **quase completamente** para uma v.a.r. X quando $\sum_{n=1}^{\infty} \mathrm{P}(\{|X_n - X| \geq \epsilon\}) < +\infty$, para todo o $\epsilon > 0$.

   (a) Prove que a convergência quase completa implica a convergência quase certa.

   (b) Mostre que se as variáveis (X_n) são independentes, as convergências quase certa e quase completa são equivalentes.
   (Sugestão: Use a lei zero-um de Borel.)

5.2 Convergência em probabilidade

Definição 5.2.1 *Dizemos que* (X_n) ***converge para*** X ***em probabilidade****, e escrevemos* $X_n \xrightarrow{p} X$, *se*

$$\forall\, \epsilon > 0 \quad \mathrm{P}(\{\omega \in \Omega : |X_n(\omega) - X(\omega)| \geq \epsilon\}) \to 0.$$

Tal como para a convergência quase certa, se X e Y são limite em probabilidade duma sucessão de variáveis aleatórias então X e Y coincidem a menos dum conjunto com probabilidade nula.

Comecemos por relacionar este modo de convergência com a convergência quase certa introduzida no parágrafo anterior.

Teorema 5.2.2 *Se $X_n \xrightarrow{qc} X$, então $X_n \xrightarrow{p} X$.*

Dem: Tendo em conta a inclusão $\{\omega : \lim X_n(\omega) = X(\omega)\} \subset \bigcup_{n\in\mathbb{N}} \bigcap_{k\geq n}\{\omega : |X_k(\omega) - X(\omega)| < \epsilon\}$, válida para todo o $\epsilon > 0$, obtemos, por hipótese, $\mathrm{P}(\bigcup_{n\in\mathbb{N}} \bigcap_{k\geq n}\{\omega : |X_k(\omega) - X(\omega)| < \epsilon\}) = 1$, ou ainda, $\lim \mathrm{P}(\bigcap_{k\geq n}\{\omega : |X_k(\omega) - X(\omega)| < \epsilon\}) = 1$. Assim $\lim \mathrm{P}(\{\omega : |X_n(\omega) - X(\omega)| < \epsilon\}) = 1$, o que permite concluir. $\square$

Apresentamos a seguir duas caracterizações importantes da convergência em probabilidade. A segunda delas permite utilizar no estudo da convergência em probabilidade resultados da convergência quase certa.

Teorema 5.2.3 *Seja (X_n) uma sucessão de variáveis aleatórias reais. As condições seguintes são equivalentes:*

(i) $X_n \xrightarrow{p} X$, para alguma variável aleatória real X;

(ii) (X_n) é de Cauchy em probabilidade, isto é,

$$\forall\, \epsilon > 0 \quad \sup_{n,m\geq k} \mathrm{P}(\{|X_n - X_m| \geq \epsilon\}) \to 0,\ k \to +\infty.$$

Dem: A implicação (i) $\Rightarrow$ (ii) é consequência imediata da inclusão $\{|X_n - X_m| \geq \epsilon\} \subset \{|X_n - X| \geq \epsilon/2\} \cup \{|X_m - X| \geq \epsilon/2\}$. Para estabelecer a implicação recíproca, comecemos por mostrar que sendo (X_n) de Cauchy em probabilidade existe uma subsucessão (X_{n_k}) que é de Cauchy quase certamente. Com efeito, sendo (X_n) de Cauchy em probabilidade, existe uma subsucessão (n_k) de (n) tal que $\mathrm{P}(\{|X_{n_{k+1}} - X_{n_k}| \geq 2^{-k}\}) < 2^{-k}$, para todo o $k \in \mathbb{N}$. Pelo teorema de Borel-Cantelli concluímos que $\mathrm{P}(N) = 0$, onde $N = \limsup\{|X_{n_{k+1}} - X_{n_k}| \geq 2^{-k}\}$. Dado $\omega \in N^c$, existe assim $\ell \in \mathbb{N}$ tal que $|X_{n_{k+1}}(\omega) - X_{n_k}(\omega)| < 2^{-k}$, para todo o $k \geq \ell$. Tomando agora $r > s \geq \ell$ obtemos $|X_{n_r}(\omega) - X_{n_s}(\omega)| \leq \sum_{j=s}^{r-1} |X_{n_{j+1}}(\omega) - X_{n_j}(\omega)| < 2^{-\ell+1}$, o que prova que (X_{n_k}) que é de Cauchy quase certamente. Finalmente, sendo X a variável aleatória real que satisfaz $X_{n_k} \xrightarrow{qc} X$, cuja existência é assegurada pelo Teorema 5.1.3, e usando uma vez mais o facto de (X_n) ser de Cauchy em probabilidade, concluímos que $X_n \xrightarrow{p} X$. $\square$

Teorema 5.2.4 *$X_n \xrightarrow{p} X$ sse toda a subsucessão de (X_n) possui uma subsucessão que converge quase certamente para X.*

DEM: Se $X_n \xrightarrow{p} X$, como toda a subsucessão de (X_n) converge em probabilidade para X, basta provar que existe uma subsucessão de (X_n) que converge quase certamente para X. Tal facto é uma consequência de (X_n) ser de Cauchy em probabilidade e do teorema anterior. Reciprocamente, suponhamos que toda a subsucessão de (X_n) possui uma subsucessão que converge quase certamente para X. Dado $\epsilon > 0$, qualquer, pretendemos provar que a sucessão $x_n = \mathrm{P}(|X_n - X| \geq \epsilon)$, converge para zero. Para tal basta provar que toda a sua subsucessão admite uma subsucessão que converge para zero. Seja então $(x_{n'})$ uma qualquer subsucessão de (x_n). Por hipótese, a subsucessão $(X_{n'})$ de (X_n) admite uma subsucessão $(X_{n''})$ que converge quase certamente, e por maioria de razão em probabilidade, para X. Assim, $\mathrm{P}(|X_{n''} - X| \geq \epsilon) \to 0$, ou seja, $x_{n''} \to 0$. $\square$

Terminamos este parágrafo com uma caracterização da convergência quase certa que nos será muito útil no próximo capítulo.

Teorema 5.2.5 (X_n) *converge quase certamente sse* $\sup_{j \geq 1} |X_{n+j} - X_n| \xrightarrow{p} 0$.

DEM: Consequência do Teorema 5.1.3 e do Exercício 5.2.4. $\square$

Exercícios

1. Se $X_n \xrightarrow{p} X$ e $X_n \xrightarrow{p} Y$, então $X = Y$ q.c..

2. Considere a sucessão (X_n) definida em $([0,1[, \mathcal{B}([0,1[), \lambda)$ por $X_n = \mathbb{1}_{[\frac{k}{2^m}, \frac{k+1}{2^m}[}$, se $n = 2^m + k$ com $m = 0, 1, 2, \ldots$ e $k \in \{0, 1, \ldots, 2^m - 1\}$. Mostre que X_n converge em probabilidade para a v.a. nula, mas não quase certamente.

3. Sendo f uma função real de variável real contínua, prove que se $X_n \xrightarrow{p} X$, então $f(X_n) \xrightarrow{p} f(X)$.

   (Sugestão: Use o Teorema 5.2.4.)

4. Seja (X_n) uma sucessão monótona de v.a. reais. Mostre que $X_n \xrightarrow{p} X$ sse $X_n \xrightarrow{qc} X$.

5.3 Convergência em média de ordem p

Definição 5.3.1 *Se $X_1, X_2, \ldots$, são variáveis aleatórias em $\mathcal{L}^p$, com $0 < p < +\infty$, dizemos que (X_n)* ***converge para a variável aleatória X em média de ordem p****, e escrevemos $X_n \xrightarrow{\mathcal{L}^p} X$, se*

$$\|X_n - X\|_p^p = \mathrm{E}|X_n - X|^p \to 0.$$

A convergência em média de ordem 2 diz-se também ***convergência em média quadrática*** *sendo denotada por $\xrightarrow{mq}$.*

Reparemos que a variável aleatória limite X está necessariamente em $\mathcal{L}^p$ pois $|X|^p \leq 2^p(|X_n - X|^p + |X_n|^p)$. O que referimos para os modos de convergência anteriores sobre a unicidade do limite, vale também para o limite em média de ordem p.

A desigualdade de Tchebychev-Markov que estabelecemos a seguir generaliza a desigualdade de Bienaymé-Tchebychev estabelecida no Exercício 4.2.7, permitindo-nos mostrar que a convergência em probabilidade é implicada pela convergência em média de ordem p.

Teorema 5.3.2 (desigualdade de Tchebychev-Markov ([1])**)** *Se X é uma variável aleatória real e $p > 0$, então para todo o $\alpha > 0$,*

$$\mathrm{P}(|X| \geq \alpha) \leq \frac{\mathrm{E}|X|^p}{\alpha^p}.$$

DEM: Como, para $\alpha > 0$, $1\!\!1_{\{|X| \geq \alpha\}} \leq |X|^p/\alpha^p$, obtemos $\mathrm{P}(|X| \geq \alpha) = \mathrm{E}(1\!\!1_{\{X \geq \alpha\}}) \leq \mathrm{E}|X|^p/\alpha^p$. $\square$

Teorema 5.3.3 *Para $0 < p < +\infty$, se $X_n \xrightarrow{\mathcal{L}^p} X$ então $X_n \xrightarrow{p} X$.*

Para diferentes valores de p, os diferentes modos de convergência em média de ordem p estão relacionados como se descreve a seguir.

Teorema 5.3.4 *Para $1 \leq p < q < +\infty$, se $X_n \xrightarrow{\mathcal{L}^q} X$, então $X_n \xrightarrow{\mathcal{L}^p} X$.*

DEM: Consequência da desigualdade $\|X\|_p \leq \|X\|_q$ que obtemos directamente da desigualdade de Hölder (cf. AMI, §5.3). $\square$

A convergência em média de ordem p não é em geral consequência das convergências quase certa ou em probabilidade. Tal ocorre, no entanto, sob certas condições sobre a sucessão de variáveis aleatórias como as que explicitamos no resultado seguinte.

Teorema 5.3.5 (da convergência dominada em $\mathcal{L}^p$) *Se*

a) $X_n \xrightarrow{qc} X$ *ou* $X_n \xrightarrow{p} X$*;*

b) $|X_n| \leq Y$, P-*q.c., para todo o n, com $Y \in \mathcal{L}^p$ para algum $0 < p < +\infty$;*

então $X \in \mathcal{L}^p$ e $X_n \xrightarrow{\mathcal{L}^p} X$.

DEM: Bastará considerar o caso em que $X_n \xrightarrow{p} X$. Provemos em primeiro lugar que $|X| \leq Y$, quase certamente. Para $\delta > 0$ temos, $\mathrm{P}(|X| > Y + \delta) \leq \mathrm{P}(|X| > |X_n| + \delta) \leq \mathrm{P}(|X_n - X| > \delta) \to 0$, quando $n \to +\infty$. Sendo $\delta > 0$ qualquer, concluímos que $\mathrm{P}(|X| \leq$

[1]Markov, A.A., Ischislenie Veroiatnostei, 1913. Este é o livro de A.A. Markov (1856-1922) sobre Cálculo de Probabilidades.

$Y) = 1$. Tomemos agora $\epsilon > 0$, qualquer. Uma vez que $\mathrm{E}(Y^p) < +\infty$, existe $M > 0$ tal que $\mathrm{E}(Y^p \mathbb{1}_{\{2Y>M\}}) < \epsilon$. Assim, $\mathrm{E}|X_n - X|^p = \mathrm{E}(|X_n - X|^p \mathbb{1}_{\{|X_n-X|\leq\epsilon\}}) + \mathrm{E}(|X_n - X|^p \mathbb{1}_{\{\epsilon<|X_n-X|\leq M\}}) + \mathrm{E}(|X_n - X|^p \mathbb{1}_{\{|X_n-X|>M\}}) < \epsilon^p + M^p\mathrm{P}(|X_n - X| > \epsilon) + 2^p\epsilon$, o que permite concluir uma vez que $\mathrm{P}(|X_n - X| > \epsilon) \to 0$. $\square$

Notemos, em particular, que sob as condições do teorema anterior com $p = 1$, vale a convergência das esperanças matemáticas respectivas, isto é, $\mathrm{E}(X_n) \to \mathrm{E}(X)$. Este resultado é o já nosso conhecido **teorema da convergência dominada de Lebesgue** (cf. AMI, §4.4).

Terminamos com uma caracterização da convergência em média de ordem p análoga às que já obtivemos para a convergência quase certa e para a convergência em probabilidade.

Teorema 5.3.6 *Seja* (X_n) *uma sucessão de variáveis aleatórias em* $\mathcal{L}^p$*, para algum* $0 < p < +\infty$*. As condições seguintes são equivalentes:*

(i) $X_n \xrightarrow{\mathcal{L}^p} X$, para alguma variável aleatória real X;

(ii) (X_n) é de Cauchy em $\mathcal{L}^p$, isto é,

$$\sup_{n,m\geq k} \|X_n - X_m\|_p \to 0,\ k \to +\infty.$$

Dem: A implicação (i) $\Rightarrow$ (ii) é consequência de $\|\cdot\|_p^p$, para $0 < p < 1$, e $\|\cdot\|_p$, para $1 \leq p < +\infty$, verificarem a desigualdade triangular (cf. §4.2). Sendo agora (X_n) de Cauchy em $\mathcal{L}^p$, da desigualdade de Tchebychev-Markov concluímos que (X_n) é de Cauchy em probabilidade. Pelo Teorema 5.2.3 existe um subsucessão (X_{n_k}) de (X_n) tal que $X_{n_k} \xrightarrow{qc} X$, para alguma variável aleatória real X. Pelo lema de Fatou (cf. AMI, §4.4) temos então $\mathrm{E}|X_n - X|^p \leq \liminf \mathrm{E}|X_n - X_{n_k}|^p$, o que permite concluir usando uma vez mais o facto de (X_n) ser de Cauchy em $\mathcal{L}^p$. $\square$

Exercícios

1. Conclua a desigualdade de Tchebychev-Markov é óptima no sentido em que para qualquer $\alpha > 0$, existe uma variável aleatória X que verifica a igualdade.

2. Considere a sucessão (X_n) definida no Exercício 5.2.2. Mostre que X_n converge em média de ordem p mas não quase certamente.

3. Considere a sucessão (X_n) definida em $([0,1], \mathcal{B}([0,1]), \lambda)$ por $X_n = n\mathbb{1}_{[0,\frac{1}{n}]}$. Mostre que X_n converge quase certamente para a função nula, mas não em média de ordem p.

4. Seja (X_n) uma sucessão de v.a. em $\mathcal{L}^p$ com $X_n \xrightarrow{\mathcal{L}^p} X$. Mostre que $\|X_n\|_p \to \|X\|_p$.

5. Seja (X_n) uma sucessão de v.a.r. de quadrado integrável. Mostre que $\mathrm{E}(X_n) \to \mu$ e $\mathrm{Var}(X_n) \to 0$ sse $X_n \xrightarrow{mq} \mu$.

6. Seja (X_n) uma sucessão de v.a. não-correlacionadas com $\mathrm{P}(X_n = 1) = \mathrm{P}(X_n = -1) = 1/2$. Mostre que $\sum_{j=1}^{n} X_j/n \xrightarrow{mq} 0$.

7. Se $\mathrm{E}|X|^p < +\infty$, para algum $p > 0$, mostre que $\lim n^p\, \mathrm{P}(|X| \geq n) = 0$.

   (Sugestão: Use o teorema da convergência dominada.)

8. Sejam (X_n) v.a.r. independentes de quadrado integrável com média zero e $\sum_{n=1}^{\infty} \mathrm{E}(X_n^2) < \infty$. Mostre que $\sum_{k=1}^{\infty} X_k$ converge em média quadrática (isto é, $S_n = \sum_{k=1}^{n} X_k$ converge em média quadrática para alguma v.a.r. S).

5.4 Convergência funcional de vectores aleatórios

Para terminar este capítulo, notemos que as noções de convergência consideradas para sucessões de variáveis aleatórias reais podem ser extendidas sem dificuldade ao caso de sucessões de vectores aleatórios definidos num mesmo espaço de probabilidade. No que se segue, denotaremos por $\|\cdot\|$ a norma euclideana de $\mathbb{R}^d$. No entanto, a definição seguinte não depende da norma considerada em $\mathbb{R}^d$.

Definição 5.4.1 *Se (X_n) e (X) são vectores aleatórios definidos num mesmo espaço de probabilidade, dizemos que (X_n)* ***converge para*** *X* P***-quase certamente*** *(resp.* ***em probabilidade*** *ou* ***em média de ordem*** *p) e escrevemos $X_n \xrightarrow{qc} X$ (resp. $\xrightarrow{p}$, $\xrightarrow{\mathcal{L}^p}$), se $\|X_n - X\| \xrightarrow{qc} 0$ (resp. $\xrightarrow{p}$, $\xrightarrow{\mathcal{L}^p}$).*

Atendendo a que a convergência duma sucessão de vectores aleatórios segundo qualquer um dos modos anteriores é equivalente à convergência das respectivas margens, versões vectoriais dos resultados apresentados nos parágrafos anteriores podem assim, sem excepção, ser obtidos.

5.5 Bibliografia

Billingsley, P. (1986). *Probability and Measure*, Wiley.

Jacod, J., Protter, P. (2000). *Probability Essentials*, Springer.

Lukacs, E. (1975). *Stochastic Convergence*, Academic Press.

Resnick, S.I. (1999). *A Probability Path*, Birkhäuser.

Capítulo 6

Leis dos grandes números e séries de variáveis aleatórias independentes

Leis dos grandes números para variáveis de quadrado integrável. Leis fracas de Kolmogorov e de Khintchine. Leis fortes e séries de variáveis aleatórias. Lei forte de Kolmogorov. O teorema das três séries.

6.1 Generalidades

Sendo $(\Omega_0, \mathcal{A}_0, \mathrm{P}_0)$ um modelo probabilístico para uma determinada experiência aleatória $\mathcal{E}$, e $A \in \mathcal{A}_0$ um acontecimento aleatório, o conceito frequencista de probabilidade a que fizemos alusão no §1.1, estabelece que a probabilidade $\mathrm{P}_0(A)$ do acontecimento A é o limite, num sentido a precisar, da frequência relativa de ocorrência do acontecimento A em sucessivas repetições, sempre nas mesmas condições, da experiência aleatória em causa.

Dito por outras palavras, para o modelo probabilístico $(\Omega, \mathcal{A}, \mathrm{P})$ com

$$\Omega = \bigotimes_{n=1}^{\infty} \Omega_0, \ \mathcal{A} = \bigotimes_{n=1}^{\infty} \mathcal{A}_0 \ \text{ e } \ \mathrm{P} = \bigotimes_{n=1}^{\infty} \mathrm{P}_0,$$

que descreve a repetição, sempre nas mesmas condições, da experiência $\mathcal{E}$, e sendo $S_n = \sum_{k=1}^{n} X_k$, onde X_k é a variável aleatória definida em $(\Omega, \mathcal{A}, \mathrm{P})$ que toma valor 1 ou 0, consoante, A ocorra ou não na k-ésima repetição da experiência, o número de ocorrências de A nas primeiras n repetições de $\mathcal{E}$, o conceito frequencista de probabilidade pode ser traduzido pela convergência

$$\frac{S_n}{n} \to \mathrm{P}_0(A),$$

segundo um modo de convergência estocástica a precisar.

Duma forma geral, sendo (X_n) uma sucessão de variáveis aleatórias reais definidas num mesmo espaço de probabilidade, um resultado que estabelece a convergência

$$\frac{S_n}{n} - \mu_n \xrightarrow{\mathcal{M}} Y$$

para alguma sucessão (μ_n) de números reais e para alguma variável aleatória Y, onde $\xrightarrow{\mathcal{M}}$ representa um dos modos de convergência em probabilidade, quase certa, ou em média de ordem p, é conhecido como **lei dos grandes números**. Quando a convergência envolvida é a convergência em probabilidade, o resultado é dito **lei fraca dos grandes números**. Quando a convergência é a convergência quase certa, o resultado é dito **lei forte dos grandes números**. Se a convergência utilizada for a convergência em média de ordem p, dizemos que temos uma **lei dos grandes números em média de ordem** p.

Com excepção do próximo parágrafo em que estabelecemos leis dos grandes números para sucessões de variáveis aleatórias não necessariamente independentes, admitiremos ao longo deste capítulo que as variáveis (X_n) são independentes mas não necessariamente identicamente distribuídas. Neste contexto, e tendo em mente a lei zero-um de Kolmogorov, sabemos que a existir o limite de S_n/n segundo um dos modos de convergência anteriores, a variável limite é necessariamente degenerada (ver Exercício 6.1.2).

Definição 6.1.1 *Dizemos que a sucessão* (X_n) **obedece a uma lei dos grandes números** *para o modo de convergência* $\mathcal{M}$ *se*

$$\frac{S_n}{n} - \mu_n \xrightarrow{\mathcal{M}} 0,$$

para alguma sucessão (μ_n) *de números reais.*

Por simplicidade, sempre que (X_n) obedeça a uma lei dos grandes números denotaremos por (μ_n) uma das sucessões que satisfaz a definição anterior.

Exercícios

1. Mostre que (X_n) obedece a uma lei dos grandes números para o modo de convergência $\mathcal{M}$ sse existe uma sucessão (ν_n) de números reais tal que $\frac{1}{n}\sum_{i=1}^{n}(X_i - \nu_i) \xrightarrow{\mathcal{M}} 0$.

2. Mostre que se a sucessão (X_n) de variáveis aleatórias independentes verifica $S_n/n - \mu_n \xrightarrow{\mathcal{M}} Y$, para alguma sucessão de números reais (μ_n) e alguma v.a.r. Y, então Y é quase certamente constante.

3. Seja (X_n) uma sucessão de v.a.r. com $|X_n| \leq M$, para todo o $n \in \mathbb{N}$. Mostre que se (X_n) obedece a uma lei fraca dos grandes números então $\mu_n - \frac{1}{n}\sum_{i=1}^n \mathrm{E}(X_i) \to 0$.
4. Considere a sucessão (X_n) satisfazendo $\mathrm{P}(X_n = n^2) = 1/n^2$ e $\mathrm{P}(X_n = -n^2/(n^2-1)) = 1 - 1/n^2$.
   (a) Mostre que $\mathrm{E}(X_n) = 0$ e $\sum_{n=1}^{\infty} \mathrm{P}(X_n = n^2) < \infty$.
   (b) Use o Lema de Borel-Cantelli para mostrar que $S_n/n \xrightarrow{qc} -1$.
   (c) Conclua que o resultado estabelecido no exercício anterior não é válido para esta sucessão.
5. Sejam (X_n) e (Y_n) sucessões de v.a.r. independentes (não necessariamente definidas num mesmo espaço de probabilidade) com $X_n \sim Y_n$. Mostre que se (X_n) obedece a uma lei dos grandes números para o modo de convergência $\mathcal{M}$, o mesmo acontece com (Y_n).

6.2 Primeiras leis dos grandes números

Neste parágrafo obtemos leis dos grandes números usando técnicas baseadas no cálculo de momentos de ordem superior ou igual à segunda. Em parágrafos posteriores, e à custa de técnicas de demonstração mais elaboradas, mostraremos que no caso das sucessões de variáveis aleatórias independentes tais leis podem ser obtidas para variáveis não necessariamente de quadrado integrável.

No resultado seguinte estabelecemos uma condição necessária e suficiente para a validade duma lei dos grandes números em média quadrática duma qualquer sucessão (X_n) de variáveis de quadrado integrável.

Teorema 6.2.1 *([1]) Seja (X_n) uma sucessão de variáveis aleatórias reais de quadrado integrável. (X_n) obedece a uma lei dos grandes números em média quadrática sse* $\mathrm{Var}(S_n)/n^2 \to 0$. *Neste caso* $\mu_n - \frac{1}{n}\sum_{i=1}^n \mathrm{E}(X_i) \to 0$.

