CAPíTULO 13

Quantificar o acaso

Ana Cristina Moreira Freitas
Centro de Matemática & Faculdade de Economia da Universidade do Porto
Rua Dr. Roberto Frias
4200-464 Porto
Portugal
e-mail: amoreira@fep.up.pt

Jorge Milhazes Freitas
Centro de Matemática da Universidade do Porto
Rua do Campo Alegre 687
4169-007 Porto
Portugal
e-mail: amoreira@fep.up.pt
url: http://www.fc.up.pt/pessoas/jmfreita

13.1. Introdução

Neste capítulo abordaremos o Teorema limite de cadeias de Markov e o Teorema de recorrência de Poincaré, assim como algumas consequências dos mesmos. Apesar de terem uma formulação simples, o contexto em se inserem é bastante abstracto, o que por vezes pode toldar a sua melhor compreensão. Isto é particularmente verdade no caso da aplicação do Teorema de recorrência de Poincaré que será apresentada, e que criou alguma celeuma por parecer desafiar a Segunda Lei da Termodinâmica. Como muitas vezes acontece nestas ocasiões, a origem do mal-entendido reside numa compreensão incompleta do resultado. Foi nosso objectivo apresentar o contexto, introduzindo as várias noções e definições que invariavelmente envolvem um grau de abstracção elevado, fugindo à tentação de evitar os conceitos mais sofisticados ou de apresentá-los simplesmente na sua forma final, elegante e depurada. Assim, tanto quanto nos foi possível, fizemos acompanhar a nudeza pura das definições e conceitos, com comentários e exemplos que ilustrassem a sua necessidade e/ou propósito.

No caminho (longo) que conduziu ao enquadramento dos dois teoremas acima mencionados, fizemos uma introdução à Teoria da Probabilidade em que lidamos com conceitos como σ-álgebra, medida de probabilidade, mensurabilidade, variável aleatória, independência. Fizemo-lo não só pela vontade de tornar o texto o mais auto-contido possível, mas também porque a nossa experiência como docentes revela a dificuldade que existe em apreender estes conceitos. De facto e a título de exemplo, observamos com frequência a dificuldade que os alunos têm em responder a questões simples como: "Conhece algum acontecimento de probabilidade zero que não seja impossível?", "O que é uma variável aleatória?" ou "Para que servem as variáveis aleatórias?", apesar destas serem usadas copiosamente nos cursos de Probabilidade e Estatística. Tivemos então a intenção de abordar estes conceitos numa perspectiva teórica mas motivada com explicações que apelassem à intuição ou que elucidassem o seu objectivo.

Na Secção 13.2 lembramos alguns conceitos preliminares de foro genérico que serão usados posteriormente. Na Secção 13.3 faremos uma introdução à T eoria da Probabilidade e apresentaremos as definições de σ-álgebra, medida de probabilidade, probabilidade condicional, independência, variável aleatória, entre outras. Como referência e fonte de leitura complementar para esta secção, aconselhamos os livros (Feller, 1950), (Kingman e Taylor, 1966), (Gonçalves e Lopes, 2000).Na Secção 13.4 faremos uma pequena exposição sobre cadeias de Markov, o seu Teorema limite e aplicações. A bibliografia recomendada, neste caso, são os livros (Anton e Rorres, 2005) e (Karlin, 1969) Finalmente, na Secção 13.5, abordaremos o Teorema de recorrência de Poincaré e uma aplicação sua à Termodinâmica. Apontamos como referências bibliográficas os livros (Petersen, 1989) e (Walters, 1982).

13.2. Noções introdutórias

13.2.1. Convergência de séries numéricas. Dada uma sucessão de números reais $\{a_n\}_{n\in\mathbb{N}}$, a sucessão $\{S_n\}_{n\in\mathbb{N}}$ definida por $S_n = a_1 + a_2 + a_3 + \ldots + a_n = \sum_{i=1}^{n} a_n$ diz-se a *sucessão das somas parciais* da série $\sum_{i=1}^{+\infty} a_i$.

Definição 13.2.1. A série $\sum_{i=1}^{+\infty} a_i$ diz-se convergente com soma S, e escreve-se

$$S = \sum_{i=1}^{+\infty} a_i,$$

se existir um número real S tal que

$$S = \lim_{n\to+\infty} S_n.$$

Caso contrário a série diz-se divergente.

Exemplo 13.2.1 (Série geométrica). Seja $0 < p < 1$. Para cada $n \in \mathbb{N}$, tomemos $a_n := p^n$. Vejamos que a série $\sum_{n=1}^{+\infty} p^n = p + p^2 + \ldots$ é convergente e

$$\sum_{n=1}^{+\infty} p^n = \frac{p}{1-p}.$$

Para verificarmos isso observe-se que

$$\begin{aligned}(1-p)S_n &= (1-p)(p + p^2 + \ldots + p^n)\\ &= p - p^2 + p^2 - p^3 + p^3 - p^4 + \ldots - p^n + p^n - p^{n+1}\\ &= p - p^{n+1}.\end{aligned}$$

Resulta então que

$$S_n = \frac{p - p^{n+1}}{1-p} \xrightarrow[n\to+\infty]{} \frac{p}{1-p}.$$

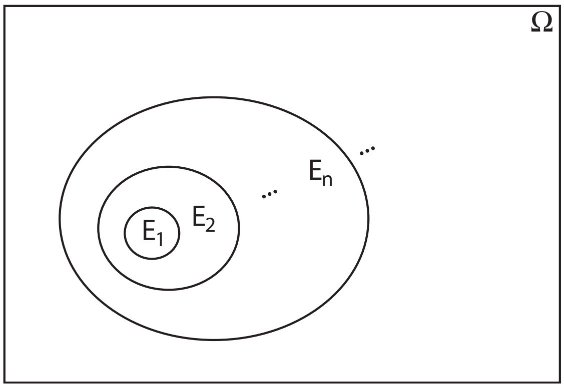

FIGURA 13.1. Sequência crescente de conjuntos

13.2.2. Operações de conjuntos. Seja Ω um conjunto e $E_1, E_2, \ldots$ uma sequência de subconjuntos de Ω. Definimos:

$$\limsup_{n\to\infty} E_n := \bigcap_{n=1}^{+\infty}\left(\bigcup_{i=n}^{+\infty} E_i\right)$$

e

$$\liminf_{n\to\infty} E_n := \bigcup_{n=1}^{+\infty}\left(\bigcap_{i=n}^{+\infty} E_i\right).$$

Se a sequência $\{E_n\}_{n\in\mathbb{N}}$ for tal que $\limsup_{n\to\infty} E_n = \liminf_{n\to\infty} E_n$, então dizemos que E_n *converge* e definimos o *conjunto limite*

$$E := \lim_{n\to\infty} E_n = \limsup_{n\to\infty} E_n.$$

Observe-se que $x \in \liminf_{n\to\infty} E_n$ significa que x pertence a todos os conjuntos E_n, à excepção de uma quantidade finita deles; de facto, $x \in \liminf_{n\to\infty} E_n$ é equivalente a existir $N \in \mathbb{N}$ tal que $x \in E_n$ para todo $n \geq N$. Por outro lado, $x \in \limsup_{n\to\infty} E_n$ significa que x pertence a uma infinidade de conjuntos E_n, isto é, existe uma sequência $\{n_i\}_{i\in\mathbb{N}} \subset \mathbb{N}$ tal que $x \in E_{n_i}$ para todo $i \in \mathbb{N}$. Observe-se que $\liminf_{n\to\infty} E_n \subset \limsup_{n\to\infty} E_n$. Fazemos ver que a notação $A \subset B$ significa que se $x \in A$ então $x \in B$, e em particular pode suceder que $A = B$.

Dizemos que uma sequência $\{E_n\}_{n\in\mathbb{N}}$ é *crescente* se $E_n \subset E_{n+1}$, $\forall n \in \mathbb{N}$, e dizemos que uma sucessão é decrescente se $E_{n+1} \subset E_n$, $\forall n \in \mathbb{N}$. Uma sequência $\{E_n\}_{n\in\mathbb{N}}$ diz-se *monótona* se for crescente ou decrescente.

OBSERVAÇÃO 13.2.1. Qualquer sequência monótona é convergente.
(i) Se $\{E_n\}_{n\in\mathbb{N}}$ for crescente, então temos que

$$\bigcup_{i=n}^{+\infty} E_i = \bigcup_{i=1}^{+\infty} E_i \qquad \text{e} \qquad \bigcap_{i=n}^{+\infty} E_i = E_n. \qquad \text{Logo} \qquad \lim_{n\to\infty} E_n = \bigcup_{i=1}^{+\infty} E_i.$$

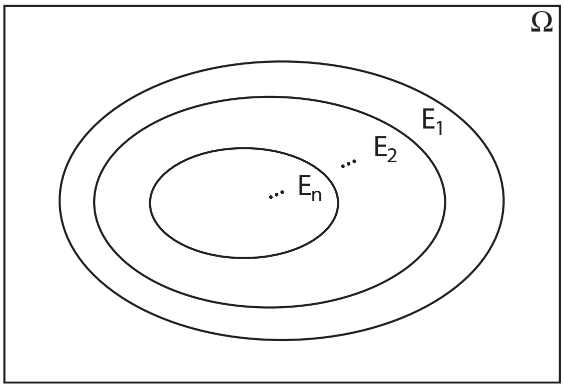

Figura 13.2. Sequência decrescente de conjuntos

(ii) Se $\{E_n\}_{n\in\mathbb{N}}$ for decrescente, então temos que

$$\bigcup_{i=n}^{+\infty} E_i = E_n \qquad \text{e} \qquad \bigcap_{i=n}^{+\infty} E_i = \bigcap_{i=1}^{+\infty} E_i. \qquad \text{Logo} \qquad \lim_{n\to\infty} E_n = \bigcap_{i=1}^{+\infty} E_i.$$

Exemplo 13.2.2. Seja $\Omega = \mathbb{R}$ e, para cada $i \in \mathbb{N}$, seja $E_i = [1/(2i), 1 + 1/i]$. Então se definirmos, para cada $n \in \mathbb{N}$, $A_n = \bigcup_{i=n}^{+\infty} E_i = (0, 1 + 1/n]$ e $B_n = \bigcap_{i=n}^{+\infty} E_i = (1/(2n), 1]$, temos que $\{A_n\}_{n\in\mathbb{N}}$ é decrescente e $\{B_n\}_{n\in\mathbb{N}}$ é crescente. Resulta então que $\limsup_{n\to+\infty} E_n = \bigcap_{j=n}^{+\infty} A_j = (0,1]$ e $\liminf_{n\to+\infty} E_n = \bigcup_{j=1}^{+\infty} B_j = (0,1]$. Em particular temos que $\lim_{n\to+\infty} E_n = (0,1]$.

Exemplo 13.2.3. Seja $\Omega = \mathbb{R}$ e para cada $i \in \mathbb{N}$, seja $E_i = [0, 2 + (-1)^n]$. Definindo, para cada $n \in \mathbb{N}$, $A_n = \bigcup_{i=n}^{+\infty} E_i = [0,3]$ e $B_n = \bigcap_{i=n}^{+\infty} E_i = [0,1]$, facilmente se observa que $\limsup_{n\to+\infty} E_n = \bigcap_{j=n}^{+\infty} A_j = [0,3]$ e $\liminf_{n\to+\infty} E_n = \bigcup_{j=1}^{+\infty} B_j = [0,1]$. Neste caso, não existe $\lim_{n\to+\infty} E_n$.

13.2.3. Funções, imagem recíproca e propriedades. Ao longo deste capítulo, sendo A e B dois subconjuntos de um conjunto X, usaremos as seguintes notações:

$$B - A := \{x \in B : x \notin A\}$$

e

$$A^C := X - A.$$

Seja $f : X \to Y$ uma função. Dado um conjunto $A \subset X$ definimos a *imagem de* A por f como

$$f(A) = \{y \in Y : f(x) = y \text{ para algum } x \in A\}.$$

Lembremos que f é *injectiva* se, para todo $x_1, x_2 \in X$ tais que $x_1 \neq x_2$, se tem que $f(x_1) \neq f(x_2)$. Quando $f(X) = Y$ então dizemos que f é *sobrejectiva.* Se f for injectiva e sobrejectiva então f é *invertível*, i.e., existe $f^{-1} : Y \to X$ tal que $f \circ f^{-1} = \mathrm{Id}_Y$ e $f^{-1} \circ f = \mathrm{Id}_X$, onde $\mathrm{Id}_X : X \to X$, $\mathrm{Id}_Y : Y \to Y$ denotam a função identidade ($\mathrm{Id}_X(x) = x$ e $\mathrm{Id}_Y(y) = y$ para todo $x \in X$ e $y \in Y$).

DEFINIÇÃO 13.2.2. Seja $D \subset Y$. Definimos a *imagem recíproca de* D por f, que denotamos por $f^{-1}(D)$, como

$$f^{-1}(D) := \{x \in X : f(x) \in D\}$$

Observe-se que a imagem recíproca de um conjunto está sempre definida, independentemente de f ser invertível ou não. Quando f é invertível, a imagem recíproca de $D \subset Y$ por f coincide com a imagem de D pela função inversa de f, o que revela a adequabilidade da notação usada para imagem recíproca.

EXEMPLO 13.2.4. Consideremos $f : \mathbb{R} \to \mathbb{R}$ dada por $f(x) = x^2$. Facilmente se verifica que:

- $f([0,1]) = f([-1,0]) = f([-1,1]) = [0,1]$,
- $f([1,2)) = [1,4)$,
- $f^{-1}(\{2\}) = \{-\sqrt{2}, \sqrt{2}\}$,
- $f^{-1}([0,1]) = f^{-1}([-3,1]) = [-1,1]$,
- $f^{-1}([1,4)) = (-2,-1] \cup [1,2)$.

