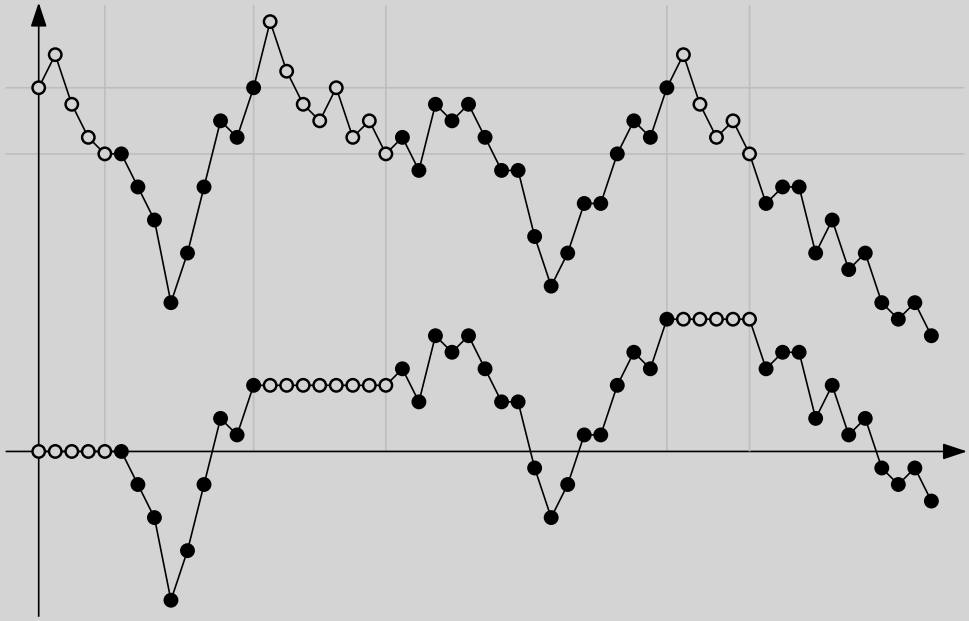
Probabilidade



Leonardo T. Rolla

Bernardo N. B. de Lima

© 2021–2025
Leonardo T. Rolla
Bernardo N. B. de Lima

Todos os direitos reservados. Permitido o uso nos termos licença Creative Commons Atribuição-SemDerivações 4.0 Internacional. Qualquer uso além daquele descrito nessa licença é expressamente proibido. Em particular, é proibido distribuir versões modificadas ou extrair trechos isolados deste material.

18 de março de 2025.

Prefácio

Este livro foi produzido a partir de apostilas e notas manuscritas acumuladas pelos autores ao longo de quase duas décadas lecionando diversas disciplinas de Probabilidade, Medida e Integração, Processos Estocásticos e Teoria Ergódica, em níveis de graduação, mestrado e doutorado, no IMPA, UFMG, PUC-Rio, NYU-Shanghai, Warwick e USP.

O objetivo principal é servir como referência para um curso de Probabilidade em nível de pós-graduação, mas procuramos torná-lo o mais flexível e autocontido possível. Dependendo das seções a serem cobertas, este livro pode ser usado em um curso de início de doutorado, mestrado, ou mesmo no fim da graduação. Em um nível mais básico, o livro pode ser estudado cobrindo-se apenas as seções iniciais dos onze primeiros capítulos.

Os capítulos mais avançados cobrem as Leis 0-1 de Kolmogorov e de Hewitt-Savage, martingales a tempo discreto com teoremas envolvendo amostragem opcional e convergência, noções básicas de Teoria Ergódica e de Grandes Desvios. Alguns tópicos fundamentais não são abordados neste livro, notadamente: convergência de medidas em espaços métricos, cadeias de Markov e processos estocásticos em tempo contínuo.

Ao escrever o livro, os autores tentaram manter um certo nível de modularidade, de forma a permitir que diversas seções possam ser saltadas sem prejudicar a leitura das que vêm depois. Para planejar um curso, pode-se montar uma lista de seções a gosto, e depois percorrer a lista de trás para

frente agregando-se os pré-requisitos indicados na Tabela 1.

Os pré-requisitos para a leitura em um nível mais básico são o cálculo de derivadas e integrais em $\mathbb{R}^n$, limites de sequências, convergência de séries, e limites laterais de funções. Há uma breve revisão desses conceitos no Apêndice A.1.

Para seguir algumas demonstrações e tópicos mais avançados, o leitor deve estar familiarizado com ideias de Análise Real, como conjuntos abertos, conjuntos compactos, conjuntos enumeráveis, supremo de conjuntos, $\limsup$ e $\liminf$. Esses conceitos são listados de forma lacônica no Apêndice A.3.

Os conceitos de Teoria da Medida que servem para dar bases sólidas à Teoria da Probabilidade serão vistos gradualmente nas Seções 1.4, 3.6, 3.7, 5.5, 11.5 e 13.1. As demonstrações mais longas são vistas no Apêndice D.

Por inúmeros comentários, correções e sugestões, agradecemos a Maria Eulália Vares, Daniel Valesin, Pablo Groisman, Rodrigo Targino, Shivakumar Mahesh, Rangel Baldasso, Henrique Felix, Tertuliano Franco, e aos vários alunos de cursos ministrados pelos autores utilizando versões preliminares deste livro. Agradecemos especialmente a Célio Terra pela leitura minuciosa do texto e dos exercícios.

18 de março de 2025.

Tabela 1. Pré-requisitos do livro

1.1 - A.1	7.2 - 7.1	13.1 - 11.4
1.2 - 1.1	7.3 - 7.2, 5.3*	13.2 - 13.1
1.3 - 1.2	7.4 - 7.1, 5.3, C.2	13.3 - 13.2
1.4 - 1.3, C.2	8.1 - 7.1	13.4 - 13.3
2.1 - 1.3	8.2 - 7.2, 7.3*	13.5 - 13.4
2.2 - 2.1	8.3 - 8.2, 5.3, C.2*	14.1 - 13.1
3.1 - 2.2	9.1 - 8.1, 4.3*	14.2 - 14.1
3.2 - 3.1	9.2 - 3.2, A.2	14.3 - 14.2
3.3 - 3.2	9.3 - 9.1, 7.4*, A.2	14.4 - 14.3
3.4 - 3.3	9.4 - 9.1, 5.5, 7.4*, A.2	14.5 - 14.3
3.5 - 3.3	9.5 - 9.1, 7.4*	15.1 - 5.5, 8.2, 10.1
3.6 - 3.1, 1.4	10.1 - 9.1, 5.3*	15.2 - 15.1
3.7 - 3.6	10.2 - 9.1, 5.3*, A.2	15.3 - 15.2
4.1 - 3.3	10.3 - 10.2, 7.4, 5.5*	A.1 - —
4.2 - 4.1	10.4 - 10.3, 5.5	A.2 - A.1
4.3 - 4.2	11.1 - 5.4, 6.3	A.3 - A.1
4.4 - 4.2	11.2 - 11.1	B - A.2
4.5 - 4.2	11.3 - 11.2	C.1 - A.1
5.1 - 4.2	11.4 - 5.5, 6.3	C.2 - C.1, A.3
5.2 - 5.1, C.1	11.5 - 11.4	D.1 - 3.6
5.3 - 5.2, C.2*	11.6 - 11.3, 11.4	D.2 - D.1
5.4 - 5.2, 3.4	12.1 - 11.4	D.3 - 3.7
5.5 - 5.3, 3.7	12.2 - 12.1	D.4 - 5.5, D.1
6.1 - 5.2	12.3 - 12.2	D.5 - 11.5
6.2 - 6.1	12.4 - 12.2	D.6 - 11.5, D.1
6.3 - 6.1	12.5 - 12.4	D.7 - 5.5
7.1 - 6.3	12.6 - 12.1	

"*" indica que a seção listada não é imprescindível, tampouco seus pré-requisitos.