Dem: Se $\mathrm{Var}(S_n)/n^2 \to 0$ então $S_n/n - \mu_n \xrightarrow{mq} 0$, com $\mu_n = \mathrm{E}(S_n/n)$, o que estabelece a suficiência da condição anterior para a validade duma lei dos grandes números em média quadrática. A condição é também necessária pois $\mathrm{Var}(S_n/n) \leq \mathrm{E}(S_n/n - \mu_n)^2$ (cf. Exercício 4.2.4). □

Atendendo ao Teorema 5.3.3, e sob as condições do teorema anterior, a condição $\mathrm{Var}(S_n)/n^2 \to 0$ é também suficiente para a validade duma lei fraca dos grandes números. No entanto, notemos que esta pode ser obtida via desigualdade de Bienaymé-Tchebychev, pois para $\epsilon > 0$,

$$\begin{aligned}\mathrm{P}(|S_n/n - \mathrm{E}(S_n/n)| \geq \epsilon) &= \mathrm{P}(|S_n - \mathrm{E}(S_n)| \geq n\epsilon) \\ &\leq \frac{1}{n^2\epsilon^2}\mathrm{Var}(S_n).\end{aligned}$$

[1] Markov, A.A., *Izv. Mat. Fiz. Ob. pri Kazanskom Univ.* (Ser. 2) 15, 135, 1906.

No caso particular em que (X_n) é uma sucessão de variáveis aleatórias reais de quadrado integrável com $\mathrm{E}(X_k) = \mu$, para todo o $k \in \mathbb{N}$, a condição $\mathrm{Var}(S_n)/n^2 \to 0$ é necessária e suficiente para que $S_n/n \xrightarrow{mq} \mu$. Além disso, se as variáveis da sucessão são duas a duas não-correlacionadas, a condição $\mathrm{Var}(S_n)/n^2 \to 0$ reduz-se a $\sum_{k=1}^n \mathrm{Var}(X_k)/n^2 \to 0$. Estas condições são, em particular, satisfeitas por uma sucessão de variáveis independentes e identicamente distribuídas de quadrado integrável.

Terminamos este parágrafo mostrando que sob condições mais restritivas que as até aqui consideradas, são também válidas leis fortes dos grandes números. Começaremos por admitir que as variáveis (X_n) são independentes e que possuem momentos de quarta ordem uniformemente limitados.

Teorema 6.2.2 *Se (X_n) é uma sucessão de variáveis aleatórias reais independentes com $\sup_{k\in\mathbb{N}} \mathrm{E}(X_k^4) < +\infty$, então (X_n) obedece a uma lei forte dos grandes números com $\mu_n - \frac{1}{n}\sum_{i=1}^n \mathrm{E}(X_i) \to 0$.*

DEM: Basta demonstrar o resultado para $\mathrm{E}(X_n) = 0$, para todo o $n \in \mathbb{N}$. Pela independência das variáveis (X_n) e da desigualdade de Hölder temos $\mathrm{E}(S_n^4) \leq n(3n - 2)\sup_{k\in\mathbb{N}} \mathrm{E}(X_k^4)$. Usando agora a desigualdade de Tchebychev-Markov obtemos $\sum_{n=1}^\infty \mathrm{P}(|S_n/n| \geq \epsilon) \leq \mathrm{E}(S_n^4)/(\epsilon^4 n^4) < +\infty$, o que, pelo Exercício 5.1.3, permite concluir. □

No resultado seguinte, utilizando uma técnica de demonstração conhecida por **método das subsucessões**, estabelecemos uma lei forte dos grandes sob condições menos restritivas que as anteriores. Admitiremos que as variáveis (X_n) são duas a duas não-correlacionadas e que possuem momentos de segunda ordem uniformemente limitados.

Teorema 6.2.3 *Seja (X_n) uma sucessão de variáveis aleatórias reais de quadrado integrável duas a duas não-correlacionadas com $\sup_{k\in\mathbb{N}} \mathrm{E}(X_k^2) < +\infty$. Então (X_n) obedece a uma lei forte dos grandes números com $\mu_n - \frac{1}{n}\sum_{i=1}^n \mathrm{E}(X_i) \to 0$.*

DEM: Sem perda de generalidade suponhamos que $\mathrm{E}(X_n) = 0$, para todo o $n \in \mathbb{N}$. Denotando $Y_n = S_n/n$, começaremos por estabelecer o resultado para a subsucessão de (Y_{n^n}) de (Y_n), Numa segunda fase extendemo-lo a toda a sucessão. temos $\mathrm{E}(Y_n^2) = \mathrm{E}(S_n^2)/n^2 = \sum_{k=1}^n \mathrm{E}(X_k^2)/n^2 \leq \gamma/n$, onde $\gamma = \sup_{k\in\mathbb{N}} \mathrm{E}(X_k^2)$. Assim, $\sum_{n=1}^\infty \mathrm{E}(Y_{n^2}^2) \leq \sum_{n=1}^\infty \gamma/n^2 < +\infty$, ou ainda, $\mathrm{E}(\sum_{n=1}^\infty Y_{n^2}^2) < +\infty$, e consequentemente $\sum_{n=1}^\infty Y_{n^2}^2 < +\infty$, quase certamente. Concluímos assim que $\lim Y_{n^2} = 0$, *q.c.*. Para demonstrar que $\lim Y_n = 0$, *q.c.*, consideremos, para $n \in \mathbb{N}$, $p(n) \in \mathbb{N}$ tal que $p(n)^2 < n \leq (p(n)+1)^2$. Assim, $\mathrm{E}(Y_n - \frac{p(n)^2}{n} Y_{p(n)^2})^2 = \mathrm{E}(\frac{1}{n}\sum_{k=p(n)^2+1}^n X_k) \leq (n - p(n)^2)\gamma/n^2 \leq (2p(n) + 1)\gamma/n^2 \leq (2\sqrt{n}+1)\gamma/n^2 \leq 3\gamma/n^{3/2}$, e tal como atrás $\mathrm{E}(\sum_{n=1}^\infty (Y_n - \frac{p(n)^2}{n} Y_{p(n)^2})^2) < +\infty$, o que implica que $\lim(Y_n - \frac{p(n)^2}{n} Y_{p(n)^2}) = 0$, *q.c.* Como $\lim Y_{p(n)^2} = 0$, *q.c.* e $p(n)^2/n \leq 1$, concluímos finalmente que $\lim Y_n = 0$, *q.c.* □

No caso particular em que (X_n) é uma sucessão de variáveis aleatórias reais de quadrado integrável duas a duas não-correlacionadas com $\mathrm{E}(X_k) = \mu$, para todo o $k \in \mathbb{N}$, concluímos que a condição $\sup_{k\in\mathbb{N}} \mathrm{E}(X_k^2) < +\infty$ é suficiente para que $S_n/n \xrightarrow{qc} \mu$. Estas condições são, em particular, satisfeitas por uma sucessão de variáveis independentes e identicamente distribuídas de quadrado integrável.

Exercícios

1. Estabeleça leis fracas e fortes dos grandes números para cada uma das seguintes sucessões de variáveis aleatórias:
   (a) (X_n) é uma sucessão de variáveis de Bernoulli de parâmetro p duas a duas não-correlacionadas ([2]).
   (b) (X_n) é uma sucessão de v.a.r. duas a duas não-correlacionadas com X_n uma variável de Bernoulli de parâmetro p_n ([3]).
   (c) (X_n) é uma sucessão de v.a.r. de quadrado integrável, duas a duas não-correlacionadas com $\mathrm{Var}(X_n) \leq \gamma$ ([4]).
2. Seja (X_n) uma sucessão de v.a.r. com $|X_n| \leq M$, para todo o $n \in \mathbb{N}$. Mostre que a condição $\mathrm{Var}(S_n)/n^2 \to 0$ é necessária para a validade duma lei fraca dos grandes números.
3. Sejam (X_n) uma qualquer sucessão de v.a.r. e $p \geq 1$. Mostre que:
   (a) $X_n \xrightarrow{qc} 0 \Rightarrow S_n/n \xrightarrow{qc} 0$;
   (b) $X_n \xrightarrow{\mathcal{L}^p} 0 \Rightarrow S_n/n \xrightarrow{\mathcal{L}^p} 0$.
   (c) Verifique que $X_n \xrightarrow{p} 0 \not\Rightarrow S_n/n \xrightarrow{p} 0$, considerando (X_n) com $\mathrm{P}(X_n = 2^n) = 1/n$ e $\mathrm{P}(X_n = 0) = 1 - 1/n$.
4. (**Velocidade de convergência em probabilidade**) Sejam (X_n) uma sucessão de v.a.r. i.i.d. de quadrado integrável e $\mu = \mathrm{E}(X_1)$.
   (a) Mostre que $b_n(S_n/n - \mu) \xrightarrow{p} 0$ (resp. $\xrightarrow{mq}$), para toda a sucessão (b_n) satisfazendo $b_n/n^{1/2} \to 0$.
   (b) Tomando $X_n \sim N(0,1)$, conclua que o resultado anterior não é, em geral, válido para $b_n = n^{1/2}$.

[2]Lei fraca de Bernoulli, J., *Ars Conjectandi*, Basel, 1713 e lei forte de Borel, E., *Rend. Circ. Mat. Palermo* 27, 247–271, 1909.

[3]Lei fraca de Poisson, S.D., *Recherches sur la Probabilité des Judgements*, Paris, 1837.

[4]Lei fraca de Tchebychev, P.L., *J. Math. Pures et Appl.* (Sér. 2) 12, 177–184, 1867 (reproduzido em *Oeuvres de P.L. Tchebychev*, Vol. 1, 28, 687–694).

6.3 Leis fracas dos grandes números

Neste parágrafo discutimos a convergência em probabilidade de S_n/n sob condições parcialmente mais fracas que as consideradas no parágrafo anterior. Em particular, verificaremos que é possível obter leis fracas dos grandes números sob condições menos restritivas sobre os momentos das variáveis em questão. No que se segue limitar-nos-emos a estabelecer condições suficientes para a validade duma lei fraca dos grandes números. No caso de existirem condições necessárias e suficientes indicá-las-emos.

Teorema 6.3.1 (Lei fraca de Kolmogorov ([5])**)** *Seja (X_n) uma sucessão de variáveis aleatórias reais independentes satisfazendo as condições seguintes para alguma sucessão (a_n) de números reais:*

a) $\sum_{k=1}^n \mathrm{P}(|X_k - a_k| > n) \to 0$;

b) $\frac{1}{n^2}\sum_{k=1}^n \mathrm{E}((X_k - a_k)^2 \mathbb{1}_{|X_k - a_k| \le n}) \to 0$.

Então, (X_n) obedece a uma lei fraca dos grandes números com $\mu_n = \frac{1}{n}\sum_{k=1}^n \{\mathrm{E}((X_k - a_k)\mathbb{1}_{|X_k-a_k|\le n}) - a_k\}$.

Dem: Basta considerar o caso $a_k = 0$, para todo o k. Para k e n naturais, consideremos as variáveis aleatórias $X'_{n,k} = X_k \mathbb{1}_{|X_k|\le n}$ e $S'_n = \sum_{k=1}^n X'_{n,k}$. Para $\epsilon > 0$, temos por a), $\mathrm{P}(|S'_n - S_n| \ge \epsilon) \le \sum_{k=1}^n \mathrm{P}(X'_{n,k} \ne X_k) = \sum_{k=1}^n \mathrm{P}(|X_k| > n) \to 0$. Como $S_n/n - \mu_n = (S_n - S'_n)/n + (S'_n - \mathrm{E}(S'_n))/n$, basta agora mostrar que $(S'_n - \mathrm{E}(S'_n))/n \xrightarrow{p} 0$. Tal facto é consequência de b) pois para $\epsilon > 0$, $\mathrm{P}(|S'_n - \mathrm{E}(S'_n)|/n \ge \epsilon) \le \mathrm{Var}(S'_n)/(\epsilon^2 n^2) = \epsilon^{-2} n^{-2} \sum_{j=1}^n \mathrm{E}(X_j^2 \mathbb{1}_{|X_j|\le n}) \to 0$. $\square$

Kolmogorov mostra ainda que as condições anteriores além de suficientes são também necessárias para a validade duma lei fraca dos grandes números quando a sucessão (a_n) é substituída por uma sucessão (m_n) de medianas de (X_n), isto é, m_n é um número real para o qual $\mathrm{P}(X_n < m_n) \le 1/2$ e $\mathrm{P}(X_n \le m_n) \ge 1/2$.

Teorema 6.3.2 ([6]) *Seja (X_n) é uma sucessão de variáveis aleatórias reais independentes e identicamente distribuídas. (X_n) obedece a uma lei fraca dos grandes números sse $n\mathrm{P}(|X_1| > n) \to 0$. Neste caso podemos tomar $\mu_n = \mathrm{E}(X_1 \mathbb{1}_{|X_1|\le n})$.*

Dem: Para estabelecer a suficiência da condição $n\mathrm{P}(|X_1| > n) \to 0$, vamos mostrar que se verifica a condição b) do teorema anterior para $a_n = 0$. Com efeito $\mathrm{E}(X_1^2 \mathbb{1}_{\{|X_1|\le n\}}) \le \sum_{k=1}^n k^2 \mathrm{P}(k-1 < |X_1| \le k) \le 2\sum_{i=1}^n i\mathrm{P}(i-1 < |X_1| \le n) \le 2\sum_{i=1}^n i\mathrm{P}(|X_1| > i-1)$, o que permite concluir. Reciprocamente, se (X_n) obedece a uma lei fraca dos grandes

[5] Kolmogorov, A.N., *Math. Ann.* 99, 309–319, 1928.

[6] Kolmogorov, A.N., *Math. Ann.* 102, 484–488, 1930.

números sabemos da observação anterior que $n\mathrm{P}(|X_1 - m| > n) \to 0$, onde m é uma mediana de X_1. Sendo esta condição equivalente a $n\mathrm{P}(|X_1| > n) \to 0$, fica concluída a demonstração. □

Notemos que as condições impostas no resultado anterior, não implicam a integrabilidade das variáveis aleatórias (X_n) (ver Exercício 6.3.2). No caso destas serem integráveis vale o resultado seguinte.

Teorema 6.3.3 (Lei fraca de Khintchine ([7])) *Se (X_n) é uma sucessão de variáveis aleatórias reais independentes, identicamente distribuídas e integráveis, então $S_n/n \xrightarrow{p} \mu$, onde $\mu = \mathrm{E}(X_1)$.*

DEM: Sendo X_1 integrável, as hipóteses do Teorema 6.3.2 são trivialmente verificadas (ver Exercício 5.3.7). □

Exercícios

1. Seja (X_n) uma sucessão de v.a.r. independentes com $\sum_{k=1}^n \mathrm{E}|X_k|^{1+\delta}/n^{1+\delta} \to 0$, para algum $0 < \delta \leq 1$. Mostre que (X_n) obedece a uma lei fraca dos grandes números com $\mu_n = \sum_{k=1}^n \mathrm{E}(X_k)/n$.

2. Seja (X_n) uma sucessão de v.a.r. i.i.d. com $\mathrm{P}(X_1 = k) = \mathrm{P}(X_1 = -k) = \frac{c}{k^2 \ln k}$, para $k = 2, 3, \ldots$, onde $c = \frac{1}{2}\big(\sum_{k=2}^{\infty} \frac{1}{k^2 \ln k}\big)^{-1}$.
   (a) Verifique que $n\mathrm{P}(|X_1| > n) \to 0$ e $\mathrm{E}|X_1| = +\infty$.
   (b) Mostre que $S_n/n \xrightarrow{p} 0$.

3. Sendo X uma variável aleatória real, mostre que:
   (a) Para $p > 0$ vale a igualdade $\mathrm{E}|X|^p = \int_{]0,+\infty[} p\, y^{p-1} \mathrm{P}(|X| > y) d\lambda(y)$.
   (Sugestão: Utilize o teorema de Fubini.)
   (b) A condição $n\,\mathrm{P}(|X| > n) \to 0$ implica que $\mathrm{E}|X|^p < +\infty$, para todo o $0 < p < 1$.

4. Se (X_n) é uma sucessão de v.a.r. i.i.d. com distribuições de Cauchy de parâmetros 0 e 1, mostre que (X_n) não obedece a uma lei fraca dos grandes números.

6.4 Leis fortes e séries de variáveis independentes

Contrariamente ao caso da lei fraca dos grandes números, não é conhecida uma condição necessária e suficiente para a validade duma lei forte dos grandes números para variáveis independentes mas não necessariamente identicamente distribuídas.

[7] Khintchine, A.Ya., *C. R. Acad. Sci. Paris* 188, 477–479, 1929.

No parágrafo 6.2 estabelecemos uma primeira lei forte para sucessões de variáveis aleatórias duas a duas não-correlacionadas com momentos de segunda ordem uniformemente limitados. Neste parágrafo vamos obter uma lei forte para sucessões de variáveis aleatórias independentes sob condições menos restritivas que as consideradas no Teorema 6.2.2. Para tal vamos utilizar a relação entre a convergência quase certa da média empírica S_n/n e a convergência da série $\sum_{k=1}^{\infty} X_k/k$ que estabelecemos no resultado seguinte.

Lema 6.4.1 (de Kronecker) *Se* (x_n) *é uma sucessão de números reais tal que* $\sum_{k=1}^{\infty} x_k/k$ *converge, então* $\sum_{k=1}^{n} x_k/n \to 0$.

Dem: Dado $\epsilon > 0$, existe por hipótese $n_0 \in \mathbb{N}$ tal que para $n \geq n_0$, $|r_n| < \epsilon$, onde $r_n = \sum_{k=n+1}^{\infty} x_k/k$. Assim, como $\sum_{k=1}^{n} x_k = \sum_{k=1}^{n}(r_{k-1} - r_k)k = \sum_{k=1}^{n-1} r_k + r_0 - nr_n$, obtemos para $n \geq n_0$, $|\sum_{k=1}^{n} x_k/n| \leq \sum_{k=1}^{n_0-1} |r_k|/n + |r_0|/n + |r_n| + \sum_{k=n_0}^{n} |r_k|/n < \epsilon(3 + (n - n_0 + 1)/n) < 4\epsilon$. □

O resultado que a seguir estabelecemos permite obter condições suficientes para a convergência quase certa duma série de variáveis aleatórias independentes e, por maioria de razão, via lema de Kronecker, condições suficientes para uma lei forte dos grandes números. Para tal necessitamos duma generalização da desigualdade

$$\mathrm{P}\Big(|S_n| \geq \epsilon\Big) \leq \frac{1}{\epsilon^2}\sum_{k=1}^{n} \mathrm{E}(X_k^2),$$

que podemos obter como aplicação directa da desigualdade Bienaymé-Tchebychev (ver Exercício 4.2.7).

Lema 6.4.2 (Desigualdade maximal de Kolmogorov ([8])) *Sejam* $X_1, \ldots, X_n$ *são variáveis aleatórias reais independentes com média zero e* $S_k = X_1 + \ldots + X_k$*, para* $k = 1, \ldots, n$*. Então, para todo o* $\epsilon > 0$,

$$\mathrm{P}\Big(\max_{1\leq k\leq n} |S_k| \geq \epsilon\Big) \leq \frac{1}{\epsilon^2}\sum_{k=1}^{n} \mathrm{E}(X_k^2).$$

Dem: Para $\epsilon > 0$, definamos os acontecimentos disjuntos $E_1 = \{|S_1| \geq \epsilon\}$ e $E_k = \{|S_1| < \epsilon, \ldots, |S_{k-1}| < \epsilon, |S_k| \geq \epsilon\}$, para $2 \leq k \leq n$, que satisfazem $\bigcup_{k=1}^{n} E_k = \{\max_{1\leq k\leq n} |S_k| \geq \epsilon\}$. Pela desigualdade de Markov temos $\mathrm{P}(E_k) \leq \epsilon^{-2}\mathrm{E}(S_k \mathbb{1}_{E_k})^2$. Usando agora a independência entre $S_k \mathbb{1}_{E_k}$ e $S_n - S_k$, podemos ainda escrever $\mathrm{E}(S_k^2 \mathbb{1}_{E_k}) \leq \mathrm{E}(S_k^2 \mathbb{1}_{E_k} + (S_n - S_k)^2 \mathbb{1}_{E_k}) = \mathrm{E}(S_k^2 \mathbb{1}_{E_k} + 2S_k(S_n - S_k)\mathbb{1}_{E_k} + (S_n - S_k)^2 \mathbb{1}_{E_k}) = \mathrm{E}(S_n \mathbb{1}_{E_k})^2$. Finalmente, $\mathrm{P}(\max_{1\leq k\leq n} |S_k| \geq \epsilon) = \sum_{k=1}^{n} \mathrm{P}(E_k) \leq \sum_{k=1}^{n} \epsilon^{-2}\mathrm{E}(S_n \mathbb{1}_{E_k})^2 \leq \epsilon^{-2}\mathrm{E}(S_n^2)$. □

[8]Kolmogorov, A.N., *Math. Ann.* 99, 309–319, 1928.

Teorema 6.4.3 (Critério da variância para séries ([9])**)** *Sejam* (X_n) *variáveis aleatórias reais independentes de quadrado integrável com* $\mathrm{E}(X_n) = 0$, *para todo o* $n \in \mathbb{N}$. *Se a série* $\sum_{n=1}^{\infty} \mathrm{Var}(X_n)$ *é convergente, então a série* $\sum_{n=1}^{\infty} X_n$ *converge quase certamente.*

Dem: Atendendo ao Teorema 5.2.5, para mostrar que $S_n = \sum_{k=1}^{n} X_k$ converge quase certamente basta mostrar que $\sup_{j\geq 1} |S_{n+j} - S_n| \xrightarrow{p} 0$. Pela desigualdade maximal de Kolmogorov e para $\epsilon > 0$, qualquer, podemos obter $\mathrm{P}(\sup_{j\geq 1} |S_{n+j} - S_n| \geq \epsilon) = \lim_{N\to+\infty} \mathrm{P}(\max_{1\leq j\leq N} |S_{n+j} - S_n| \geq \epsilon) \leq \lim_{N\to+\infty} \mathrm{P}(\max_{1\leq j\leq N} |\sum_{k=n+1}^{n+j} X_k| \geq \epsilon) \leq \sum_{k=n+1}^{\infty} \mathrm{Var}(X_k)/\epsilon^2$, o que permite concluir. □

Como aplicação directa do critério anterior obtemos um primeiro conjunto de condições suficientes para a convergência duma série de variáveis aleatórias independentes de quadrado integrável.

Teorema 6.4.4 *Sejam* (X_n) *variáveis aleatórias reais independentes de quadrado integrável. Se as séries* $\sum_{n=1}^{\infty} \mathrm{E}(X_n)$ *e* $\sum_{n=1}^{\infty} \mathrm{Var}(X_n)$ *são convergentes então a série* $\sum_{n=1}^{\infty} X_n$ *converge quase certamente.*

Uma segunda consequência do critério da variância para séries é uma lei forte geral para variáveis independentes de quadrado integrável mas não necessariamente identicamente distribuídas, sob condições menos restritivas que as do Teorema 6.2.3.

Teorema 6.4.5 ([10]) *Sejam* (X_n) *variáveis aleatórias reais independentes de quadrado integrável. Se a série* $\sum_{k=1}^{\infty} \mathrm{Var}(X_k)/k^2$ *é convergente, então* $S_n/n - \mu_n \xrightarrow{qc} 0$, *onde* $\mu_n = \frac{1}{n}\sum_{k=1}^{n} \mathrm{E}(X_k)$.

Dem: Como por hipótese $\sum_{k=1}^{\infty} \mathrm{Var}(X_k/k) < +\infty$, pelo critério da variância para séries concluímos que $\sum_{k=1}^{\infty}(X_k - \mathrm{E}(X_k))/k$ converge quase certamente. Do Lema 6.4.1 deduzimos o pretendido. □

Exercícios

1. Seja (X_n) uma sucessão de v.a.r. satisfazendo $\mathrm{P}(X_n = n^2) = \mathrm{P}(X_n = -n^2) = 1/(2n^2)$ e $\mathrm{P}(X_n = 0) = 1 - 1/n^2$. Conclua que a condição estabelecida no Teorema 6.4.5 não é necessária para a validade duma lei forte dos grandes números.