PROPOSIÇÃO 13.2.1. Para todo $A, B \subset X$ e qualquer sequência $\{A_i\}_{i\in\mathbb{N}}$ com $A_i \subset X$, para todo $i \in \mathbb{N}$, temos:

(1) $f(\emptyset) = \emptyset$
(2) $f(\bigcup_{i\in\mathbb{N}} A_i) = \bigcup_{i\in\mathbb{N}} f(A_i)$
(3) $A \subset B \Rightarrow f(A) \subset f(B)$
(4) $f(A \cap B) \subset f(A) \cap f(B)$
(5) $f(A) - f(B) \subset f(A - B)$
(6) $A \cap B \neq \emptyset \Rightarrow f(A) \cap f(B) \neq \emptyset$
(7) $A \cap B = \emptyset$ e f é injectiva $\Rightarrow f(A) \cap f(B) = \emptyset$
(8) f é sobrejectiva $\Rightarrow (f(A))^C \subset f(A^C)$

A prova desta proposição decorre da definição da imagem de um conjunto por f, é simples e como tal deixamo-la como exercício. Prova-se também facilmente que as desigualdades que aparecem podem ser estritas bastando para isso exemplificar com funções não injectivas.

PROPOSIÇÃO 13.2.2. Para todo $A \subset X$, $D, E \subset Y$ e qualquer sequência $\{D_i\}_{i\in\mathbb{N}}$ com $D_i \subset Y$, para todo $i \in \mathbb{N}$, temos:

(1) $f^{-1}(\emptyset) = \emptyset$
(2) $f^{-1}(\bigcup_{i\in\mathbb{N}} D_i) = \bigcup_{i\in\mathbb{N}} f^{-1}(D_i)$
(3) $D \subset E \Rightarrow f^{-1}(D) \subset f^{-1}(E)$
(4) $f^{-1}(D \cap E) = f^{-1}(D) \cap f^{-1}(E)$
(5) $f^{-1}(D) - f^{-1}(E) = f^{-1}(D - E)$
(6) $(f^{-1}(D))^C = f^{-1}(D^C)$
(7) $A \subset f^{-1}(f(A))$ e se f é injectiva então $A = f^{-1}(f(A))$
(8) $f(f^{-1}(D)) \subset D$ e se f é sobrejectiva então $f(f^{-1}(D)) = D$.

Esta proposição afirma que o operador f^{-1}, ao contrário do operador f, comuta com as operações de conjunto e a sua prova decorre das definições de imagem recíproca e imagem de conjuntos por f. A título ilustrativo fazemos apenas a prova da afirmação (4).

Demonstração. Mostremos primeiro que $f^{-1}(D \cap E) \subset f^{-1}(D) \cap f^{-1}(E)$. Se $x \in f^{-1}(D \cap E)$ então $f(x) \in D \cap E$, o que implica que $f(x) \in D$ e $f(x) \in E$, ou seja, $x \in f^{-1}(D)$ e $x \in f^{-1}(E)$.

No segundo passo, mostremos que $f^{-1}(D) \cap f^{-1}(E) \subset f^{-1}(D \cap E)$. Se $x \in f^{-1}(D) \cap f^{-1}(E)$, então $f(x) \in D$ e $f(x) \in E$; logo $f(x) \in D \cap E$, ou seja, $x \in f^{-1}(D \cap E)$. □

13.3. Probabilidades e variáveis aleatórias

13.3.1. Noção de medida de probabilidade. No que se segue não falaremos de probabilidades sem estar subentendido um espaço de acontecimentos associado a uma experiência aleatória conceptual. Assim sendo, comecemos por considerar uma experiência aleatória e um conjunto Ω, que modela a experiência, no sentido em que todo o possível resultado da experiência é completamente descrito por um e um só elemento de Ω. Designamos os subconjuntos $A \subset \Omega$ por *acontecimentos* e Ω por *espaço de acontecimentos*. Uma vez realizada a experiência, dizemos que o acontecimento $A \subset \Omega$ *ocorreu* se o elemento de Ω que representa o resultado obtido estiver contido em A. O conjunto vazio $\emptyset$, que nunca ocorre, é designado por *acontecimento impossível*. Já o conjunto de todos os possíveis resultados da experiência Ω, que ocorre sempre que se realiza a mesma, diz-se *acontecimento certo*.

Para exemplificar, consideremos a experiência que consiste pegar numa moeda, que num dos lados tem uma face representada e no outro uma coroa, lançá-la ao ar e depois desta cair, observar qual das superfícies fica voltada para cima. Escolhemos o conjunto $\Omega = \{F, C\}$ para espaço de acontecimentos, em que o elemento F representa o resultado da experiência em que a superfície da moeda com uma face representada fica voltada para cima e C representa o caso em que a superfície voltada para cima ostenta uma coroa. Os conjuntos $\emptyset, \{F\}, \{C\}$ e $\{F, C\}$ representam os acontecimentos realizáveis. Se ao realizarmos a experiência, a superfície com uma coroa representada fica voltada para cima dizemos que os acontecimentos $\{C\}$ e $\{F, C\}$ ocorreram (pois C é elemento de ambos) enquanto que os acontecimentos $\{F\}$ e $\emptyset$ não ocorreram (porque C não é elemento de qualquer um dos dois).

Definição 13.3.1. Consideremos uma classe de subconjuntos de Ω, que denotamos por $\mathcal{B}$, com as seguintes propriedades:

(a) $\emptyset \in \mathcal{B}$;
(b) Se $A_1, A_2, \ldots$ é uma sequência de acontecimentos de $\mathcal{B}$ então $\bigcup_{i=1}^{+\infty} A_i \in \mathcal{B}$;
(c) Se $A \in \mathcal{B}$ então $\Omega - A \in \mathcal{B}$.

Qualquer classe de subconjuntos de Ω que satisfaça as propriedades (a), (b) e (c) diz-se uma *σ-álgebra*.

Verifica-se facilmente que se $\mathcal{B}$ for uma σ-álgebra, então

(1) $\Omega \in \mathcal{B}$;
(2) Se $A_1, \ldots, A_n \in \mathcal{B}$, então $\bigcup_{i=1}^{n} A_i \in \mathcal{B}$ e $\bigcap_{i=1}^{n} A_i \in \mathcal{B}$;
(3) Se $A, B \in \mathcal{B}$ então $A - B \in \mathcal{B}$;
(4) Se $A_1, A_2, \ldots$ é uma sequência de acontecimentos de $\mathcal{B}$, então $\bigcap_{i=1}^{+\infty} A_i \in \mathcal{B}$;
(5) Se $A_1, A_2, \ldots$ é uma sequência de acontecimentos de $\mathcal{B}$, então $\limsup_{n\to\infty} A_n \in \mathcal{B}$ e $\liminf_{n\to\infty} A_n \in \mathcal{B}$.

A noção de σ-álgebra poderá parecer introduzir abstracção despropositada para o objectivo de definir medida de probabilidade. Na realidade, pretendemos definir uma função que permita medir subconjuntos de Ω atribuindo-lhes um "peso" que expresse a "probabilidade" de, ao realizarmos a experiência aleatória, se observar um resultado que pertença ao subconjunto em causa, ou seja, que permita quantificar a probabilidade de ocorrer um determinado acontecimento. A σ-álgebra providencia o habitat natural para definir esta função peso. A razão é que queremos que esta função peso tenha propriedades algébricas relativamente às operações de conjuntos. Por exemplo, pretendemos que o peso da união de dois acontecimentos disjuntos seja a soma dos pesos individuais de cada um. Ora a questão é que não é claro que seja sempre possível "pesar" a união de acontecimentos cujo peso conhecemos. Esta afirmação perde significado se considerarmos que a função peso está definida na classe de todos os subconjuntos de Ω. Porém, há espaços de acontecimentos de tal forma grandes para os quais não é possível definir uma função peso em todos os seus subconjuntos. É nestes casos que a σ-álgebra se revela fundamental pois dá-nos a estrutura algébrica necessária para garantir que é sempre possível quantificar a probabilidade de acontecimentos que resultam da aplicação das operações de conjuntos a acontecimentos cuja probabilidade já sabemos aferir. Na maior parte dos exemplos dados, e sempre que possível, tomaremos a classe de todos os subconjuntos de Ω como a nossa σ-álgebra.

Para concretizar, analisemos os dois casos que se seguem.

Deixando os detalhes para o Exemplo 13.3.2, consideremos o lançamento de um dado de 6 faces numeradas de 1 a 6. Neste caso, tomamos para espaço de acontecimentos $\Omega = \{1, 2, 3, 4, 5, 6\}$. São exemplos de acontecimentos os subconjuntos: $\{3\}$, $\{1, 3, 5\}$, que representam, respectivamente, os resultados do lançamento de um dado em que a face com o número 3 ficou voltada para cima e o acontecimento correspondente a um lançamento em que o número da face voltada para cima é ímpar. A σ-álgebra natural nesta situação é a classe de todos os subconjuntos de Ω ilustrados no Exemplo 13.3.2.

Quando o espaço de acontecimentos é $\Omega = \mathbb{R}$ (ou qualquer intervalo de números reais) estamos perante um daqueles casos em que a classe de todos os seus subconjuntos é demasiado grande. Nestes casos, começa-se por construir uma classe de subconjuntos de forma simples, mas com uma estrutura mais fraca, que depois estendemos para que sejam satisfeitas as propriedades definidoras de σ-álgebra. Considere-se a classe de subconjuntos de $\mathbb{R}$ dada pelos intervalos semi-abertos à esquerda:

$$\mathcal{C} = \{I \subset \mathbb{R} : I = (a, b],\ a < b \in \mathbb{R}\}. \tag{13.1}$$

Observe-se que $\mathcal{C}$ não é uma σ-álgebra porque não é fechada para a união (uma classe $\mathcal{C}$ diz-se fechada para a união se a união de quaisquer dois elementos de $\mathcal{C}$ for ainda um elemento de $\mathcal{C}$). Por exemplo, $(0, 1/3] \in \mathcal{C}$ e $(2/3, 1] \in \mathcal{C}$ mas $(0, 1/3] \cup (2/3, 1] \notin \mathcal{C}$. Contudo, $\mathcal{C}$ é fechada para a intersecção e, para todo $A, B \in \mathcal{C}$, existem $C_1, C_2 \in \mathcal{C}$ disjuntos tais que $A - B = C_1 \cup C_2$. Esta propriedade continuaria válida se tivéssemos escolhido intervalos semi-abertos à direita. Se Ω fosse um intervalo $J \subset \mathbb{R}$ de números reais poderíamos considerar a classe dos intervalos que resultam da intersecção dos elementos de $\mathcal{C}$ com J, pois esta continuaria a ter as duas propriedades de que $\mathcal{C}$ goza.

Definição 13.3.2. Definimos a σ-álgebra de Borel $\mathcal{B}$ em $\mathbb{R}$ como sendo a menor classe de subconjuntos que contém $\mathcal{C}$ definido em (13.1) e satisfaz as propriedades (a), (b) e (c) da Definição 13.3.1. No caso em que $\Omega = J$, onde J é um intervalo de qualquer tipo de $\mathbb{R}$, a σ-álgebra de Borel define-se como a anterior tomando-se apenas $\mathcal{C} \cap J = \{I \cap J : I \in \mathcal{C}\}$ no lugar de $\mathcal{C}$.

Note-se que há subconjuntos de $\Omega = \mathbb{R}$ que não pertencem à σ-álgebra de Borel.

Uma vez definida a estrutura onde pretendemos definir uma função que permita "pesar" ou "medir" acontecimentos, definamos então o que entendemos por medida de probabilidade e as propriedades que queremos que ela tenha.

Definição 13.3.3. Uma *medida de probabilidade* em Ω é uma função $\mu : \mathcal{B} \to [0, 1]$ definida numa σ-álgebra $\mathcal{B}$ de subconjuntos de Ω com as seguintes propriedades:

(i) Se $A_1, A_2, \ldots$ é uma sequência de acontecimentos de $\mathcal{B}$ tais que $A_i \cap A_j = \emptyset \; \forall i \neq j$, então

$$\mu\left(\bigcup_{i=1}^{+\infty} A_i\right) = \sum_{i=1}^{+\infty} \mu(A_i);$$

(ii) $\mu(\Omega) = 1$.

De seguida listamos algumas propriedades que decorrem da definição de medida de probabilidade.

Proposição 13.3.1. As seguintes afirmações são verdadeiras.

(1) $\mu(\emptyset) = 0$

(2) Se $A_1, \ldots, A_n \in \mathcal{B}$ são disjuntos dois a dois, então

$$\mu\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} A_i$$

(3) $\mu(A - B) = \mu(A) - \mu(A \cap B)$. Em particular, $\mu(\Omega - B) = 1 - \mu(B)$

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

Demonstração. (1) Notemos que podemos escrever

$$\Omega = A_1 \cup A_2 \cup \ldots,$$

onde $A_1 = \Omega$ e $A_j = \emptyset \; \forall j \geq 2$. Portanto, por (i),

$$\mu(\Omega) = \mu(\Omega) + \sum_{i=2}^{+\infty} \mu(\emptyset),$$

donde se conclui que $\mu(\emptyset) = 0$.

(2) Para provar o pretendido basta considerar a sequência disjunta

$$B_1, \ldots, B_n, B_{n+1}, \ldots,$$

em que $B_i = A_i \; \forall i = 1, \ldots, n$, e $B_{n+j} = \emptyset \; \forall j = 1, \ldots$. Desta forma, o resultado segue de (i) e de (1).

(3) Observemos que os acontecimentos $A - B$ e $A \cap B$ são disjuntos e que podemos escrever

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

Por isso, de (2) segue que

$$\mu(A) = \mu(A - B) + \mu(A \cap B),$$

ou equivalentemente,

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

Deste facto e de (ii) conclui-se em particular que $\mu(\Omega - B) = 1 - \mu(B)$.

(4) Como observamos na prova de (3), podemos escrever

$$A = (A - B) \cup (A \cap B),$$

e da mesma forma,

$$B = (B - A) \cup (A \cap B)$$

e

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

O resultado segue de (2) e de (3).

□

O próximo resultado dá-nos a continuidade das medidas de probabilidade, o que, em sentido lacto, significa que a operação "limite" comuta com a medida de probabilidade.