Sumário

Prefácio **5**

1 Espaços de Probabilidade **17**

1.1 Alguns modelos probabilísticos 18

1.1.1 Espaço amostral 21

1.1.2 Eventos aleatórios 22

1.1.3 Medida de probabilidade 24

1.2 Contagem e simetria 31

1.3 Formulação axiomática de Kolmogorov 41

1.4 Espaços de medida 44

1.4.1 σ-álgebras e conjuntos borelianos 44

1.4.2 Medidas 47

1.5 Exercícios 48

2 Probabilidade Condicional e Independência **55**

2.1 Probabilidade condicional 55

2.1.1 Regra do Produto 57

2.1.2 Lei da Probabilidade Total 59

2.1.3 Fórmula de Bayes 60
2.2 Independência 62
2.3 Exercícios 65

3 Variáveis Aleatórias 71
3.1 Variáveis aleatórias 71
3.1.1 Espaço induzido e lei de uma variável aleatória 73
3.1.2 Função de distribuição 74
3.1.3 Função quantil 78
3.2 Variáveis aleatórias discretas 80
3.3 Variáveis aleatórias absolutamente contínuas 86
3.4 Distribuição condicional dado um evento 92
3.5 Distribuições mistas e singulares 93
3.6 Existência e unicidade de distribuições 96
3.7 Funções mensuráveis 97
3.8 Exercícios 101

4 Vetores Aleatórios 109
4.1 Vetores aleatórios 109
4.2 Vetores aleatórios discretos e contínuos 115
4.3 Soma de variáveis independentes 119
4.4 Método do jacobiano 123
4.5 Sequência de variáveis independentes 126
4.6 Exercícios 129

5 Esperança Matemática 135

5.1 Variáveis aleatórias simples 136
5.2 Esperança matemática 142
5.3 Aproximação e convergência da esperança 152
5.3.1 Teorema da Convergência Monótona 153
5.3.2 Teorema da Convergência Dominada 156
5.4 Esperança condicional dado um evento 158
5.5 Integral de Lebesgue 161
5.5.1 Construção 162
5.5.2 Principais propriedades 167
5.5.3 Convergência 170
5.5.4 Integral de Riemann e integral imprópria 172
5.5.5 Densidade de medidas 174
5.5.6 Espaços produto e integrais iteradas 177
5.6 Exercícios 183

6 Momentos e Desigualdades 189
6.1 Momentos e variância 189
6.2 Correlação 193
6.3 Desigualdades básicas 198
6.4 Exercícios 204

7 Convergência de Variáveis Aleatórias 207
7.1 Modos de convergência 207
7.2 Lema de Borel-Cantelli 212
7.3 Relações entre os modos de convergência 217
7.4 Mais sobre a convergência em distribuição 222

7.5 Exercícios 225

8 Lei dos Grandes Números 231

8.1 Lei Fraca dos Grandes Números 232

8.2 Lei Forte dos Grandes Números 234

8.3 Leis Fortes de Kolmogorov 237

8.4 Algumas aplicações 240

8.5 Exercícios 246

9 Teorema do Limite Central 251

9.1 Teorema do Limite Central 251

9.2 Teorema de De Moivre-Laplace 254

9.3 Teorema do Limite Central de Lyapunov 258

9.4 Teorema do Limite Central de Lindeberg 261

9.5 Teorema de Slutsky e Método Delta 269

9.6 Exercícios 274

10 Transformadas 279

10.1 Função geradora de momentos 279

10.2 Função característica 284

10.3 Unicidade e convergência 291

10.4 Fórmula de inversão 296

10.5 Exercícios 299

11 Esperança Condicional 303

11.1 Esperança condicional dada uma partição 303

11.2 Função de probabilidade condicional 308

11.3 Densidade condicional 314
11.4 Esperança condicional dada uma σ-álgebra 319
11.5 Teorema de Radon-Nikodým 325
11.6 Distribuição condicional regular 328
11.7 Exercícios 334

12 Martingales **341**
12.1 Definições e exemplos 342
12.2 Tempos de parada 346
12.3 Amostragem opcional 351
12.4 Convergência quase certa de martingales 355
12.5 Convergência de martingales em $\mathcal{L}^p$ 361
12.6 Decomposição de Doob 367
12.7 Exercícios 369

13 Leis 0-1 e Séries Aleatórias **375**
13.1 Álgebras e espaços de sequências 376
13.2 Independência de σ-álgebras 378
13.3 Lei 0-1 de Kolmogorov 380
13.4 Lei 0-1 de Hewitt-Savage 382
13.5 Convergência de séries aleatórias 385
13.6 Exercícios 390

14 Teoria Ergódica **393**
14.1 Transformações que preservam medida 393
14.2 Teorema Ergódico de Birkhoff 397

14.3 Transformações ergódicas e misturadoras 401
14.4 Rotação do círculo 408
14.5 Teorema Ergódico de von Neumann 412
14.6 Exercícios 413

15 Grandes Desvios **417**
15.1 Desigualdade de concentração 418
15.2 Princípio dos Grandes Desvios 420
15.3 Demonstração do Teorema de Cramér 424

A Preliminares **429**
A.1 Cálculo 429
A.2 Expansão de Taylor 433
A.3 Análise Real 435

B Fórmula de Stirling **437**

C A Reta Real e o Infinito **443**
C.1 Reta estendida 443
C.2 Supremo e limite superior 445

D Elementos de Teoria da Medida **447**
D.1 Teorema π-λ de Dynkin 447
D.2 Teorema de Extensão de Carathéodory 451
D.3 Operações com funções mensuráveis 457
D.4 Teoremas de Fubini e de Tonelli 459
D.5 Teorema de Radon-Nikodým 464

D.6 Distribuição condicional regular 468
D.7 Desigualdades de Hölder e de Minkowski 473

Lista de Figuras **477**

Notação **479**

Bibliografia **481**

Tabela Normal **483**

Índice Remissivo **485**

Capítulo 1

Espaços de Probabilidade

A Teoria da Probabilidade estuda eventos *aleatórios*, isto é, eventos que não possuem *regularidade determinística*, mas possuem *regularidade estatística*. A ausência de regularidade determinística significa que observações feitas nas mesmas condições não dão o mesmo resultado, enquanto a regularidade estatística se manifesta na estabilidade estatística de frequências.

Por exemplo, no lançamento de uma moeda, apesar de a trajetória da moeda ser determinística do ponto de vista da mecânica Newtoniana, é impraticável tentar prever seu resultado: este experimento não possui regularidade determinística. No entanto, esse experimento possui regularidade estatística e o tratamento probabilístico é o mais adequado.

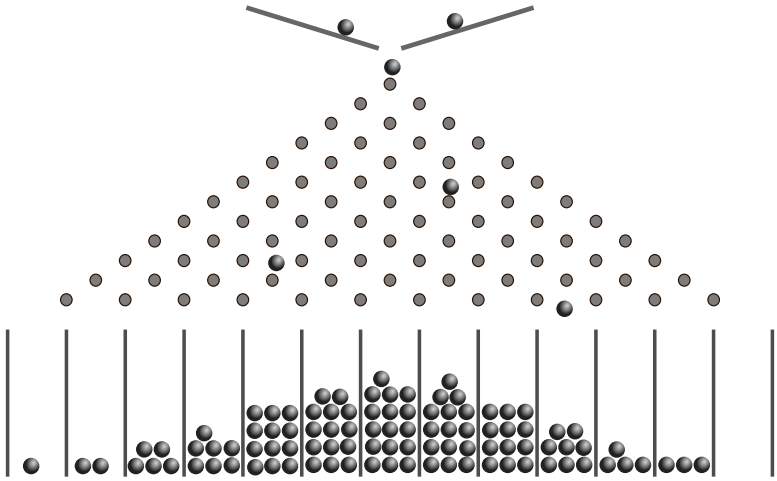
Um exemplo de regularidade estatística extremamente simples e ao mesmo tempo fantástico é o engenhoso Tabuleiro de Galton, ilustrado na Figura 1.1. Cada bolinha faz sua trajetória sem interagir com as demais e, apesar disso, cada vez que repetimos o experimento observamos o mesmo padrão.

Um modelo probabilístico é uma idealização que descreve a realização de um experimento cujo resultado é aleatório. O objeto matemático que representa modelos probabilísticos é um *espaço de probabilidade*. Este capítulo introduz o estudo de espaços de probabilidade e suas principais propriedades.

1.1 Alguns modelos probabilísticos

Antes de darmos qualquer definição formal, vamos tentar analisar algumas situações simples e tentar capturar se há feitos em comum nos exemplos a seguir:

Exemplo 1.1. Sorteamos uma carta de um baralho comum (52 cartas, numeradas $A, 2, 3, \ldots, 10, J, Q, K$ e de naipes $\clubsuit, \heartsuit, \spadesuit, \diamondsuit$). Qual a probabilidade de a carta sorteada ser $4\clubsuit$, $7\heartsuit$, $A\spadesuit$ ou $7\diamondsuit$? Nossa primeira tentativa de formalizar este problema seria reescrever a primeira frase da seguinte maneira: sorteamos um elemento do conjunto $\Omega = \{A\diamondsuit, 2\diamondsuit, \ldots, K\diamondsuit, A\heartsuit, \ldots, A\spadesuit, \ldots, A\clubsuit, \ldots, K\clubsuit\}$. Assim a pergunta passaria a ser: qual a probabilidade deste elemento pertencer ao subconjunto $A = \{4\clubsuit, 7\heartsuit, A\spadesuit, 7\diamondsuit\}$? Melhor ainda, qual a proporção que o conjunto A ocupa dentro de Ω? A princípio, é razoável que esta "proporção" entre os



Crédito: Merlijn van Deen em Wikimedia Commons, licença CC-BY, com modificações pelos autores.