2. Sejam (X_n) e (Y_n) sucessões de v.a.r. independentes (não necessariamente definidas num mesmo espaço de probabilidade) com $X_n \sim Y_n$. Mostre que $\sum X_n$ converge quase certamente sse $\sum Y_n$ converge quase certamente.

[9] Khintchine, A.Ya., Kolmogorov, A.N., *Mat. Sb.* 32, 668–676, 1925.
[10] Kolmogorov, A.N., *C. R. Acad. Sci. Paris* 191, 910–912, 1930.

3. (**Velocidade de convergência quase certa**) Sejam (X_n) uma sucessão de v.a.r. i.i.d. de quadrado integrável e $\mu = \mathrm{E}(X_1)$.

   (a) Mostre que se $\sum a_n^2/n^2 < \infty$ para alguma sucessão de números reais (a_n) então $a_n(S_n/n - \mu) \xrightarrow{qc} 0$.

   (b) Conclua que $n^{1/2}(\ln n)^{-1/2-\epsilon}(S_n/n - \mu) \xrightarrow{qc} 0$, para todo o $\epsilon > 0$.

6.5 Lei forte dos grandes números de Kolmogorov

Mostramos neste parágrafo que se (X_n) é uma sucessão de variáveis aleatórias reais e independentes e identicamente distribuídos, a condição $\mathrm{E}|X_1| < +\infty$ é necessária e suficiente para que S_n/n convirja quase certamente para um valor real μ, ou de forma equivalente, para que (X_n) obedeça a uma lei forte dos grandes números com $\mu_n = \mu$. Trata-se da lei forte dos grandes números de Kolmogorov.

6.5.1 Necessidade da condição de integrabilidade

A necessidade da condição de integrabilidade para a validade duma lei forte dos grandes números cuja variável limite não é constantemente infinita, é estabelecida à custa dos resultados seguintes.

Lema 6.5.1 *Se Y é uma variável aleatória real então*

$$\sum_{n=1}^{\infty} \mathrm{P}(|Y| > n) \leq \mathrm{E}|Y| \leq 1 + \sum_{n=1}^{\infty} \mathrm{P}(|Y| > n).$$

Dem: Pelo Exercício 6.3.3 temos $\mathrm{E}|Y| = \int_{[0,+\infty[} \mathrm{P}(|Y| > y)dy = \sum_{n=0}^{\infty} \int_{[n,n+1[} \mathrm{P}(|Y| > y)dy$, o que permite concluir. $\square$

Lema 6.5.2 *Sejam (X_n) variáveis aleatórias reais independentes e identicamente distribuídas. As condições seguintes são equivalentes:*

i) $\mathrm{E}|X_1| < +\infty$;

ii) $\lim X_n/n = 0$, *q.c.*;

iii) $\forall \epsilon > 0 \ \sum_{n=1}^{\infty} \mathrm{P}(|X_1| > \epsilon n) < +\infty$.

Dem: Para $\epsilon > 0$, tomando $Y = X_1/\epsilon$ no lema anterior obtemos a equivalência entre as condições *i*) e *iii*). A equivalência entre as condições *ii*) e *iii*) é uma consequência imediata da equivalência entre as convergências quase certa e quase completa para zero da sucessão (X_n/n) (ver Exercício 5.1.3). $\square$

Teorema 6.5.3 *Sejam (X_n) variáveis aleatórias reais independentes e identicamente distribuídas e $\mu \in \mathbb{R}$ tais que $S_n/n \xrightarrow{qc} \mu$. Então $\mathrm{E}|X_1| < +\infty$.*

DEM: Como por hipótese, $X_n/n = (S_n - S_{n-1})/n \xrightarrow{qc} 0$, o resultado é consequência do Lema 6.5.2. □

6.5.2 Suficiência da condição de integrabilidade

Estamos agora em condições de estabelecer o principal resultado deste capítulo.

Teorema 6.5.4 (Lei forte de Kolmogorov ([11])) *Seja (X_n) uma sucessão de variáveis aleatórias reais independentes e identicamente distribuídas. Então, existe $\mu \in \mathbb{R}$ tal que $S_n/n \xrightarrow{qc} \mu$ sse $\mathrm{E}|X_1| < +\infty$. Nesse caso $\mu = \mathrm{E}(X_1)$.*

DEM: Atendendo ao Teorema 6.5.3 basta mostrar que $S_n/n \xrightarrow{qc} \mathrm{E}(X_1)$, quando $\mathrm{E}|X_1| < +\infty$. Sem perda de generalidade vamos admitir que $\mathrm{E}(X_1) = 0$. Consideremos as variáveis $X_n' = X_n \mathbb{1}_{\{|X_n| \leq n\}}$, para $n \geq 1$. Pelo Lema 6.3.1, $\sum_{n=1}^{\infty} \mathrm{P}(X_n \neq X_n') = \sum_{n=1}^{\infty} \mathrm{P}(|X_n| > n) < +\infty$, e assim, pelo teorema de Borel-Cantelli, $\mathrm{P}(X_n \neq X_n'\ i.o.) = 1$. Concluímos assim que existe $N \in \mathcal{A}$ com $\mathrm{P}(N) = 0$ tal que para todo o $\omega \in N^c$ as sucessões $(X_n(\omega))$ e $(X_n'(\omega))$ coincidem para n suficientemente grande. Bastará assim provar que $S_n'/n \xrightarrow{qc} 0$, onde $S_n' = \sum_{k=1}^{n} X_n'$. Para tal vamos lançar mão do Teorema 6.4.5, mostrando que a série $\sum_{k=1}^{\infty} \mathrm{Var}(X_k')/k^2$ é convergente. Ora $\sum_{k=1}^{\infty} \mathrm{Var}(X_k')/k^2 \leq \sum_{k=1}^{\infty} \mathrm{E}(X_1^2 \mathbb{1}_{\{|X_1| \leq k\}})/k^2 = \sum_{k=1}^{\infty} \sum_{j=1}^{k} \mathrm{E}(X_1^2\ \mathbb{1}_{\{j-1<|X_1| \leq j\}})/k^2 = \sum_{j=1}^{\infty} \sum_{k=j}^{\infty} \mathrm{E}(X_1^2 \mathbb{1}_{\{j-1<|X_1| \leq j\}})/k^2$, onde $\sum_{k=1}^{\infty} \frac{1}{k^2} \leq 2$, $\sum_{k=j}^{\infty} \frac{1}{k^2} \leq \frac{1}{j-1}$, para $j \geq 2$, e $\mathrm{E}(X_1^2 \mathbb{1}_{\{j-1<|X_1| \leq j\}}) \leq j\mathrm{E}(|X_1| \mathbb{1}_{\{j-1<|X_1| \leq j\}})$. Assim, $\sum_{k=1}^{\infty} \mathrm{Var}(X_k')/k^2 \leq 2\mathrm{E}(|X_1| \mathbb{1}_{\{|X_1| \leq 1\}}) + \sum_{j=2}^{\infty} \frac{j}{j-1}\ \mathrm{E}(|X_1| \mathbb{1}_{\{j-1<|X_1| \leq j\}}) \leq 2 \sum_{j=1}^{\infty} \mathrm{E}(|X_1| \mathbb{1}_{\{j-1<|X_1| \leq j\}}) = 2\mathrm{E}|X_1| < +\infty$. □

Exercícios

1. Sejam (X_n) v.a.r. i.i.d. em $\mathcal{L}^p$. Mostre que $\frac{1}{n} \sum_{i=1}^{n} X_i^p \xrightarrow{qc} E(X_1^p)$.

2. Denotemos por $\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$ e $\widehat{\sigma}_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$, a **média empírica** e **variância empírica**, das v.a.r. $X_1, \ldots, X_n$. Mostre que se (X_n) são variáveis i.i.d. de quadrado integrável com variância σ^2, então $\mathrm{E}(\widehat{\sigma}_n^2) = \sigma^2$ e $\widehat{\sigma}_n^2 \xrightarrow{qc} \sigma^2$.

3. Retome os Exercícios 1.8.4 e 2.1.6. Conclua que $S_n \xrightarrow{qc} -\infty$.

4. (**Integração pelo método de Monte Carlo, I**) Sejam (U_n) uma sucessão de v.a. i.i.d. uniformemente distribuídas sobre o intervalo $[0, 1]$, e f uma função real mensurável definida em $[0, 1]$ tal que $\int_{[0,1]} |f| d\lambda < +\infty$. Mostre que $\frac{1}{n} \sum_{i=1}^{n} f(U_i) \xrightarrow{qc} \int_{[0,1]} f d\lambda$.

5. (**Integração pelo método de Monte Carlo, II**) Sejam $U_1, V_1, U_2, V_2, \ldots$ v.a. i.i.d. uniformemente distribuídas sobre o intervalo $[0, 1]$, e $f : [0, 1] \to [0, 1]$ uma função mensurável. Para $n \in \mathbb{N}$, defina $Z_n = \mathbb{1}_{\{f(U_n)>V_n\}}$, e mostre que $\frac{1}{n} \sum_{i=1}^{n} Z_i \xrightarrow{qc} \int_{[0,1]} f d\lambda$.

[11] Kolmogorov, A.N., *Grundbegriffe der Wahrscheinlichkeitrechnung*, Berlin, 1933.

6. (**Velocidade de convergência quase certa** ([12])) Sejam (X_n) uma sucessão de v.a.r. i.i.d. e $p \in]1,2[$. Mostre que $n^{1-1/p}(S_n/n - \mu) \xrightarrow{qc} 0$ para algum $\mu \in \mathbb{R}$ sse $\mathrm{E}|X|^p < \infty$. Neste caso $\mu = \mathrm{E}(X_1)$.

   (Sugestão: Retome as demonstrações dos Teoremas 6.5.3 e 6.5.4, mostrando no primeiro caso que $X_n/n^{1/p} \xrightarrow{qc} 0$ e no segundo que $S'_n/n^{1/p} \xrightarrow{qc} 0$, onde $S'_n = \sum_{k=1}^n X_k \mathbb{I}_{\{|X_k| \le k^{1/p}\}}$.)

6.6 O teorema das três séries

No Teorema 6.4.4 obtivemos condições suficientes para a convergência quase certa duma série de variáveis aleatórias independentes. De seguida aprofundamos este assunto começando por mostrar que no caso das variáveis aleatórias serem limitadas as condições anteriores são também necessárias. Para tal lançamos mão da desigualdade seguinte devida a Kolmogorov.

Lema 6.6.1 *Sejam $X_1, \ldots, X_n$ variáveis aleatórias reais independentes com média zero, $S_k = X_1 + \ldots + X_k$, e suponhamos que existe $\gamma > 0$ tal que $|X_k| \le \gamma$ q.c., para $k = 1, \ldots, n$. Então, para todo o $\epsilon > 0$,*

$$\mathrm{P}\Big(\max_{1 \le k \le n} |S_k| \ge \epsilon\Big) \ge 1 - \frac{(\epsilon + \gamma)^2}{\sum_{k=1}^n \mathrm{E}(X_k^2)}.$$

Dem: Sejam E_k, para $1 \le k \le n$, os acontecimentos definidos na demonstração da desigualdade maximal de Kolmogorov, e D_k, para $0 \le k \le n$, os acontecimentos $D_0 = \Omega$ e $D_k = \{|S_1| < \epsilon, \ldots, |S_{k-1}| < \epsilon, |S_k| < \epsilon\}$, para $1 \le k \le n$. Claramente $\{\max_{1 \le k \le n} |S_k| \ge \epsilon\} = \sum_{k=1}^n E_k = D_n^c$. Para $k \ge 1$, D_k e E_k são disjuntos e $D_k + E_k = D_{k-1}$, o que permite escrever $S_{k-1}\mathbb{I}_{D_{k-1}} + X_k \mathbb{I}_{D_{k-1}} = S_k \mathbb{I}_{D_{k-1}} = S_k \mathbb{I}_{D_k} + S_k \mathbb{I}_{E_k}$, onde $S_0 = 0$. Usando a independência entre $S_{k-1}\mathbb{I}_{D_{k-1}}$ e X_k e entre $\mathbb{I}_{D_{k-1}}$ e X_k temos $\mathrm{E}(S_{k-1}^2 \mathbb{I}_{D_{k-1}}) + \mathrm{E}(X_k^2)\mathrm{P}(D_{k-1}) = \mathrm{E}(S_k^2 \mathbb{I}_{D_k}) + \mathrm{E}(S_k^2 \mathbb{I}_{E_k})$. Além disso, como $\mathrm{P}(D_{k-1}) \ge \mathrm{P}(D_n)$ e $|S_k \mathbb{I}_{E_{k-1}}| \le (\epsilon + \gamma)\mathbb{I}_{E_k}$, obtemos $\mathrm{E}(S_{k-1}^2 \mathbb{I}_{D_{k-1}}) + \mathrm{E}(X_k^2)\mathrm{P}(D_n) \le \mathrm{E}(S_k^2 \mathbb{I}_{D_k}) + (\epsilon+\gamma)^2 \mathrm{P}(E_k)$. Finalmente, somando todas as inequações anteriores obtemos $\sum_{k=1}^n \mathrm{E}(X_k^2)\mathrm{P}(D_n) \le \mathrm{E}(S_n^2 \mathbb{I}_{D_n}) + (\epsilon+\gamma)^2 \mathrm{P}(D_n^c) \le (\epsilon+\gamma)^2$, o que permite concluir. □

Estabelecemos em primeiro lugar a recíproca do critério de Kolmogorov para variáveis uniformemente limitadas.

Teorema 6.6.2 *Sejam (X_n) variáveis aleatórias reais independentes tais que $\sup_{k \in \mathbb{N}} |X_k| \le \gamma$ q.c., para alguma constante $\gamma > 0$, e $\mathrm{E}(X_k) = 0$ para todo o $k \in \mathbb{N}$. Então $\sum_{n=1}^\infty X_n$ converge quase certamente sse a série $\sum_{n=1}^\infty \mathrm{Var}(X_n)$ é convergente.*

[12] Marcinkiewicz, J., Zygmund, A., *Fund. Math.* 29, 60–90, 1937.

Dem: Tendo em conta o Teorema 6.4.3, basta mostrar que $\sum_{n=1}^{\infty} \mathrm{Var}(X_n)$ é convergente quando $\sum_{n=1}^{\infty} X_n$ converge quase certamente. Neste caso, para todo o $\epsilon > 0$ $\mathrm{P}(\sup_{j\geq 1} |S_{n+j} - S_n| \geq \epsilon) \to 0$ (cf. Teorema 5.2.5). Ora, pelo Lema 6.6.1, $\mathrm{P}(\sup_{j\geq 1} |S_{n+j} - S_n| \geq \epsilon) = \lim_{N\to+\infty} \mathrm{P}(\max_{1\leq j\leq N} |S_{n+j} - S_n| \geq \epsilon) \geq 1 - (\epsilon + 2\gamma)^2 / \sum_{k=n+1}^{\infty} \mathrm{Var}(X_k)$, obtendo-se uma contradição se $\sum_{n=1}^{\infty} \mathrm{Var}(X_n) = +\infty$. □

Passemos agora ao estudo da série não centrada no caso das variáveis da sucessão serem uniformemente limitadas.

Teorema 6.6.3 *Sejam (X_n) variáveis aleatórias reais independentes tais que $\sup_{k\in\mathbb{N}} |X_k| \leq \gamma$ q.c., para alguma constante $\gamma > 0$. Então a série $\sum_{n=1}^{\infty} X_n$ converge quase certamente sse as séries $\sum_{n=1}^{\infty} \mathrm{E}(X_n)$ e $\sum_{n=1}^{\infty} \mathrm{Var}(X_n)$ são convergentes.*

Dem: Pelo Teorema 6.4.3 basta mostrar que a convergência quase certa da série $\sum X_n$ implica a convergência das séries $\sum \mathrm{E}(X_n)$ e $\sum \mathrm{Var}(X_n)$. Sabemos do Exercício 3.2.6 que existem variáveis aleatórias reais independentes $Y_1, Z_1, Y_2, Z_2, \ldots$ definidas num mesmo espaço de probabilidade com $X_n \sim Y_n \sim Z_n$, para todo o $n \in \mathbb{N}$. Além disso, se $\sum X_n$ é quase certamente convergente, também o são as séries $\sum Y_n$ e $\sum Z_n$ (cf. Exercício 6.1.5). Consideremos agora as variáveis $U_n = Y_n - Z_n$, para $n \in \mathbb{N}$ (notemos que $U_n \sim -U_n$, pelo que esta técnica é conhecida por **simetrização**). Tais variáveis são independentes, com $\mathrm{E}(U_n) = 0$, $|U_n| \leq 2\gamma$, *q.c.* e além disso $\sum U_n$ é quase certamente convergente. Pelo Teorema 6.6.2 concluímos que $\sum \mathrm{Var}(U_n) < +\infty$, ou ainda $\sum \mathrm{Var}(X_n) < +\infty$, uma vez que $\mathrm{Var}(U_n) = \mathrm{Var}(Y_n) + \mathrm{Var}(Z_n) = 2\mathrm{Var}(X_n)$. Novamente pelo Teorema 6.6.2, $\sum (X_n - \mathrm{E}(X_n))$ converge quase certamente, o que implica a convergência da série $\sum \mathrm{E}(X_n)$, pois $\mathrm{E}(X_n) = X_n - (X_n - \mathrm{E}(X_n))$, para $n \in \mathbb{N}$. □

Finalmente, no caso geral das variáveis não serem uniformemente limitadas é válido o seguinte resultado.

Teorema 6.6.4 (das três séries [13]**)** *Se (X_n) é uma sucessão de variáveis aleatórias reais independentes então $\sum_{n=1}^{\infty} X_n$ converge quase certamente sse para algum $c > 0$ as três séries seguintes são convergentes:*

$$a)\ \sum_{n=1}^{\infty} \mathrm{P}(|X_n| > c); \quad b)\ \sum_{n=1}^{\infty} \mathrm{E}(X_n \mathbb{1}_{|X_n|\leq c}); \quad c)\ \sum_{n=1}^{\infty} \mathrm{Var}(X_n \mathbb{1}_{|X_n|\leq c}).$$

Dem: Começamos por notar que a convergência da série a) é, pela lei zero-um de Borel, equivalente à condição $\mathrm{P}(|X_n| > c \ i.o.) = 0$, ou ainda a $\mathrm{P}(X_n \neq X_n \mathbb{1}_{|X_n|\leq c} \ i.o.) = 0$.

[13] Kolmogorov, A.N., *Math. Ann.* 99, 309–319, 1928.

Assim, a menos dum conjunto de pontos ω com probabilidade nula as sucessões $(X_n(\omega))$ e $(X_n(\omega)\mathbb{1}_{\{|X_n|\leq c\}}(\omega))$ coincidem para n suficientemente grande, o que implica que a convergência quase certa de $\sum X_n$ é equivalente à convergência quase certa da série $\sum X_n\mathbb{1}_{\{|X_n|\leq c\}}$. Por outro lado, a convergência das séries b) e c) é, pelo Teorema 6.6.3, equivalente à convergência quase certa de $\sum X_n\mathbb{1}_{\{|X_n|\leq c\}}$. Concluímos assim que a convergência das séries a), b) e c) implica a convergência quase certa de $\sum X_n$. Reciprocamente, se $\sum X_n$ converge quase certamente, então como $\{|X_n| > c\ i.o.\} \subset \{\limsup X_n \neq 0\}$, para $c > 0$ qualquer, concluímos que $\mathrm{P}(|X_n| > c\ i.o.) = 0$, o que, como já referimos é equivalente à convergência da série a). Repetindo o raciocínio anterior, concluímos que a convergência quase certa de $\sum X_n$ é equivalente à convergência quase certa da série $\sum X_n\mathbb{1}_{\{|X_n|\leq c\}}$, o que, por sua vez, é equivalente às convergência das séries b) e c). $\square$

Terminamos este parágrafo mostrando que as condições necessárias e suficientes anteriores para a convergência quase certa da série $\sum_{n=1}^{\infty} X_n$, são também necessárias e suficientes para a sua convergência em probabilidade.

Lema 6.6.5 (Desigualdade de Lévy) *Sejam $X_1,\ldots,X_n$ variáveis aleatórias reais e independentes, $S_k = X_1+\ldots+X_k$, para $k=1,\ldots,n$, e $\epsilon,\delta>0$. Se*

$$\max_{1\leq i\leq n} \mathrm{P}(|X_i+\ldots+X_n| \geq \epsilon/2) \leq \delta,$$

então

$$\mathrm{P}\Big(\max_{1\leq k\leq n} |S_k| \geq \epsilon\Big) \leq \frac{\delta}{1-\delta}.$$

DEM: Sejam E_k, $k \geq 1$, os conjuntos definidos na demonstração da desigualdade maximal de Kolmogorov. Pela independência dos acontecimentos E_k e $|S_n - S_k| \geq \epsilon/2$ temos $\mathrm{P}(\max_{1\leq k\leq n}|S_k| \geq \epsilon, |S_n| \leq \epsilon/2) = \sum_{k=1}^n \mathrm{P}(E_k, |S_n| \leq \epsilon/2) \leq \sum_{k=1}^n \mathrm{P}(E_k, |S_n - S_k| \leq \epsilon/2) = \sum_{k=1}^n \mathrm{P}(E_k)\mathrm{P}(|S_n-S_k| \leq \epsilon/2) \leq \delta\mathrm{P}(\max_{1\leq k\leq n}|S_k|\geq\epsilon)$. Por outro lado, $\mathrm{P}(\max_{1\leq k\leq n}|S_k| \geq \epsilon, |S_n| > \epsilon/2) \leq \mathrm{P}(|S_n| > \epsilon/2) \leq \delta$, o que permite concluir. $\square$

Teorema 6.6.6 (de Lévy ([14])) *Se (X_n) é uma sucessão de variáveis aleatórias reais e independentes então $S_n = \sum_{k=1}^n X_k$ converge quase certamente sse converge em probabilidade.*

DEM: Consequência imediata do Teorema 5.2.5 e da desigualdade de Lévy. $\square$

[14] Lévy, P., *Théorie de l'Addition des Variables Aléatoires*, Paris, 1937.

Exercícios

1. Recorde a natureza das séries $\sum 1/n$ e $\sum(-1)^n/n$. Considere uma sucessão (X_n) de v.a.r. i.i.d. com $P(X_n = -1) = P(X_n = 1) = 1/2$. Estude a convergência da série $\sum X_n/n$.
2. Sendo (X_n) uma qualquer sucessão de v.a.r., mostre que se $\sum_{n=1}^{\infty} \mathrm{E}(|X_n|) < \infty$, então $\sum_{n=1}^{\infty} X_n$ converge quase certamente.
3. Sejam $\ldots, Y_1, Y_0, Y_{-1}, \ldots$ uma sucessão de v.a.r. i.i.d. com $\mathrm{E}(Y_n) = 0$ e $\alpha_0, \alpha_1, \ldots$ uma sucessão de números reais com $\sum_{j=0}^{\infty} |\alpha_j| < \infty$.
   (a) Para $n \in \mathbb{N}$, mostre que $\sum_{j=0}^{\infty} \alpha_j Y_{n-j}$ converge quase certamente.
   (b) Definindo $X_n = \sum_{j=0}^{\infty} \alpha^j Y_{n-j}$, para $n \in \mathbb{N}$ e $|\alpha| < 1$, mostre que $X_n = \alpha X_{n-1} + Y_n$.

6.7 Bibliografia

Chow, Y.S., Teicher, H. (1997). *Probability Theory: Independence, Interchangeability, Martingales*, Springer.

Chung, K.L. (1974). *A Course in Probability Theory*, Academic Press.