Teorema 13.3.1. Seja $\{A_n\}_{n\in\mathbb{N}}$ uma sequência convergente de subconjuntos de Ω tal que $A_n \in \mathcal{B}$, para todo $n \in \mathbb{N}$. Então

$$\mu(\lim_{n\to+\infty} A_n) = \lim_{n\to+\infty} \mu(A_n)$$

Observação 13.3.1. Note-se que o limite que figura no primeiro membro da última igualdade é o limite de conjuntos definido na secção 13.2.2, enquanto que o limite do segundo membro é o limite usual de sucessões de números reais.

Demonstração. Para cada $n \in \mathbb{N}$ seja $B_n = \bigcap_{i=n}^{+\infty} A_i$. É fácil verificar que B_n é uma sequência monótona crescente de subconjuntos de Ω pertencentes a $\mathcal{B}$. Para além disso, $A := \lim_{n\to+\infty} A_n = \lim_{n\to+\infty} B_n = \bigcup_{n=1}^{+\infty} B_n \in \mathcal{B}$. Definamos agora $E_1 = B_1$ e para cada $n \in \mathbb{N}$, tomemos $E_{n+1} = B_{n+1} - B_n$. Observe-se que para todo $n \in \mathbb{N}$ vale que $E_n \in \mathcal{B}$, $A = \bigcup_{n=1}^{+\infty} B_n = \bigcup_{n=1}^{+\infty} E_n$ e, para todo $i \neq j$, $E_i \cap E_j = \emptyset$, ou seja, os conjuntos E_n dão-nos uma partição de A em subconjuntos disjuntos.

$$\begin{aligned}
\mu(A) &= \sum_{i=1}^{+\infty} \mu(E_i) && \text{por (i) da Definição 13.3.3} \\
&= \lim_{n\to+\infty} \sum_{i=1}^{n} \mu(E_i) && \text{pela Definição 13.2.1} \\
&= \lim_{n\to+\infty} \mu\left(\bigcup_{i=1}^{n} E_i\right) && \text{por (2) da Proposição 13.3.1} \\
&= \lim_{n\to+\infty} \mu\left(B_n\right) && \text{por definição dos conjuntos } E_i.
\end{aligned}$$

□

Observação 13.3.2. O terno $(\Omega, \mathcal{B}, \mu)$ é habitualmente designado por *espaço de medida*. Referir-nos-emos a $\mu(A)$ como a *medida de* A ou a probabilidade do acontecimento A ocorrer. A σ-álgebra $\mathcal{B}$ determina os acontecimentos mensuráveis, ou, por outras palavras, os acontecimentos cuja probabilidade de ocorrerem sabemos medir. Assim, A é *mensurável* se e só se $A \in \mathcal{B}$. A estrutura que (a), (b) e (c) impõem a $\mathcal{B}$ destina-se a garantir que a colecção dos conjuntos mensuráveis contém o $\emptyset$ e o Ω e é fechada para as operações de conjuntos: $\cup$, $\cap$, $-$, Δ, ^{C}, $\limsup$ e $\liminf$. As propriedades que definem a medida de probabilidade e que delas emanam reflectem aquilo que intuitivamente associamos à noção de medir probabilidade. Essencialmente, formalizam que a medida do espaço todo totaliza 1 e que se repartirmos o acontecimento A em pedaços disjuntos, então a medida de A é a soma das medidas das suas partes.

Quando a σ-álgebra é muito grande, como acontece no caso $\Omega = \mathbb{R}$ em que se considera a σ-álgebra de Borel $\mathcal{B}$, uma forma de construirmos a medida de probabilidade é defini-la primeiro numa classe mais pequena (sem a estrutura de σ-álgebra), como é o caso de $\mathcal{C}$ (ver definição em (13.1)) dos intervalos semi-abertos, e depois estendê-la[1] à σ-álgebra $\mathcal{B}$ usando o Teorema da extensão de Carathéodory (veja-se, por exemplo, (Kingman e Taylor, 1966). Essencialmente, é importante reter que, uma vez que se saiba como "medir" todos os intervalos semi-abertos, é possível "medir" em toda a σ-álgebra de Borel de forma coerente.

Exemplo 13.3.1. Lançamento de uma moeda ao ar

O espaço de acontecimentos associado a esta experiência é o seguinte:

$$\Omega = \{F, C\},$$

em que $\{F\}$ corresponde ao acontecimento "sair face" e $\{C\}$ corresponde ao acontecimento "sair coroa".

Podemos tomar para σ-álgebra:

$$\mathcal{B} = \{\emptyset, \{F\}, \{C\}, \{F, C\}\}.$$

A medida de probabilidade pode definir-se como

$$\mu(\emptyset) = 0, \quad \mu(\{F\}) = p, \quad \mu(\{C\}) = q, \quad \mu(\{F, C\}) = 1,$$

em que $p \geq 0$, $q \geq 0$ e $p + q = 1$. No caso particular de se tratar de uma moeda equilibrada, temos que $p = q = 1/2$.

Exemplo 13.3.2. Lançamento de um dado

O espaço de acontecimentos associado a esta experiência é dado por

$$\Omega = \{1, 2, 3, 4, 5, 6\},$$

em que $\{i\}$ corresponde ao acontecimento "número de pontos da face voltada para cima ser igual a i", $i = 1, 2, 3, 4, 5, 6$.

Tomamos para σ-álgebra $\mathcal{B}$ a classe formada por todos os subconjuntos de Ω, cujos elementos escrevemos na Tabela 13.1.

Notemos que o número de elementos da σ-álgebra é igual 2^6. Em geral, se o espaço de acontecimentos Ω for finito e tiver n elementos, o número de subconjuntos diferentes que

[1]Se $\mathcal{C} \subset \mathcal{B}$ são classes de subconjuntos de Ω, $\mu : \mathcal{C} \to [0, 1]$ uma função de conjuntos e $\nu : \mathcal{B} \to [0, 1]$ uma medida de probabilidade tal que $\nu(A) = \mu(A)$ para todo $A \in \mathcal{C}$ então ν diz-se uma extensão de μ a $\mathcal{B}$.

$\emptyset$	$\{1,3\}$	$\{3,4\}$	$\{1,2,5\}$	$\{2,3,4\}$	$\{3,5,6\}$	$\{1,3,4,5\}$	$\{3,4,5,6\}$
$\{1\}$	$\{1,4\}$	$\{3,5\}$	$\{1,2,6\}$	$\{2,3,5\}$	$\{4,5,6\}$	$\{1,3,4,6\}$	$\{1,2,3,4,5\}$
$\{2\}$	$\{1,5\}$	$\{3,6\}$	$\{1,3,4\}$	$\{2,3,6\}$	$\{1,2,3,4\}$	$\{1,3,5,6\}$	$\{1,2,3,4,6\}$
$\{3\}$	$\{1,6\}$	$\{4,5\}$	$\{1,3,5\}$	$\{2,4,5\}$	$\{1,2,3,5\}$	$\{1,4,5,6\}$	$\{1,2,3,5,6\}$
$\{4\}$	$\{2,3\}$	$\{4,6\}$	$\{1,3,6\}$	$\{2,4,6\}$	$\{1,2,3,6\}$	$\{2,3,4,5\}$	$\{1,2,4,5,6\}$
$\{5\}$	$\{2,4\}$	$\{5,6\}$	$\{1,4,5\}$	$\{2,5,6\}$	$\{1,2,4,5\}$	$\{2,3,4,6\}$	$\{1,3,4,5,6\}$
$\{6\}$	$\{2,5\}$	$\{1,2,3\}$	$\{1,4,6\}$	$\{3,4,5\}$	$\{1,2,4,6\}$	$\{2,3,5,6\}$	$\{2,3,4,5,6\}$
$\{1,2\}$	$\{2,6\}$	$\{1,2,4\}$	$\{1,5,6\}$	$\{3,4,6\}$	$\{1,2,5,6\}$	$\{2,4,5,6\}$	$\{1,2,3,4,5,6\}$

TABELA 13.1. Subconjuntos do espaço de acontecimentos $\{1,2,3,4,5,6\}$.

é possível formar com os elementos de Ω é igual a 2^n. A razão para esse valor prende-se com o facto de que ao construirmos um subconjunto de Ω, para cada elemento de Ω temos 2 possibilidades: ele pertence ao subconjunto ou não pertence. Como há n elementos em Ω então há 2^n possibilidades distintas de construirmos um subconjunto de Ω.

A medida de probabilidade pode definir-se como

$$\mu(\emptyset) = 0, \qquad \mu(\{i\}) = p_i,\ i = 1,\ldots,6 \qquad \text{e} \qquad \mu(A) = \sum_{i\in A} p_i,$$

em que $p_i \geq 0$ e $\sum_{i=1}^{6} p_i = 1$. No caso particular de se tratar de um dado equilibrado, temos que $p_i = 1/6$, para todo $i \in \{1,2,3,4,5,6\}$.

Mais geralmente, se $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$ for um espaço de acontecimentos finito com n elementos, os acontecimentos $\{\omega_1\}, \{\omega_2\}, \ldots, \{\omega_n\}$ são ditos *elementares*. Tomando a σ-álgebra habitual $\mathcal{B}$ como sendo o conjunto de todos os subconjuntos de Ω, podemos definir a medida μ atribuindo a cada acontecimento elementar probabilidade $p_i = \mu(\{\omega_i\})$, para $i = 1,2,\ldots,n$, em que $\sum_{i=1}^{n} p_i = 1$. A probabilidade de qualquer outro acontecimento $A \in \mathcal{B}$ é calculada pela fórmula:

$$\mu(A) = \sum_{\omega_i \in A} p_i.$$

EXEMPLO 13.3.3. Lançamentos independentes de uma moeda ao ar.

Consideremos primeiro o caso de dois lançamentos consecutivos de uma moeda ao ar. O espaço de acontecimentos associado a esta experiência é o conjunto $\Omega = \{F,C\} \times \{F,C\} = \{F,C\}^2$ cujos elementos são os pares: FF, FC, CF, CC, sendo que, por exemplo, o acontecimento $\{FC\}$ representa a saída de face no primeiro lançamento e coroa no segundo. É habitual neste caso escolher a σ-álgebra $\mathcal{B}$ como sendo a colecção de todos os subconjuntos de Ω e a medida produto μ dada por

$$\mu(\{FF\}) = \mu(\{FC\}) = \mu(\{CF\}) = \mu(\{CC\}) = \frac{1}{4},$$

que reflecte o facto de a probabilidade do acontecimento $\{FC\}$, por exemplo, ser o produto da probabilidade de sair face no primeiro lançamento pela probabilidade de sair coroa no segundo, admitindo que a moeda é equilibrada.

Agora consideremos o caso genérico de $n \in \mathbb{N}$ lançamentos consecutivos da mesma moeda equilibrada ao ar. O espaço de acontecimentos é o conjunto $\Omega = \{F,C\}^n$ cujos elementos são palavras com n letras do alfabeto $\{F,C\}$. Assim podemos representar cada elemento $\omega \in \Omega$ como uma sequência de n símbolos, i.e., $\omega = \omega_1\omega_2\ldots\omega_n$ em que cada $\omega_i \in \{F, C\}$ para todo $i = 1, 2, \ldots, n$. Por exemplo, o acontecimento elementar

$$\{\underbrace{FFC\ldots C}_{n \text{ símbolos}}\}$$

representa a saída de face nos dois primeiros lançamentos e coroa do terceiro ao n-ésimo. Para σ-álgebra $\mathcal{B}$ podemos tomar novamente a classe de todos os subconjuntos de Ω. Neste caso, Ω tem 2^n elementos e portanto $\mathcal{B}$ terá 2^{2^n} elementos. Admitindo novamente que a moeda é equilibrada, basta definirmos a medida produto μ em cada acontecimento elementar atribuindo a cada um probabilidade $(1/2)^n$.

Finalmente, podemos considerar a experiência de lançar indefinidamente a mesma moeda ao ar. Neste caso, o espaço amostral é o conjunto $\Omega = \{F, C\}^{\mathbb{N}}$ cujos elementos são palavras com um número infinito de letras do alfabeto $\{F, C\}$. Podemos representar cada elemento de $\omega \in \Omega$ como uma sequência infinita de símbolos da forma

$$\omega = \omega_1\omega_2\omega_3\ldots,$$

em que $\omega_i \in \{F, C\}$ para todo $i \in \mathbb{N}$. Neste caso, acontece que Ω é demasiado grande para tomarmos a classe de todos os seus subconjuntos para σ-álgebra. À semelhança do que fizemos quando definimos a σ-álgebra de Borel, teremos que escolher uma colecção de subconjuntos de Ω mais pequena, sem a estrutura de σ-álgebra e depois estendê-la. Dados $n, k \in \mathbb{N}$ e n símbolos $a_1, a_2, \ldots, a_n \in \{F, C\}$ definimos o cilindro

$$C(n, k, [a_1, a_2, \ldots, a_n]) = \{\omega \in \Omega : \omega_k = a_1, \omega_{k+1} = a_2, \ldots, \omega_{k+n-1} = a_n\}, \tag{13.2}$$

que representa o conjunto de todas as sequências ω em que o bloco de símbolos $[a_1, a_2, \ldots, a_n]$ ocorre entre as posições k e $k+n-1$. Podemos pensar nos cilindros como segmentos finitos de história na realização da nossa experiência. Ou seja, o cilindro $C(n, k, [a_1, a_2, \ldots, a_n])$ corresponde a todas as realizações em que entre o k-ésimo lançamento e o $k+n-1$-ésimo lançamento se observa a sequência $[a_1, a_2, \ldots, a_n]$. Denotamos por $\mathcal{C}$ a classe de todos os cilindros. A σ-álgebra $\mathcal{B}$ é a σ-álgebra gerada pelos cilindros, i.e., a menor classe de subconjuntos que contém $\mathcal{C}$ e satisfaz as propriedades (a), (b) e (c) da Definição 13.3.1. Começamos por definir a medida de probabilidade μ como uma medida produto nos cilindros que, no caso da moeda ser equilibrada, para $n, k \in \mathbb{N}$ e $a_0, \ldots, a_n \in \{F, C\}$ é dada por:

$$\mu(C(n, k, [a_1, a_2, \ldots, a_n])) = (1/2)^n. \tag{13.3}$$

Uma vez definida nos cilindros, é possível usar o Teorema da extensão de Carathéodory para estendê-la à σ-álgebra $\mathcal{B}$. Para ilustrar os conceitos introduzidos consideremos o cilindro

$$C(3, 2, [F, F, C]) = \{\omega \in \Omega : \omega_2 = F, \omega_3 = F, \omega_4 = C\}$$

que representa o conjunto de todas as realizações da experiência em que no segundo lançamento se obtém face, no terceiro sai face e no quarto sai coroa. A probabilidade deste acontecimento é

$$\mu(C(3, 2, [F, F, C])) = (1/2)^3 = 1/8,$$

que é a probabilidade de sair face no segundo lançamento vezes a probabilidade de sair face no terceiro vezes a probabilidade de sair coroa no quarto lançamento.