Figura 1.1. Tabuleiro de Galton. As bolinhas colidem com pinos, dispostos em 12 níveis, que as defletem para direita ou esquerda. Quando há muitas bolinhas, o padrão observado é sempre o mesmo. Esse padrão é descrito (quando há muitos níveis de pregos) pela chamada curva gaussiana, exibida na página 254.

"tamanhos" dos conjuntos seja dada pela razão[1] entre o número de elementos de A e de Ω, isto é:

$$\mathbb{P}(A) = \frac{\#A}{\#\Omega} = \frac{4}{52} = \frac{1}{13}. \qquad \triangle$$

A notação $\#A$ se refere à quantidade de elementos que pertencem ao conjunto A.

Exemplo 1.2. Um baile ocorre em um grande salão, cujo piso é formado por longas tábuas corridas de madeira, todas elas de largura igual a 20 cm. Uma moeda cai do bolso furado de um dos dançarinos. Após a queda, qual a probabilidade de a distância entre o centro da moeda e a linha mais próxima que separa duas das tábuas ser no máximo de 3 cm? Como a largura das tábuas é de 20 cm, a menor distância entre o centro da moeda e as linhas do piso é um número real não-negativo limitado à metade da largura das tábuas. Portanto, a distância em que estamos interessados corresponde ao sorteio de um número real no intervalo $\Omega = [0, 10]$. Então nossa pergunta pode ser reescrita como: se sortearmos um número real no conjunto $\Omega = [0, 10]$, qual a probabilidade de o número sorteado pertencer ao conjunto $A = [0, 3]$? Novamente voltamos à pergunta: qual proporção que o conjunto A ocupa dentro de Ω? Neste caso, nos parece que a melhor resposta seria a razão entre os comprimentos dos intervalos A e Ω, ou seja:

$$\mathbb{P}(A) = \frac{\text{comprimento de } A}{\text{comprimento de } \Omega} = \frac{\text{comprimento de } [0,3]}{\text{comprimento de } [0,10]} = \frac{3}{10}. \qquad \triangle$$

Exemplo 1.3. No mês seguinte ao baile anterior, um novo baile é realizado em outro grande salão, cujo piso é formado por azulejos quadrados de 10 cm de lado. Todos os azulejos são idênticos e como na Figura 1.2 abaixo.

Novamente, nosso distraído dançarino deixa outra moeda cair de seu bolso furado. Qual a probabilidade de o centro da moeda cair sobre um ponto preto de um dos azulejos? Já que agora temos uma compreensão melhor, podemos

[1] Esse raciocínio, hoje considerado mais ou menos óbvio, foi formalizado na célebre troca de cartas entre Pascal e Fermat ocorrida em 1654. Estas cartas são consideradas um marco na história da Probabilidade.

intuir que uma boa resposta seria a proporção que a região de pontos pretos ocupa em um azulejo, isto é, a razão entre essas áreas. Denotemos por Ω o conjunto dos pontos do quadrado que representa um azulejo e por A o subconjunto de pontos de cor preta. Sendo assim,

$$\mathbb{P}(A) = \frac{\text{área de } A}{\text{área de } \Omega} = \frac{4}{9}. \qquad \triangle$$

Um fato comum nos três exemplos acima é que a ideia de probabilidade de um conjunto passa pela ideia de proporção, medida que o conjunto A ocupa dentro de um conjunto maior Ω, conjunto este que congrega todos os resultados possíveis no nosso experimento aleatório.

A questão é que em cada um destes exemplos foi diferente o conceito que utilizamos para "medir" o quanto o conjunto A ocupa dentro do conjunto Ω. Nos exemplos acima, nossas medidas foram as razões entre o número de elementos, o comprimento e a área. Definir probabilidade será um modo de medir conjuntos.

Um modelo probabilístico tem três componentes básicas:

1. Um conjunto Ω formado por todos os resultados possíveis do experimento, chamado *espaço amostral*.
2. Uma coleção apropriada $\mathcal{F}$ de subconjuntos do espaço amostral, chamados *eventos aleatórios*. São os conjuntos desta coleção que

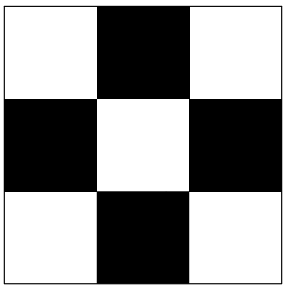


Figura 1.2. Modelo do azulejo usado no Exemplo 1.3.

gostaríamos de "medir", ou atribuir probabilidade, e também realizar operações elementares de conjuntos.

3. Uma função $\mathbb{P}$ que associa a cada evento aleatório um número real, que representa a ideia de chance, verossimilhança, confiança, ou credibilidade. Esta função é chamada de *probabilidade* ou *medida de probabilidade*.

No restante desta seção, discutiremos com mais detalhes cada um dos três objetos que definem os espaços de probabilidade.

1.1.1 Espaço amostral

Um conjunto não-vazio Ω, cujos elementos representam todos os resultados possíveis de um determinado experimento, é chamado de *espaço amostral*. Uma *realização* do experimento é representada pela escolha de algum dos possíveis $\omega \in \Omega$, e às vezes nos referimos ao próprio ω como sendo a realização do experimento.

Exemplo 1.4. Se o experimento consiste em lançar uma moeda, então $\Omega = \{\text{Cara, Coroa}\}$ ou $\Omega = \{0, 1\}$, se convencionarmos que 1 representa a face "cara" e 0 representa a face "coroa". $\triangle$

Exemplo 1.5. Se o experimento consiste em lançar um dado e observar a face superior, então

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

onde cada número representa o possível valor da face observada. $\triangle$

Exemplo 1.6. Se o experimento consiste em lançar uma moeda duas vezes, então

$$\Omega = \{0, 1\}^2 = \{0, 1\} \times \{0, 1\} = \{(0, 0), (0, 1), (1, 0), (1, 1)\},$$

onde a primeira coordenada representa o valor observado no primeiro lançamento, e a segunda coordenada, o do segundo lançamento. $\triangle$

Exemplo 1.7. Se o experimento consiste em lançar um dado duas vezes e

observar a face superior, então

$$\Omega = \{1,2,3,4,5,6\}^2 = \{\omega = (\omega_1, \omega_2) : \omega_1, \omega_2 \in \{1,2,3,4,5,6\}\}. \qquad \triangle$$

Exemplo 1.8. Lançar uma moeda infinitas vezes, em sequência. Se $\omega_1 \in \{0,1\}$ denota o resultado do primeiro lançamento da moeda, $\omega_2 \in \{0,1\}$ o da segunda, e assim por diante, então uma realização desse experimento equivale a sortear um $\omega = (\omega_1, \omega_2, \omega_3, \dots)$ do conjunto

$$\Omega = \{0,1\}^{\mathbb{N}} = \{\omega = (\omega_n)_{n\in\mathbb{N}} : \omega_n \in \{0,1\} \text{ para todo } n\}. \qquad \triangle$$

A propósito, neste livro, $\mathbb{N} = \{1,2,3,\dots\}$ e $\mathbb{N}_0 = \{0,1,2,3,\dots\}$.

Exemplo 1.9. Se o experimento consiste em medir a duração de uma lâmpada, então um possível espaço amostral é dado por $\Omega = [0,\infty)$. $\triangle$

1.1.2 Eventos aleatórios

Eventos são caracterizados por condições que podem ser cumpridas ou não. Nos exemplos vistos no início deste capítulo, o observador sempre era capaz de responder às seguintes perguntas. A carta sorteada foi uma dentre 4♣, 7♡, A♠ ou 7♢? A moeda caiu a menos de 3 cm de distância das linhas do chão? A moeda caiu sobre a parte do azulejo pintada de preto? Em um modelo probabilístico, a condição a ser observada é representada pelo conjunto A dos elementos ω para os quais a condição é cumprida.

Um *evento aleatório*, ou simplesmente *evento*, é um conjunto $A \subseteq \Omega$ tal que o observador sempre é capaz de dizer, ao final do experimento, se $\omega \in A$ ou $\omega \notin A$. Denotaremos por $\mathcal{F}$ a coleção formada pelos eventos aleatórios, chamada *espaço de eventos*. Na Seção 1.3 iremos pedir que a coleção $\mathcal{F}$ satisfaça certas propriedades de modo a nos permitir realizar operações com os conjuntos pertencentes a $\mathcal{F}$.

Dizemos que o evento A *ocorre* se a realização ω é tal que $\omega \in A$. Vamos traduzir algumas operações sobre conjuntos para a linguagem de eventos.