Durrett, R. (1996). *Probability: Theory and Examples*, Duxbury Press.

Kallenberg, O. (1997). *Foundations of Modern Probability*, Springer.

Kolmogorov, A.N. (1950). *Foundations of the Theory of Probability*, Chelsea Publishing Company.

Loève, M. (1977). *Probability Theory I*, Springer.

Resnick, S.I. (1999). *A Probability Path*, Birkhäuser.

Révész, P. (1968). *The Laws of Large Numbers*, Academic Press.

Parte III

Teorema do limite central

Capítulo 7

Função característica

Integração de variáveis aleatórias complexas. Função característica dum vector aleatório. Derivadas e momentos. Injectividade. Fórmulas de inversão. Aplicações à caracterização da independência e ao estudo da distribuição da soma de vectores aleatórios.

7.1 Integração de variáveis aleatórias 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 $z = x + i\,y$ 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 $\mathcal{B}(\mathbb{C})$ e $\mathcal{B}(\mathbb{R}^2)$.

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 $Im(Z)(\omega) = Im(Z(\omega))$. As observações preliminares anteriores implicam que uma função Z definida num espaço de probabilidade $(\Omega, \mathcal{A}, \mathrm{P})$ com valores em $(\mathbb{C}, \mathcal{B}(\mathbb{C}))$ é uma variável aleatória sse a função de $(\Omega, \mathcal{A}, \mathrm{P})$ em $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$ definida por $(Re(Z), Im(Z))$ é também uma variável aleatória, ou ainda, sse $Re(Z)$ e $Im(Z)$ são variáveis aleatórias reais. Neste caso dizemos que Z é uma **variável aleatória complexa**.

Tendo em conta o que atrás foi dito, a definição de esperança matemática duma variável aleatória complexa surge agora de forma natural.

Definição 7.1.1 *Uma variável aleatória complexa Z diz-se* ***integrável*** *se $Re(Z)$ e $Im(Z)$ o forem, e nesse caso, a sua* ***esperança matemática*** *é dada por*

$$\mathrm{E}(Z) = \mathrm{E}(Re(Z)) + \mathrm{i}\,\mathrm{E}(Im(Z)).$$

Teorema 7.1.2 *a) O conjunto das variáveis aleatórias complexas integráveis é um espaço vectorial complexo (com a soma e produto escalar definidos da forma habitual).*

b) A aplicação $Z \to \mathrm{E}(Z)$ *desse espaço em* $\mathbb{C}$ *é linear.*

Dem: Basta ter em conta que o conjunto das variáveis aleatórias reais integráveis é um espaço vectorial real e a linearidade da esperança matemática para variáveis aleatórias reais. □

Teorema 7.1.3 *Uma variável aleatória complexa* Z *é integrável sse* $|Z|$ *o for, e nesse caso* $|\mathrm{E}(Z)| \leq \mathrm{E}(|Z|)$.

Dem: A primeira afirmação resulta das desigualdades $|Re(Z)| \leq |Z|$, $|Im(Z)| \leq |Z|$ e $|Z| \leq |Re(Z)| + |Im(Z)|$. A desigualdade $|\mathrm{E}(Z)| \leq \mathrm{E}(|Z|)$ é válida se $\mathrm{E}(Z) = 0$. Se $\mathrm{E}(Z) \neq 0$, seja $w = \mathrm{E}(Z)/|\mathrm{E}(Z)|$. Então $|\mathrm{E}(Z)| = w^{-1}\mathrm{E}(Z) = \mathrm{E}(w^{-1}Z) = \mathrm{E}(Re(w^{-1}Z))$ (pois $|\mathrm{E}(Z)|$ é real) $\leq \mathrm{E}(|w^{-1}Z|) = \mathrm{E}(|Z|)$. □

Antes de terminarmos este curto parágrafo sobre a integração de variáveis aleatórias complexas, 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.

7.2 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 dum vector aleatório. Para $x = (x_1, \ldots, x_d)$ e $y = (y_1, \ldots, y_d)$ em $\mathbb{R}^d$, denotaremos por $\langle x, y\rangle$ o produto interno usual em $\mathbb{R}^d$, isto é, $\langle x, y\rangle = \sum_{j=1}^d x_j y_j$.

Definição 7.2.1 *Chamamos **função característica** dum vector aleatório* X *em* $\mathbb{R}^d$ *(ou função característica de* P_X*), à função de* $\mathbb{R}^d$ *em* $\mathbb{C}$ *definida por*

$$\phi_X(t) = \mathrm{E}(e^{\mathrm{i}\langle t, X\rangle}), \text{ para } t \in \mathbb{R}^d.$$

Notemos que como $|e^{\mathrm{i}\langle t, X\rangle}| = 1$, a esperança matemática anterior está bem definida.

Teorema 7.2.2 *Se* ϕ_X *é a função característica dum vector aleatório* 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.*

DEM: As alíneas a), b) e c) são consequência imediata da definição de função característica. A continuidade de ϕ_X resulta da continuidade sob o sinal de integral. □

Atendendo à alínea c) anterior, a função característica duma variável aleatória simétrica relativamente à origem é uma função real. Neste caso $\phi_X(t) = \mathrm{E}(\cos(\langle t, X\rangle))$, para $t \in \mathbb{R}^d$.

O cálculo da função característica duma variável aleatória pode revelar-se um trabalho árduo. Tal é o caso do segundo dos exemplos seguintes.

Exemplos 7.2.3 1. Se X é uma v.a. de Bernoulli de parâmetro p, então $\phi_X(t) = e^{\mathrm{i}\,t.1}p + e^{\mathrm{i}\,t.0}(1-p) = 1 - p(1 - e^{\mathrm{i}\,t})$, para $t \in \mathbb{R}$.

2. Se X é uma v.a. normal centrada e reduzida, então $\phi_X(t) = e^{-t^2/2}$, para $t \in \mathbb{R}$. Com efeito, como $\phi_X(t) = \mathrm{E}(\cos(tX)) = \frac{1}{\sqrt{2\pi}} \int \cos(tx)e^{-x^2/2}\, d\lambda(x)$ e $\phi'_X(t) = \frac{-1}{\sqrt{2\pi}} \int x \sin(tx)e^{-x^2/2}\, d\lambda(x) = -t\phi_X(t)$, obtemos a equação diferencial $\phi'_X(t)/\phi_X(t) = -t$, que possui como solução $\phi_X(t) = e^c e^{-t^2/2}$, ou ainda, $\phi_X(t) = e^{-t^2/2}$, uma vez que $\phi_X(0) = 1$.

As funções características de subvectores dum vector X podem ser obtidas facilmente a partir de ϕ_X. Fáceis de obter são também as funções características de transformações afins dum vector X.

Teorema 7.2.4 *Se $X = (X_1, X_2)$ é um vector aleatório em $\mathbb{R}^{p+q}$, então*

$$\phi_{X_1}(t_1) = \phi_X(t_1, 0) \quad e \quad \phi_{X_2}(t_2) = \phi_X(0, t_2),$$

para todo o $t_1 \in \mathbb{R}^p$ e $t_2 \in \mathbb{R}^q$.

Teorema 7.2.5 *Sejam X um vector aleatório sobre $\mathbb{R}^{p+q}$, A uma matriz real de tipo $p \times q$ e $b \in \mathbb{R}^p$. Então $\phi_{AX+b}(t) = e^{\mathrm{i}\,\langle t,b\rangle}\phi_X(A^T t)$, para $t \in \mathbb{R}^p$.*

Como aplicação deste último resultado, podemos obter a função característica duma variável $Y \sim N(m, \sigma^2)$, pois $Y \sim \sigma X + m$, com $X \sim N(0, 1)$, e assim

$$\phi_Y(t) = e^{\mathrm{i}\,tm}\phi_X(\sigma t) = e^{\mathrm{i}\,tm}e^{-\sigma^2 t^2/2},\ t \in \mathbb{R}.$$

Exercícios

1. Demonstre os Teoremas 7.2.4 e 7.2.5.

2. Para as seguintes v.a. calcule a sua função característica:

    (a) Variável constantemente igual a m;

(b) Binomial de parâmetros n e p;

(c) Poisson de parâmetro λ;

(d) Exponencial de parâmetro λ;

(e) Uniforme sobre o intervalo $[-a, a]$.

3. Seja (X, Y) o vector aleatório com densidade

$$f(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}}\; e^{-(x^2-2\rho xy+y^2)/(2(1-\rho^2))},$$

onde $\rho \in]-1, 1[$ (ver Exemplo 1.3.5). Calcule $\phi_{(X,Y)}$ e ϕ_Y.

(Sugestão: Use o Exercício 1.7.3.)

4. Mostre que são equivalentes as seguintes proposições: i) $\mathrm{P}(X \in \mathbb{Z}) = 1$; ii) ϕ_X é periódica de período 2π; iii) $\phi_X(2\pi) = 1$.

5. Prove que se P_X é difusa, então $\phi_X(t) < 1$, para todo o $t \in \mathbb{R} \setminus \{0\}$.

7.3 Derivadas e momentos

Uma aplicação importante das funções características é agora abordada. Trata-se do cálculo dos momentos dum vector aleatório.

Teorema 7.3.1 *Se X é um vector aleatório sobre $\mathbb{R}^d$ com $\mathrm{E}\|X\|^m < +\infty$, para algum $m \in \mathbb{N}$, então ϕ_X possui derivadas parciais de ordem m e, para $t \in \mathbb{R}^d$,*

$$\frac{\partial^m \phi_X}{\partial t_{j_1} \dots \partial t_{j_m}}(t) = \mathrm{i}^m \mathrm{E}(X_{j_1} \dots X_{j_m} e^{\mathrm{i}\langle t, X\rangle}).$$

Dem: Comecemos por estabelecer o resultado para $m = 1$. Sendo e_i o i-ésimo vector da base canónica de $\mathbb{R}^d$, temos, para $t \in \mathbb{R}^d$ e $h \in \mathbb{R}$, $(\phi_X(t + he_{j_1}) - \phi_X(t))/h = \mathrm{E}(e^{\mathrm{i}\langle t,X\rangle}(e^{\mathrm{i}hX_{j_1}} - 1)/h)$, onde $|e^{\mathrm{i}\langle t,X\rangle}(e^{\mathrm{i}hX_{j_1}} - 1)/h| \le |(e^{\mathrm{i}hX_{j_1}} - 1)/h| \le |X_{j_1}| \le \|X\|$, uma vez que $|e^{\mathrm{i}x} - 1| \le |x|$, para todo o $x \in \mathbb{R}$. Como $e^{\mathrm{i}\langle t,X\rangle}(e^{\mathrm{i}hX_{j_1}} - 1)/h \to e^{\mathrm{i}\langle t,X\rangle}\,\mathrm{i}\,X_{j_1}$ e X é integrável, do teorema da convergência dominada concluímos que $\frac{\partial \phi_X}{\partial t_j}(t) = \mathrm{i}\,\mathrm{E}(X_{j_1} e^{\mathrm{i}\langle t,X\rangle})$. Suponhamos agora que o resultado é verdadeiro para o natural k e provemos que ainda válido para $k+1$. Para $t \in \mathbb{R}^d$ e $h \in \mathbb{R}$, temos $(\frac{\partial^k \phi_X}{\partial t_{j_1} \dots \partial t_{j_k}}(t + he_{j_{k+1}}) - \frac{\partial^k \phi_X}{\partial t_{j_1} \dots \partial t_{j_k}}(t))/h = \mathrm{i}^k \mathrm{E}(X_{j_1} \dots X_{j_k} e^{\mathrm{i}\langle t,X\rangle}(e^{\mathrm{i}hX_{j_{k+1}}} - 1)/h)$. Uma nova aplicação do teorema da convergência dominada permite concluir. □

Tendo em conta o resultado sobre a derivação sob o sinal de integral, concluímos, do resultado anterior, que as derivadas parciais de ordem m de ϕ_X são contínuas.

No caso das variáveis aleatórias reais obtemos o corolário seguinte:

Corolário 7.3.2 *Se X é uma variável aleatória real com* $\mathrm{E}|X|^m < +\infty$*, para algum* $m \in \mathbb{N}$*, então*

$$\phi_X^{(k)}(0) = \mathrm{i}^k \mathrm{E}(X^k), \text{ para } k = 1, \ldots, m.$$

A não existência da derivada de ordem k de ϕ_X na origem, implica assim a não integrabilidade de X^k. Ainda no contexto real, é possível provar que a existência da derivada de ordem m de ϕ_X na origem, implica a existência do momento de ordem m de X quando m é par, e do momento de ordem $m-1$ de X quando m é ímpar (ver Métivier, 1972, pg. 157 e seguintes).

Exercícios

1. Utilize o Corolário 7.3.2 para calcular a média e variância das seguintes variáveis:
   (a) Binomial de parâmetros n e p;
   (b) Poisson de parâmetro λ;
   (c) Exponencial de parâmetro λ.
2. Se $X \sim N(0,1)$, mostre que $\mathrm{E}(X^{2n-1}) = 0$ e $\mathrm{E}(X^{2n}) = (2n)!/(2^n n!)$, para todo o $n \in \mathbb{N}$.
3. Retome o Exercício 7.2.3 e calcule $\mathrm{C}_{(X,Y)}$.

7.4 Injectividade

Neste parágrafo mostraremos que a função característica dum vector aleatório caracteriza a sua distribuição de probabilidade. Fá-lo-emos a partir dos dois resultados auxiliares seguintes, onde por N_σ denotaremos um vector aleatório sobre $\mathbb{R}^d$ de densidade

$$g_\sigma(u) = \frac{1}{(\sqrt{2\pi\sigma^2}\,)^d}\, e^{-\|u\|^2/(2\sigma^2)} = \prod_{j=1}^{d} \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-u_j^2/(2\sigma^2)}, \tag{7.4.1}$$

para $u = (u_1, \ldots, u_d)$, onde $\|\cdot\|$ denota a norma euclideana em $\mathbb{R}^d$. Atendendo à forma da sua densidade, N_σ é um vector aleatório com margens independentes que seguem distribuições normais de média zero e variância σ^2. Comecemos por determinar a função característica deste vector.

Lema 7.4.2 *Para $t \in \mathbb{R}^d$,*

$$\phi_{N_\sigma}(t) = e^{-\sigma^2\|t\|^2/2}.$$

DEM: Atendendo à forma produto (7.4.1) da densidade de N_σ podemos dizer que $N_\sigma \sim (N_{1\sigma}, \ldots, N_{d\sigma})$, onde, para $i = 1, \ldots, d$, $N_{i\sigma}$ é uma variável aleatória normal de média zero e variância σ^2, e além disso, tais variáveis são independentes. Assim, para $t \in$

$\mathbb{R}^d$, $\phi_{N_\sigma}(t) = \mathrm{E}(e^{\mathrm{i}\,\langle t,N_\sigma\rangle}) = \mathrm{E}(e^{\mathrm{i}\,\sum_{j=1}^d t_j N_{j\sigma}}) = \mathrm{E}(\prod_{j=1}^d e^{\mathrm{i}\,t_j N_{j\sigma}}) = \prod_{j=1}^d \mathrm{E}(e^{\mathrm{i}\,t_j N_{j\sigma}}) = \prod_{j=1}^d \phi_{N_{j\sigma}}(t_j) = \prod_{j=1}^d e^{-\sigma^2 t_j^2/2} = e^{-\sigma^2\|t\|^2/2}$. $\square$

Lema 7.4.3 *Se X é um vector aleatório em $\mathbb{R}^d$ e h é uma função limitada e contínua de $\mathbb{R}^d$ em $\mathbb{R}$, então*

$$\mathrm{E}(h(X)) = \frac{1}{(2\pi)^d}\lim_{\sigma\to 0}\int h(x)\int \phi_X(u)e^{-\mathrm{i}\,\langle u,x\rangle-\sigma^2\|u\|^2/2}d\lambda(u)d\lambda(x).$$

Dem: Comecemos por notar que $\mathrm{E}(h(X)) = \lim_{\sigma\to 0}\int h(x)\int g_\sigma(x-y)\,d\mathrm{P}_X(y)d\lambda(x)$. Com efeito, pelo teorema da mudança de variável e pelo teorema da convergência dominada, temos $\int h(x)\int g_\sigma(x-y)\,d\mathrm{P}_X(y)d\lambda(x) = \int\int h(x)\sigma^{-d}g_1(\sigma^{-1}(x-y))\,d\mathrm{P}_X(y)d\lambda(x)$ $= \int\int h(y+u\sigma)g_1(u)\,d(\mathrm{P}_X\otimes\lambda)(y,u) \to \int\int h(y)g_(u)\,d(\mathrm{P}_X\otimes\lambda)(y,u) = \int h(y)\,d\mathrm{P}_X(y) = \mathrm{E}(h(X))$. Para concluir vamos agora mostrar que $\int g_\sigma(x-y)\,d\mathrm{P}_X(y) = (2\pi)^{-d}\int\phi_X(u)$ $e^{-\mathrm{i}\,\langle x,u\rangle-\sigma^2\|u\|^2/2}d\lambda(u)$. Para tal, notemos que as funções g_σ e ϕ_{N_σ} estão relacionadas pela igualdade $g_\sigma(x) = \phi_{N_{1/\sigma}}(-x)/(\sigma\sqrt{2\pi})^d$, para $x\in\mathbb{R}^d$, o que permite escrever $\int g_\sigma(x-y)\,d\mathrm{P}_X(y) = \int\phi_{N_{1/\sigma}}(y-x)/(\sigma\sqrt{2\pi})^d\,d\mathrm{P}_X(y) = (\sigma\sqrt{2\pi})^{-d}\int\int e^{\mathrm{i}\,\langle y-x,u\rangle}g_{1/\sigma}(u)$ $d\lambda(u)\,d\mathrm{P}_X(y) = (2\pi)^{-d}\int e^{-\mathrm{i}\,\langle x,u\rangle-\sigma^2\|u\|^2/2}\int e^{\mathrm{i}\,\langle y,u\rangle}\,d\mathrm{P}_X(y)\,d\lambda(u) = (2\pi)^{-d}\int\phi_X(u)$ $e^{-\mathrm{i}\,\langle x,u\rangle-\sigma^2\|u\|^2/2}\,d\lambda(u)$. $\square$

Teorema 7.4.4 *Se X e Y são vectores aleatórios em $\mathbb{R}^d$ (não necessariamente definidos sobre o mesmo espaço de probabilidade), então $\phi_X = \phi_Y$ sse $X\sim Y$.*

Dem: Provaremos que o conhecimento de ϕ_X implica o conhecimento de $\mathrm{P}_X(A)$ para todo o $A\in\mathcal{B}(\mathbb{R}^d)$, ou equivalentemente, para todo o rectângulo A semi-aberto à esquerda. Conhecendo ϕ_X, sabemos pelo lema anterior que conhecemos $\mathrm{E}(h(X))$ para toda a função limitada e contínua em $\mathbb{R}^d$. Dado agora um rectângulo A semi-aberto à esquerda, sabemos que existe uma sucessão (h_n) de funções contínuas e limitadas com $0\le h_n\le 1$ e $h_n\to\mathbb{1}_A$, o que, pelo teorema da convergência dominada, implica que $\mathrm{E}(h_n(X))\to\mathrm{E}(\mathbb{1}_A(X)) = \mathrm{P}_X(A)$. $\square$

Exercícios

1. Sendo X um vector aleatório em $\mathbb{R}^d$, mostre que ϕ_X é uma função real sse X é simétrico relativamente à origem (i.e. $X\sim -X$).

2. Sendo X e Y vectores aleatórios em $\mathbb{R}^d$, mostre que $X\sim Y$ sse $\langle a,X\rangle = \langle a,Y\rangle$, para todo o $a\in\mathbb{R}^d$.

7.5 Fórmulas de inversão

Dos resultados anteriores, sabemos que para $A \in \mathcal{B}(\mathbb{R}^d)$ e sendo (h_n) uma sucessão de funções contínuas e uniformemente limitadas com $h_n \to \mathbb{1}_A$, vale a igualdade

$$\mathrm{P}_X(A) = \lim_{n\to+\infty}\lim_{\sigma\to 0}\frac{1}{(2\pi)^d}\int h_n(x)\int \phi_X(u)e^{-\mathrm{i}\,\langle u,x\rangle-\sigma^2\|u\|^2/2}d\lambda(u)d\lambda(x).$$

Esta igualdade dá-nos uma primeira fórmula de inversão da função característica de X, permitindo explicitar P_X em função de ϕ_X.

Apesar de existirem outras fórmulas de inversão mais expeditas que a anterior em termos de cálculo efectivo, limitar-nos-emos, no que se segue, a apresentar uma fórmula de inversão da função característica no caso desta ser integrável à Lebesgue.

Teorema 7.5.1 *Seja X um vector aleatório em $\mathbb{R}^d$. Se ϕ_X é integrável à Lebesgue, então X é absolutamente contínuo e admite uma densidade de probabilidade contínua e limitada dada, para $x \in \mathbb{R}^d$, por*

$$g(x) = \frac{1}{(2\pi)^d}\int \phi_X(u)e^{-\mathrm{i}\,\langle u,x\rangle}d\lambda(u).$$

DEM: Comecemos por notar que sendo ϕ_X integrável, a função g dada pela fórmula anterior é limitada e contínua. Além disso, g é real pois $g = \overline{g}$. Tendo em conta o Lema 7.4.3 e o teorema da convergência dominada, $\mathrm{E}(h(X)) = \frac{1}{(2\pi)^d}\int h(x)\int \phi_X(u)e^{-\mathrm{i}\,\langle u,x\rangle}$ $d\lambda(u)d\lambda(x) = \int h(x)g(x)\,d\lambda(x)$, para todo a função h contínua e limitada em $\mathbb{R}^d$ de suporte compacto. Dado agora um rectângulo A semi-aberto à esquerda, existe uma sucessão de funções (h_n) contínuas de suporte compacto com $h_n \to \mathbb{1}_A$ e $0 \le h_n \le \mathbb{1}_E$, onde E é um rectângulo fechado que contém A (esta majoração é essencial para podermos aplicar o teorema da convergência dominada, uma vez que não provámos ainda que g é λ-integrável). Pelo teorema da convergência dominada, obtemos $\mathrm{P}_X(A) = \mathrm{E}(\mathbb{1}_A(X)) = \lim \mathrm{E}(h_n(X)) = \lim\int h_n(x)g(x)\,d\lambda(x) = \int_A g(x)\,d\lambda(x)$, o que permite concluir que $\mathrm{P}_X = g\,\lambda$, como pretendíamos (a integrabilidade e não-negatividade de g é consequência desta igualdade). $\square$

Exercícios

1. Se X é uma v.a. de Cauchy de parâmetros 0 e 1, mostre que $\phi_X(t) = e^{-|t|}$, para $t \in \mathbb{R}$. Conclua que $\mathrm{E}|X| = +\infty$.

2. Se X é tal que $\mathrm{P}(X \in \mathbb{Z}) = 1$, mostre que, para todo o $n \in \mathbb{Z}$,

$$\mathrm{P}(X = n) = \frac{1}{2\pi}\int_0^{2\pi} e^{-\mathrm{i}\,tn}\phi_X(t)\,d\lambda(t).$$

7.6 Independência e soma de vectores aleatórios

Iniciamos este parágrafo apresentando uma caracterização da independência das margens dum vector aleatório em termos da sua função característica.

Teorema 7.6.1 *Seja $X = (X_1, X_2)$ um vector aleatório sobre $\mathbb{R}^{p+q}$. X_1, X_2 são independentes sse*

$$\phi_X(t_1, t_2) = \phi_{X_1}(t_1)\phi_{X_2}(t_2),$$

para todo o $t_1 \in \mathbb{R}^p$ e $t_2 \in \mathbb{R}^q$.