Exemplo 13.3.4. Medida de Lebesgue e comprimento generalizado.

Suponhamos que queremos escolher aleatoriamente um número no intervalo $[0,1]$ de forma que a probabilidade de ele cair num determinado intervalo seja proporcional ao tamanho deste. Uma forma de formalizarmos esta experiência é considerarmos o espaço amostral $\Omega = [0,1]$ e a classe de subconjuntos $\mathcal{C}$ de $[0,1]$ da forma:

$$\mathcal{C} = \{I \subset [0,1] : I = (a,b],\ a < b \in [0,1] \text{ ou } I = [0,b],\ b \in (0,1]\}.$$

Para σ-álgebra $\mathcal{B}$ tomamos a σ-álgebra gerada por $\mathcal{C}$, ou seja, a menor classe de subconjuntos de $[0,1]$ que contém $\mathcal{C}$ e satisfaz as propriedades (a), (b) e (c) da Definição 13.3.1. Notemos que há subconjuntos de $[0,1]$ que não fazem parte de $\mathcal{B}$; no entanto, $\mathcal{C}$ já é suficientemente rica para estudarmos esta experiência. Definimos a medida de Lebesgue em $\mathcal{B}$, começando por defini-la primeiro em $\mathcal{C}$:

$$\mu(I) = \begin{cases} b - a & \text{se } I = (a,b],\ a < b \in [0,1] \\ b & \text{se } I = [0,b],\ b \in (0,1], \end{cases}$$

que, essencialmente, a cada intervalo atribui um peso igual ao seu comprimento. Uma vez definida em $\mathcal{C}$, podemos usar o Teorema de extensão de Carathéodory para estendê-la a $\mathcal{B}$.

13.3.2. Acontecimentos possíveis de probabilidade zero. Um assunto que costuma gerar alguma controvérsia é a distinção entre acontecimento impossível e acontecimento de probabilidade zero.

Qualquer acontecimento que, após uma análise mais cuidada, se traduza no conjunto vazio $\emptyset$ diz-se impossível. Por exemplo, no lançamento de uma moeda ao ar, o acontecimento em que a moeda não cai, à luz do modelo que assumimos (em que $\Omega = \{F, C\}$), só pode ser representado pelo conjunto vazio $\emptyset$, tratando-se assim de um acontecimento impossível. Da mesma forma sair um 7 no lançamento de um dado de 6 faces numeradas de 1 a 6 (em que tomamos $\Omega = \{1,2,3,4,5,6\}$) é também um acontecimento impossível.

Decorre de forma óbvia da definição da medida de probabilidade que um acontecimento impossível tem probabilidade zero de ocorrer. Contudo, nem todo o acontecimento de probabilidade zero é impossível. Ilustraremos este facto usando os espaços de medida introduzidos nos Exemplos 13.3.3 e 13.3.4.

Exemplo 13.3.4 (Continuação). Consideremos o acontecimento $\{1/2\} \in \mathcal{B}$, que corresponde à escolha aleatória recair exactamente sobre o número $1/2$. Claramente este acontecimento é possível pois o número $1/2 \in [0,1]$ é um dos potenciais alvos de uma escolha aleatória, facto que assinalamos escrevendo a trivialidade $\{1/2\} \neq \emptyset$. Vejamos que, no entanto, o acontecimento $\{1/2\}$ tem probabilidade zero, i.e., $\mu(\{1/2\}) = 0$. Para tal, considere-se a sucessão $\{I_n\}_{n\in\mathbb{N}}$ de elementos de $\mathcal{B}$ dada, para cada $n \in \mathbb{N}$, pelo intervalo:

$$I_n = \left[\frac{1}{2} - \frac{1}{n+1}, \frac{1}{2} + \frac{1}{n+1}\right].$$

Por definição de μ temos que $\mu(I_n) = \frac{2}{n+1}$, para todo $n \in \mathbb{N}$. Para além disso, $\{I_n\}_{n\in\mathbb{N}}$ é monótona decrescente, converge e

$$\lim_{n\to+\infty} I_n = \{1/2\},$$

já que $1/2$ é o único elemento comum a todos os intervalos I_n, $n = 1, 2, \ldots$. Pelo Teorema 13.3.1 temos que

$$\mu(\{1/2\}) = \mu\left(\lim_{n\to+\infty} I_n\right) = \lim_{n\to+\infty} \mu(I_n) = \lim_{n\to+\infty} \frac{2}{n+1} = 0.$$

Concluímos assim que a probabilidade de o número $1/2$ ser escolhido aleatoriamente entre todos os números de $[0,1]$ é zero. A ideia intuitiva é que podemos ver o conjunto $\{1/2\}$ como o limite dos intervalos I_n que são pequenas vizinhanças de $1/2$ cuja amplitude decai para zero. Como a medida de Lebesgue atribui a cada intervalo um peso igual ao da sua amplitude, não é de estranhar que o peso atribuído a $\{1/2\}$ seja efectivamente zero.

Na realidade, seja qual for o número $x \in [0,1]$, a probabilidade de se escolher exactamente x é zero. De facto, podemos até afirmar algo mais surpreendente: a probabilidade de ser escolhido um número racional é zero. A razão prende-se com o facto de ser possível definir uma sequência de conjuntos $\{A_n\}_{n\in\mathbb{N}_0}$ tal que cada A_n possui apenas um elemento de $\mathbb{Q}\cap[0,1]$ e, para todo o número racional $z \in \mathbb{Q}\cap[0,1]$, existe $n \in \mathbb{N}_0$ tal que $z \in A_n$, o que implica que $\mathbb{Q}\cap[0,1] \subset \cup_{n=0}^{+\infty} A_n$. Uma vez construída tal sequência, pelo facto de cada A_n conter apenas um elemento segue pelo argumento anterior que $\mu(A_n) = 0$ para todo $n \in \mathbb{N}$. Logo

$$\mu(\mathbb{Q}\cap[0,1]) \leq \mu(\cup_{n=0}^{+\infty} A_n) \leq \sum_{n=0}^{+\infty} \mu(A_n) = 0.$$

Para vermos que é possível construir uma tal sequência de conjuntos, lembremos que todo o número racional em $[0,1]$ pode ser escrito na forma p/q onde $p \in \mathbb{N}_0$, $q \in \mathbb{N}$ e $p \leq q$. Podemos definir $A_0 = \{0\}$ e, usando a figura como motivação, $A_n = \{p/q\}$ onde p e q são as coordenadas do nódulo n da Figura 13.3. Assim por exemplo, $A_1 = A_2 = A_6 = A_7 = A_{15} = \ldots = \{1\}$, $A_{23} = \{2/7\}$, $A_{24} = \{3/7\}$ e $A_{40} = \{4/9\}$.

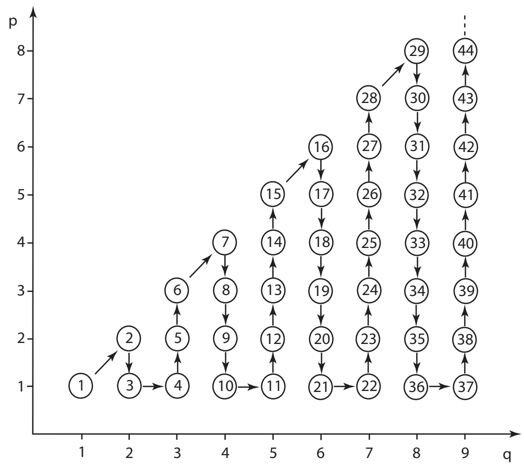

Figura 13.3. Ordenação de $\mathbb{Q}\cap[0,1]$

Exemplo 13.3.3 (Continuação). No caso do lançamento indefinido de uma moeda ao ar consideremos o acontecimento $\{FFF\ldots\} \in \mathcal{B}$ que corresponde a sair face sempre que se atira a moeda ao ar. Claramente este acontecimento é possível. No entanto, a probabilidade de este se verificar, ou seja, a probabilidade de nunca sair coroa, é zero. Para vermos isso, considere-se a sequência de cilindros $\{A_n\}_{n\in\mathbb{N}}$, dada por

$$A_n = C(n, 1, [F, F, \dots, F]), \quad \text{para cada } n \in \mathbb{N}.$$

É claro que $\{A_n\}_{n\in\mathbb{N}}$ é monótona decrescente, converge e

$$\lim_{n\to+\infty} A_n = \{FFF\dots\},$$

já que $FFF\dots$ é o único elemento comum a todos os cilindros A_n, $n = 1, 2, \dots$ Para além disso, por definição da medida de probabilidade μ, segue que

$$\mu(A_n) = \mu(C(n, 1, [F, F, \dots, F])) = 1/2^n.$$

Pelo Teorema 13.3.1 temos que

$$\mu(\{FFF\dots\}) = \mu\left(\lim_{n\to+\infty} A_n\right) = \lim_{n\to+\infty} \mu(A_n) = \lim_{n\to+\infty} \frac{1}{2^n} = 0.$$

Obviamente que a probabilidade de se observar qualquer $\omega \in \Omega = \{F, C\}^{\mathbb{N}}$ específico é zero pelo mesmo argumento.

13.3.3. Probabilidade condicional.

Definição 13.3.4. Dado um espaço de medida $(\Omega, \mathcal{B}, \mu)$, considerem-se dois acontecimentos A e B mensuráveis tais que $\mu(B) > 0$. Podemos então definir a probabilidade condicional de A dado B (ou a probabilidade de A condicionada a B) como

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

Uma das propriedades fundamentais que a probabilidade condicional possui é que, quando encarada como uma função definida em $\mathcal{B}$, ela é uma medida de probabilidade em $(\Omega, \mathcal{B})$ para a qual o conjunto B tem probabilidade um.

Teorema 13.3.2. Sejam $(\Omega, \mathcal{B}, \mu)$ um espaço de medida de probabilidade e $B \in \mathcal{B}$ tal que $\mu(B) > 0$. Então, a função $\mu_B : \mathcal{B} \to \mathbb{R}$ dada por $\mu_B(A) = \mu(A|B)$, para cada $A \in \mathcal{B}$, é uma medida de probabilidade em $(\Omega, \mathcal{B})$ com $\mu_B(B) = 1$.

Demonstração. Pela Definição 13.3.4 decorre imediatamente que $\mu_B(A) \geq 0$ e $\mu_B(\emptyset) = 0$. Sejam $A_1, A_2, \dots$ uma sequência de elementos de $\mathcal{B}$ disjuntos dois a dois cuja união é $A := \bigcup_{n=1}^{+\infty} A_n$. Deste modo,

$$\begin{aligned}\mu_B(A) &= \frac{\mu(A \cap B)}{\mu(B)} = \frac{\mu(\bigcup_{n=1}^{+\infty} A_n \cap B)}{\mu(B)} \\ &= \frac{\sum_{n=1}^{+\infty} \mu(A_n \cap B)}{\mu(B)} = \sum_{n=1}^{+\infty} \mu_B(A_n).\end{aligned}$$

Finalmente, $\mu_B(\Omega) = \mu(B)/\mu(B) = 1$ e $\mu_B(B) = \mu(B)/\mu(B) = 1$. □

Podemos dizer que a probabilidade condicional permite reduzir o universo Ω a B. De facto, $\mathcal{B}_B := \{A \in \mathcal{B} : A \subset B\}$ é uma σ-álgebra e se μ^B denotar a restrição de μ_B a $\mathcal{B}_B$ então podemos considerar o espaço de medida de probabilidade $(B, \mathcal{B}_B, \mu^B)$.

O próximo resultado compila algumas propriedades úteis da probabilidade condicional.

Teorema 13.3.3. Seja $(\Omega, \mathcal{B}, \mu)$ um espaço de medida de probabilidade.

(1) Se $A, B \in \mathcal{B}$ e $\mu(B) > 0$, então $A \subset B \Rightarrow \mu(A|B) = \mu(A)/\mu(B)$ e $A \supset B \Rightarrow \mu(A|B) = 1$.
(2) Se $A, B, C \in \mathcal{B}$ e $\mu(B \cap C) > 0$, então $\mu(A|B \cap C) = \mu(A \cap B|C)/\mu(B|C)$.
(3) Se $D_1, D_2, \dots$ é uma sequência disjunta de elementos de $\mathcal{B}$ cuja união é Ω e $\mu(D_n) > 0$, para todo $n \in \mathbb{N}$, então para todo $A \in \mathcal{B}$,
$$\mu(A) = \sum_{n=1}^{+\infty} \mu(A|D_n)\mu(D_n),$$
e se $\mu(A) > 0$ então $\mu(D_n|A) = \mu(A|D_n)\mu(D_n)/\mu(A)$.
(4) Se $C_1, C_2, \dots$ é uma sequência monótona de elementos de $\mathcal{B}$, cujo limite é C e $\mu(C) > 0$, então, para todo $A \in \mathcal{B}$,
$$\mu(A|C) = \lim_{n\to+\infty} \mu(A|C_n).$$

13.3.4. Independência. Sejam $(\Omega, \mathcal{B}, \mu)$ um espaço de probabilidade, A e B dois acontecimentos mensuráveis com $0 < \mu(B) < 1$. Então podemos calcular as duas probabilidades condicionais:
$$\alpha := \mu(A|B), \qquad \beta := \mu(A|B^C),$$
em que α representa a probabilidade de A ocorrer dado que B ocorre e β a probabilidade de A ocorrer dado que B não ocorre. Suponhamos que $\alpha > \beta$, então A tem mais chances de ocorrer se B acontecer do que se B não acontecer. Conclui-se assim que a ocorrência de B influencia a probabilidade de A acontecer, isto é, o facto de B ocorrer ou não dá-nos informação extra relativamente à probabilidade de A ocorrer. O mesmo acontece quando $\alpha < \beta$; simplesmente, desta vez, é menos provável que A ocorra se B aconteceu do que ao contrário.