A união $A \cup B$ é o conjunto $\{\omega \in \Omega : \omega \in A \text{ ou } \omega \in B\}$, ou seja, é o conjunto das realizações ω tais que pelo menos um dos eventos A ou B ocorre, portanto $A \cup B$ é o evento "A ou B".

Exemplo 1.10. No lançamento de um dado ($\Omega = \{1, 2, 3, 4, 5, 6\}$) considere os eventos $A =$ "par" $= \{2, 4, 6\}$ e $B =$ "múltiplo de 3" $= \{3, 6\}$. O evento "A ou B" contém todos os resultados que sejam pares ou múltiplos de 3 (ou ambos!), e é dado por $C = A \cup B = \{2, 3, 4, 6\}$. $\triangle$

Analogamente, a interseção $A \cap B$, que é dada por $\{\omega \in \Omega : \omega \in A \text{ e } \omega \in B\}$, é o conjunto das realizações ω tais que ambos os eventos A e B ocorrem, portanto $A \cap B$ é o evento "A e B".

Exemplo 1.11. Considerando os mesmo eventos do Exemplo 1.10, o evento "A e B" contém todos os resultados que sejam ao mesmo tempo pares e múltiplos de 3, e é dado por $A \cap B = \{6\}$. $\triangle$

Denotamos por A^c o complementar do conjunto A, dado por $A^c = \{\omega \in \Omega : \omega \notin A\}$, ou seja, o conjunto das realizações ω para as quais o evento A não ocorre, portanto A^c é o evento "não A". Geralmente é óbvio no contexto qual é o espaço amostral Ω, e por isso optamos por A^c como notação mais compacta que $\Omega \setminus A$.

Exemplo 1.12. Continuando o Exemplo 1.11, considere o evento $A =$ "par" $= \{2, 4, 6\}$. O evento "não A" contém todos os resultados que não sejam pares, ou seja, que são ímpares, e é dado por $C = A^c = \{1, 3, 5\}$. $\triangle$

O conjunto vazio $\emptyset$ é denominado *evento impossível*. O conjunto Ω também é um evento, denominado *evento certo*. Dois eventos A e B são ditos *mutuamente exclusivos*, *incompatíveis* ou *disjuntos* se $A \cap B = \emptyset$, isto é, se o evento "A e B" for impossível. De forma mais geral, dizemos que $(A_n)_n$ são disjuntos se A_j e A_k são disjuntos para todos $j \neq k$.

Exemplo 1.13. Continuando o exemplo anterior, considere os eventos $A =$ "par" $= \{2, 4, 6\}$ e $B =$ "ímpar" $= \{1, 3, 5\}$. O evento "A e B" é o evento impossível porque nenhum número é par e ímpar ao mesmo tempo. Em

termos de conjuntos, $A \cap B = \emptyset$. O evento "A ou B" é o evento certo, porque todo número é par ou ímpar. Em termos de conjuntos, $A \cup B = \Omega$. $\triangle$

A relação $A \subseteq B$ significa que $\omega \in A$ sempre implica $\omega \in B$, ou seja, para qualquer realização ω, se o evento A ocorre então necessariamente o evento B ocorre. Portanto, $A \subseteq B$ significa que a ocorrência do evento A implica a ocorrência do evento B.

Mencionamos uma questão técnica. Em princípio gostaríamos de atribuir probabilidade a qualquer subconjunto de Ω, o que equivale a tomar $\mathcal{F} = \mathcal{P}(\Omega)$, isto é, o conjunto de todos os subconjuntos de Ω, chamado de *conjunto das partes de* Ω. Geralmente, isso é possível quando o espaço amostral Ω é um conjunto enumerável.[2] Entretanto, existem subconjuntos de $\mathbb{R}$, $\mathbb{R}^2$, $\mathbb{R}^3$ aos quais não é possível atribuir uma medida de comprimento, área ou volume. A solução para esse problema envolve considerar um espaço de eventos $\mathcal{F}$ que, apesar de não conter todos os subconjuntos de Ω, contém todos os subconjuntos nos quais estaremos interessados.

1.1.3 Medida de probabilidade

Para um determinado experimento aleatório, após definidos o espaço amostral Ω e a coleção $\mathcal{F}$ de subconjuntos de Ω aos quais gostaríamos de atribuir uma probabilidade, falta definir a medida de probabilidade propriamente dita. Isto é, precisamos especificar uma função $\mathbb{P} : \mathcal{F} \to \mathbb{R}$ que atribua a cada evento $A \in \mathcal{F}$ a sua respectiva probabilidade $\mathbb{P}(A)$.

Na Seção 1.3, definiremos de modo preciso quais propriedades a função de probabilidade $\mathbb{P} : \mathcal{F} \to \mathbb{R}$ deve satisfazer. Porém, antes de tratarmos este problema em total generalidade, gostaríamos de apresentar abaixo

[2] Um conjunto Ω é dito *enumerável* se existe uma função injetiva de Ω em $\mathbb{N}$. Isso quer dizer que os elementos de Ω podem ser indexados por $\mathbb{N}$, ou seja, listados em uma sequência. Por exemplo, $\mathbb{N} = \{1, 2, 3, \dots\}$, $\mathbb{Z} = \{0, 1, -1, 2, -2, \dots\}$, números pares $\{0, 2, 4, 6, 8, \dots\}$, primos $\{2, 3, 5, 7, 11, 13, \dots\}$, e $\mathbb{Q} = \{0, \frac{1}{1}, \frac{-1}{1}, \frac{1}{2}, \frac{-1}{2}, \frac{2}{1}, \frac{-2}{1}, \frac{1}{3}, -\frac{1}{3}, \frac{2}{2}, -\frac{2}{2}, \frac{3}{1}, \frac{-3}{1}, \frac{1}{4}, \dots\}$. O conjunto dos números reais não é enumerável, pois dada qualquer sequência $(x_n)_{n=1,2,3,\dots}$ de números reais, sempre existirá um número $z \in \mathbb{R}$ que não estará nessa sequência.

algumas situações relativamente simples onde podemos especificar a medida de probabilidade explicitamente.

Caso equiprovável

Em alguns experimentos, há um número finito de resultados possíveis e estes são todos idênticos no sentido de que trocar a forma como objetos são etiquetados não afeta as chances de cada um deles. Havendo esse tipo de *simetria*, vale a hipótese de *equiprobabilidade*, isto é, todos os elementos $\omega \in \Omega$ têm a mesma chance de ocorrer. Neste caso, a probabilidade de um evento A é simplesmente a razão entre o número de elementos de A e o número de elementos do espaço amostral Ω:

$$\mathbb{P}(A) = \frac{\#A}{\#\Omega}.$$

Foi exatamente isto que fizemos no Exemplo 1.1. Ou seja, quando há equiprobabilidade, a tarefa de calcular a probabilidade resume-se a um problema de contagem: quantos são os elementos de A e de Ω? Abordaremos a questão de contagem com mais profundidade na próxima seção.

Exemplo 1.14. Um dado comum é lançado. Qual a probabilidade de o valor exibido na face superior ser múltiplo de 3? Podemos modelar esse experimento tomando $\Omega = \{1, 2, 3, 4, 5, 6\}$, e o evento de interesse é $B = \{3, 6\}$. Como as faces do dado podem ser reetiquetadas sem que isso afete as chances de cada resultado, vale a hipótese de equiprobabilidade e, portanto,

$$\mathbb{P}(B) = \frac{\#B}{\#\Omega} = \frac{2}{6} = \frac{1}{3}. \qquad \triangle$$

Exemplo 1.15. Lançamos dois dados, um azul e um branco. Qual a probabilidade de que a soma dos valores observados seja igual a 3? Neste caso, podemos tomar $\Omega = \{1, 2, 3, 4, 5, 6\}^2$ e $A = \{(1, 2), (2, 1)\}$. Como as faces de cada dado podem ser reetiquetadas sem que isso afete as chances de

cada resultado, vale a hipótese de equiprobabilidade neste espaço, e

$$\mathbb{P}(A) = \frac{\#A}{\#\Omega} = \frac{2}{36} = \frac{1}{18}. \qquad \triangle$$

Exemplo 1.16. Retiramos uma carta de um baralho comum, observamos sua face e a retornamos ao baralho, que é novamente embaralhado. Em seguida retiramos outra carta. Qual a probabilidade de as cartas retiradas serem um rei e uma dama do mesmo naipe? Neste caso, podemos tomar

$$\Omega = \{A\diamondsuit, 2\diamondsuit, \ldots, K\diamondsuit, A\heartsuit, \ldots, A\spadesuit, \ldots, A\clubsuit, \ldots, K\clubsuit\}^2$$

e

$$\begin{aligned} A = \{&(Q\diamondsuit, K\diamondsuit), (Q\heartsuit, K\heartsuit), (Q\spadesuit, K\spadesuit), (Q\clubsuit, K\clubsuit), \\ &(K\diamondsuit, Q\diamondsuit), (K\heartsuit, Q\heartsuit), (K\spadesuit, Q\spadesuit), (K\clubsuit, Q\clubsuit)\}. \end{aligned}$$

Como as cartas podem ser reetiquetadas cada vez que embaralhamos o maço, sem que isso afete as chances de cada possível resultado, vale a hipótese de equiprobabilidade nesse espaço Ω. Logo,

$$\mathbb{P}(A) = \frac{\#A}{\#\Omega} = \frac{8}{52^2} = \frac{1}{338}. \qquad \triangle$$

Espaços amostrais enumeráveis

Um contexto que vai além do caso equiprovável e que podemos tratar sem mais complicações técnicas, é quando o espaço amostral Ω é um conjunto enumerável.