Dem: Procedendo como na demonstração do Lema 7.4.2, concluímos facilmente que a independência dos vectores X_1 e X_2 implica a forma produto anterior para a função característica de X. Reciprocamente, sejam Y_1 e Y_2 vectores independentes definidos num espaço de probabilidade $(\Omega', \mathcal{A}', \mathrm{P}')$ com $Y_i \sim X_i$, para $i = 1, 2$, e $Y = (Y_1, Y_2)$. Pela primeira parte da demonstração e por hipótese, $\phi_Y(t_1, t_2) = \phi_{Y_1}(t_1)\phi_{Y_2}(t_2) = \phi_{X_1}(t_1)\phi_{X_2}(t_2) = \phi_X(t_1, t_2)$, para todo o $t_1 \in \mathbb{R}^p$ e $t_2 \in \mathbb{R}^q$. Assim, $X \sim Y$, ou ainda, $\mathrm{P}_X = \mathrm{P}'_Y = \mathrm{P}'_{Y_1} \otimes \mathrm{P}'_{Y_2} = \mathrm{P}_{X_1} \otimes \mathrm{P}_{X_2}$, o que permite concluir que X_1 e X_2 são independentes. $\square$

O resultado seguinte, tem um papel importante no estudo da distribuição duma soma de vectores aleatórios independentes. A sua demonstração é deixada ao cuidado do aluno.

Teorema 7.6.2 *Sejam $X_1, \ldots, X_n$ vectores aleatórios com valores em $\mathbb{R}^d$ definidos num mesmo espaço de probabilidade. Se $X_1, \ldots, X_n$ são independentes , então*

$$\phi_{\sum_{j=1}^n X_j}(t) = \prod_{j=1}^n \phi_{X_j}(t),$$

para todo o $t \in \mathbb{R}^d$.

Usando este resultado, concluímos facilmente que qualquer combinação linear não-nula de variáveis aleatórias normais independentes $X_1, \ldots, X_n$, com $X_j \sim N(m_j, \sigma_j^2)$, é ainda uma variável aleatória normal, uma vez que, para $t \in \mathbb{R}$, e $a_1, \ldots, a_n \in \mathbb{R}$,

$$\phi_{\sum_{j=1}^n a_j X_j}(t) = e^{\mathrm{i}\, t \sum_{j=1}^n a_j m_j} e^{-\sum_{j=1}^n \sigma_j^2 a_j^2 t^2/2},$$

que não é mais do que a função característica duma variável aleatória normal de média $\sum_{j=1}^n a_j m_j$ e variância $\sum_{j=1}^n \sigma_j^2 a_j^2$, sempre que pelo menos um dos a_j seja diferente de zero.

Exercícios

1. Verifique que o recíproco do Teorema 7.6.2 é falso, considerando $X_1 = \ldots = X_n = X$, com X uma variável de Cauchy de parâmetros 0 e 1.

2. Use o Teorema 7.6.2 para calcular a função característica duma v.a. binomial de parâmetros n e p.

3. Se $X_1, \ldots, X_n$ são v.a. independentes com distribuições de Poisson de parâmetros $\lambda_1, \ldots, \lambda_n$, mostre que $\sum_{j=1}^n X_j$ é ainda uma v.a. de Poisson de parâmetro $\sum_{j=1}^n \lambda_j$.

4. Dizemos que uma v.a. real X tem uma distribuição **Gama de parâmetros** $\alpha > 0$ **e** $\beta > 0$, e escrevemos $X \sim Gama(\alpha, \beta)$, se admite uma densidade de probabilidade da forma
$$f(x) = \begin{cases} \frac{\beta^\alpha}{\Gamma(\alpha)}\, x^{\alpha-1} e^{-x\beta}, & \text{se } x \geq 0 \\ 0, & \text{se } x < 0, \end{cases}$$
onde Γ é a função Gama (ver Exercício 3.3.4).

   (a) Sabendo que uma v.a. X com uma distribuição Gama de parâmetros $\alpha > 0$ e $\beta > 0$, tem por função característica
   $$\phi_X(t) = \frac{\beta^\alpha}{(\beta - \mathrm{i}\, t)^\alpha},$$
   mostre que se $X_1, \ldots, X_n$ são v.a.r. i.i.d. com $X_j \sim Gama(\alpha_j, \beta)$, então $\sum_{j=1}^n X_j \sim Gama(\sum_{j=1}^n \alpha_j, \beta)$.

   (b) Verifique que as distribuições exponencial e do qui-quadrado são casos particulares da distribuição Gama. Mais precisamente $\chi_n^2 = Gama(n/2, 1/2)$ e $E(\lambda) = Gama(1, \lambda)$.

7.7 Bibliografia

Jacod, J., Protter, P. (2000). *Probability Essentials*, Springer.

Kallenberg, O. (1997). *Foundations of Modern Probability*, Springer.

Lukacs, E. (1964). *Fonctions Caractéristiques*, Dunod, Paris.

Métivier, M. (1972). *Notions Fondamentales de la Théorie des Probabilités*, Dunod.

Capítulo 8

Vectores aleatórios normais

Definição de vector aleatório normal. Função característica e independência das margens. Continuidade absoluta.

8.1 Definição e existência

Como sabemos, uma variável aleatória real diz-se **normal centrada e reduzida**, se é absolutamente contínua relativamente à medida de Lebesgue sobre $\mathbb{R}$ e admite uma versão da densidade de probabilidade da forma

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\Big(-\frac{x^2}{2}\Big),\ x \in \mathbb{R}.$$

A noção de variável aleatória normal que a seguir introduzimos, é, como veremos, mais geral do que a que considerámos nos capítulos anteriores.

Definição 8.1.1 *Dizemos que uma variável aleatória real X é* ***normal****, se*

$$X \sim \sigma U + m,$$

para algum $\sigma, m \in \mathbb{R}$, onde U é uma variável aleatória normal centrada e reduzida.

Claramente $\mathrm{E}(X) = m$ e $\mathrm{Var}(X) = \sigma^2$. Se $\sigma \neq 0$, a noção de variável normal agora introduzida é precisamente a noção anteriormente considerada, uma vez que neste caso X possui uma densidade de probabilidade dada por

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{(x-m)^2}{2\sigma^2}\Big),\ x \in \mathbb{R}.$$

Se $\sigma = 0$, X é degenerada. Estamos assim a incluir na família das variáveis aleatórias normais as variáveis degeneradas. Tal como atrás, indicaremos $X \sim N(m, \sigma^2)$, e facilmente se deduz que a função característica de X é dada por

$$\phi_X(t) = \exp(\mathrm{i}\, tm) \exp(-t^2\sigma^2/2),\ t \in \mathbb{R}.$$

Definição 8.1.2 *Um vector aleatório X em $\mathbb{R}^d$ diz-se* ***normal****, ou que possui uma* **distribuição normal**, *se $\langle a, X\rangle = \sum_{i=1}^d a_i X_i$ é uma variável aleatória normal, para todo o $a \in \mathbb{R}^d$.*

Por outras palavras, um vector aleatório diz-se normal se qualquer combinação linear das suas margens for uma variável aleatória normal. Se $X_1, \dots, X_d$ são variáveis aleatórias normais independentes e não-degeneradas, sabemos do capítulo anterior que qualquer combinação linear delas é ainda uma variável aleatória normal. Nesse caso $(X_1, \dots, X_d)$ é um vector aleatório normal. Como podemos concluir do Exercício 3 seguinte, um vector aleatório com margens normais não é necessariamente normal.

Exercícios

1. Mostre que as margens dum vector aleatório normal são normais.

2. Mostre que o vector N_σ com densidade de probabilidade dada por (7.4.1) é normal.

3. Considere o vector aleatório (U, V) definido no Exercício 2.2.3. Prove que $U + V$ não é uma v.a. normal, apesar de U e V o serem.

4. Sejam X um vector aleatório normal em $\mathbb{R}^p$, A uma matriz real de tipo $d \times p$, e $m \in \mathbb{R}^d$. Prove que $AX + m$ é um vector aleatório normal em $\mathbb{R}^d$.

8.2 Função característica e independência das margens

Se X é um vector aleatório de quadrado integrável com margens independentes, sabemos já que a sua matriz de covariância C_X é diagonal. Mostramos a seguir que no caso dos vectores aleatórios normais, a condição recíproca é também verdadeira.

Comecemos por determinar a função característica dum vector aleatório normal.

Teorema 8.2.1 *Se X é um vector aleatório normal em $\mathbb{R}^d$, a sua função característica é dada por*

$$\phi_X(t) = \exp(\mathrm{i}\,\langle t, \mathrm{E}(X)\rangle) \exp(-\langle t, \mathrm{C}_X t\rangle/2),\ t \in \mathbb{R}^d.$$

Dem: Sendo X normal, $\langle t, X\rangle$ é uma variável normal para $t \in \mathbb{R}^d$. Assim, $\phi_X(t) = \phi_{\langle t,X\rangle}(1) = \exp(\mathrm{i}\,\mathrm{E}(\langle t, X\rangle)) \exp(-\mathrm{Var}(\langle t, X\rangle)/2)$. Para concluir basta agora notar que $\mathrm{E}(\langle t, X\rangle) = \langle t, \mathrm{E}(X)\rangle$ e $\mathrm{Var}(\langle t, X\rangle) = \langle t, \mathrm{C}_X t\rangle$. $\square$

Concluímos do resultado anterior que, analogamente ao caso real, a distribuição dum vector aleatório normal é caracterizada pela sua esperança matemática e pela sua matriz de covariância. A notação $X \sim N(m, \Sigma)$, indica assim que X é um vector aleatório normal de média m e matriz de covariância Σ.

Estamos agora em condições de estabelecer a caracterização já anunciada da independência das margens dum vector aleatório normal.

Teorema 8.2.2 *Se $X = (X_1, \ldots, X_d)$ é um vector aleatório normal em $\mathbb{R}^d$, então $X_1, \ldots, X_d$ são variáveis aleatórias reais independentes sse $\mathrm{Cov}(X_i, X_j) = 0$ para todo o $i \neq j$.*

DEM: Sendo $X_1, \ldots, X_d$ variáveis independentes, sabemos já que são duas a duas não correlacionadas. Reciprocamente, se $\mathrm{Cov}(X_i, X_j) = 0$, para $i \neq j$, então $\phi_X(t) = \exp(\mathrm{i} \sum_{j=1}^d \mathrm{E}(X_j)t_j) \exp(-\sum_{j=1}^d t_j^2 \mathrm{Var}(X_j)/2) = \prod_{j=1}^d \exp(\mathrm{i}\, \mathrm{E}(X_j)t_j) \exp(-t_j^2 \mathrm{Var}(X_j) /2) = \prod_{j=1}^d \phi_{X_j}(t_j)$, para $t \in \mathbb{R}^d$. O Teorema 7.6.1 permite agora concluir. $\square$

Exercícios

1. Seja (X, Y) um ve.a. absolutamente contínuo de densidade

$$f(x,y) = \frac{1}{2\pi}\Big((\sqrt{2}\, e^{-x^2/2} - e^{-x^2})e^{-y^2} + (\sqrt{2}\, e^{-y^2/2} - e^{-y^2})e^{-x^2}\Big),$$

para $(x, y) \in \mathbb{R}^2$. Prove que:

(a) X e Y são v.a. normais;

(b) $\mathrm{Cov}(X, Y) = 0$;

(c) X e Y não são v.a. independentes.

2. Utilizando o Teorema 8.2.1:

(a) resolva o Exercício 8.1.4;

(b) mostre que $(X_1, \ldots, X_d)$ é normal quando $X_1, \ldots, X_d$ são v.a.r. normais e independentes.

8.3 Continuidade absoluta

Neste parágrafo apresentamos uma caracterização da continuidade absoluta dum vector aleatório normal em termos da sua matriz de covariância.

Lema 8.3.1 *Sejam X um vector aleatório normal sobre $\mathbb{R}^d$ não-degenerado com média m e matriz de covariância Σ, e $k = \mathrm{car}(\Sigma)$. Então existe uma matriz A de tipo $d \times k$ com $AA^T = \Sigma$, tal que $X \sim AY + m$, onde $Y \sim N(0, I_k)$.*

DEM: Sendo Σ a matriz de covariância de X, Σ é simétrica e semi-definida positiva. Existe então uma matriz ortogonal P ($P^T = P^{-1}$) que diagonaliza Σ, isto é, $P^T \Sigma P = D$, com $D = diag(\lambda_1, \ldots, \lambda_d)$, onde $\lambda_i > 0$, para $i = 1, \ldots, k$, e $\lambda_i = 0$, para $i = k+1, \ldots, d$, são os valores próprios de Σ. Tomando agora

$$A = P \begin{bmatrix} \sqrt{\lambda_1} & & 0 \\ & \ddots & \\ 0 & & \sqrt{\lambda_k} \\ 0 & 0 & 0 \end{bmatrix}, \tag{8.3.2}$$

temos $\Sigma = AA^T$, com A uma matriz de tipo $d \times k$. Além disso, se $Y \sim N(0, I_k)$, é fácil verificar que $X \sim AY + m$. $\square$

Teorema 8.3.3 *Seja X um vector aleatório normal sobre $\mathbb{R}^d$ com matriz de covariância C_X. Então:*

a) Se $\mathrm{car}(\mathrm{C}_X) = 0$, X é degenerado.

b) Se $0 < \mathrm{car}(\mathrm{C}_X) < d$, X é singular e P_X está concentrada num subespaço afim de dimensão k.

c) Se $\mathrm{car}(\mathrm{C}_X) = d$, X é absolutamente contínuo e tem por versão da densidade de probabilidade

$$f_X(x) = \frac{1}{\sqrt{(2\pi)^d \det(\mathrm{C}_X)}} \exp\Big(-\frac{1}{2}\langle x - \mathrm{E}(X), \mathrm{C}_X^{-1}(x - \mathrm{E}(X))\rangle\Big),\ x \in \mathbb{R}^d.$$

Dem: a) Se $\mathrm{car}(\mathrm{C}_X) = 0$, temos $\mathrm{Var}(X_i) = 0$, para todo o $i = 1, \ldots, d$, e portanto todas variáveis X_i são degeneradas. b) Se $0 < \mathrm{car}(\mathrm{C}_X) = k < d$, pelo lema anterior existe A de tipo $d \times k$ dada por (8.3.2), tal que $X \sim AY + \mathrm{E}(X)$, onde $Y \sim N(0, I_k)$. Para $S = \{Ay + \mathrm{E}(X) : y \in \mathbb{R}^k\}$, temos $\mathrm{P}_X(S) = \mathrm{P}_{AY+\mathrm{E}(X)}(S) = \mathrm{P}_Y(\mathbb{R}^k) = 1$ e $\lambda_d(S) = 0$. X está assim concentrada no subespaço afim S de dimensão k e é alheia relativamente à medida de Lebesgue sobre $\mathbb{R}^d$. Além disso, P_X é difusa pois, para $x = Ay + \mathrm{E}(X) \in S$, $\mathrm{P}_X(\{x\}) = \mathrm{P}(AY = Ay) = \mathrm{P}(Y = y) = 0$. X é assim um vector difuso. c) Pelo Lema 8.3.1, existe A invertível de tipo $d \times d$ tal que $AA^T = \mathrm{C}_X$ e $X \sim AY + \mathrm{E}(X)$, com $Y \sim N(0, I_d)$. Utilizando agora a fórmula de transformação de vectores aleatórios absolutamente contínuos, obtemos $f_X(x) = f_Y(A^{-1}(x - \mathrm{E}(X)))|\det(A^{-1})| = |\det(A)|^{-1}(2\pi)^{-d/2}\exp(-(A^{-1}(x - \mathrm{E}(X)))^T(A^{-1}(x - \mathrm{E}(X)))/2) = ((2\pi)^d \det(\mathrm{C}_X))^{-1/2}\exp(-\langle x - \mathrm{E}(X), \mathrm{C}_X^{-1}(x - \mathrm{E}(X))\rangle/2)$. $\square$

Exercícios

1. O vector (X, Y) segue uma distribuição normal sobre $\mathbb{R}^2$ de densidade

$$f(x, y) = k\ \exp(-(x^2 - xy + y^2/2)/2),\ (x, y) \in \mathbb{R}^2.$$

   (a) Determine k e o coeficiente de correlação de (X, Y).

   (b) Sejam U e V as v.a.r. definidas, para $a \in \mathbb{R}$, por $U = 3X + aY$ e $V = aX - Y$. Determine a de modo que U e V sejam independentes e nesse caso calcule $\mathrm{E}(UV)^2$.

2. Mostre que o vector (X, Y) definido no Exemplo 2.1.9 é um vector aleatório normal.

3. Sejam $X_1, \ldots, X_n$ v.a.r. independentes com distribuição normal de média 0 e variância $\sigma^2 > 0$, e Y o vector aleatório sobre $\mathbb{R}^n$ definido por $Y = AX$, com $X = (X_1, \ldots, X_n)^T$ e A uma matriz ortogonal de ordem n (note que A possui por linhas (resp. colunas) vectores ortonormados). Sejam ainda $\bar{X}_n$ e $\widehat{\sigma}_n^2$ as média e variância empíricas das variáveis $X_1, \ldots, X_n$ (ver Exercício 6.5.2).

   (a) Mostre que $Y \sim X$.

   (b) Se a primeira linha de A é igual a $(1/\sqrt{n}, \ldots, 1/\sqrt{n})$, mostre que $\sum_{k=2}^n Y_k^2 = \sum_{i=1}^n (X_i - \bar{X}_n)^2$.

   (c) Conclua que:

      i. $\bar{X}_n$ e $\widehat{\sigma}_n^2$ são variáveis independentes;

      ii. $\dfrac{(n-1)\,\widehat{\sigma}_n^2}{\sigma^2} \sim \chi^2_{n-1}$.

8.4 Bibliografia

Jacod, J., Protter, P. (2000). *Probability Essentials*, Springer.

Métivier, M. (1972). *Notions Fondamentales de la Théorie des Probabilités*, Dunod.

Monfort, A. (1980). *Cours de Probabilités*, Economica.

Capítulo 9

Convergência em distribuição

Convergência em distribuição de vectores aleatórios. Algumas caracterizações. Relações com os outros modos de convergência. Os teoremas da selecção de Helly, de Prohorov. e da continuidade de Lévy–Bochner. O teorema de Cramér–Wold.

9.1 Definição e unicidade do limite

A noção de convergência duma sucessão (X_n) de vectores aleatórios para um vector aleatório X que estudamos neste capítulo é de natureza distinta das convergências funcionais consideradas no Capítulo 5. Para tais modos de convergência interessam os valores particulares que tomam os vectores X_n e X em pontos do conjunto onde estão definidos. Para a noção de convergência que a seguir introduzimos, interessam apenas as probabilidades com que esses vectores tomam tais valores.

Se X é uma variável aleatória em $\mathbb{R}^d$, denotaremos por F_X a sua função de distribuição e por $C(F_X)$ o conjunto dos pontos de continuidade de F_X. Salvo indicação em contrário, ao longo deste capítulo (X_n) e X são vectores aleatórios em $\mathbb{R}^d$. Como já sabemos, e com excepção do caso real, o conjunto dos pontos de descontinuidade de F_X pode ser não-numerável. No entanto, tal como no caso real, $C(F_X)$ é denso em $\mathbb{R}^d$, uma vez que $\prod_{i=1}^d C(F_{X_i}) \subset C(F_X)$, onde $X = (X_1, \ldots, X_d)$.

Definição 9.1.1 *Dizemos que uma sucessão (X_n) de vectores aleatórios, não necessariamente definidos num mesmo espaço de probabilidade,* ***converge em distribuição*** *(ou em lei) para 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 $\mathbb{R}^d$ como ilustra o exemplo da sucessão $X_n = 1/n$ que, segundo um qualquer modo de convergência aceitável, deverá convergir para $X = 0$. Reparemos que $F_{X_n}(x)$ converge para $F_X(x)$, para todo o $x \in \mathbb{R}$, com excepção 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 $X_n \sim X$, e no entanto X_n não converge em probabilidade para X.

Terminamos este parágrafo estabelecendo a unicidade do limite em distribuição no sentido seguinte:

Proposição 9.1.2 *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 $\mathbb{R}^d$ (porquê?) e que F_X e F_Y são contínuas à direita, concluímos que $F_X = F_Y$, ou seja, $X \sim Y$. $\square$

Exercícios

1. Se $X = (X_1, \ldots, X_d)$ é ve.a. em $\mathbb{R}^d$, mostre que $\prod_{i=1}^d C(F_{X_i}) \subset C(F_X)$.
   (Sugestão: Tenha em conta o Teorema 2.4.3.)
2. Sejam (X_n) e X v.a. definidas por $X_n = \alpha_n$ e $X = \alpha$, onde (α_n) e α, são números reais. Mostre que $X_n \xrightarrow{d} X$ sse $\alpha_n \to \alpha$.
3. Sejam (X_n) uma sucessão de v.a. independentes com distribuição exponencial de parâmetro 1 e $M_n = \bigvee_{i=1}^n X_i$, para $n \in \mathbb{N}$. Mostre que $M_n - \ln n \xrightarrow{d} Y$, onde $\mathrm{P}(Y \leq x) = \exp(-e^{-x})$, para $x \in \mathbb{R}$.
4. Sejam (X_n) e X ve.a. em $\mathbb{R}^d$ com densidades de probabilidade (f_n) e f, respectivamente, tais que: a) $|f_n| \leq |g|$, λ-q.c., para alguma função integrável g; b) $\lim f_n = f$, λ-q.c.. Mostre que $X_n \xrightarrow{d} X$.

9.2 Caracterizações e primeiras propriedades

Estabelecemos neste parágrafo caracterizações importantes e algumas propriedades da convergência em distribuição. Qualquer uma destas caracterizações pode ser usada para definir convergência em distribuição para variáveis aleatórias com valores em espaços métricos gerais nos quais a noção de função de distribuição se revela desprovida de sentido.