Se $\alpha = \beta$ a ocorrência do acontecimento B ou do seu complementar não alteram a probabilidade de A ocorrer. Digamos que não ganhamos informação alguma se soubermos de antemão se B ocorreu ou não. Por outras palavras ainda, podemos dizer que A é independente da ocorrência de B. Observemos que neste caso (em que $\alpha = \beta$) temos que:
$$\begin{aligned}\mu(A) &= \mu(A|B)\mu(B) + \mu(A|B^C)\mu(B^C) && \text{por (3) do Teorema 13.3.3}\\ &= \alpha(\mu(B) + \mu(B^C))\\ &= \alpha.\end{aligned}$$

Resulta daqui que $\mu(A) = \mu(A|B) = \mu(A \cap B)/\mu(B)$ e, consequentemente, $\mu(A \cap B) = \mu(A)\mu(B)$. Isto motiva a definição de independência estatística que se segue.

Definição 13.3.5. Sejam $(\Omega, \mathcal{B}, \mu)$ um espaço de probabilidade, A e B dois acontecimentos mensuráveis. Dizemos que A e B são *independentes* se
$$\mu(A \cap B) = \mu(A)\mu(B).$$

Estamos agora em condições de escrever aquelas que são conhecidas como leis aditiva e multiplicativa da probabilidade:

$$\mu(A \cup B) = \mu(A) + \mu(B) \qquad \text{se } A \text{ e } B \text{ forem disjuntos;}$$
$$\mu(A \cap B) = \mu(A)\mu(B) \qquad \text{se } A \text{ e } B \text{ forem independentes.}$$

De seguida, estendemos o conceito de independência a qualquer colecção de acontecimentos.

DEFINIÇÃO 13.3.6. Se $(\Omega, \mathcal{B}, \mu)$ é um espaço de probabilidade, uma colecção de acontecimentos $\mathcal{C}$ é dita independente se para qualquer subcolecção finita $\{E_1, E_2, \ldots, E_n\} \subset \mathcal{C}$ valer

$$\mu\left(\bigcap_{j=1}^{n} E_j\right) = \prod_{j=1}^{n} \mu(E_j).$$

13.3.5. Variáveis aleatórias. Até aqui introduzimos um modelo matemático para lidar com uma situação aleatória: o espaço de medida de probabilidade $(\Omega, \mathcal{B}, \mu)$. Os elementos de Ω representam possíveis resultados de uma experiência. Os elementos da σ-álgebra $\mathcal{B}$ representam os conjuntos mensuráveis, ou seja, acontecimentos cuja probabilidade de ocorrerem pode ser quantificada pela medida de probabilidade μ.

Consideremos um exemplo muito simples: suponhamos que são lançados dois dados equilibrados. Esta experiência tem 36 resultados possíveis que listamos abaixo. Definamos

$i \setminus j$	1	2	3	4	5	6
1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

TABELA 13.2. Espaço de acontecimentos

$\omega_{i,j} = (i, j)$, para $i, j \in \{1, 2, 3, 4, 5, 6\}$, onde i é o número representado na face voltada para cima do primeiro dado e j o número representado na face do segundo. O espaço de acontecimentos é então dado por $\Omega = \bigcup_{i,j\in\{1,2,3,4,5,6\}}\{\omega_{i,j}\}$. Cada uma das possíveis realizações da experiência tem probabilidade $1/36$ de ocorrer, quando supomos que o dado é equilibrado. Podemos tomar para σ-álgebra $\mathcal{B}$ a classe de todos os subconjuntos de Ω. A medida de probabilidade μ atribui a cada acontecimento A um peso igual a $n/36$ onde n é o número de elementos de Ω contidos em A $(n = 0, 1, \ldots, 36)$.

Pode acontecer, como é o caso de muitos jogos de tabuleiro, que não estejamos interessados em toda a informação resultante da realização da experiência. Suponhamos, por exemplo, que estamos apenas interessados na soma X dos pontos observados nas faces voltadas para cima dos dados. Esta quantidade pode tomar os valores $2, 3, \ldots, 12$, dependendo do resultado da realização da experiência. Este é um exemplo do que designamos por *variável aleatória*. Na realidade, X não é uma variável, mas antes uma função que a cada realização $\omega \in \Omega$ associa o número $X(\omega)$, que representa a soma dos pontos observados no lançamento dos dados. Mais precisamente, $X : \Omega \to \{2, 3, \dots, 12\}$ é dada por $X(\omega_{i,j}) = i + j$.

Uma característica fundamental de uma variável aleatória genérica X é o facto de ser sempre possível determinar a probabilidade de X tomar certos valores. Por exemplo, no caso anterior:

$$\mu(\{\omega \in \Omega : X(\omega) = 5\}) = \mu(\{\omega_{1,4}, \omega_{2,3}, \omega_{3,2}, \omega_{4,1}\}) = 4/36 = 1/9,$$

afirmação esta que habitualmente se abrevia escrevendo:

$$\mu(X = 5) = 1/9,$$

que é a probabilidade de obter soma dos pontos igual a 5 no lançamento de dois dados.

Um caso mais complicado é, por exemplo, a descrição do estado do tempo num determinado local específico a um dado instante. Claro que descrever em detalhe um acontecimento deste tipo é claramente impossível devido ao número elevado de factores que o influencia. Indicar, por exemplo, a posição e velocidade de todas as moléculas de ar num determinado local da atmosfera é irrealizável. Podemos contudo estar apenas interessados em determinadas quantidades, como a temperatura ou a pressão atmosférica médias, que obviamente estão sujeitas às arbitrariedades do sistema de partículas, por dependerem de todas as possíveis configurações das moléculas do ar nesse determinado local. Na realidade, não interessa conhecer exactamente a configuração do sistema, mas apenas uma quantidade observável tal como a temperatura correspondente a essa configuração, que certamente está sujeita a variação aleatória. À semelhança do que fizemos atrás, podemos definir uma função X de Ω (espaço de todas as configurações possíveis das moléculas de ar num determinado local à face da terra) em $\mathbb{R}$ que a cada observação instantânea da configuração ω (realização da experiência) associa a temperatura $X(\omega)$. O que pretendemos com esta definição é poder calcular, por exemplo, a probabilidade de X estar enquadrada entre determinados valores.

Mais geralmente podemos definir:

Definição 13.3.7. Dado um espaço de medida de probabilidade $(\Omega, \mathcal{B}, \mu)$, uma *variável aleatória* X é uma função $X : \Omega \to \mathbb{R}$ tal que, para todo o subconjunto $B \subset \mathbb{R}$ pertencente à σ-álgebra de Borel $\mathcal{F}$, se tem que $X^{-1}(B) \in \mathcal{B}$.

Esta definição implica que, para todo o conjunto $B \in \mathcal{F}$, podemos sempre calcular

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

uma vez que $X^{-1}(B)$ é mensurável por definição de variável aleatória. Habitualmente, costumamos omitir a dependência de ω na definição dos acontecimentos e, por exemplo, no caso particular em que $B = [a, b]$, com $a < b \in \mathbb{R}$, em vez de escrevermos $\mu(\{\omega \in \Omega : a \le X(\omega) \le b\})$, abreviamos escrevendo $\mu(a \le X \le b)$.

Observação 13.3.3. Toda a função contínua $f : J \to \mathbb{R}$, onde $J \subset \mathbb{R}$ é um intervalo equipado com a σ-álgebra de Borel, é uma variável aleatória.

Como variáveis aleatórias são funções, há uma série de operações que podemos fazer com elas. Por exemplo, se X e Y são variáveis aleatórias definidas em Ω, podemos definir

$X+Y$ e XY através de $(X+Y)(\omega) = X(\omega)+Y(\omega)$ e $XY(\omega) = X(\omega)Y(\omega)$. Para além disso, se $f:\mathbb{R}\to\mathbb{R}$ for mensurável então $Z:\Omega\to\mathbb{R}$ dada por $Z(\omega) = f(X(\omega))$ é também uma variável aleatória, se X o for.

DEFINIÇÃO 13.3.8. Uma sequência de variáveis aleatórias $X_1, X_2, \ldots$ definidas num espaço de medida de probabilidade $(\Omega, \mathcal{B}, \mu)$ diz-se um *processo estocástico*.

Um caso particular de um processo estocástico é aquele em que as variáveis aleatórias $X_1, X_2, \ldots$ são independentes.

DEFINIÇÃO 13.3.9. As variáveis aleatórias $X_1, X_2, \ldots$ definidas num espaço de medida de probabilidade $(\Omega, \mathcal{B}, \mu)$ dizem-se independentes se para toda a sequência de conjuntos $B_1, B_2, \ldots \in \mathcal{F}$ a colecção de acontecimentos $\{X_1^{-1}(B_1), X_2^{-1}(B_2), \ldots\}$ for independente.

EXEMPLO 13.3.5. Consideremos a experiência do lançamento de uma moeda ao ar, como no Exemplo 13.3.1, sendo $\Omega = \{F, C\}$ o espaço de acontecimentos. Podemos definir uma variável aleatória:

$$\begin{array}{rccc} X: & \Omega & \to & \mathbb{R} \\ & F & \mapsto & 0 \\ & C & \mapsto & 1. \end{array} \tag{13.4}$$

Os valores observáveis de X são $\{0,1\}$ e obviamente: $\mu(X=0) = \mu(\{F\}) = p$, $\mu(X = 1) = \mu(\{C\}) = 1-p$, $\mu(0 < X < 2) = \mu(X=1) = 1-p$, etc.

EXEMPLO 13.3.6. Tal como no Exemplo 13.3.3, consideremos a experiência que consiste em três lançamentos consecutivos de uma moeda equilibrada ao ar. Consideramos o espaço de acontecimentos $\Omega = \{FFF, FFC, FCF, FCC, CFF, CFC, CCF, CCC\}$, a σ-algebra $\mathcal{B}$ dada pela colecção de subconjuntos de Ω e a medida de probabilidade produto μ definida como no Exemplo 13.3.3, que a cada acontecimento elementar de Ω atribui probabilidade $1/2^3 = 1/8$. Tal como no Exemplo 13.3.3 usaremos a notação $\omega = \omega_1\omega_2\omega_3$, onde $\omega_i \in \{F, C\}$, $i = 1,2,3$, para escrever um elemento genérico de Ω. Suponhamos que, da realização da experiência em causa, estamos apenas interessados em contar o número de vezes que sai coroa. A melhor maneira de formalizarmos o problema, de modo a concentrarmonos na informação que pretendemos estudar, é introduzir uma variável aleatória que conte o número de coroas cada vez que se realiza a experiência, ou seja, definir: $S_3 : \Omega \to \mathbb{R}$, de tal forma que $S_3(\omega)$ é o número de ocorrências de C em ω. Os valores possíveis para S_3 são $\{0,1,2,3\}$, sendo que:

$$\begin{aligned} \mu(S_3 = 0) &= \mu(\{FFF\}) = 1/8 \\ \mu(S_3 = 1) &= \mu(\{FFC, FCF, CFF\}) = 3/8 \\ \mu(S_3 = 2) &= \mu(\{CCF, CFC, FCC\}) = 3/8 \\ \mu(S_3 = 1) &= \mu(\{CCC\}) = 1/8. \end{aligned}$$

Uma forma mais elegante de definirmos S_3 é usarmos a variável aleatória X definida em (13.4). Para cada $i = 1,2,3$, definamos a variável aleatória $X_i : \Omega \to \{0,1\}$ através de $X_i(\omega) = X_i(\omega_1\omega_2\omega_3) = X(\omega_i)$. Digamos que cada X_i é uma variável aleatória de contagem: se o i-ésimo lançamento da moeda corresponder a coroa, conta 1; enquanto que, se o i-ésimo lançamento da moeda corresponder a face, conta 0. Podemos então definir agora

$S_3 : \Omega \to \mathbb{R}$ tomando

$$S_3 = X_1 + X_2 + X_3.$$

Mais geralmente, no caso da experiência que consiste em lançamentos independentes e indefinidos de uma moeda equilibrada, tomamos, tal como no Exemplo 13.3.3, o espaço de acontecimentos $\Omega = \{F, C\}^{\mathbb{N}}$, a σ-algebra $\mathcal{B}$ gerada pelos cilindros definidos em (13.2) e a medida de probabilidade produto definida em (13.3). Consideremos a sequência de variáveis aleatórias $X_1, X_2, X_3, \ldots$ onde $X_i : \Omega \to \{0, 1\}$, para cada $i \in \mathbb{N}$, é dada por:

$$X_i(\omega) = X_i(\omega_1\omega_2\omega_3\ldots) = X(\omega_i).$$

O processo estocástico $X_1, X_2, X_3, \ldots$ assim definido é de facto uma sequência de variáveis aleatórias independentes identicamente distribuídas pela estrutura que a medida produto μ impõe. Efectivamente, a título de exemplo, observe-se que

$$\begin{aligned}\mu(X_1 = 0 \text{ e } X_2 = 1) &= \mu(C(2, 1, [F, C]) = 1/4 \\ &= \mu(C(1, 1, [F])\mu(C(1, 2, [C]) = \mu(X_1 = 0)\mu(X_2 = 1).\end{aligned}$$

Se estivermos interessados em contar o número de vezes que ocorrem coroas nos n primeiros lançamentos podemos definir, à semelhança do que fizemos atrás, $S_n : \Omega \to \{0, 1, \ldots, n\}$ dado por:

$$S_n(\omega) = \sum_{i=1}^{n} X_i(\omega).$$

Exemplo 13.3.7. Consideremos a experiência que consiste em lançamentos independentes e consecutivos de uma moeda equilibrada ao ar. Suponhamos que estamos interessados em contar o número de lançamentos necessários até sair coroa. Podemos formalizar considerando o espaço de todas as realizações possíveis que resultam do lançamentos consecutivos e indefinidos, i.e, $\Omega = \{F, C\}^{\mathbb{N}}$, munido com a σ-algebra gerada pelos cilindros definidos em (13.2) e com a medida produto μ definida por (13.3). Definamos agora a variável aleatória $Y : \Omega \to \mathbb{N} \cup \{+\infty\}$ dada por

$$\begin{aligned}&Y(\omega) = i \qquad \text{se } \omega \in C(i, 1, [F, F, \ldots, F, C]), \text{ para } i \in \mathbb{N} \\ &Y(\omega) = +\infty \qquad \text{se } \omega = FFF\ldots\end{aligned}$$