Se $\Omega = \{\omega_1, \omega_2, \omega_3, \ldots\}$, então a cada possível resultado ω_n é associada uma probabilidade $p(\omega_n)$ de forma que

$$\sum_{n=1}^{\infty} p(\omega_n) = 1.$$

Para um subconjunto $B \subseteq \Omega$ definimos

$$\mathbb{P}(B) = \sum_{\omega \in B} p(\omega). \tag{1.17}$$

Exemplo 1.18. Lançamos um dado sucessivamente e contamos o número de lançamentos necessários até obtermos o número 3 pela primeira vez. Então podemos tomar como espaço amostral $\Omega = \mathbb{N}$ e $p(n) = \frac{1}{6}\,(\frac{5}{6})^{n-1}$. Se $A =$ "obter um 3 em no máximo 5 tentativas" e $B =$ "não se obter o 3 nas primeiras 10 tentativas", então

$$\mathbb{P}(A) = \tfrac{1}{6} + \tfrac{1}{6} \times \tfrac{5}{6} + \cdots + \tfrac{1}{6} \times (\tfrac{5}{6})^4 = \frac{\frac{1}{6} - (\frac{5}{6})^5 \frac{1}{6}}{1 - \frac{5}{6}} = 1 - (\tfrac{5}{6})^5 \approx 0{,}598.$$

e

$$\mathbb{P}(B) = \tfrac{1}{6}(\tfrac{5}{6})^{10} + \tfrac{1}{6}(\tfrac{5}{6})^{11} + \tfrac{1}{6}(\tfrac{5}{6})^{12} + \cdots = \frac{\frac{1}{6}(\frac{5}{6})^{10}}{1 - (\frac{5}{6})} = (\tfrac{5}{6})^{10} \approx 0{,}161. \qquad \triangle$$

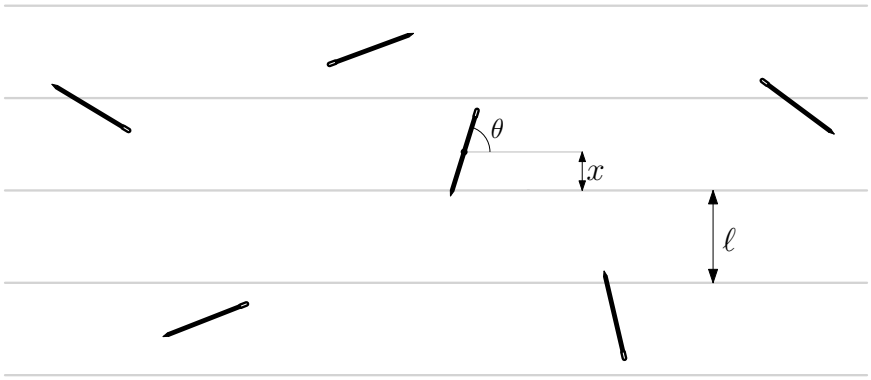


Figura 1.3. Piso indicando várias agulhas de Buffon e as coordenadas θ e x.

A agulha de Buffon

Uma agulha é lançada de modo aleatório para cima e cai sobre um piso que é cortado por um feixe de retas paralelas, todas elas espaçadas por uma mesma distância igual ao comprimento da agulha. Qual a probabilidade de a agulha cruzar uma das retas do piso?

Seja ℓ o comprimento da agulha. Inspecionando a Figura 1.3, podemos verificar que a agulha está perfeitamente localizada em relação às retas do piso se conhecermos as variáveis

$$\theta = \text{menor ângulo formado entre a agulha e as retas do piso}$$

e

$$x = \text{distância entre o ponto médio da agulha e a reta mais próxima.}$$

Observe que $0 \leqslant x \leqslant \frac{\ell}{2}$ e $0 \leqslant \theta \leqslant \frac{\pi}{2}$, ou seja, o lançamento da agulha corresponde a um sorteio de um ponto no retângulo $\Omega = [0, \frac{\pi}{2}] \times [0, \frac{\ell}{2}]$. Recorrendo novamente à Figura 1.3, podemos verificar que a agulha cruza uma das retas do piso se, e somente se, é satisfeita a condição $x < \frac{\ell}{2} \operatorname{sen} \theta$. Sendo assim, nossa pergunta agora é quanto vale $\mathbb{P}(A)$, onde $A = \{(\theta, x) \in \Omega : x < \frac{\ell}{2} \operatorname{sen} \theta\}$. Assim como agimos intuitivamente no Exemplo 1.3, por uma questão de simetria (invariância por rotação e translação), a probabilidade do conjunto A deve ser a razão entre as áreas dos conjuntos A e Ω:

$$\mathbb{P}(A) = \frac{\text{Área } A}{\text{Área } \Omega} = \frac{\int_0^{\frac{\pi}{2}} \frac{\ell}{2} \operatorname{sen} \theta \, \mathrm{d}\theta}{\frac{\ell\pi}{4}} = \frac{2}{\pi}.$$

Lançando a agulha muitas vezes, esperamos que a proporção de lançamentos em que a agulha intersecta as linhas do piso se aproxime de $\frac{2}{\pi}$. Este é, pois, um método probabilístico para calcular as casas decimais de π! Na verdade não é um método muito eficiente, mas impressiona pela audácia.

Espaços contínuos

Observamos que, no exemplo acima, o espaço amostral $\Omega = [0, \frac{\pi}{2}] \times [0, \frac{\ell}{2}]$ não é enumerável. Por isso, precisamos de uma medida de probabilidade diferente da dos casos anteriores. Intuitivamente, dissemos que a probabilidade do evento A era a razão entre as áreas de A e Ω. Mas quais são os conjuntos aos quais podemos atribuir área? Como se calcula a área nestes casos?

Neste mesmo problema, qual é a probabilidade de a agulha cair paralela às retas do piso? Ou seja, queremos saber quanto vale $\mathbb{P}(B_0)$, onde $B_0 = \{0\} \times [0, \frac{\ell}{2}]$. Como B_0 é um segmento de reta dentro do retângulo Ω, a área de B_0 é zero, logo $\mathbb{P}(B_0) = 0$.

De modo análogo, o evento $B_\alpha = \{\alpha\} \times [0, \frac{\ell}{2}]$ em que a agulha forma um ângulo exatamente igual a α com o feixe de retas do piso, também tem a propriedade de que $\mathbb{P}(B_\alpha) = 0$, e isso vale para todo $\alpha \in [0, \frac{\pi}{2}]$. Em particular, $\mathbb{P}(\{\omega\}) = 0$ para todo $\omega \in \Omega$.

Acabamos de ver algo bastante curioso: há eventos de probabilidade zero que não são o evento impossível! Mais do que isso, cada ponto do espaço amostral tem probabilidade zero. Mas isso quer dizer que qualquer evento é composto por eventos menores, cada um com probabilidade zero, incluindo o evento A estudado acima (a agulha cortar uma das retas do piso). Daí vemos que

$$\mathbb{P}(A) \neq \sum_{\omega \in A} \mathbb{P}(\{\omega\}),$$

pois $\mathbb{P}(A) = \frac{2}{\pi}$ e $\sum_{\omega \in A} \mathbb{P}(\{\omega\}) = 0$. Observe o contraste com (1.17). Quando um evento pode ser decomposto em uma coleção enumerável de partes menores, o uso de somatórios é adequado. Quando a decomposição não é enumerável, necessitamos modelos em que a passagem da parte ao todo se faz através de integração ao invés de soma.

Simetria e equiprobabilidade

Concluímos esta seção com uma discussão mais minuciosa a respeito da hipótese de equiprobabilidade.