Teorema 9.2.1 *As proposições seguintes são equivalentes:*

i) $X_n \xrightarrow{d} X$;

ii) $\mathrm{E}(f(X_n)) \to \mathrm{E}(f(X))$*, para toda a função* f *contínua e limitada de* $\mathbb{R}^d$ *em* $\mathbb{R}$*.*

iii) $\mathrm{E}(f(X_n)) \to \mathrm{E}(f(X))$*, para toda a função* f *uniformemente contínua e limitada de* $\mathbb{R}^d$ *em* $\mathbb{R}$*.*

iv) $\mathrm{P}_{X_n}(A) \to \mathrm{P}_X(A)$*, para todo o* $A \in \mathcal{B}(\mathbb{R}^d)$*, com* $\mathrm{P}_X(fr(A)) = 0$*.*

Dem: As implicações *ii)* $\Rightarrow$ *iii)* e *iv)* $\Rightarrow$ *i)*, são claramente verdadeiras. Para estabelecer *iii)* $\Rightarrow$ *iv)*, consideremos $A \in \mathcal{B}(\mathbb{R}^d)$, com $\mathrm{P}_X(fr(A)) = 0$, e consideremos a função uniformemente contínua

$$\varphi(t) = \begin{cases} 1, & \text{se } t \leq 0 \\ 1 - t, & \text{se } 0 < t < 1 \\ 0, & \text{se } t \geq 1. \end{cases}$$

Para $p \in \mathbb{N}$, tomemos as funções $f_p(y) = \varphi(p\, d(y, A))$ e $g_p(y) = \varphi(1 - p\, d(y, A^c))$, definidas para $y \in \mathbb{R}^d$, onde $d(y, A)$ denota a distância de y a A. Para $p \in \mathbb{N}$, temos $\mathrm{E}(g_p(X_n)) \leq \mathrm{E}(\mathbb{I}_A(X_n)) \leq \mathrm{E}(f_p(X_n))$, para $n \in \mathbb{N}$, e por hipótese $\mathrm{E}(g_p(X)) \leq \liminf \mathrm{P}_{X_n}(A) \leq \liminf \mathrm{P}_{X_n}(A) \leq \mathrm{E}(f_p(X))$, uma vez que g_p e f_p são uniformemente contínuas. Pelo teorema da convergência dominada, $g_p \to \mathbb{I}_{\mathrm{int}(A)}$ e $f_p \to \mathbb{I}_{\overline{A}}$, o que implica $\mathrm{P}_X(\mathrm{int}(A)) \leq \liminf \mathrm{P}_{X_n}(A) \leq \liminf \mathrm{P}_{X_n}(A) \leq \mathrm{P}_X(\overline{A})$, ou ainda, $\mathrm{P}_X(A) \leq \liminf \mathrm{P}_{X_n}(A) \leq \liminf \mathrm{P}_{X_n}(A) \leq \mathrm{P}_X(A)$, uma vez que $\mathrm{P}_X(fr(A)) = 0$. Finalmente, e no caso $d = 1$, vamos estabelecer a implicação *i)* $\Rightarrow$ *ii)*. Sejam $a, b \in C(F_X)$ tais que $\mathrm{P}_X(]a, b]) > 1 - \epsilon$, com $\epsilon > 0$ fixo à partida. Por hipótese, e para $n \geq n_1$, temos $\mathrm{P}_{X_n}(]a, b]) = F_{X_n}(b) - F_{X_n}(a) = (F_{X_n}(b) - F_X(b)) + (F_X(b) - F_X(a)) + (F_X(a) - F_{X_n}(a)) > 1 - 2\epsilon$, ou ainda, $\mathrm{P}(X_n \notin]a, b]) < 2\epsilon$. Seja agora f uma função contínua e limitada em $\mathbb{R}$. Sendo f uniformemente contínua em $[a, b]$ existe um conjunto finito de pontos $a = a_0 < a_1 < \ldots < a_k = b$ tal que $|f(x) - f(a_j)| < \epsilon$, para $x \in [a_{j-1}, a_j[$, onde os a_j podem ser tomados em $C(F_X)$. A função escalonada $g = \sum_{j=1}^k f(a_j)\mathbb{I}_{]a_{j-1}, a_j]}$ satisfaz $|f(x) - g(x)| < \epsilon$, para todo o $x \in]a, b]$. Assim, $|\mathrm{E}(f(X_n)) - \mathrm{E}(g(X_n))| \leq \mathrm{E}(|f(X_n) - g(X_n)|\mathbb{I}_{X_n \in]a,b]}) + \mathrm{E}(|f(X_n) - g(X_n)|\mathbb{I}_{X_n \notin]a,b]}) \leq \epsilon + \sup_{x\in\mathbb{R}} |f(x)|\, \mathrm{P}(X_n \notin]a, b]) < \epsilon(1 + 2\sup_{x\in\mathbb{R}} |f(x)|)$. De forma análoga, $|\mathrm{E}(f(X)) - \mathrm{E}(g(X))| < \epsilon(1 + \sup_{x\in\mathbb{R}} |f(x)|)$. Tendo agora em conta a definição de g, $\mathrm{E}(g(X_n)) = \sum_{j=1}^k f(a_j)(F_{X_n}(a_j) - F_{X_n}(a_{j-1}))$, obtendo-se uma expressão análoga para $\mathrm{E}(g(X))$. Existe então $n_2 \in \mathbb{N}$, tal que $|E(g(X_n)) - \mathrm{E}(g(X))| < \epsilon$, para $n \geq n_2$. Finalmente, para $n \geq \max(n_1, n_2)$, obtemos $|\mathrm{E}(f(X_n)) - \mathrm{E}(f(X))| < 3\epsilon(1 + \sup_{x\in\mathbb{R}} |f(x)|)$. $\square$

Tal como para os outros modos de convergência estudados, a convergência em distribuição é preservada por transformações contínuas.

Teorema 9.2.2 *Se* $X_n \xrightarrow{d} X$ *então* $g(X_n) \xrightarrow{d} g(X)$*, para toda a função contínua de* $\mathbb{R}^d$ *em* $\mathbb{R}^k$*.*

Dem: Sendo $f : \mathbb{R}^k \to \mathbb{R}$ contínua e limitada, temos por hipótese $\mathrm{E}((f \circ g)(X_n)) \to \mathrm{E}((f \circ g)(X))$, ou ainda $\mathrm{E}(f(g(X_n))) \to \mathrm{E}(f(g(X)))$. Tendo em conta teorema anterior concluímos que $g(X_n) \xrightarrow{d} g(X)$. $\square$

No caso dos vectores aleatórios (X_n) e X serem absolutamente contínuos, se as densidades de probabilidade de f_{X_n} de X_n são uniformemente limitadas por uma função integrável, a convergência λ-quase em todo o ponto de f_{X_n} para f_X, implica a convergência em distribuição de X_n para X (ver Exercício 9.1.4). Como se mostra a seguir, esta convergência em distribuição pode ser obtida sob condições menos restritivas.

Teorema 9.2.3 (de Scheffé (1)**)** *Sejam (X_n) e X vectores aleatórios absolutamente contínuos em $\mathbb{R}^d$ com densidades (f_{X_n}) e f_X, respectivamente. Se $f_{X_n} \to f_X$, λ-q.t.p., então $X_n \xrightarrow{d} X$.*

Dem: Para $x \in \mathbb{R}^d$, temos $|F_{X_n}(x) - F_X(x)| \leq \int_{]-\infty,x]} |f_n(t) - f(t)| d\lambda(t) \leq \int |f_n(t) - f(t)| d\lambda(t)$. Ora $\int |f_n - f| d\lambda = \int (f - f_n)^+ d\lambda + \int (f - f_n)^- d\lambda$, e como $0 = \int (f - f_n) d\lambda = \int (f - f_n)^+ d\lambda - \int (f - f_n)^- d\lambda$, concluímos que $\int |f_n - f| d\lambda = 2 \int (f - f_n)^+ d\lambda$. O resultado é agora consequência do teorema da convergência dominada, pois $(f - f_n)^+ \leq f$ e $(f - f_n)^+ \to 0$, λ.*q.c.* $\square$

Exercícios

1. Retome a demonstração, feita no caso real, da implicação $i) \Rightarrow ii)$ do Teorema 9.2.1. Adapte-a ao caso multidimensional.

2. Sejam X_n, para $n \in \mathbb{N}$, uma v.a. uniforme sobre o conjunto $\{i/n : i = 1, \ldots, n\}$. Mostre que $X_n \xrightarrow{d} U([0,1])$.

3. Se (X_n) é uma sucessão de v.a.r. com $X_n \sim N(m_n, \sigma_n^2)$, onde $m_n \to m$ e $\sigma_n \to \sigma > 0$, mostre que $X_n \xrightarrow{d} N(m, \sigma^2)$.

4. Para $n \in \mathbb{N}$, seja X_n uma v.a. uniformemente distribuída sobre o intervalo $[a_n, b_n]$, onde $a_n \to a$ e $b_n \to b$, com $a \leq b$. Mostre que $X_n \xrightarrow{d} U([a,b])$.

5. Se (X_n) e X são v.a. que tomam valores em $\mathbb{N}_0$, mostre que $X_n \xrightarrow{d} X$ sse $\mathrm{P}(X_n = j) \to \mathrm{P}(X = j)$, para todo o $j \in \mathbb{N}_0$.

6. (**Convergência da binomial para a Poisson**) Sejam $X_n \sim B(n, p_n)$ com $np_n \to \lambda \in]0, +\infty[$, e X v.a. de Poisson de parâmetro λ. Mostre que $X_n \xrightarrow{d} X$.

   (Sugestão: Use o Exercício 2.1.10.)

7. Verifique que o recíproco do teorema de Scheffé não é verdadeiro, mostrando que a sucessão (X_n) de v.a.r. absolutamente contínuas com densidades $f_{X_n}(x) = (1 - \cos(2n\pi x))\, \mathbb{1}_{[0,1]}(x)$, satisfaz $X_n \xrightarrow{d} U([0,1])$, e no entanto f_{X_n} não converge λ-q.t.p. para $\mathbb{1}_{[0,1]}$.

[1] Scheffé, H., *Ann. Math. Statist.* 28, 434–458, 1947.

8. **(Teorema de Scheffé para variáveis discretas)** Sejam (X_n) e X v.a. que tomam valores num conjunto finito ou numerável S. Mostre que se $\mathrm{P}(X_n = j) \to \mathrm{P}(X = j)$, para todo o $j \in S$, então $X_n \xrightarrow{d} X$. Verifique que a recíproca não é em geral verdadeira considerando $X_n = 1/n$ e $X = 0$.

9.3 Relações com os outros modos de convergência

Como decorre da própria definição, quando falamos em convergência em distribuição de X_n para X os vectores aleatórios $X, X_1, X_2, \ldots$ não necessitam de estar definidos num mesmo espaço de probabilidade. No entanto, quando tal acontece a convergência em distribuição pode ser relacionada com os outros tipos de convergência já estudados.

Teorema 9.3.1 *Se $X_n \xrightarrow{p} X$, então $X_n \xrightarrow{d} X$.*

Dem: Consequência da caracterização ii) dada no Teorema 9.2.1 e do teorema da convergência dominada. □

Recordemos que a convergência em probabilidade é a mais fraca das convergências funcionais estudadas. Assim qualquer das convergências $\xrightarrow{qc}$ ou $\xrightarrow{\mathcal{L}^p}$, implica a convergência em distribuição.

No caso particular da variável limite ser degenerada, mostramos a seguir que a convergência em distribuição é equivalente à convergência em probabilidade.

Teorema 9.3.2 *Se $X_n \xrightarrow{d} a$, com $a \in \mathbb{R}^d$, então $X_n \xrightarrow{p} a$.*

Dem: Comecemos por estabelecer o resultado para $d = 1$. Neste caso, se $X = a$, $F_X = 1\!\!1_{[a,+\infty[}$, e assim $\lim F_{X_n}(x) = 0$, se $x < a$, e $\lim F_{X_n}(x) = 1$, se $x > a$. Dado $\epsilon > 0$, temos $\mathrm{P}(|X_n - a| < \epsilon) = \mathrm{P}(a - \epsilon < X_n < a + \epsilon) \geq F_{X_n}(a + \epsilon/2) - F_{X_n}(a - \epsilon) \to 1$. Para $d > 1$, basta ter em conta que se $X_n \xrightarrow{d} a$, então $\pi_i(X_n) \xrightarrow{d} \pi_i(a)$, para $i = 1, \ldots, d$, e pela primeira parte da demonstração obtemos $\pi_i(X_n) \xrightarrow{p} \pi_i(a)$, para $i = 1, \ldots, d$, ou equivalentemente, $X_n \xrightarrow{p} a$. □

9.4 O teorema de Prohorov

O objectivo principal deste parágrafo é a obtenção do teorema de Prohorov sobre a caracterização da compacidade sequencial duma sucessão de vectores aleatórios. Por outras palavras, pretendemos caracterizar as sucessões de vectores aleatórios para as quais toda a sua subsucessão possui uma subsucessão convergente em distribuição.

A importância dum resultado deste tipo será clara quando, no próximo parágrafo, caracterizarmos a convergência em distribuição duma sucessão de vectores aleatórios a partir das respectivas funções características. No entanto, e para já, o resultado seguinte, cuja demonstração deixamos ao cuidado do aluno, indica-nos que a compacidade sequencial duma sucessão de vectores aleatórios é uma propriedade necessária, mas não suficiente, para a sua convergência em distribuição. Ele é consequência do seguinte facto sobre sucessões de números reais: uma sucessão (x_n) converge para $x \in \mathbb{R}$ sse toda a subsucessão de (x_n) admite uma subsucessão que converge para x.

Teorema 9.4.1 *Sejam (X_n) e X vectores aleatórios em $\mathbb{R}^d$. $X_n \xrightarrow{d} X$ sse toda a subsucessão de (X_n) admite uma subsucessão que converge em distribuição para X.*

O teorema de Prohorov estabelece que as sucessões de vectores aleatórios cujas subsucessões admitem uma subsucessão convergente, são precisamente as sucessões limitadas em probabilidade no sentido da definição seguinte.

Definição 9.4.2 *Uma sucessão (X_n) de vectores aleatórios em $\mathbb{R}^d$ diz-se **limitada em probabilidade** se para todo o $\epsilon > 0$, existe $M > 0$ tal que*

$$\mathrm{P}_{X_n}(]-M,M]) = F_{X_n}]-M,M] > 1-\epsilon,\ \forall n \in \mathbb{N}.$$

Notemos mais uma vez a analogia com o caso das sucessões de números reais: uma sucessão (x_n) é limitada sse toda a sua subsucessão admite uma subsucessão convergente. Reparemos também que impor que uma sucessão de vectores aleatórios seja limitada em probabilidade quando estudamos a sua convergência em distribuição não é demasiadamente restritivo, uma vez que (X_n) é limitada em probabilidade sempre que $X_n \xrightarrow{d} X$, para algum vector aleatório X. No entanto, o facto de (X_n) ser limitada em probabilidade não implica só por si a convergência em distribuição da sucessão para algum vector aleatório. Um exemplo disso é o da sucessão $X_n = X$, se n é par, e $X_n = Y$, se n é ímpar, com $X \not\sim Y$.

O teorema da selecção de Helly que estabelecemos a seguir é de importância fundamental na demonstração do teorema de Prohorov. A notação que usamos sobre a função de distribuição dum vector aleatório foi introduzida no Exemplo 1.4.3.

Lema 9.4.3 *Sejam $D_1, \ldots, D_d$ subconjuntos numeráveis e densos em $\mathbb{R}$ e (X_n) uma sucessão de vectores aleatórios tais que $\lim F_{X_n}(y)$ existe para todo o $y \in \prod_{i=1}^d D_i$. Então existe uma função F_∞ não-decrescente, contínua à direita, com $0 \le F_\infty \le 1$, tal que $\lim F_{X_n}(x) = F_\infty(x)$, para todo o $x \in C(F_\infty)$.*

Dem: Para $x \in D = \prod_{i=1}^d D_i$, definamos $F_\infty(x) = \lim F_{X_n}(x)$. Claramente, $0 \le F_\infty(x) \le 1$, para todo o $x \in D$. Para $x \in \mathbb{R}^d \setminus D$, definamos $F_\infty(x) = \inf_{y>x, y\in D} F_\infty(y)$.

Como $\{F_\infty(y) : y > x, y \in D\}$ é limitado em $\mathbb{R}$, o ínfimo anterior é um elemento do intervalo $[0,1]$. Assim, $0 \leq F_\infty \leq 1$, e $F_\infty(x_1) \leq F_\infty(x_2)$, se $x_1 \leq x_2$. i) Verifiquemos que F_∞ é contínua à direita em todo o ponto $x \in \mathbb{R}^d$. Dado $\epsilon > 0$, tomemos $x' > x$ com $x' \in D$ tal que $F_\infty(x) + \epsilon \geq F_\infty(x')$. Dado agora $y \in]x, x']$ temos $F_\infty(y) \leq F_\infty(x')$, e portanto $F_\infty(x) + \epsilon \geq F_\infty(y) \geq \inf_{y>x} F_\infty(y)$. Fazendo tender ϵ para zero, obtemos $F_\infty \geq \inf_{y>x} F_\infty(y)$, ou ainda, $F_\infty = \inf_{y>x} F_\infty(y)$. ii) Verifiquemos que F_∞ é não-decrescente. Se $a, b \in D$ são tais que $a < b$, e sendo V o conjunto dos vértices de $]a,b]$, temos $0 \leq F_{X_n}]a,b] = \sum_{x\in V} \mathrm{sgn}(x) F_{X_n}(x) \to \sum_{x\in V} \mathrm{sgn}(x) F_\infty(x) = F_\infty]a,b]$. Dados agora $a, b \in \mathbb{R}^d$ com $a < b$, tomemos $a_n \geq a$ e $b_n \geq b$, com $a_n, b_n \in D$, $a_n \to a$ e $b_n \to b$. Denotando por V_n o conjunto dos vértices de $]a_n, b_n]$, temos $0 \leq F_\infty]a_n, b_n] = \sum_{x_n \in V_n} \mathrm{sgn}(x_n) F_\infty(x_n) \to \sum_{x_n\in V} \mathrm{sgn}(x) F_\infty(x) = F_\infty]a,b]$. iii) Verifiquemos finalmente que $\lim F_{X_n}(x) = F_\infty(x)$, para todo o $x \in C(F_\infty)$. Sejam então $x \in C(F_\infty)$ e (a_i) e (b_i) em D tais que $a_i \uparrow x$ e $b_i \downarrow x$. Assim, $F_{X_n}(a_i) \leq F_{X_n}(x) \leq F_{X_n}(b_i)$ e $F_\infty(a_i) = \liminf F_{X_n}(a_i) \leq \liminf F_{X_n}(x) \leq \limsup F_{X_n}(x) \leq \limsup F_{X_n}(b_i) = F_\infty(b_i)$. Tomando agora limite em i quando i tende para $+\infty$ e tendo em conta que $x \in C(F_\infty)$, obtemos $F_\infty(x) \leq \liminf F_{X_n}(x) \leq \limsup F_{X_n}(x) \leq F_\infty(x)$, o que prova o pretendido. □

Teorema 9.4.4 (da selecção de Helly ([2])) *Se (X_n) é uma sucessão de vectores aleatórios em $\mathbb{R}^d$, então existem uma subsucessão (X_{n_k}) de (X_n) e uma função $F_\infty : \mathbb{R}^d \to \mathbb{R}$ contínua à direita, não-decrescente com $0 \leq F_\infty \leq 1$, tais que*

$$\lim F_{X_{n_k}}(x) = F_\infty(x),\ \forall x \in C(F_\infty).$$

Dem: Tendo em conta o Teorema 9.4.3, e sendo $D = \mathbb{Q}^d = \{a_i : i \in \mathbb{N}\}$, basta mostrar que existe uma subsucessão (X_{n_k}) para a qual existe o limite $\lim F_{X_{n_k}}(a_i)$, para todo o $i \in \mathbb{N}$. Sendo $(F_{X_n}(a_1))$ limitada, comecemos por tomar uma sua subsucessão $(F_{X_{n(1,k)}}(a_1))$ convergente. De forma análoga seja $(F_{X_{n(2,k)}}(a_2))$ uma subsucessão convergente da sucessão limitada $(F_{X_{n(1,k)}}(a_2))$. As sucessões $(F_{X_{n(2,k)}}(a_1))$ e $(F_{X_{n(2,k)}}(a_2))$ são ambas convergentes. Repetindo este processo, determinamos $(F_{X_{n(i,k)}}(a_i))$ convergente tal que as sucessões $(F_{X_{n(i,k)}}(a_1))$,...,$(F_{X_{n(i,k)}}(a_{i-1}))$ são convergentes. Tomemos então a sucessão diagonal $(F_{X_{n(k,k)}})$. Para cada $i \in \mathbb{N}$, $(F_{X_{n(k,k)}}(a_i))$ é convergente, pois $\{F_{X_{n(k,k)}}(a_i) : k \geq i\} \subset \{F_{X_{n(i,k)}}(a_i) : k \geq i\}$, e $(F_{X_{n(i,k)}}(a_i))$ é convergente. Basta então tomar $n_k = n(k,k)$. □

Sendo a função F_∞, cuja existência é estabelecida no resultado anterior, não-decrescente e contínua à direita, é possível associar-lhe uma e uma só medida μ_∞

[2] Helly, E., *Sitzungsber. Nat. Kais. Akad. Wiss.* 121, 265–297, 1912.

sobre $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ tal que

$$\mu_\infty(]a,b]) = F_\infty]a,b] = \sum_{x\in V} \operatorname{sgn}(x) F_\infty(x),$$

para todo o $a, b \in \mathbb{R}^d$, onde V é o conjunto dos vértices de $]a,b]$ (cf. Billingsley, 1986, pg. 177–180). Sempre que $\mu_\infty(\mathbb{R}^d) = 1$, μ_∞ é uma probabilidade, e nesse caso $X_n \xrightarrow{d} X$, onde X é um qualquer vector aleatório que tenha μ_∞ como distribuição de probabilidade. Caso contrário, temos $\mu_\infty(\mathbb{R}^d) < 1$ não existindo por isso o limite em distribuição da sucessão (X_n) (ver Exercício 9.4.6). Dizemos neste caso que ocorre uma "perda de probabilidade no infinito". Um exemplo simples de tal situação é o da sucessão $X_n = n$.

Teorema 9.4.5 (de Prohorov ([3])) *Seja (X_n) é uma sucessão de vectores aleatórios em $\mathbb{R}^d$. (X_n) é limitada em probabilidade sse toda a subsucessão de (X_n) possui uma subsucessão convergente em distribuição.*

Dem: Suponhamos que (X_n) é limitada em probabilidade,e provemos que toda a sua subsucessão possui uma subsucessão convergente em distribuição. Como qualquer subsucessão duma sucessão limitada em probabilidade é ainda limitada em probabilidade, basta que mostremos que (X_n) possui uma subsucessão convergente em distribuição. Pelo teorema da selecção de Helly, existe uma subsucessão (X_{n_k}) de (X_n) e uma função $F_\infty : \mathbb{R}^d \to \mathbb{R}$ contínua à direita, não-decrescente com $0 \leq F_\infty \leq 1$, tais que $\lim F_{X_{n_k}}(x) = F_\infty(x)$, $\forall x \in C(F_\infty)$. Para concluir basta provar que a medida finita μ_∞ associada a F_∞ é uma probabilidade. Para $\epsilon > 0$, existe $M > 0$ tal que $\mathrm{P}_{X_{n_k}}(]-M, M]) > 1-\epsilon$, $\forall k \in \mathbb{N}$. Tomando agora $a < -M$ e $b > M$ tais que $V \subset C(F_\infty)$ onde V é o conjunto dos vértices do rectângulo $]a,b]$, temos $\mu_\infty(]a,b]) = \sum_{x\in V} \operatorname{sgn}(x) F_\infty(x) = \lim_k \sum_{x\in V} \operatorname{sgn}(x) F_{X_{n_k}}(x) = \lim_k \mathrm{P}_{X_{n_k}}(]a,b]) \geq \lim_k \mathrm{P}_{X_{n_k}}(]-M, M]) \geq 1-\epsilon$. Sendo $\epsilon > 0$ qualquer concluímos que $\mu_\infty(R^d) = 1$. Reciprocamente, suponhamos por absurdo que (X_n) não é limitada em probabilidade. Tendo em conta o Exercício 9.4.4, existem $\epsilon > 0$ e uma sucessão (n_k) de números naturais estritamente crescente tais que $\mathrm{P}_{X_{n_k}}(]-K, K]) \leq 1-\epsilon$, para todo o $k \in \mathbb{N}$, onde $K = (k, \ldots, k)$. Por hipótese, existe $(X_{n_{k'}})$ subsucessão de (X_{n_k}) tal que $X_{n_{k'}} \xrightarrow{d} X$, para algum vector aleatório X em $\mathbb{R}^d$. Para quaisquer $a, b \in \mathbb{R}^d$ tais que $V \subset C(F_X)$, onde V é o conjunto dos vértices do rectângulo $]a,b]$, temos $\mathrm{P}_X(]a,b]) = \sum_{x\in V} \operatorname{sgn}(x) F_X(x) = \lim \sum_{x\in V} \operatorname{sgn}(x) F_{X_{n_{k'}}}(x) = \lim \mathrm{P}_{X_{n_{k'}}}(]a,b]) \leq 1-\epsilon$, o que é falso quando fazemos $\max_i a_i \to -\infty$ e $\min_i b_i \to +\infty$. $\square$

Exercícios

1. Se $X_n = \alpha_n$, com $\alpha_n \in \mathbb{R}$, mostre que (X_n) é limitada em probabilidade sse (α_n) é limitada.

[3] Prohorov, Yu.V., *Theory Probab. Appl.* 1, 157–214, 1956.

2. Mostre que se $X_n \xrightarrow{d} X$ então (X_n) é limitada em probabilidade.

3. Prove que (X_n) é limitada em probabilidade sse cada uma das sucessões coordenadas de (X_n) é limitada em probabilidade.