Observemos que a variável aleatória Y induz uma medida de probabilidade μ^* em $\Omega^* = \mathbb{N} \cup \{+\infty\}$ equipado com a σ-algebra constituída por todos os subconjuntos de Ω^*. Seja $B \subset \Omega^*$; definimos

$$\mu^*(B) = \mu(Y \in B) = \sum_{j \in B} \mu(Y = j) = \sum_{j \in B} \mu(C(j, 1, [F, F, \ldots, F, C])) = 1/2^j.$$

Por exemplo, $\mu^*(\{4\}) = \mu(Y = 4) = 1/2^4$, $\mu^*(\{+\infty\}) = \mu(Y = +\infty) = 0$ (veja-se Secção 13.3.2, Exemplo 13.3.3) e

$$\mu^*(\Omega^*) = \sum_{j=0}^{+\infty} \mu(Y = j) + \mu(Y = +\infty) = \sum_{j=0}^{+\infty} 1/2^j = 1.$$

Exemplo 13.3.8. Consideremos um arame de comprimento 1 metro que parametrizamos usando a distância a uma das extremidades que fixamos para origem. Desta forma, podemos identificar os pontos do arame com os pontos de $[0,1]$. Nesta identificação, $1/3$, por exemplo, é conotado com o ponto do arame que dista $1/3$ metro da extremidade fixada. Suponhamos que a temperatura em cada ponto do arame, num determinado momento (medida em ºC), é dada pela função $T:[0,1]\to\mathbb{R}$, tal que $T(x)=(x-1/2)^2$. Tal como no Exemplo 13.3.4, consideramos que $[0,1]$ está equipado com a σ-álgebra de Borel $\mathcal{B}$ e a medida de Lebesgue μ. Pela Observação 13.3.3, T é uma variável aleatória. Podemos então escolher aleatoriamente um ponto do intervalo $[0,1]$ tal como no Exemplo 13.3.4 e calcular a probabilidade da sua temperatura se enquadrar entre certos valores. Por exemplo, a probabilidade de a temperatura ser superior a $1/9$ é

$$\mu(T>1/9)=\mu\left(\{x:T(x)>1/9\}\right)=\mu\left([0,1/6)\cup(5/6,1]\right)=2/6.$$

Já a probabilidade da temperatura ser 0 é $\mu(T=0)=\mu(\{1/2\})=0$ (veja-se Secção 13.3.2, Exemplo 13.3.4).

13.4. Cadeias de Markov

Consideremos uma sequência de variáveis aleatórias $X_0,X_1,\ldots$ definidas num espaço de medida de probabilidade $(\Omega,\mathcal{B},\mathbb{P})$. Nesta secção, o conjunto de valores que as variáveis aleatórias $X_0,X_1,\ldots$ podem assumir será designado por *espaço de estados*. Consideraremos apenas o caso em que o espaço de estados é finito. A variável aleatória X_n denota o estado do sistema (ou do processo) no instante n.

Dizemos que o processo estocástico $X_0,X_1,\ldots$ é uma *cadeia de Markov* em tempo discreto se tiver a seguinte propriedade:

$$\mathbb{P}\left(X_n=i_n|X_0=i_0,\ldots,X_{n-1}=i_{n-1}\right)=\mathbb{P}\left(X_n=i_n|X_{n-1}=i_{n-1}\right).$$

Traduzindo por palavras, estamos perante uma cadeia de Markov se: para determinar probabilidade de, no instante n, o sistema se encontrar no estado i_n ($X_n=i_n$), o conhecimento de toda a história anterior, i.e., a sequência de estados pelos quais as v.a. X_j com $j<n$ passaram, nada mais acrescenta à informação resultante do conhecimento do estado do sistema no momento imediatamente anterior, i.e., ao conhecimento do estado em que a v.a. X_{n-1} se encontrava. Para ilustrar, imaginemos um jogador que faz apostas relativamente à realização de lançamentos independentes de um dado de 6 faces equilibrado e suponhamos que $X_0,X_1,\ldots$ é um processo estocástico em que X_n representa a riqueza do jogador após o n-ésimo lançamento. Pretendendo calcular a probabilidade de a riqueza do jogador ser um determinado valor após o n-ésimo lançamento, é fácil observar que, uma vez que os lançamentos são independentes, o conhecimento da evolução da sua riqueza desde que começou a jogar não acrescenta mais informação ao conhecimento da sua riqueza momentos antes de se realizar o dito lançamento.

A probabilidade do sistema estar no instante n no estado j, sabendo que no instante $n-1$ esteve no estado i é denotada por

$$p_{ij,n}:=\mathbb{P}\left(X_n=j|X_{n-1}=i\right),$$

e é chamada *probabilidade de transição de i para j no tempo n*. Se, $\forall k\in\mathbb{N}$ $p_{ij,k}=p_{ij,1}$, isto é, $\mathbb{P}\left(X_k=j|X_{k-1}=i\right)=\mathbb{P}\left(X_1=j|X_0=i\right)$, então dizemos que a cadeia de

Markov tem *probabilidades de transição estacionárias* e passamos a usar a notação $p_{ij} := \mathbb{P}(X_1 = j | X_0 = i)$. Neste secção, iremos tratar apenas cadeias de Markov com probabilidades de transição estacionárias.

Se uma cadeia de Markov tiver k estados possíveis que enumeramos por $\{1, 2, \ldots, k\}$, então as probabilidades de transição podem ser apresentadas numa matriz $P = [p_{ij}]$, dita *matriz de transição*, e que é tal que

(i) $p_{ij} \geq 0$, $\forall i, j \in \{1, \ldots, k\}$

(ii) $\sum_{j=1}^{k} p_{ij} = 1$, $\forall i \in \{1, \ldots, k\}$.

Por exemplo, uma cadeia de Markov de três estados tem matriz de transição da forma

$$\begin{pmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{pmatrix},$$

em que $p_{ij} \geq 0$, $\forall i, j \in \{1, 2, 3\}$, e $\sum_{j=1}^{3} p_{ij} = 1$, $\forall i \in \{1, 2, 3\}$. Nesta matriz, p_{32} representa a probabilidade do sistema mudar do estado 3 para o estado 2 e p_{11} é a probabilidade do sistema permanecer no estado 1 sabendo que no momento anterior já se encontrava no estado 1.

No caso geral, observa-se que o processo fica completamente determinado conhecidas a matriz de transição P e a distribuição de X_0, ou seja, conhecidas P e $p_i = \mathbb{P}(X_0 = i)$ para cada $i \in \{1, \ldots, k\}$. De facto,

$$\begin{aligned} &\mathbb{P}(X_0 = i_0, \ldots, X_n = i_n) \\ &= \mathbb{P}(X_n = i_n | X_0 = i_0, \ldots, X_{n-1} = i_{n-1}) . \mathbb{P}(X_0 = i_0, \ldots, X_{n-1} = i_{n-1}) \\ &= p_{i_{n-1} i_n} . \mathbb{P}(X_0 = i_0, \ldots, X_{n-1} = i_{n-1}) . \end{aligned}$$

Por indução, concluímos que

$$\begin{aligned} &\mathbb{P}(X_0 = i_0, \ldots, X_n = i_n) \\ &= p_{i_{n-1} i_n} \ldots p_{i_1 i_2} p_{i_0 i_1} \mathbb{P}(X_0 = i_0) \\ &= p_{i_{n-1} i_n} \ldots p_{i_1 i_2} p_{i_0 i_1} p_{i_0} \end{aligned}$$

Exemplo 13.4.1. Uma empresa de aluguer de automóveis tem três agências: 1, 2 e 3. Um cliente pode alugar um carro em qualquer uma das agências e devolvê-lo também a qualquer das três agências. Através de um estudo efectuado, estima-se que os clientes devolvam os automóveis às diferentes agências, de acordo com as seguintes probabilidades, que dependem do local onde o automóvel foi alugado:

$$\begin{pmatrix} 0.8 & 0.1 & 0.1 \\ 0.3 & 0.2 & 0.5 \\ 0.2 & 0.6 & 0.2 \end{pmatrix}.$$

Esta matriz é a matriz de transição da cadeia de Markov considerada. Nesta matriz, o valor 0.5 corresponde à probabilidade de um carro que é alugado na agência 2 ser devolvido à agência 3.

Exemplo 13.4.2. Um treinador de futebol verificou que a probabilidade de um jogador marcar golo na cobrança de um penalti depende do seu sucesso na marcação do penalti anterior, da forma que é descrita de seguida. A probabilidade do jogador marcar golo na cobrança de um penalti sabendo que marcou golo no penalti anterior é igual a 0.7, e a probabilidade de marcar golo sabendo que falhou o penalti anterior é igual a 0.2. A matriz de transição associada é:

$$\begin{pmatrix} 0.7 & 0.3 \\ 0.2 & 0.8 \end{pmatrix}.$$

Neste exemplo, a cadeia de Markov tem dois estados. Quando o jogador marca um golo de penalti podemos dizer que o sistema se encontra no estado 1, e quando falha o sistema encontra-se no estado 2.

Definamos agora *probabilidade de transição a n passos de i para j* do seguinte modo:

$$p_{ij}^{(n)} := \mathbb{P}\left(X_n = j | X_0 = i\right).$$

Usaremos a notação P^n para designar o produto matricial de P por si mesma repetido n vezes, i.e, $P^n = \underbrace{P \cdot P \cdot \ldots \cdot P}_{n \text{ vezes}}$. Por exemplo: $P^2 = P \cdot P$ onde $\cdot$ denota o produto de matrizes. (Veja-se Secção 1.3 de (Anton e Rorres, 2005)).

Teorema 13.4.1. Seja $P = [p_{ij}]$ a matriz de transição de uma cadeia de Markov. Então, a probabilidade de transição a n passos de i para j, $p_{ij}^{(n)}$, coincide com a entrada i, j da matriz P^n.

Consideremos uma cadeia de Markov com k estados. Chamamos *vector de estados da cadeia no momento n* ao vector linha de dimensão k, cuja i-ésima componente é a probabilidade do sistema estar, no momento n, no estado i. Notamos que todas as entradas de um vector de estados são não negativas e que a sua soma é igual a 1. O teorema anterior permite-nos obter os vectores de estados nos momentos 1,...,n,...:

$$x^{(1)}, \ldots, x^{(n)}, \ldots,$$

a partir do vector de estados no momento inicial $x^{(0)}$. O teorema anterior pode pois ser reescrito da seguinte forma:

Teorema. Seja $P = [p_{ij}]$ a matriz de transição de uma cadeia de Markov. Então, o vector de estados no tempo n, $x^{(n)}$, pode ser obtido a partir do vector de estados $x^{(0)}$, da seguinte forma:

$$x^{(n)} = x^{(0)} P^n.$$

Exemplo 13.4.1 (Continuação). Relembremos que, neste caso, da empresa de aluguer de automóveis, a matriz de transição é:

$$P = \begin{pmatrix} 0.8 & 0.1 & 0.1 \\ 0.3 & 0.2 & 0.5 \\ 0.2 & 0.6 & 0.2 \end{pmatrix}.$$

Suponhamos que, num determinado momento, um automóvel se encontra na agência 2. O vector de estados inicial é então $x^{(0)} = (0 \;\; 1 \;\; 0)$. A probabilidade de, no terceiro aluguer consecutivo, o automóvel ser devolvido à agência 1 é dada por $x_1^{(3)}$ (ou, equivalentemente, igual a $p_{21}^{(3)}$), em que

$$\begin{aligned} x^{(3)} &= x^{(0)}P^3 \\ &= (0\ \ 1\ \ 0)\cdot\begin{pmatrix}0.8 & 0.1 & 0.1\\ 0.3 & 0.2 & 0.5\\ 0.2 & 0.6 & 0.2\end{pmatrix}\cdot\begin{pmatrix}0.8 & 0.1 & 0.1\\ 0.3 & 0.2 & 0.5\\ 0.2 & 0.6 & 0.2\end{pmatrix}\cdot\begin{pmatrix}0.8 & 0.1 & 0.1\\ 0.3 & 0.2 & 0.5\\ 0.2 & 0.6 & 0.2\end{pmatrix} \\ &= (0.447\quad 0.252\quad 0.271), \end{aligned}$$

isto é, a probabilidade de, no terceiro aluguer consecutivo, o automóvel ser devolvido à agência 1 é igual a 0.447.

Exemplo 13.4.2 (Continuação). A matriz de transição deste exemplo é:

$$P = \begin{pmatrix}0.7 & 0.3\\ 0.2 & 0.8\end{pmatrix}.$$

Suponhamos que o jogador começa por marcar golo na cobrança do primeiro penalti. O vector de estados inicial é então $x^{(0)} = (1\quad 0)$. A probabilidade de, após mais quatro penaltis, o jogador falhar o último é dada por $x_2^{(4)}$ (ou, equivalentemente, igual a $p_{12}^{(4)}$), em que

$$\begin{aligned} x^{(4)} &= x^{(0)}P^4 \\ &= (1\ \ 0)\cdot\begin{pmatrix}0.7 & 0.3\\ 0.2 & 0.8\end{pmatrix}\cdot\begin{pmatrix}0.7 & 0.3\\ 0.2 & 0.8\end{pmatrix}\cdot\begin{pmatrix}0.7 & 0.3\\ 0.2 & 0.8\end{pmatrix}\cdot\begin{pmatrix}0.7 & 0.3\\ 0.2 & 0.8\end{pmatrix} \\ &= (0.438\quad 0.562), \end{aligned}$$

isto é, a probabilidade de, após mais quatro penaltis, o jogador falhar o último é igual a 0.562.