Um dado pode ter suas faces etiquetadas com $\{1, 2, 3, 4, 5, 6\}$ ou $\{2, 3, 4, 5, 6, 1\}$. As chances de se observar a etiqueta "2" são as mesmas para ambos os casos, pois as faces do dado que seriam etiquetadas com "2" são idênticas e, uma vez que o dado tenha girado ao redor de seus eixos várias vezes, não há resquícios da posição inicial que possam fazer com que uma das faces tenha mais chances que as outras. Por outro lado, sair "2" no segundo caso é o mesmo que sair "1" no primeiro caso, e portanto a probabilidade de se observar a face "2" é a mesma que a de se observar a face "1". Seguindo o mesmo argumento, a chance de sair "2" é a mesma que de sair "3" e assim por diante.

Certamente, há exemplos com finitos resultados possíveis em que a hipótese de equiprobabilidade não é válida. Considere o lançamento simultâneo de dois dados de mesma cor (mais precisamente, idênticos, se é que isso é possível) sobre o mesmo tabuleiro. Qual a probabilidade de que a soma dos valores exibidos em cada face seja 3? Um erro sutil, que inclusive grandes matemáticos já cometeram no passado, é usar um modelo equiprovável sobre os 21 pares não-ordenados possíveis. Neste caso, teríamos a mesma probabilidade de que a soma seja 2 ou 3, pois 2 somente pode ser obtido de uma forma (com um par de "1"s) e 3 também pode ser obtido apenas de uma forma (com um "1" e um "2"). A ideia de considerar os dados como sendo idênticos e definir um espaço amostral com apenas 21 elementos, ao invés de ajudar, atrapalhou. Isso porque não há simetria (isto é, invariância pela forma como as faces são etiquetadas) que nos permita justificar que $\omega = \{1, 1\}$ e $\omega = \{1, 2\}$ tenham a mesma chance. Pelo contrário, podemos pensar que dois dados nunca são idênticos e, ainda que fossem, um deles vai cair mais à esquerda e, ainda que caiam perfeitamente alinhados, é possível escolher um dos dois dados no início e seguir sua trajetória de forma a diferenciá-lo do outro. Felizmente, mesmo que o observador insista na sua incapacidade de

distinguir os dados, já sabemos que o modelo com 36 resultados possíveis nos permitiu resolver o problema usando simetria. Portanto, se queremos muito modelar o experimento com um espaço amostral de 21 elementos, a medida de probabilidade deve atribuir peso $\frac{1}{36}$ a cada um dos 6 pares de números idênticos e $\frac{2}{36}$ a cada um dos 15 pares de números diferentes, totalizando $\frac{36}{36}$, como esperado.[3]

1.2 Contagem e simetria

O problema de contar quantos são os elementos de um dado conjunto é quase sempre baseado no uso de duas regras básicas. Apesar de simples de serem entendidas, seu uso pode se dar de maneira bastante sofisticada. Estas regras são conhecidas como os *princípios fundamentais da contagem.*

O princípio aditivo diz o seguinte:

Se uma coleção de objetos pode ser decomposta em duas subcoleções disjuntas, a primeira com m objetos e a segunda com n objetos, então o número total de objetos é $m + n$.

[3]Quem disse que dois dados podem somar 2 ou 3 com a mesma probabilidade foi ninguém menos que G. Leibniz (o co-inventor do Cálculo) em 1666. Claramente, Leibniz não tinha muita experiência prática lançando dados. Meio século antes, teve lugar uma versão um pouco mais complicada desse problema, o *Problema do Grão-Duque da Toscana.* Considere o lançamento de três dados, e observe que a soma 9 pode ser obtida como $\{1, 2, 6\}, \{1, 3, 5\}, \{1, 4, 4\}, \{2, 2, 5\}$, $\{2, 3, 4\}$ ou $\{3, 3, 3\}$, enquanto a soma 10 pode ser obtida como $\{1, 3, 6\}$, $\{1, 4, 5\}$, $\{2, 2, 6\}$, $\{2, 3, 5\}$, $\{2, 4, 4\}$ ou $\{3, 3, 4\}$. Apesar de que tanto 9 quanto 10 podem ser obtidos a partir de seis listas diferentes, por que somas 10 são mais frequentes que somas 9? Indagado por seu patrono, Galileu responde de modo correto que essas seis listas não são equiprováveis. Analogamente ao lançamento de dois dados, a lista $\{3, 3, 3\}$ é menos provável que $\{3, 3, 4\}$, que por sua vez é menos provável que $\{2, 3, 5\}$. A escolha correta para a medida de probabilidade deve atribuir peso 1/216 a cada uma das triplas com números idênticos, 3/216 a cada tripla formada por dois números idênticos e outro distinto e 6/216 a cada tripla de três números distintos. Portanto, a soma 9 ocorre com probabilidade 25/216, enquanto uma soma 10 ocorre com probabilidade 27/216. É muito curioso que o Grão-Duque tenha tido tanta experiência com lançamentos de dados a ponto de perceber empiricamente uma diferença relativa entre as probabilidades inferior a um décimo!

O princípio multiplicativo diz o seguinte:

Suponha que cada objeto de uma coleção pode ser especificado em duas etapas, de forma que na primeira etapa há m opções e, independentemente do resultado da primeira etapa, na segunda etapa há n opções. Suponha também que escolhas diferentes sempre resultem em objetos diferentes. Então a coleção tem $m \times n$ elementos.

Exemplo 1.19. Fernanda tem quatro saias: preta, cinza, azul e vermelha, e três camisetas: branca, azul e vermelha. Quantas peças de roupa tem Fernanda? Pelo princípio aditivo, Fernanda tem $4 + 3 = 7$ peças de roupa. De quantas formas distintas Fernanda pode se vestir? Escolhendo primeiro a saia e depois a camiseta, pelo princípio multiplicativo vemos que Fernanda tem $4 \times 3 = 12$ formas de se vestir. △

Exemplo 1.20. Severino vai a um restaurante onde há duas opções de prato: peixe e carne (bovina). Como opções de vinho, há Merlot e Malbec (tintos), Sauvignon Blanc e Pinot Grigio (brancos). Além disso, há uma opção de cerveja de trigo. Severino não gosta de combinar carne com vinho branco nem peixe com vinho tinto. Escolhendo primeiro o prato e depois a bebida, vemos, pelo princípio multiplicativo, que Severino tem $2 \times 3 = 6$ possibilidades para pedir. Essa solução é mais simples do que se escolhêssemos primeiro a bebida. Se Severino toma cerveja, há 2 opções de prato. Se Severino toma vinho, há 4 opções de vinho, e para cada opção de vinho há 1 opção de prato. Combinando os princípios aditivo e multiplicativo, Severino tem $1 \times 2 + 4 \times 1 = 6$ possibilidades para fazer o pedido. △

Exemplo 1.21 (Sorteio com reposição). Retiramos uma carta do baralho, anotamos o valor, devolvemos a carta, voltamos a embaralhar e voltamos a retirar uma carta. Qual a probabilidade de que pelo menos uma das duas cartas retiradas seja o 4♣? Pelo princípio multiplicativo, o número de resultados possíveis desse experimento é 52^2. Para o evento em questão, há duas possibilidades disjuntas, ou a primeira carta retirada é o 4♣, e a segunda carta é qualquer das 52 cartas do baralho, ou a primeira carta retirada é qualquer das outras 51 diferentes e a segunda é um 4♣. Combinando os

princípios aditivo e multiplicativo, obtemos $\#A = 1\times 52+51\times 1 = 103$. Como os números das cartas podem ser permutados antes de cada retirada sem que isso afete as chances de cada resultado, vale a hipótese de equiprobabilidade e, portanto, $\mathbb{P}(A) = \frac{103}{2.704}$. $\triangle$

Exemplo 1.22 (Sorteio sem reposição). Retiramos, sem reposição, duas cartas de um baralho. Qual a probabilidade de que uma das duas cartas retiradas seja o 4♣? Para facilitar a contagem, vamos supor que as cartas são retiradas de forma ordenada. Ou seja, primeiro retiramos uma carta do baralho e depois retiramos a outra carta. Pelo princípio multiplicativo, o número de resultados possíveis desse experimento é 52×51, pois, independentemente de qual seja a primeira carta retirada, para a segunda retirada sempre haverá 51 cartas restantes no baralho. O evento em questão também pode ser especificado em duas etapas: primeiro escolhemos se o 4♣ sai na primeira ou na segunda retirada e, independentemente disso, escolhemos depois qual das outras 51 cartas sai na outra retirada. Pelo princípio multiplicativo, temos $\#A = 2 \times 51 = 102$. Como os números das cartas podem ser permutados antes de cada retirada sem que isso afete as chances de cada resultado, vale a hipótese de equiprobabilidade e, portanto, $\mathbb{P}(A) = \frac{102}{51\times 52} = \frac{1}{26}$. $\triangle$

O princípio multiplicativo pode ser estendido para k etapas por indução, se consideramos as etapas $2, \ldots, k$ como uma única etapa que consiste em $k-1$ subetapas.