4. Prove que (X_n) é limitada em probabilidade sse $\lim_{k\to+\infty} \limsup_n \mathrm{P}_{X_n}(]-K,K]^c) = 0$, com $K = (k,\ldots,k)$.

5. Mostre que se (X_n) e (Y_n) são limitadas em probabilidade, então (X_nY_n) é limitada em probabilidade.

6. Sejam (X_n) é uma sucessão de vectores aleatórios em $\mathbb{R}^d$, F_∞ a função cuja existência é assegurada pelo Teorema 9.4.4 e μ_∞ a medida sobre $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ que lhe está associada. Para $i = 1,\ldots,d$, consideremos as funções coordenada
$$F_{\infty,i}(x_i) = \lim_{\substack{x_j\to+\infty\\ j\neq i}} F_\infty(x_1,\ldots,x_{i-1},x_i,x_{i+1},\ldots,x_d).$$
   (a) Conclua que o conjunto E_i dos pontos de descontinuidade de $F_{\infty,i}$ é quando muito numerável.
   (b) Mostre que $E_1^c \times \ldots \times E_d^c \subset C(F_\infty)$.
   (c) Prove que se $\mu_\infty(\mathbb{R}^d) < 1$, então (X_n) não converge em distribuição.

9.5 O teorema da continuidade de Lévy–Bochner

Como veremos neste parágrafo, o teorema de Prohorov permite-nos caracterizar a convergência em distribuição duma sucessão de vectores aleatórios apenas em termos das funções características respectivas. Uma tal caracterização será de grande utilidade no estudo da distribuição assintótica da soma de vectores aleatórios independentes uma vez que, como vimos anteriormente, a função característica é bem mais útil para esse efeito do que a função de distribuição.

Teorema 9.5.1 *Seja (X_n) uma sucessão de vectores aleatórios em $\mathbb{R}^d$.*
a) Se $X_n \xrightarrow{d} X$, então $\phi_{X_n}(t) \to \phi_X(t)$, para todo o $t \in \mathbb{R}^d$.
b) Se (X_n) é limitada em probabilidade e $\phi_{X_n}(t) \to \phi_\infty(t)$, para todo o $t \in \mathbb{R}^d$, então $\phi_\infty = \phi_X$ para algum vector aleatório X em $\mathbb{R}^d$ e $X_n \xrightarrow{d} X$.

Dem: a) Para $t \in \mathbb{R}^d$ fixo, sendo as funções $x \to \sin(\langle t,x\rangle)$ e $x \to \cos(\langle t,x\rangle)$, contínuas e limitadas em $\mathbb{R}^d$, concluímos pelo Teorema 9.2.1 que $\mathrm{E}(\sin(\langle t,X_n\rangle) \to \mathrm{E}(\sin(\langle t,X\rangle)$ e $\mathrm{E}(\cos(\langle t,X_n\rangle) \to \mathrm{E}(\cos(\langle t,X\rangle)$, uma vez que $X_n \xrightarrow{d} X$, ou ainda, $\phi_{X_n}(t) \to \phi_X(t)$.
b) Comecemos por mostrar que $\phi_X = \phi_\infty$. Sendo (X_n) limitada em probabilidade, existe, pelo teorema de Prohorov, uma subsucessão (X_{n_k}) de (X_n) tal que $X_{n_k} \xrightarrow{d} X$, para algum vector aleatório X. Pela alínea a) obtemos $\phi_{X_{n_k}}(t) \to \phi_X(t)$, para todo o $t \in \mathbb{R}^d$, e portanto $\phi_X = \phi_\infty$. Mostremos agora que $X_n \xrightarrow{d} X$. Para tal, consideremos

uma qualquer subsucessão $(X_{n'})$ de (X_n), e provemos que ela admite uma subsucessão convergente para X. Com efeito, sendo $(X_{n'})$ limitada em probabilidade, existe $(X_{n''})$ subsucessão de $(X_{n'})$ com $X_{n''} \to Y$, para algum vector aleatório Y, o que implica que $\phi_{X_{n''}}(t) \to \phi_Y(t)$, para todo o $t \in \mathbb{R}^d$. Assim, $\phi_Y = \phi_\infty = \phi_X$, ou ainda, $X \sim Y$. $\square$

Mostramos agora que a condição de (X_n) ser limitada em probabilidade pode ser substituída por uma hipótese de continuidade na origem da função limite ϕ_∞. Um tal resultado é conhecido como teorema da continuidade de Lévy–Bochner.

Lema 9.5.2 *Se X é uma variável aleatória real, então para todo o $r > 0$,*

$$\mathrm{P}(|X| \geq 2r) \leq r \int_{-1/r}^{1/r} (1 - \phi_X(t)) d\lambda(t).$$

DEM: Para $r > 0$ temos, $\int_{-1/r}^{1/r}(1 - \phi_X(t))d\lambda(t) = \int_{-1/r}^{1/r} \int (1 - e^{\,\mathrm{i}\, tx})\, d\mathrm{P}_X(x) d\lambda(t) = \int \int_{-1/r}^{1/r} (1 - e^{\,\mathrm{i}\, tx})\, d\lambda(t) d\mathrm{P}_X(x) = \int_{\mathbb{R}\setminus\{0\}} 2(1 - \sin(x/r)/(x/r))/r d\mathrm{P}_X(x) \geq \int_{|x|\geq 2r} 1/r d\mathrm{P}_X$ $= \mathrm{P}(|X| \geq 2r)/r$, pois $1 - \sin(x/r)/(x/r) \geq 1/2$, se $|x| \geq 2r$. $\square$

Teorema 9.5.3 (de Lévy ([4])–Bochner ([5])) *Seja (X_n) uma sucessão de vectores aleatórios em $\mathbb{R}^d$. Se $\phi_{X_n}(t) \to \phi_\infty(t)$, para todo o $t \in \mathbb{R}^d$, onde ϕ_∞ é contínua na origem, então $X_n \xrightarrow{d} X$ para algum vector aleatório X em $\mathbb{R}^d$ e $\phi_X = \phi_\infty$.*

DEM: Atendendo ao Teorema 9.5.1, basta demonstrar que se $\phi_{X_n}(t) \to \phi_\infty(t)$, para todo o $t \in \mathbb{R}^d$, onde ϕ_∞ é contínua na origem, então a sucessão (X_n) é limitada em probabilidade. Comecemos por demonstrar tal facto no caso real utilizando o Exercício 9.4.4. Pelo Lema 9.5.2 e para $k > 0$, temos $\mathrm{P}_{X_n}(]-k,k]^c) \leq \mathrm{P}(|X_n| \geq k) \leq (k/2) \int_{-2/k}^{2/k} (1 - \phi_{X_n}(t)) d\lambda(t)$, onde $1 - \phi_{X_n}(t) \to 1 - \phi_\infty(t)$ e $|1 - \phi_{X_n}(t)| \leq 2$. Pelo teorema da convergência dominada obtemos $\limsup \mathrm{P}_{X_n}(]-k,k]^c) \leq (k/2) \int_{-2/k}^{2/k} (1 - \phi_\infty(t))\, d\lambda(t) = \int_{-1}^{1} (1 - \phi_\infty(2t/k))\, d\lambda(t)$. Pela continuidade de ϕ_∞ na origem, uma nova aplicação do teorema da convergência dominada permite finalmente concluir que $\lim_k \limsup_n \mathrm{P}_{X_n}(]-k,k]^c) = 0$. Para estabelecer o resultado no caso multivariado, vamos lançar mão do Exercício 9.4.3. Tendo em conta a primeira parte da demonstração, bastará demonstrar que para cada uma das sucessões coordenadas $(X_{n,i})$ de (X_n), a sucessão das funções características $(\phi_{X_{n,i}})$ converge pontualmente para uma função contínua na origem. Tal é com efeito verdade uma vez que $\phi_{X_{n,i}}(s) = \phi_{X_n}(se_i) \to \phi_\infty(se_i) =: \phi_{\infty,i}(s)$, para $s \in \mathbb{R}$, onde e_i representa o i-ésimo vector da base canónica de $\mathbb{R}^d$, e $\phi_{\infty,i}$ e contínua na origem pela continuidade na origem de ϕ_∞. $\square$

---

[4]Lévy, P., *C. R. Acad. Sci. Paris* 175, 854–856, 1922.

[5]Bochner, S., *Math. Ann.* 108, 378–410, 1933.

Notemos que a continuidade na origem da função limite é essencial para a validade do resultado como o comprova o exemplo da sucessão $X_n \sim U([-n,n])$. Atendendo a que a função característica dum vector aleatório é uma função contínua, concluímos do resultado anterior que o limite ϕ_∞ duma sucessão de funções características é uma função contínua se o for na origem.

Corolário 9.5.4 *$X_n \xrightarrow{d} X$ sse $\phi_{X_n}(t) \to \phi_X(t)$, para todo o $t \in \mathbb{R}^d$.*

Sabemos já que a distribuição dum vector aleatório é caracterizada pelas distribuições de probabilidade das variáveis aleatórias reais $\langle a, X\rangle$, para todo o $a \in \mathbb{R}^d$ (ver Exercício 7.4.2). O resultado seguinte aponta no mesmo sentido relativamente à convergência em distribuição, sendo importante no estudo da convergência em distribuição de sucessões de vectores aleatórios, pois permite fazê-lo a partir da convergência em distribuição de variáveis aleatórias reais.

Teorema 9.5.5 (de Cramér–Wold ([6])) *Sejam (X_n) e X vectores aleatórios em $\mathbb{R}^d$. Então $X_n \xrightarrow{d} X$ sse $\langle a, X_n\rangle \xrightarrow{d} \langle a, X\rangle$, para todo o $a \in \mathbb{R}^d$.*

DEM: Se $X_n \xrightarrow{d} X$, então sendo $g(x) = \langle a, x\rangle$ contínua, para a fixo em $\mathbb{R}^d$, concluímos, pelo Teorema 9.2.1 que $g(X_n) \xrightarrow{d} g(X)$, isto é, $\langle a, X_n\rangle \xrightarrow{d} \langle a, X\rangle$. Reciprocamente, dado $t \in \mathbb{R}^d$, temos $\phi_{X_n}(t) = \phi_{\langle t, X_n\rangle}(1) \to \phi_{\langle t, X\rangle}(1) = \phi_X(t)$, e portanto $X_n \xrightarrow{d} X$. $\square$

Exercícios

1. Sejam (X_n) e X ve.a. normais. Mostre que $X_n \xrightarrow{d} X$ sse $\mathrm{E}(X_n) \to \mathrm{E}(X)$ e $\mathrm{C}_{X_n} \to \mathrm{C}_X$.
2. **(Teorema de Slutsky ([7]))** Sejam (X_n), (Y_n) e X ve.a. em $\mathbb{R}^d$ com $X_n \xrightarrow{d} X$ e $X_n - Y_n \xrightarrow{p} 0$. Prove que $Y_n \xrightarrow{d} X$.
3. Sejam (X_n), (Y_n) e X v.a.r. tais que $X_n \xrightarrow{d} X$ e $Y_n \xrightarrow{p} c$, com $c \in \mathbb{R}$. Prove que: a) $X_n + Y_n \xrightarrow{d} X + c$; b) $Y_n X_n \xrightarrow{d} cX$.
4. **(Método delta)** Sejam $X_1, X_2, \ldots$ ve.a. em $\mathbb{R}^d$ tais que
$$\sqrt{n}\,(X_n - \mu) \xrightarrow{d} N(0, \Sigma),$$
com $\mu \in \mathbb{R}^d$, Σ uma matriz de covariância e $g : \mathbb{R}^d \to \mathbb{R}^p$.
   (a) Se $\lim_{x \to \mu} g(x) = \alpha \in \mathbb{R}^p$, prove que $g(X_n) \xrightarrow{p} \alpha$.
   (b) Se g é diferenciável em μ com derivada $g\prime(\mu)$, mostre que
$$\sqrt{n}\,(g(X_n) - g(\mu)) \xrightarrow{d} N(0, g\prime(\mu)\Sigma\, g\prime(\mu)^T).$$
(Sugestão: Tenha em conta que se g é diferenciável em μ, então para $h \in \mathbb{R}^d$, $g(\mu + h) = g(\mu) + g\prime(\mu)h + r(h)$, onde $\lim_{h\to 0} r(h)/\|h\| = 0$.)

[6] Cramér, H., Wold, H., *J. London Math. Soc.* 11, 290–295, 1936.
[7] Slutsky, E., *Metron* 5, 1–90, 1925.

9.6 Bibliografia

Billingsley, P. (1968). *Convergence of Probability Measures*, Wiley.

Billingsley, P. (1986). *Probability and Measure*, Wiley.

Jacod, J., Protter, P. (2000). *Probability Essentials*, Springer.

Kallenberg, O. (1997). *Foundations of Modern Probability*, Springer.

Resnick, S.I. (1999). *A Probability Path*, Birkhäuser.

Capítulo 10

O teorema do limite central

O teorema do limite central clássico e de Lindeberg. A condição de Liapounov. O teorema do limite central multidimensional.

10.1 Preliminares

Se $X_1, \ldots, X_n, \ldots$ são variáveis aleatórias independentes e identicamente distribuídas com distribuições normais de média μ e variância σ^2, sabemos pela lei fraca dos grandes números que

$$\frac{1}{n} S_n \xrightarrow{p} \mu,$$

onde

$$S_n = X_1 + \ldots + X_n.$$

Sendo a convergência em distribuição implicada pela convergência em probabilidade, a distribuição assintótica de S_n/n é assim degenerada. 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,

$$\frac{S_n/n - \mu}{\sqrt{\sigma^2/n}} \sim N(0,1).$$

Concluímos assim que apesar de S_n/n possuir uma distribuição assintótica degenerada, S_n/n convenientemente normalizada (centragem e redução) possui uma distribuição assintótica não-degenerada:

$$S_n^\star = \frac{S_n - \mathrm{E}(S_n)}{\sqrt{\mathrm{Var}(S_n)}} \xrightarrow{d} N(0,1). \tag{10.1.1}$$

O facto de uma tal distribuição assintótica ser normal, não é, como veremos neste capítulo, uma propriedade exclusiva das variáveis normais. Indícios de tal facto são já nossos conhecidos (ver, por exemplo, o §3.3). Para algumas distribuições de probabilidade já estudadas, apresentamos a seguir, para alguns valores de n, os gráficos da densidade ou da função de probabilidade da variável $S_n^\star$, esta última convenientemente calibrada. A tracejado surge também o gráfico da densidade normal centrada e reduzida.

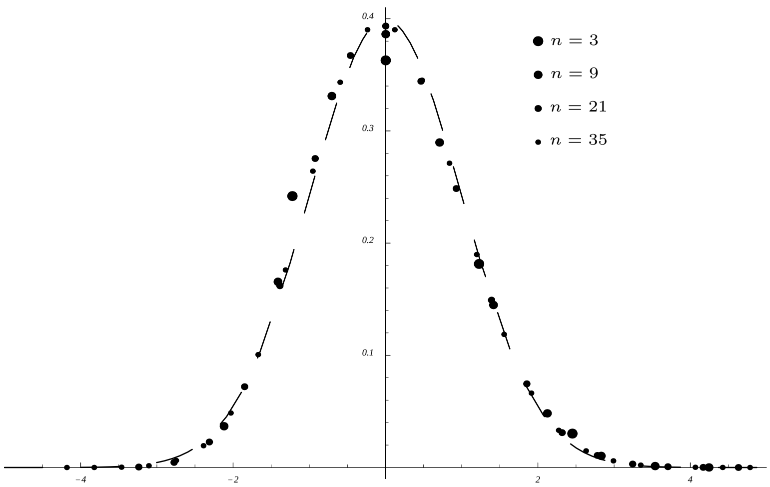


Figura 9.1: Distribuição de $S_n^\star$ quando $X_1, \ldots, X_n \sim B(1/3)$

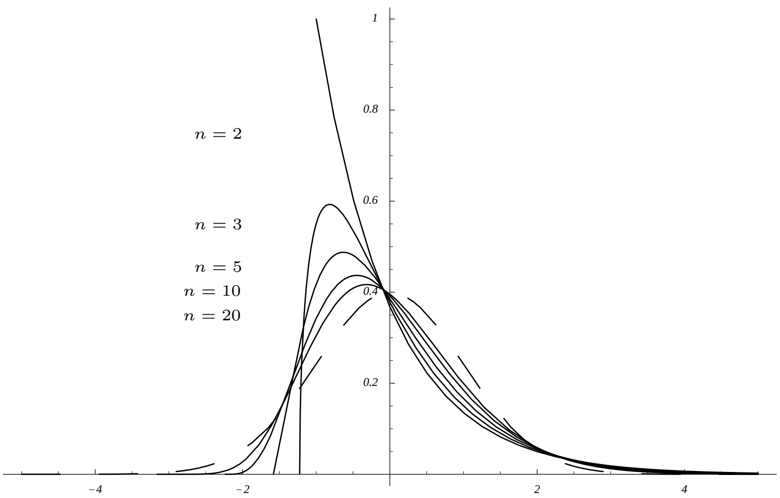


Figura 9.2: Distribuição de $S_n^\star$ quando $X_1, \ldots, X_n \sim \chi_1^2$

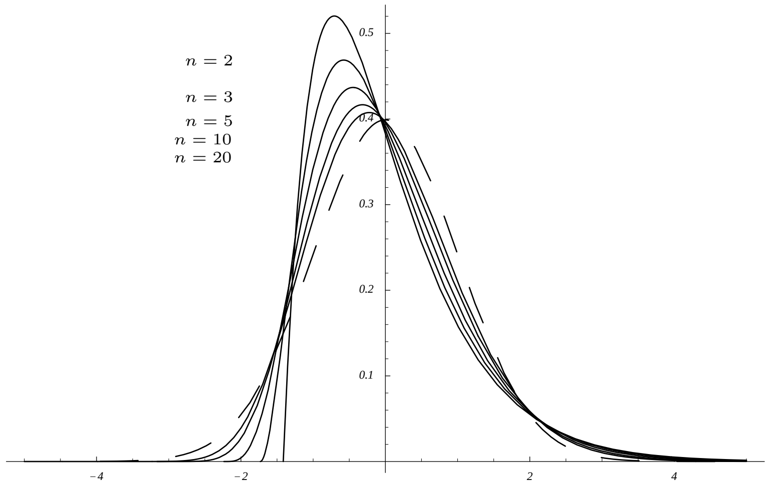


Figura 9.3: Distribuição de $S_n^\star$ quando $X_1, \ldots, X_n \sim E(1)$

No caso das variáveis independentes $X_1, \ldots, X_n$ serem exponenciais de parâmetro $\lambda > 0$, podemos confirmar de forma simples o comportamento sugerido pelos gráficos da Figura 9.3. Para tais variáveis sabemos que $\mathrm{E}(X_k) = 1/\lambda$, $\mathrm{Var}(X_k) = 1/\lambda^2$ e $\phi_{X_k}(t) = \lambda/(\lambda - \mathrm{i}\, t)$, para $t \in \mathbb{R}$. Assim, pela independência das variáveis $X_1, \ldots, X_n$,

$$\begin{aligned} \phi_{S_n^\star}(t) &= e^{-\mathrm{i}\, t\sqrt{n}}\, \phi_{S_n}(t\lambda/\sqrt{n}) \\ &= e^{-\mathrm{i}\, t\sqrt{n}} \left(\frac{1}{1 - \mathrm{i}\, t/\sqrt{n}} \right)^n \\ &= \left(1 + \frac{x_n(t)}{n} \right)^n, \end{aligned}$$

onde

$$\begin{aligned} x_n(t) &= n\left(e^{-\mathrm{i}\, t/\sqrt{n}} - \left(1 - \frac{\mathrm{i}\, t}{\sqrt{n}}\right)\right) \Big/ \left(1 - \frac{\mathrm{i}\, t}{\sqrt{n}}\right) \\ &= n\left(1 - \frac{\mathrm{i}\, t}{\sqrt{n}} - \frac{t^2}{2n} + \ldots - \left(1 - \frac{\mathrm{i}\, t}{\sqrt{n}}\right)\right) \Big/ \left(1 - \frac{\mathrm{i}\, t}{\sqrt{n}}\right) \\ &\to -\frac{t^2}{2}. \end{aligned}$$

Concluímos assim que

$$\phi_{S_n^\star}(t) \to e^{-t^2/2} = \phi_{N(0,1)}(t),$$

para todo o $t \in \mathbb{R}$ (note que se $x_n \to x$ então $(1 + x_n/n)^n \to e^x$), o que, pelo teorema de Lévy–Bochner, permite concluir que

$$S_n^\star \xrightarrow{d} N(0,1).$$

Nos próximos parágrafos mostraremos que a convergência em distribuição (10.1.1) ocorre para uma vasta família de variáveis aleatórias. Um resultado deste tipo é conhecido como **teorema do limite central** ou **teorema central do limite**, designação esta devida a G. Pólya (1887–1985) ([1]), onde a palavra "central" realça a importância que um tal resultado teve na investigação em probabilidades até meados do século XX.

Exercícios

1. Sejam $X_1, X_2, \ldots$ variáveis i.i.d. com $\mathrm{P}(X_i = \pm 1) = 1/2$. Mostre que $S_n/\sqrt{n} \xrightarrow{d} N(0,1)$. Suponha agora que, partindo dum ponto inicial, uma partícula se desloca uma unidade para a esquerda ou para a direita com probabilidade 0.5, em cada segundo. Dê uma aproximação para a probabilidade de ao fim de uma hora a partícula se encontrar a uma distância superior a 200 unidades do ponto inicial.

2. Sejam $X_1, X_2, \ldots$ variáveis aleatórias independentes com distribuições de Poisson de parâmetro $\lambda > 0$. Prove que $(S_n - n\lambda)/\sqrt{n\lambda} \xrightarrow{d} N(0,1)$.

3. Sejam (Y_n) uma sucessão de v.a.r. e (a_n) uma sucessão de números reais tais que $a_n(Y_n - \mu) \xrightarrow{d} Y$, com $\mu \in \mathbb{R}$ e Y uma v.a.r.. Mostre que $b_n(Y_n - \mu) \xrightarrow{p} 0$, para toda a sucessão de números reais (b_n) com $b_n/a_n \to 0$.

4. Seja (X_n) uma sucessão de v.a.r. de quadrado integrável satisfazendo (10.1.1). Mostre que se $n/\sqrt{\mathrm{Var}(S_n)} \to +\infty$, então (X_n) obedece a uma lei fraca dos grandes números com $\mu_n = \sum_{i=1}^n \mathrm{E}(X_i)/n$.

5. Seja (X_n) uma sucessão de v.a.r. i.i.d. de quadrado integrável com média μ satisfazendo (10.1.1). Mostre que $b_n(S_n/n - \mu) \xrightarrow{p} 0$, para toda a sucessão de números reais (b_n) com $b_n/n^{1/2} \to 0$ (ver Exercício 6.2.4), mas que $n^{1/2}(S_n/n - \mu) \not\xrightarrow{p} 0$.

10.2 O teorema do limite central clássico

Neste parágrafo estabelecemos a convergência em distribuição (10.1.1) para variáveis aleatórias independentes e identicamente distribuídas de quadrado integrável.

Para que possamos generalizar os argumentos utilizados no parágrafo anterior a outras distribuições, é essencial o resultado seguinte que não é mais do que um desenvolvimento de Taylor duma função característica em que o resto é apresentado numa forma que nos será útil.