Uma questão que surge naturalmente é a seguinte: Sob que condições é que $x^{(n)}$ converge quando $n \to +\infty$? A resposta a esta questão é o propósito do Teorema Limite de Cadeias de Markov que enunciamos em seguida. Antes, porém, apresentamos a seguinte definição:

Dizemos que uma matriz é *regular* se alguma potência inteira dela mesma tiver todas as entradas estritamente positivas, isto é, se existir um número inteiro positivo m tal que P^m tenha todas as entradas estritamente positivas.

As matrizes dos Exemplos 13.4.1 e 13.4.2 são regulares. Basta tomar na definição anterior, em ambos os casos, $m = 1$.

Exemplo 13.4.3. A matriz de transição

$$P = \begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}$$

não é regular. De facto, nenhuma potência de P tem todas as entradas estritamente positivas uma vez que, para todo k ímpar temos

$$P^k = \begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}$$

e, para todo k par,

$$P^k = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Notamos que, neste caso, se tomarmos $x^{(0)} = (1 \ \ 0)$, temos

$$x^{(0)} = (1 \ \ 0), \ \ x^{(1)} = (0 \ \ 1), \ \ x^{(2)} = (1 \ \ 0), \ \ x^{(3)} = (0 \ \ 1), \ \ \ldots$$

e portanto $x^{(n)}$ não converge quando $n \to +\infty$.

O Teorema limite para Cadeias de Markov descreve o comportamento de P^n quando $n \to +\infty$ para matrizes de transição regulares P.

Teorema 13.4.2. Se P for uma matriz de transição regular com k estados, então, quando $n \to +\infty$,

$$P^n \to Q = \begin{pmatrix} q_1 & q_2 & \ldots & q_k \\ q_1 & q_2 & \ldots & q_k \\ . & . & \ldots & . \\ . & . & \ldots & . \\ . & . & \ldots & . \\ q_1 & q_2 & \ldots & q_k \end{pmatrix},$$

onde os q_i $(i = 1, \ldots, k)$ são números positivos tais que $q_1 + q_2 + \ldots, q_k = 1$.

Com base neste teorema concluímos que, para uma matriz de transição regular, a probabilidade de transição a n passos de i para j, $p_{ij}^{(n)}$, tende para q_j quando $n \to +\infty$. Em particular, observamos que $\lim_{n\to+\infty} p_{ij}^{(n)}$ não depende de i, isto é, não depende do estado inicial do sistema.

Notemos agora que, sendo Q a matriz limite do teorema anterior e x um vector de estados, temos que

$$\begin{aligned} xQ &= (x_1 \ \ x_2 \ \ \ldots \ \ x_k) \begin{pmatrix} q_1 & q_2 & \ldots & q_k \\ q_1 & q_2 & \ldots & q_k \\ . & . & \ldots & . \\ . & . & \ldots & . \\ . & . & \ldots & . \\ q_1 & q_2 & \ldots & q_k \end{pmatrix} \\ &= ((x_1 + x_2 + \ldots + x_k)q_1 \quad (x_1 + x_2 + \ldots + x_k)q_2 \quad \ldots \quad (x_1 + x_2 + \ldots + x_k)q_k) \\ &= (x_1 + x_2 + \ldots + x_k)(q_1 \ \ q_2 \ \ \ldots \ \ q_k) \\ &= q, \end{aligned}$$

onde $q = (q_1 \ \ q_2 \ \ \ldots \ \ q_k)$. Pelo Teorema 13.4.2 e observações anteriores obtemos o seguinte:

Teorema 13.4.3. Se P for uma matriz de transição regular e x um vector de estados, então, quando $n \to +\infty$,

$$xP^n \to q,$$

onde $q = (q_1 \ \ q_2 \ \ \dots \ \ q_k)$ é um vector de probabilidade fixo, que não depende de n nem de x, e cujas entradas são todas estritamente positivas.

Ao vector q do teorema anterior chamamos *vector de estado de equilíbrio*. Este vector pode ser calculado usando o teorema que se segue.

Teorema 13.4.4. *O vector de estado de equilíbrio q de uma matriz de transição regular P é o único vector de probabilidade que satisfaz $qP = q$.*

Este teorema decorre da igualdade $P^nP = P^{n+1}$ e do facto de que P^n e P^{n+1} convergem para Q quando $n \to +\infty$. Deste modo temos $QP = Q$, donde se conclui que $qP = q$. Para mostrar que q é o único vector de probabilidade que satisfaz a equação, suponhamos que existe um outro vector de probabilidade r tal que $rP = r$. Então temos $rP^n = r$ para $n = 1, 2, \dots$. Fazendo $n \to +\infty$, pelo Teorema 13.4.3 temos que $q = r$.

Exemplo 13.4.1 (Continuação). Vamos determinar o vector de estado de equilíbrio. Determinemos q tal que $qP = q$, ou, equivalentemente, $q(P - I) = 0$, que corresponde a determinar q_1, q_2 e q_3 tais que

$$(q_1 \ \ q_2 \ \ q_3)\begin{pmatrix} -0.2 & 0.1 & 0.1 \\ 0.3 & -0.8 & 0.5 \\ 0.2 & 0.6 & -0.8 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$$\Leftrightarrow \begin{cases} -0.2q_1 + 0.3q_2 + 0.2q_3 = 0 \\ 0.1q_1 - 0.8q_2 + 0.6q_3 = 0 \\ 0.1q_1 + 0.5q_2 - 0.8q_3 = 0 \end{cases} .$$

A solução deste sistema é $\begin{cases} q_1 = \frac{34}{13}q_3 \\ q_2 = \frac{14}{13}q_3 \end{cases}$. Impondo agora que o vector q seja um vector de probabilidade e, portanto, $q_1 + q_2 + q_3 = 1$, obtemos

$$q = \begin{pmatrix} \frac{34}{61} & \frac{14}{61} & \frac{13}{61} \end{pmatrix} = (0.5573\dots \quad 0.2295\dots \quad 0.2131\dots) .$$

Concluímos assim que, por exemplo, $\lim_{n\to+\infty} p_{i1}^{(n)} = \frac{34}{61}$, isto é, a probabilidade de um automóvel ser devolvido à agência 1 a longo prazo é $\frac{34}{61}$ (independentemente da agência onde foi alugado da primeira vez).

Este tipo de conclusão pode ser deveras relevante para tomar certas decisões. Por exemplo, se a empresa de aluguer tiver uma frota de 1000 viaturas, as suas instalações devem ser concebidas ou reestruturadas para que existam pelo menos 558 lugares de estacionamento na agência 1, 230 na agência 2 e 214 na agência 3.

Exemplo 13.4.2 (Continuação). Vamos calcular, para este exemplo, o vector de estado de equilíbrio. Determinemos, como no exemplo anterior, q tal que $qP = q$, ou, equivalentemente, $q(P - I) = 0$, que corresponde a determinar q_1 e q_2 tais que

$$(q_1 \quad q_2)\begin{pmatrix} 0.7 & 0.3 \\ 0.2 & 0.8 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\Leftrightarrow \begin{cases} -0.3q_1 + 0.2q_2 = 0 \\ 0.3q_1 - 0.2q_2 = 0 \end{cases}.$$

A solução deste sistema é $q_1 = \frac{2}{3}q_2$. Impondo agora que o vector q seja um vector de probabilidade e, portanto, $q_1 + q_2 = 1$, obtemos $q = (0.4 \quad 0.6)$.

Temos, por exemplo, que $\lim_{n\to+\infty} p_{i2}^{(n)} = 0.6$, isto é, a probabilidade do jogador não marcar golo de penalti a longo prazo é 0.6 (independentemente de ter marcado golo ou falhado no primeiro penalti).

13.5. Teorema de Recorrência de Poincaré

Seja $(\Omega, \mathcal{B}, \mu)$ um espaço de probabilidade e $T : \Omega \to \Omega$ uma aplicação mensurável, isto é, uma aplicação tal que, para todo $B \in \mathcal{B}$, $T^{-1}(B) \in \mathcal{B}$.

Definição 13.5.1. A aplicação $T : \Omega \to \Omega$ diz-se μ-invariante se para todo $B \in \mathcal{B}$ tivermos que $\mu(T^{-1}(B)) = \mu(B)$.

Por outras palavras, se assumirmos que T representa o efeito da passagem do tempo sobre Ω, então T ser μ invariante diz-nos que a maneira de medir a probabilidade de acontecimentos permanece inalterada com a passagem do tempo. Elaborando sobre esta analogia, se assumirmos que $T(x)$ representa a posição do ponto x após a passagem de um dia, então T é μ-invariante se a probabilidade de amanhã estarmos em B ($x \in T^{-1}(B)$) for igual à probabilidade de hoje estarmos em B ($x \in B$).

Usaremos a notação T^k para designar a composição de T consigo mesma k vezes, i.e, $T^k = \underbrace{T \circ T \circ \ldots \circ T}_{k \text{ vezes}}$. Por exemplo: $T^2 = T \circ T$ onde $\circ$ denota a composição de funções.

Definição 13.5.2. Seja $B \in \mathcal{B}$. Um ponto $x \in B$ diz-se *recorrente* com respeito a B se existir $k \in \mathbb{N}$ tal que $T^k(x) \in B$.

Um ponto de B é recorrente se em algum instante futuro retornar a B. Um dos primeiros e mais simples resultados de Teoria Ergódica afirma que quase todo o ponto é recorrente. Foi provado por Poincaré, que é considerado o percursor do estudo de Sistemas Dinâmicos.

Teorema 13.5.1 (Teorema de Recorrência de Poincaré – 1899). Consideremos um espaço de probabilidade $(\Omega, \mathcal{B}, \mu)$ e uma aplicação $T : \Omega \to \Omega$ mensurável e μ-invariante. Para todo $B \in \mathcal{B}$, se F denotar o conjunto dos pontos de B não recorrentes, então $\mu(F) = 0$.

Demonstração. Comecemos por escrever

$$F = B - \left(\bigcup_{k=1}^{+\infty} T^{-k}B\right) = B \cap T^{-1}(\Omega - B) \cap T^{-2}(\Omega - B) \cap \ldots$$

e observemos que se $x \in F \subset B$ então $T^n(x) \notin B$ para todo $n \in \mathbb{N}$. Em particular, $F \cap T^{-n}(F) = \emptyset$, para todo $n \in \mathbb{N}$, uma vez que se $x \in F \cap T^{-n}(F)$ então $x \in F \subset B$ e $T^n(x) \in F \subset B$. Aplicando o operador T^{-k} a ambos os conjuntos da última igualdade, concluímos que $T^{-k}(F) \cap T^{-(n+k)}(F) = \emptyset$, para todo $n \in \mathbb{N}$ e $k \in \mathbb{N}$. Significa então que os conjuntos $F, T^{-1}(F), T^{-2}(F), T^{-3}(F), \ldots$ são disjuntos e, por invariância da medida de probabilidade, para todo $n \in \mathbb{N}$, temos $\mu(T^{-n}(F)) = \mu(F)$. Se supusermos que $p = \mu(F) > 0$ então facilmente chegamos a um absurdo pois $\bigcup_{k=0}^{+\infty} T^{-k}(F) \subset \Omega$ e

$$\mu\left(\bigcup_{k=0}^{+\infty} T^{-k}(F)\right) = \sum_{k=0}^{+\infty} p = +\infty > 1.$$

Consequentemente, $p = \mu(F) = 0$. $\square$

Podemos fazer a seguinte interpretação física deste resultado. Consideremos que Ω contém os possíveis estados de um sistema, que a σ-álgebra $\mathcal{B}$ representa a colecção de acontecimentos e que μ é a medida de probabilidade definida em $\mathcal{B}$ que permite especificar a probabilidade de se observar cada um dos diferentes estados. Suponhamos que o sistema dinâmico evolui em tempo discreto (podemos pensar que fazemos medições uma vez por unidade de tempo) e que $T : \Omega \to \Omega$ é a aplicação que descreve a evolução do sistema, indicando como se processa a transição de um estado para o sucedâneo ao fim de uma unidade de tempo. Numa situação de equilíbrio, T preserva a medida μ o que significa que a probabilidade de observar um dado estado não muda com o tempo. Nestas condições, o Teorema de Recorrência de Poincaré diz-nos que, se no instante inicial o sistema se encontrar num estado observável $E \in \mathcal{B}$ (com $\mu(E) > 0$), então, com probabilidade 1, o sistema retorna ao estado inicial E.

Consideremos a experiência em que montamos dois recipientes com uma ligação entre os dois, que pode ser aberta ou fechada através de uma válvula. Enchemos um dos recipientes com um gás e deixamos o outro vazio. Abrimos a válvula e deixamos o sistema evoluir. O Teorema de Recorrência permite-nos concluir que, quase certamente (com probabilidade 1), num instante futuro, o sistema voltará ao estado inicial em que todas as moléculas do gás estarão num dos recipientes.

O facto do Teorema implicar que um acontecimento tão improvável vai acontecer quase certamente parece desafiar a Segunda Lei da Termodinâmica que pode ser enunciada da seguinte forma:

"A quantidade de entropia de qualquer sistema isolado termodinamicamente tende a incrementar-se com o tempo, até alcançar um valor máximo. Mais sensivelmente, quando uma parte de um sistema fechado interage com outra parte, a energia tende a dividir-se por igual, até que o sistema alcance um equilíbrio térmico."

Observe-se que o Teorema de Recorrência de Poincaré nada afirma acerca do tempo que é necessário esperar até haver um retorno ao estado inicial. Apenas assevera que, quase certamente, haverá uma recorrência, sendo que esta poderá ocorrer apenas num instante futuro muito longínquo, o que dissipa a aparente incompatibilidade com a Segunda Lei da Termodinâmica.