Exemplo 1.23. Sortear 4 cartas de um baralho comum, com reposição. Neste caso, podemos tomar $\Omega = \big(\{A, 2, 3, \ldots, 9, 10, J, Q, K\} \times \{\clubsuit, \heartsuit, \spadesuit, \diamondsuit\}\big)^4$ e $\#\Omega = 52^4$. Como os números das cartas podem ser permutados antes de cada retirada sem que isso afete as chances de cada resultado, vale a hipótese de equiprobabilidade e, portanto, $\mathbb{P}(D) = \frac{\#D}{52^4}$ para todo $D \subseteq \Omega$.

Qual a probabilidade do evento A = "as quatro cartas são valetes"? Escrevendo $A = (\{J\} \times \{\heartsuit, \diamondsuit, \spadesuit, \clubsuit\})^4$, temos $\#A = 4^4$ e $\mathbb{P}(A) = \frac{4^4}{52^4} = \frac{1}{13^4}$.

Qual a probabilidade do evento B = "todas as cartas têm o mesmo naipe"? Temos 4 escolhas para o naipe, e 13 escolhas para cada uma das cartas retiradas, logo $\#B = 4 \times 13^4$ e portanto $\mathbb{P}(B) = \frac{4\times 13^4}{52^4} = \frac{1}{4^3}$. $\triangle$

O princípio aditivo pode ser utilizado de trás para frente.

Exemplo 1.24. No Exemplo 1.19, de quantas formas Fernanda pode se vestir sem que a saia e a camiseta tenham a mesma cor? As formas de Fernanda se vestir, que são 12, podem ser decompostas em dois tipos: com cores iguais, que são 2, ou cores diferentes. Portanto, há 10 combinações que envolvem cores diferentes. △

Um objeto considerado frequentemente em probabilidade são as *permutações*. Comecemos por um assunto familiar: os anagramas.

Exemplo 1.25. Quantos anagramas tem a palavra M-A-R-C-E-L-O? Podemos construir um anagrama escolhendo uma letra de cada vez, e vemos que, independentemente das escolhas anteriores, a k-ésima letra escolhida terá $8 - k$ opções. Pelo princípio multiplicativo, a quantidade de anagramas é dada por $7 \times 6 \times \cdots \times 1 = 7! = 5.040$. Este é o número de *permutações* das letras que compõem a palavra. △

O princípio multiplicativo também pode ser usado de trás para frente, ou seja, quando conhecemos o número total de objetos e queremos saber quantas opções há na primeira etapa.

Exemplo 1.26. Quantos anagramas tem a palavra V-L-A-D-A-S? A mesma solução que usamos para M-A-R-C-E-L-O já não funciona, porque o número de escolhas para a segunda letra depende da primeira letra (se foi A ou não). Para resolver esse problema, vamos supor, artificialmente, que os dois A's da palavra V-L-A-D-A-S sejam distintos, ou seja, vamos calcular os anagramas de V-L-A_1-D-A_2-S. Pelo método anterior, essa palavra tem 6! anagramas. Agora veja que um anagrama de V-L-A_1-D-A_2-S pode ser construído em duas etapas: primeiro escolhemos um anagrama de V-L-A-D-A-S e depois escolhemos uma permutação de A_1A_2. O número de opções da primeira etapa é justamente o que queremos calcular, e a segunda tem 2! opções. Portanto, V-L-A-D-A-S tem $\frac{6!}{2!} = 360$ anagramas. △

Exemplo 1.27. Quantos anagramas tem a palavra M-I-S-S-I-S-S-I-P-P-I? Se as letras fossem todas diferentes, seriam 11!. Podemos proceder como no exemplo

anterior, mas agora em quatro etapas: primeiro tomando um anagrama de M-I-S-S-I-S-S-I-P-P-I, depois uma permutação de S_1-S_2-S_3-S_4, depois de I_1-I_2-I_3-I_4, e finalmente uma de P_1-P_2. Portanto, M-I-S-S-I-S-S-I-P-P-I tem $\frac{11!}{4!4!2!} = 34.650$ anagramas. △

Estudaremos agora um objeto tão importante que tem um nome e notação próprios.

Exemplo 1.28. Daniela tem 13 camisas de cores diferentes. Para sua próxima viagem de negócios, ela vai levar 5 camisas. De quantas formas possíveis Daniela pode escolhê-las? Usaremos novamente o princípio multiplicativo de trás para frente. Uma permutação das 13 camisas pode ser especificada em três etapas: primeiro escolhemos 5 camisas das 13 para ficarem nas 5 primeiras posições, depois permutamos essas 5 camisas, e finalmente permutamos as 8 camisas restantes. Deduzimos que, na primeira etapa, que é a escolha de 5 camisas entre as 13 disponíveis, há $\frac{13!}{5!8!}$ opções. Portanto, Daniela pode fazer a mala de $\frac{13!}{5!8!} = 1.287$ formas diferentes. △

O exemplo acima é um caso particular de um objeto combinatorial muito importante. Dados $n \in \mathbb{N}_0$ e $k \in \mathbb{Z}$, denotamos por $\binom{n}{k}$ o número de subconjuntos de k elementos que um conjunto qualquer de n elementos possui, e lê-se "combinações de n elementos, k a k" ou mais vulgarmente "n escolhe k". Analisaremos agora propriedades desses coeficientes a partir dessa definição.

Estes números especiais são também chamados de *coeficientes binomiais*. Podemos dispor os coeficientes binomiais em uma grande tabela onde na n-ésima linha e k-ésima coluna escrevemos $\binom{n}{k}$. Essa tabela infinita se chama Triângulo de Pascal.[4] Vejamos algumas das principais propriedades desses números. Primeiro, dado $n \in \mathbb{N}_0$, temos $\binom{n}{k} = 0$ para $k > n$ ou $k < 0$ pois, dado um conjunto com n elementos, não há subconjuntos com k elementos para esses valores de k. Ademais, $\binom{n}{0} = \binom{n}{n} = 1$, pois o único subconjunto com zero elementos é o conjunto vazio e o único subconjunto com n elementos é todo o conjunto.

[4] O Triângulo de Pascal foi um objeto recorrente em diversos trabalhos da Matemática

Uma propriedade fundamental é a *Relação de Stifel*: $\binom{n+1}{k+1} = \binom{n}{k} + \binom{n}{k+1}$. Com efeito, um subconjunto de $\{1, \ldots, n+1\}$ com $k+1$ elementos pode ser obtido escolhendo-se $k+1$ elementos de $\{1, \ldots, n\}$, ou escolhendo-se o elemento $n+1$ e outros k elementos em $\{1, \ldots, n\}$.

As propriedades acima nos dizem que, se ignorarmos os termos nulos da nossa tabela de coeficientes binomiais, esta assumirá uma forma triangular, e encontramos o número 1 no início e no final de cada linha do triângulo. Esse formato triangular combinado com a Relação de Stifel nos permite encontrar todos os valores de uma linha a partir da linha anterior.

Os coeficientes binomiais também têm a propriedade de simetria: $\binom{n}{k} = \binom{n}{n-k}$. Com efeito, basta ver que há uma bijeção entre o conjunto de subconjuntos de $\{1, \ldots, n\}$ contendo k elementos e o de subconjuntos contendo $n-k$ elementos. Um exemplo de bijeção é simplesmente tomar o complementar de cada conjunto.

O *Teorema das Linhas* diz que $\sum_{k=0}^{n} \binom{n}{k} = 2^n$, o que é justificado observando-se que ambos os lados da igualdade acima correspondem à quantidade de subconjuntos de $\{1, \ldots, n\}$. Para o lado esquerdo, classificamos os subconjuntos pelo seu tamanho e observamos que, para cada k, há $\binom{n}{k}$ subconjuntos com exatamente k elementos e, somando sobre os possíveis valores de k, concluímos pelo princípio aditivo que o número total de subconjuntos é $\sum_{k=0}^{n} \binom{n}{k}$. Para o lado direito, observamos que um subconjunto $A \subseteq \{1, \ldots, n\}$ pode ser especificado em n etapas, onde na k-ésima etapa dizemos se $k \in A$ ou $k \notin A$, totalizando 2^n possibilidades pelo princípio

antiga. Suas origens remontam ao tratado de metrificação em sânscrito *Chandas sutra* associado a Pingala, por volta dos séculos III-II a.C., onde havia inclusive uma versão da Relação de Stifel. Posteriormente, aparece também na obra do matemático e poeta persa Omar Khayyan, séculos XI-XII d.C, bem como na Grécia Antiga, na China Antiga, dentre outras manifestações. Mesmo o uso do triângulo no cálculo de probabilidades é anterior a Pascal. No século XVI, Tartaglia já estudava as possíveis combinações de resultados em lançamentos sucessivos de um dado, relacionando-as com esse triângulo. O uso sistemático que Pascal fez do triângulo e suas propriedades na solução de problemas de probabilidade relacionados a jogos de azar, dentre eles o famoso problema dos pontos, fez seu nome definitivamente associado ao triângulo.

multiplicativo. Portanto, $\sum_{k=0}^{n} \binom{n}{k} = 2^n$, como queríamos demonstrar.