Lema 10.2.1 *Se* $\mathrm{E}|X|^n < +\infty$*, para algum* $n \in \mathbb{N}$*, então para todo o* $t \in \mathbb{R}$*,*

$$\phi_X(t) = \sum_{k=0}^{n} \frac{(\mathrm{i}\,t)^k}{k!}\,\mathrm{E}(X^k) + u_n(t),$$

[1] Pólya, G., *Math. Z.* 8, 171–180, 1920.

onde

$$|u_n(t)| \leq \mathrm{E}\Big(\frac{|tX|^{n+1}}{(n+1)!} \wedge \frac{2|tX|^n}{n!}\Big).$$

Dem: Para $n \geq 0$ vale a igualdade

$$\int_0^x (x-s)^n e^{\mathrm{i}\, s} ds = \frac{x^{n+1}}{n+1} + \frac{\mathrm{i}}{n+1}\int_0^x (x-s)^{n+1} e^{\mathrm{i}\, s} ds.$$

Por indução podemos então obter

$$e^{\mathrm{i}\, x} = \sum_{k=0}^{n} \frac{(\mathrm{i}\, x)^k}{k!} + \frac{\mathrm{i}^{\,n+1}}{n!}\int_0^x (x-s)^n e^{\mathrm{i}\, s} ds,$$

para $n \geq 1$. Por um lado, a última parcela do segundo membro da igualdade anterior é, em módulo, majorada por $\int_0^x |x-s|^n ds/n! \leq |x|^{n+1}/(n+1)!$. Por outro lado, e atendendo à primeira das igualdades anteriores, é majorada por $|\int_0^x (x-s)^{n-1} e^{\mathrm{i}\, s} ds - x^n/n|/(n-1)! \leq 2|x|^n/n!$. Assim, integrando ambos os membros da segunda igualdade depois de tomar $x = tX$, obtemos o pretendido. □

Teorema 10.2.2 (do limite central clássico (2)**)** *Sejam* (X_n) *variáveis aleatórias independentes e identicamente distribuídas de quadrado integrável, com* $\mathrm{E}(X_1) = \mu$ *e* $\mathrm{Var}(X_1) = \sigma^2 > 0$. *Então*

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} N(0,1).$$

Dem: Basta considerar o caso em que $\mu = 0$ e $\sigma = 1$. Denotemos por ϕ_n a função característica de $S_n/\sqrt{n}$ e por ϕ a função característica de X_1. Para $t \in \mathbb{R}$, temos $\phi_n(t) = \phi_{S_n}(t/\sqrt{n}) = \phi^n(t/\sqrt{n})$, onde pelo Lema 10.2.1, $\phi(t/\sqrt{n}) = 1 + \mathrm{i}\, t\mathrm{E}(X_1)/\sqrt{n} + \mathrm{i}^{\,2} t^2 \mathrm{E}(X_1)^2/(2n) + v_n(t) = 1 - t^2/(2n) + v_n(t)$, com $n|v_n(t)| \leq \mathrm{E}(|tX_1|^3/(6n^{1/2}) \wedge |tX_1|^2) \to 0$ (porquê?). Assim, $\phi_n(t) = (1 + (-t^2/2 + nv_n(t))/n)^n \to e^{-t^2/2} = \phi_{N(0,1)}(t)$, o que permite concluir. □

Reescrevendo a variável aleatória $(S_n - n\mu)/\sqrt{n}$ na forma $\sqrt{n}\,(S_n/n - \mu)$, o teorema anterior estabelece que $\sqrt{n}\,(S_n/n - \mu) \xrightarrow{d} N(0, \sigma^2)$. Em particular $S_n/n \xrightarrow{p} \mu$ (cf. Exercício 10.2.5), isto é, o teorema do limite central clássico implica a lei fraca dos grandes números. Além disso, estabelecendo a forma da distribuição assintótica de S_n, o teorema do limite central dá-nos uma informação mais precisa sobre o comportamento assintótico de S_n do que a lei fraca dos grandes números.

[2] Laplace, P.S., *Mém. Acad. Sci. Paris* 10, 353–415 e 559–565, 1810 (reproduzidos em *Oeuvres de Laplace* 12, 301–345 e 349–353); Liapounov, A., *Bull. Acad. Sci. St. Petersbourg* 13, 359–386, 1900, e *Mem. Acad. Sci. St. Petersbourg* 12, 1–24, 1901.

Exercícios

1. **(Convergência da binomial para a normal** ([3])**)** Para $n \in \mathbb{N}$, Seja Y_n uma v.a. binomial de parâmetros (n, p) com $0 < p < 1$. Mostre que

$$\frac{Y_n - np}{\sqrt{np(1-p)}} \xrightarrow{d} N(0,1).$$

Determine $K \in \mathbb{N}$, de modo que a probabilidade de em 1000 lançamentos duma moeda equilibrada obter entre $500-K$ e $500+K$ caras, seja aproximadamente 0.99. Se em 1000 lançamento duma moeda forem observadas 455 caras, poderemos considerar essa moeda equilibrada?

2. Retome os Exercícios 1.8.4 e 2.1.6. Mostre que

$$\sqrt{n}\,(S_n/n + 1/37) \xrightarrow{d} N(0, \sigma^2),$$

onde $\sigma^2 = (37^2 - 1)/37^2$. Obtenha uma aproximação para $\mathrm{P}(S_n \geq 0)$, quando $n = 200, 1000$ e 2000. Compare os resultados com os obtidos nos exercícios referidos.

3. **(Convergência do χ^2 para a normal)** Se Y_n é uma variável com uma distribuição do qui-quadrado com n graus de liberdade, mostre que $(Y_n - n)/\sqrt{2n} \xrightarrow{d} N(0,1)$.

4. Sejam (X_n) uma sucessão de v.a.r. i.i.d. com momentos finitos de quarta ordem, $\mu = \mathrm{E}(X_1)$, $\sigma^2 = \mathrm{Var}(X_1)$ e $\tau = \mathrm{E}(X_1 - \mu)^4$.
   (a) Mostre que $\sqrt{n}(\frac{1}{n}\sum_{i=1}^n (X_i - \mu)^2 - \sigma^2) \xrightarrow{d} N(0, \tau - \sigma^4)$.
   (b) Conclua que $\sqrt{n}(\widehat{\sigma}_n^2 - \sigma^2) \xrightarrow{d} N(0, \tau - \sigma^4)$, onde $\widehat{\sigma}_n^2$ é a variância empírica das variáveis $X_1, \ldots, X_n$ (ver Exercício 6.5.2).

5. Utilizando a técnica das funções características demonstre a lei fraca dos grande números de Khintchine (ver Teorema 6.3.3).

10.3 O teorema do limite central de Lindeberg

Vamos neste parágrafo generalizar o Teorema 10.2.2 ao caso em que as variáveis aleatórias $X_1, X_2, \ldots$, apesar de independentes e de quadrado integrável não são necessariamente identicamente distribuídas. Denotaremos $\mu_k = \mathrm{E}(X_k)$, $\sigma_k^2 = \mathrm{Var}(X_k)$ e $s_n^2 = \mathrm{Var}(S_n) = \sigma_1^2 + \ldots + \sigma_n^2$.

Definição 10.3.1 *Dizemos que a sucessão (X_n) de variáveis aleatórias independentes e de quadrado integrável satisfaz a **condição de Lindeberg** ([4]) se*

$$\forall\, \epsilon > 0 \quad \frac{1}{s_n^2}\sum_{k=1}^n \mathrm{E}((X_k - \mu_k)^2 \mathbb{1}_{\{|X_k - \mu_k| > \epsilon s_n\}}) \to 0.$$

[3] de Moivre, A., *Approximatio ad Summam Terminorum Binomii* $(a+b)^n$ *in Seriem Expansi*, 1733, e de Moivre, A., *The Doctrine of Chances*, 1738.

[4] Lindeberg, J.W., *Math. Z.* 15, 211–225, 1922.

Comecemos por notar que uma sucessão de variáveis aleatórias independentes e identicamente distribuídas de quadrado integrável satisfaz a condição de Lindeberg. Para $\epsilon > 0$, e pelo teorema da convergência dominada, temos

$$\begin{aligned}
&\frac{1}{s_n^2}\sum_{k=1}^{n} \mathrm{E}((X_k-\mu_k)^2 \mathbb{1}_{\{|X_k-\mu_k|>\epsilon s_n\}}) \\
&\qquad = \frac{1}{n\sigma^2}\sum_{k=1}^{n} \mathrm{E}((X_k-\mu_k)^2 \mathbb{1}_{\{|X_k-\mu_k|>\epsilon\sigma\sqrt{n}\}}) \\
&\qquad = \frac{1}{\sigma^2}\mathrm{E}((X_1-\mu_1)^2 \mathbb{1}_{\{|X_1-\mu_1|>\epsilon\sigma\sqrt{n}\}}) \to 0.
\end{aligned}$$

A condição de Lindeberg impõe que para cada k, a variável aleatória X_k deve estar concentrada num intervalo centrado na sua média e cuja amplitude deve ser pequena quando comparada com s_n. A proposição seguinte dá ênfase a esta interpretação, expremindo-a em termos de variâncias.

Proposição 10.3.2 *Se (X_n) satisfaz a condição de Lindeberg então*

$$\frac{\bigvee_{k=1}^{n}\sigma_k^2}{s_n^2} \to 0.$$

DEM: Para $\epsilon > 0$, basta notar que $\sigma_k^2/s_n^2 = \mathrm{E}((X_k-\mu_k)^2 \mathbb{1}_{\{|X_k-\mu_k|\le\epsilon s_n\}})/s_n^2 + \mathrm{E}((X_k-\mu_k)^2 \mathbb{1}_{\{|X_k-\mu_k|>\epsilon s_n\}})/s_n^2 \le \epsilon^2 + \sum_{k=1}^{n}\mathrm{E}((X_k-\mu_k)^2 \mathbb{1}_{\{|X_k-\mu_k|>\epsilon s_n\}})/s_n^2$. $\square$

Para que possamos generalizar os argumentos utilizados na demonstração do teorema de limite central clássico a variáveis aleatórias não são necessariamente identicamente distribuídas é importante e lema seguinte sobre a comparação de produtos de números complexos.

Lema 10.3.3 *Para $n \in \mathbb{N}$, sejam $a_1,\ldots,a_n,b_1,\ldots,b_n$ números complexos em módulo inferiores ou iguais a 1. Então*

$$\Big|\prod_{i=1}^{n} a_i - \prod_{i=1}^{n} b_i\Big| \le \sum_{i=1}^{n}|a_i - b_i|.$$

DEM: Basta ter em conta que o resultado é válido para $n = 2$ e que $|\prod_{i=1}^{n} a_i - \prod_{i=1}^{n} b_i| = |a_1\prod_{i=2}^{n} a_i - b_1\prod_{i=2}^{n} b_i| \le |a_1 - b_1| + |\prod_{i=2}^{n} a_i - \prod_{i=2}^{n} b_i|$. $\square$

Teorema 10.3.4 (de Lindeberg) *Sejam (X_n) variáveis aleatórias reais independentes e de quadrado integrável com $\mathrm{Var}(X_n) > 0$ para n suficientemente grande. Se (X_n) satisfaz a condição de Lindeberg, então*

$$\frac{S_n - \mathrm{E}(S_n)}{s_n} \xrightarrow{d} N(0,1).$$

Dem: Basta demonstrar o resultado para variáveis centradas. Sendo ϕ_k a função característica de X_k, pela independência das variáveis $X_1, \ldots, X_n$, obtemos, $\phi_{S_n/s_n}(t) = \prod_{k=1}^n \phi_{X_k/s_n}(t) = \prod_{k=1}^n \phi_k(t/s_n)$, para $t \in \mathbb{R}$. Com o objectivo de mostrar que $\prod_{k=1}^n \phi_k(t/s_n) \to e^{-t^2/2}$, para todo o $t \in \mathbb{R}$, provaremos que $A_n = |\prod_{k=1}^n \phi_k(t/s_n) - \exp(\sum_{k=1}^n(\phi_k(t/s_n) - 1))| \to 0$ e que $B_n = |\sum_{k=1}^n(\phi_k(t/s_n) - 1)) + t^2/2| \to 0$. Pelo Lema 10.3.3, $A_n \leq \sum_{k=1}^n |\phi_k(t/s_n) - \exp(\phi_k(t/s_n) - 1))| = \sum_{k=1}^n |\exp(\phi_k(t/s_n) - 1) - 1 - (\phi_k(t/s_n) - 1)|$, uma vez que $|\exp(z-1)| \leq 1$, quando $|z| \leq 1$. Pelo Lema 10.2.1 e pela Proposição 10.3.2, obtemos ainda $|\phi_k(t/s_n) - 1| \leq \mathrm{E}(|tX_k|^2/(2s_n) \wedge 2|tX_k|/s_n) \leq t^2\mathrm{E}(X_k^2)/(2s_n^2) \leq (t^2/2) \vee_{k=1}^n \sigma_k^2/s_n^2 \to 0$. Assim, e tendo agora em conta que $|\exp(z) - 1 - z| \leq 2|z|^2$, quando $|z| \leq 1/2$, obtemos finalmente, $A_n \leq \sum_{k=1}^n 2|\phi_k(t/s_n) - 1|^2 \leq \sum_{k=1}^n 2|\phi_k(t/s_n) - 1|(t^2/2) \vee_{k=1}^n \sigma_k^2/s_n^2 \leq t^2(\vee_{k=1}^n \sigma_k^2/s_n^2) \sum_{k=1}^n (t^2/2)\sigma_k^2/s_n^2 = (t^4/2) \vee_{k=1}^n \sigma_k^2/s_n^2 \to 0$. Pelo Lema 10.2.1 temos agora, para $\epsilon > 0$, $B_n = \sum_{k=1}^n \mathrm{E}(|t|^3|X_k|^3/(6s_n^3) \wedge t^2X_k^2/s_n^2) \leq |t|^2 \sum_{k=1}^n \mathrm{E}(X_k^2 \mathbb{I}_{\{|X_k|>\epsilon s_n\}})/s_n^2 + |t|^3 \sum_{k=1}^n \mathrm{E}(|X_k|^3 \mathbb{I}_{\{|X_k| \leq \epsilon s_n\}})/(6s_n^3) \leq |t|^2 \sum_{k=1}^n \mathrm{E}(X_k^2 \mathbb{I}_{\{|X_k|>\epsilon s_n\}})/s_n^2 + |t|^3\epsilon/6$. Sendo $\epsilon > 0$ qualquer, a condição de Lindeberg permite agora concluir. □

Em 1935, William Feller (1906–1970) ([5]) e Paul Lévy (1886-1971) ([6]), trabalhando independentemente, estabelecem condições necessárias para a validade do teorema do limite central mostrando que, na presença da condição apresentada na Proposição 10.3.2, a condição de Lindeberg é também necessária para que se tenha $\frac{S_n - \mathrm{E}(S_n)}{s_n} \xrightarrow{d} N(0,1)$ (ver Feller, 1971, pg. 518–521; sobre a prioridade da descoberta ver Le Cam, 1986.).

A condição que a seguir apresentamos, apesar de mais restrictiva que a condição de Lindeberg, é normalmente simples de utilizar, em particular para $\delta = 1$.

Proposição 10.3.5 *Se (X_n) é uma sucessão de variáveis aleatórias reais independentes que, para algum $\delta > 0$, satisfaz a condição*

$$\frac{1}{s_n^{2+\delta}} \sum_{k=1}^n \mathrm{E}|X_k - \mu_k|^{2+\delta} \to 0,$$

dita de **condição de Liapounov** ([7]) *então (X_n) satisfaz a condição de Lindeberg.*

Exercícios

1. Mostre que $\bigvee_{k=1}^n \sigma_k^2/s_n^2 \to 0$ sse $s_n^2 \to \infty$ e $\sigma_n^2/s_n^2 \to 0$.

[5] Feller, W., *Math. Z.* 40, 521–559, 1935.

[6] Lévy, P., *J. Math. Pures Appli.* 14, 347–402, 1935.

[7] Liapounov, A., *Bull. Acad. Sci. St. Petersbourg* 13, 359–386, 1900, e *Mem. Acad. Sci. St. Petersbourg* 12, 1–24, 1901.

2. Demonstre a Proposição 10.3.5.

3. Sejam $X_1, X_2, \ldots$ v.a. independentes com $X_n \sim U([-n,n])$. Mostre que $S_n/\mathrm{Var}(S_n) \xrightarrow{d} N(0,1)$, onde

   (Sugestão: Use o facto de $\frac{1}{n^{\lambda+1}} \sum_{k=1}^n k^\lambda \to \frac{1}{\lambda+1}$.)

10.4 O teorema do limite central multidimensional

Neste parágrafo obtemos, via teorema de Cramér–Wold, versões multivariadas dos teoremas do limite central clássico e de Lindeberg.

Teorema 10.4.1 *Se (X_n) é uma sucessão de vectores aleatórios independentes e identicamente distribuídos de quadrado integrável com média μ e matriz de covariância Σ, então*

$$\frac{S_n - \mathrm{E}(S_n)}{\sqrt{n}} \xrightarrow{d} N(0, \Sigma).$$

Dem: Sem perda de generalidade supomos que os vectores X_k são centrados. Pelo Teorema 9.5.5, basta mostrar que, para todo o $a \in \mathbb{R}^d$, $\langle a, S_n/\sqrt{n}\rangle \xrightarrow{d} \langle a, X\rangle$, onde $X \sim N(0, \Sigma)$, ou de forma equivalente, $\langle a, S_n/\sqrt{n}\rangle \xrightarrow{d} N(0, a^T\Sigma a)$. Ora, $\langle a, S_n/\sqrt{n}\rangle = \sum_{k=1}^n \langle a, X_k\rangle/\sqrt{n}$, onde $\langle a, X_k\rangle$, $k = 1, 2, \ldots$, são variáveis reais independentes com média 0 e variância $a^T\Sigma a$. Se $a^T\Sigma a > 0$, o resultado é assim consequência do Teorema 10.2.2. Se $a^T\Sigma a = 0$, $\langle a, X_k\rangle = 0$, *q.c.*, para $k = 1, 2, \ldots$, e $\langle a, S_n/\sqrt{n}\rangle \sim N(0,0) = N(0, a^T\Sigma a)$. □

Teorema 10.4.2 *Seja (X_n) uma sucessão de vectores aleatórios independentes de quadrado integrável com médias μ_n e matrizes de covariância Σ_n. Se*

$$\frac{1}{n}(\Sigma_1 + \ldots + \Sigma_n) \to \Sigma,$$

e

$$\forall\, \epsilon > 0 \quad \frac{1}{n}\sum_{k=1}^n \mathrm{E}(\|X_k - \mu_k\|^2 \mathbb{I}_{\{\|X_k - \mu_k\| > \epsilon\sqrt{n}\}}) \to 0,$$

então

$$\frac{S_n - \mathrm{E}(S_n)}{\sqrt{n}} \xrightarrow{d} N(0, \Sigma).$$

Exercícios

1. Demonstre o Teorema 10.4.2.

2. Para $n \in \mathbb{N}$, seja $X_n \sim M(n, p_1, \ldots, p_k)$ com $\sum_{i=1}^k p_i = 1$. Mostre que $(X_n - \mathrm{E}(X_n))/\sqrt{n}$ é assintoticamente normal.

3. Seja (X_n) uma sucessão de ve.a. i.i.d. com momentos de ordem $2k$, para $k \in \mathbb{N}$ fixo.
   (a) Estabeleça a normalidade assintótica do vector dos k primeiros momentos empíricos $(\sum_{i=1}^{n} X_i^\ell / n; \ell = 1, \dots, k)$.
   (b) Usando o Exercício 9.5.4 e a normalidade assintótica estabelecida na alínea anterior, resolva novamente a alínea (b) do Exercício 10.3.4.

10.5 Bibliografia

Araújo, A., Giné, E. (1980). *The Central Limit Theorem for Real and Banach Valued Random Variables*, Wiley.

Feller, W. (1971). *An Introduction to Probability Theory and its Applications*, Vol. 2, Wiley.

James, B.R. (1981). *Probabilidades: um curso de nível intermediário*, IMPA.

Le Cam, L. (1986). The central limit theorem around 1935, *Statistical Science*, 1, 78–96.

Resnick, S.I. (1999). *A Probability Path*, Birkhäuser.

Tabela 1

Valores da função de distribuição normal standard

Tabela 1: Valores da função de distribuição normal

$$x \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt$$

	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

	0,00	0,02	0,04	0,06	0,08
3,0	0,998650	0,998736	0,998817	0,998893	0,998965
3,1	0,999032	0,999096	0,999155	0,999211	0,999264
3,2	0,999313	0,999359	0,999402	0,999443	0,999481
3,3	0,999517	0,999550	0,999581	0,999610	0,999638
3,4	0,999663	0,999687	0,999709	0,999730	0,999749
3,5	0,999767	0,999784	0,999800	0,999815	0,999828
3,6	0,999841	0,999853	0,999864	0,999874	0,999883
3,7	0,999892	0,999900	0,999908	0,999915	0,999922
3,8	0,999928	0,999933	0,999938	0,999943	0,999948
3,9	0,999952	0,999956	0,999959	0,999963	0,999966
4,0	0,999968	0,999971	0,999973	0,999975	0,999977

Bibliografia Geral

Sobre Teoria das Probabilidades:

Billingsley, P. (1986). *Probability and Measure*, Wiley.

Chow, Y.S., Teicher, H. (1997). *Probability Theory: Independence, Interchangeability, Martingales*, Springer.

Chung, K.L. (1974). *A Course in Probability Theory*, Academic Press.

Durrett, R. (1996). *Probability: Theory and Examples*, Duxbury Press.

Feller, W. (1971). *An Introduction to Probability Theory and its Applications*, Vol. 2, Wiley.

Hennequin, P.L., Tortrat, A. (1965). *Théorie des Probabilités et Quelques Applications*, Masson.

James, B.R. (1981). *Probabilidades: um curso de nível intermediário*, IMPA.

Jacod, J., Protter, P. (2000). *Probability Essentials*, Springer.

Kallenberg, O. (1997). *Foundations of Modern Probability*, Springer.

Kolmogorov, A.N. (1950). *Foundations of the Theory of Probability*, Chelsea Publishing Company (tradução do original *Grundbegriffe der Wahrscheinlichkeitrechnung* datado de 1933).

Laha, R.G., Rohatgi, V.K. (1979). *Probability Theory*, Wiley.

Loève, M. (1977). *Probability Theory I*, Springer.

Métivier, M. (1972). *Notions Fondamentales de la Théorie des Probabilités*, Dunod.

Monfort, A. (1980). *Cours de Probabilités*, Economica.

Resnick, S.I. (1999). *A Probability Path*, Birkhäuser.

Sobre alguns temas específicos:

Araújo, A., Giné, E. (1980). *The Central Limit Theorem for Real and Banach Valued Random Variables*, Wiley.

Billingsley, P. (1968). *Convergence of Probability Measures*, Wiley.

Gnedenko, B.V., Kolmogorov, A.N. (1968). *Limit Distributions for Sums of Independent Random Variables*, Addison-Wesley.

Le Cam, L. (1986). The central limit theorem around 1935, *Statistical Science*, 1, 78–96.

Lukacs, E. (1964). *Fonctions Caractéristiques*, Dunod.

Lukacs, E. (1975). *Stochastic Convergence*, Academic Press.

Révész, P. (1968). *The Laws of Large Numbers*, Academic Press.

Williams, D. (1991). *Probability with Martingales*, Cambridge University Press.

Sobre Teoria da Medida e Integração:

Cohn, D.L. (1980). *Measure Theory*, Birkhäuser.

Fernandez, P.J. (1976). *Medida de Integração*, IMPA.

Halmos, P.R. (1950). *Measure Theory*, D. Van Nostrand Company.

Rudin, W. (1974). *Real and Complex Analysis*, McGraw-Hill.

Sobre a história das Probabilidades (e não só):

Borel, E. (1950). *Éléments de la Théorie des Probabilités*, Éditions Albin Michel.

Hald, A. (1990). *A History of Probability and Statistics and their applications before 1750*, Wiley.

Hald, A. (1998). *A History of Mathematical Statistics from 1759 to 1930*, Wiley.

Sobre simulação de experiências aleatórias:

Grycko, E., Pohl, C., Steinert, F. (1998). *Experimental Stochastics*, Springer.

Knuth, D.E. (1981). *The Art of Computer Programming*, vol. II, Addison-Wesley.

Tompson, J.R. (2000). *Simulation: a Modeler's Approach*, Wiley.

Índice Remissivo