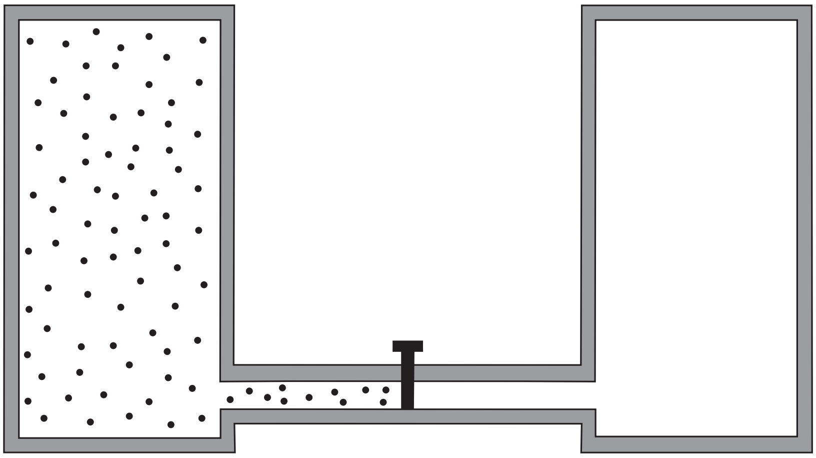

FIGURA 13.4. Experiência dos recipientes

Para melhor ilustrarmos a aplicação do Teorema de Recorrência de Poincaré a esta experiência dos recipientes e do gás, estudaremos o exemplo sugerido pelos Ehrenfests (1957), que envolve um número muito mais pequeno de partículas, mas que se revela adequado para criar entropia suficiente para entendermos a relação entre os dois princípios que aparentam estar em conflito. Um ingrediente importante e que será usado no final é o Teorema de Kac que, em linhas gerais, afirma o seguinte:

Se o sistema dinâmico tiver boas propriedades, então o valor esperado do tempo que o sistema demora a retornar a um dado estado observável é o inverso da probabilidade de ocorrer esse estado.

Consideremos o jogo (sistema) em que existem duas urnas: a urna 1 que contém 100 bolas numeradas de 1 a 100 e a urna 2 que se encontra vazia. Num saco colocam-se 100 pedaços de papel numerados de 1 a 100. Em cada unidade de tempo retiramos um papel do saco, lemos o número nele inscrito, recolocamo-lo no saco e movemos a bola que tem esse número da urna onde se encontra para a outra. A Segunda Lei da Termodinâmica, assim como a nossa intuição, indicam-nos que o sistema evoluirá para o estado de equilíbrio que maximiza a entropia, em que existem 50 bolas em cada urna. Certamente, haverão flutuações aleatórias em torno da divisão $50-50$, mas parece altamente improvável que a flutuação seja tão grande que as 100 bolas retornem todas à urna 1. O Teorema de Recorrência afirma que, apesar de parecer muito pouco verosímel, ela acontecerá quase certamente.

Podemos descrever o estado do sistema no instante $k \in \mathbb{N}_0$ especificando o número $\omega_k \in \{0, 1, \dots, 100\}$ de bolas na urna 1 nesse instante. Se no instante $k = 0$ existirem ω_0 bolas na urna 1 (na experiência atrás descrita foi considerado que $\omega_0 = 100$), se prosseguirmos com a extracção de papéis do saco e procedermos em conformidade com as regras do jogo, o sistema passará sucessivamente pelos estados $\omega_0, \omega_1, \omega_2, \dots$ sujeitos às seguintes condições:

$$\omega_k \in \{0, 1, \dots, 100\}, \qquad |\omega_k - \omega_{k+1}| = 1, \qquad \forall k = 0, 1, 2, \dots$$

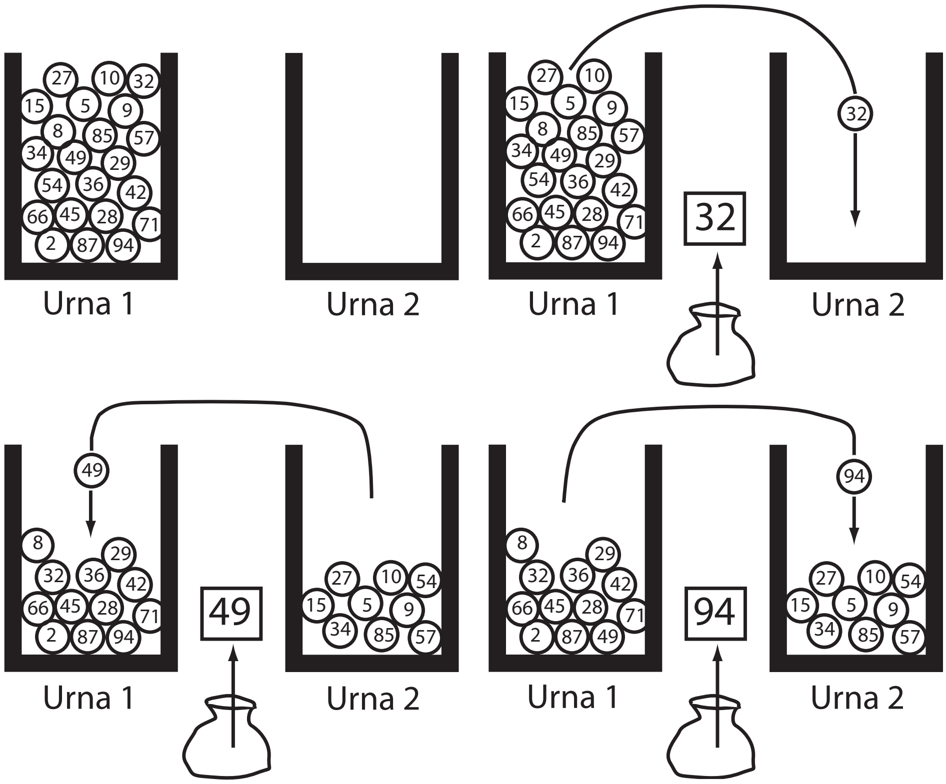

FIGURA 13.5. Exemplo dos Ehrenfests

Vamos considerar, mais geralmente, que o jogo começou num instante passado remoto indefinido e se perpetua indefinidamente para o futuro. Tal como no exemplo 13.3.3 consideramos o espaço produto $\Omega = \{0, 1, , \dots 100\}^{\mathbb{Z}}$ constituído pelas palavras

$$\omega = \dots \omega_{-2}\omega_{-1}\omega_0\omega_1\omega_2 \dots$$

compostas por sequências bilaterais infinitas de símbolos do alfabeto $\{0, 1, \dots 100\}$. Podemos identificar cada palavra ω com uma função $\omega : \mathbb{Z} \to \{0, 1, \dots, 100\}$ em que $\omega(n) = \omega_n$. Seja $\sigma : \Omega \to \Omega$ a transformação *shift* que actua em cada palavra deslocando os símbolos uma casa para a esquerda, $(\sigma(\omega))(n) = \omega(n + 1)$ para cada $n \in \mathbb{Z}$. Tal como no Exemplo 13.3.3, seja $\mathcal{B}$ a σ-algebra gerada pelos cilindros compostos por segmentos finitos de história:

$$C(n, k, [i_1, i_2, \dots, i_n]) = \{\omega \in \Omega : \omega_k = i_1, \omega_{k+1} = i_2, \dots, \omega_{k+n-1} = i_n\}$$

Para definirmos uma medida de probabilidade invariante para o shift em $\mathcal{B}$ compatível com as transições do jogo, consideremos a cadeia de Markov $\dots, X_{-1}, X_0, X_1, \dots$ onde, para cada $l \in \mathbb{Z}$, a variável aleatória $X_l : \Omega \to \{0, 1, \dots, 100\}$ é tal que $X_l(\omega) = \omega_l$. Se o sistema estiver no estado i significa que existem i bolas na urna 1; então só há duas transições possíveis: o sistema passa para o estado $i - 1$ ou para o estado $i + 1$, consoante o número no papel retirado do saco corresponda a uma bola da urna 1 ou da urna 2, respectivamente. Claro que se o sistema se encontrar no estado i, a probabilidade de se retirar um papel com um número de uma bola que se encontra na urna 1 é $\frac{i}{100}$ e na urna

2 é $\frac{100-i}{100}$. É assim claro que

$$P(X_1 = j|X_0 = i) = \begin{cases} i/100 & \text{se } j = i-1 \\ (100-i)/100 & \text{se } j = i+1 \\ 0 & \text{caso contrário} \end{cases}.$$

Obtemos desta forma a matriz de transição:

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & \cdots & \cdots & 0 \\ \frac{1}{100} & 0 & \frac{99}{100} & 0 & 0 & \cdots & \cdots & 0 \\ 0 & \frac{2}{100} & 0 & \frac{98}{100} & 0 & \cdots & \cdots & 0 \\ 0 & 0 & \frac{3}{100} & 0 & \frac{97}{100} & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \cdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \frac{99}{100} & 0 & \frac{1}{100} \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}.$$

Seja p o vector de dimensão 101, cuja i-ésima componente é dada por

$$p_i = C_i^{100}\frac{1}{2^{100}}, \qquad i \in \{0, 1, \dots, 100\}, \tag{13.5}$$

onde $C_i^{100} = \frac{100!}{i!(100-i)!}$, ou seja, em que p_i corresponde à probabilidade de calharem exactamente i bolas na urna 1 depois de as distribuirmos aleatoriamente pelas duas urnas, atribuindo peso 1/2 a cada urna. É fácil ver que p é um estado de equilíbrio do processo de Markov, i.e., $pA = p$. Então, podemos definir uma medida μ nos cilindros da seguinte forma:

$$\mu(C(n, k, [i_1, i_2, \dots, i_n]) = p_{i_1} a_{i_1 i_2} a_{i_2 i_3} \dots a_{i_{n-1} i_n}, \tag{13.6}$$

para todo $k \in \mathbb{Z}$, $n \in \mathbb{N}$ e $i_1, \dots, i_n \in \{0, 1, \dots, 100\}$.

Notemos que μ tem que ser em particular σ-aditiva, o que é consequência de $pA = p$. Para ilustrar este facto, consideremos por exemplo o cilindro $C(1, 1, [j])$, para algum $j \in \{0, 1, \dots, 100\}$, e observemos que $C(1, 1, [j]) = \bigcup_{i=0}^{100} C(2, 0, [i, j])$ e $C(2, 0, [i, j]) \cap C(2, 0, [\ell, j]) = \emptyset$ para todo $\ell \neq i$. Para termos σ-aditividade temos que ter $\mu(C(1, 1, [j])) = \sum_{i=0}^{100} \mu(C(2, 0, [i, j]))$, para todo $j \in \{0, 1, \dots 100\}$, ou seja, $p_j = \sum_{i=0}^{100} p_i a_{ij}$. Resulta assim a necessidade de $pA = p$ para definirmos a medida μ. Uma vez definida a medida nos cilindros é possível estendê-la univocamente à σ-álgebra gerada pelos mesmos. Para além disso, a medida de probabilidade é invariante para o "shift" já que por definição da medida μ nos cilindros temos que (13.6) vale para todo $k \in \mathbb{Z}$.

Consideremos agora o acontecimento $E = \{\omega \in \Omega : X_0(\omega) = 100\} = C(1, 0, [100])$, que consiste nas realizações do jogo em que estamos interessados, pois corresponde às experiências em que a primeira urna se encontra cheia no início da contagem do tempo. A probabilidade de E ocorrer é $1/2^{100}$, uma vez que por definição: $\mu(E) = \mu(C(1, 0, [100])) = p_{100} = \binom{100}{100} \cdot 1/2^{100} = 1/2^{100} > 0$. Pelo Teorema de Recorrência de Poincaré temos que a probabilidade de um ponto de E ser não recorrente a E é 0. Por outras palavras:

$$\mu(\{\omega \in E : \exists k \in \mathbb{N} \quad \sigma^k(\omega) \in E\}) = \mu(\{\omega \in E : \exists k \in \mathbb{N} \quad X_k(\omega) = 100\}) = \mu(E),$$

o que significa que, com probabilidade 1 (probabilidade condicionada ao facto de sabermos que no instante inicial a urna 1 está cheia), o sistema volta ao estado inicial em que existem 100 bolas na urna 1!

Esta aparente dissonância com a Segunda Lei da Termodinâmica (e com a nossa intuição de que o sistema tende a evoluir para um "equilíbrio" compreendendo pequenas oscilações em torno do estado 50) dissipa-se à luz do Teorema de Kac, que afirma que o tempo médio de retorno à posição inicial é $1/\mu(E) = 2^{100}$ unidades de tempo. Se supusermos que cada transição ocorre ao fim de um milésimo de segundo, então o tempo médio de retorno ao estado em que a urna 1 está cheia é 2^{100} milésimo de segundo que, após uma pequena conta, se verifica ser mais que o tempo do Universo, estimado em $13,8 \times 10^9$ ano! Se agora imaginarmos o caso em que em vez de 100 bolas temos uma mole delas (aproximadamente 6×10^{23}), então o tempo médio de espera, até se verificar uma recorrência ao estado em que a urna 1 está cheia, seria ainda muito maior.

Como é fácil verificar em (13.5), o estado mais provável é o estado 50, i.e., aquele em que as duas urnas têm o mesmo número de bolas. Efectivamente, $\mu(\{\omega \in \Omega : X_0(\omega) = 50\}) = 7.96 \times 10^{-2}$, o que significa que, se o sistema se encontrar no estado 50, então o tempo médio de retorno a esse mesmo estado é aproximadamente 13 unidades de tempo. Isto explica a nossa intuição ao acharmos que o sistema entra num equilíbrio aparente em torno do estado em que as duas urnas têm o mesmo número de bolas.

Bibliografia

[1] Anton, H.; Rorres, C., *Elementary Linear Algebra with Applications*, John Wiley & Sons, Inc. 9th Edition, 2005.

[2] Feller, W., *An introduction to Probability Theory and Applications*, John Wiley & Sons, Inc., New York and London, 1950.

[3] Gonçalves, E.; Lopes, N.M., *Probabilidades. Princípios teóricos.*, Escolar Editora, 2000.

[4] Karlin, S., *A first course in stochastic processes*, Academic Press, New York and London, 1969.

[5] Kingman, J.F.C.; Taylor, S. J., *Introduction to measure and probability*, Cambridge University Press, London-New York-Ibadan, 1966.

[6] Petersen, K., *Ergodic Theory*, Cambridge Studies in Advanced Mathematics, 2, Cambridge University Press, Cambridge, 1989.

[7] Walters, P., *An introduction to ergodic theory*, Graduate Texts in Mathematics, 79, Springer-Verlag, New York-Berlin, 1982.