O *Teorema Binomial* generaliza o Teorema das Linhas, e diz que

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$$

para todos $a, b \in \mathbb{R}$ e $n \in \mathbb{N}_0$. Com efeito, ao expandir o binômio $(a+b)^n = (a+b) \times \cdots \times (a+b)$, a operação a ser feita é escolher a ou b em cada um dos n parêntesis acima, multiplicar os valores escolhidos e depois somar sobre todas as possíveis escolhas. Ao multiplicar, obtemos produtos da forma $a^k b^{n-k}$, onde k é o número de parêntesis na expansão acima em que escolhemos o símbolo a. Ao somar todas as possíveis escolhas, podemos agrupá-las em função dos distintos valores de k. Cada termo da forma $a^k b^{n-k}$ aparecerá uma determinada quantidade de vezes, dada pelo número de possíveis escolhas de k parêntesis dentre os n disponíveis, isto é, $\binom{n}{k}$. Isso prova o teorema.

Para fins de cálculos de probabilidades, há uma fórmula explícita para o coeficiente binomial, que podemos extrair do Exemplo 1.28. Naquele exemplo tínhamos $n = 13$ e $k = 5$ e, aplicando exatamente o mesmo raciocínio para $n \geqslant k \geqslant 0$ inteiros, obtemos

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!}.$$

Exemplo 1.29. No Exemplo 1.23, qual a probabilidade do evento $C =$ "há um par de cartas de um naipe e um par de cartas de um outro naipe"? Temos $\binom{4}{2}$ escolhas para os naipes. Escolhidos os naipes, temos $\binom{4}{2}$ combinações para quais retiradas correspondem a cada naipe. Escolhidos os naipes e as posições, há 13 escolhas de cartas para cada retirada. Assim, $\#C = \binom{4}{2}\binom{4}{2}13^4 = (\frac{4!}{2!\,2!})^2 13^4 = 6^2 13^4$ e, portanto, $\mathbb{P}(C) = \frac{6^2 13^4}{52^4} = \frac{6^2}{4^4} = \frac{9}{64}$. $\triangle$

Exemplo 1.30. No Exemplo 1.28, se Daniela escolhe as camisas ao acaso, qual a probabilidade de que ela leve as camisas azul e vermelha em sua mala? Para isso temos que considerar de quantas formas é possível levar ambas

as camisas azul e vermelha. Especificar uma configuração cumprindo essas condições é o mesmo que especificar um conjunto de 3 camisas dentre as 11 que não são nem azul nem vermelha. Novamente, por razões de simetria, a medida de probabilidade correta é $\mathbb{P}(D) = \frac{\#D}{\#\Omega}$ para todo $D \subseteq \Omega$ e, se A denota o evento acima, então

$$\mathbb{P}(A) = \frac{\binom{11}{3}}{\binom{13}{5}} = \frac{\frac{11!}{3!8!}}{\frac{13!}{5!8!}} = \frac{5 \times 4}{13 \times 12} = \frac{5}{39}. \qquad \triangle$$

Observação 1.31. O Exemplo 1.22 foi resolvido sem usar o coeficiente binomial, supondo-se que as cartas eram retiradas de forma ordenada. Se considerássemos as cartas retiradas como um subconjunto do baralho contendo dois elementos, sem distinção de ordem, a solução seria

$$\mathbb{P}'(A') = \frac{\#A'}{\#\Omega'} = \frac{51}{\binom{52}{2}} = \frac{1}{26}. \qquad \triangle$$

O uso de bijeções pode ser muito poderoso, como veremos a seguir.

Exemplo 1.32 (Princípio da reflexão e números de Catalan). Eva teve quatro filhas, a primeira e a terceira não tiveram filhas, a segunda teve duas filhas (a primeira das quais não teve filhas e a segunda teve uma filha que não teve filhas), e a quarta teve três filhas, que não tiveram filhas. Veja que Eva teve 10 descendentes por linhagem feminina (consideramos apenas a linhagem feminina). Na Figura 1.4, ilustramos a árvore de descendência de Eva, que é uma das possíveis árvores com 10 descendentes. Quantas árvores distintas existem com 10 descendentes? Esse número é conhecido como o décimo *número de Catalan*. Se consideramos o mesmo problema para n descendentes, a solução é o n-ésimo número de Catalan, que denotaremos por C_n.

Para deixar claro como as árvores são contadas, na Figura 1.4 mostramos todas as árvores com 3 descendentes, e há 5 delas: filha, neta e bisneta; filha e duas netas; três filhas; duas filhas e uma neta através da filha mais velha; duas filhas e uma neta através da filha mais jovem. Portanto, $C_3 = 5$.

Dada a árvore de descendência de Eva, podemos definir um caminho que contorna os elos da árvore no sentido anti-horário. Tal caminho é formado por 10 passos para cima, 10 passos para baixo, termina na altura inicial e nunca fica abaixo desta. Na Figura 1.4 mostramos como construir esse caminho. Observe que é possível reconstruir a árvore a partir do caminho, portanto esses objetos estão em bijeção.

De modo mais geral, quantas são as árvores com n descendentes? Dado $m \in \mathbb{N}$, dizemos que a m-upla $s = (s_1, \ldots, s_m) \in \mathbb{Z}^m$ é *um caminho simples de duração* m se $s_k - s_{k-1} = +1$ ou -1 para $k = 1, \ldots, m$, com $s_0 = 0$. Assumindo duração sempre de $2n$, temos que C_n é igual ao número de caminhos que voltam à origem no tempo $2n$ e nunca passam por -1.

Definimos $\tau = \tau(s) = \min\{k \geqslant 1 : s_k = -1\}$, onde $\min(\emptyset) = +\infty$, como o instante da primeira passagem por -1, ou seja, $s_j \geqslant 0$ para $j < \tau$ e $s_j = -1$ para $j = \tau$ caso $\tau < \infty$. Nessa notação, $C_n = \#\{s : s_{2n} = 0, \tau(s) > 2n\}$.

Observe que, se $s_{2n} = 2k$, há exatamente $n + k$ passos para cima e $n - k$ passos para baixo, e a quantidade de tais caminhos é $\binom{2n}{n+k}$. Por outro lado, o conjunto de todas as trajetórias que terminam em $s_{2n} = 0$ pode ser particionado pelos conjuntos daquelas que nunca passam por -1 e daquelas

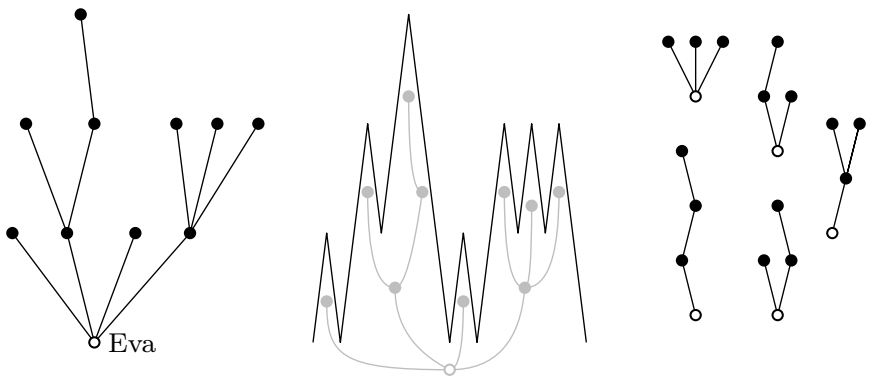


Figura 1.4. A árvore genealógica de Eva e o respectivo caminho gerado. À direita temos os cinco exemplos possíveis de árvores com três descendentes.

que sim passam por -1. Pelo princípio aditivo, conclui-se que

$$\binom{2n}{n} = C_n + \#\{s : s_{2n} = 0, \tau(s) \leqslant 2n\}.$$

Portanto, nossa tarefa agora consiste em calcular o último termo da equação acima, o que faremos usando um truque chamado *princípio da reflexão*.

Dado um caminho s de duração $2n$ tal que $\tau(s) \leqslant 2n$, definimos o caminho s' refletindo, em relação à reta $y = -1$, a parte do caminho posterior à primeira visita a -1. Ou seja, $s'_j = s_j$ para $j \leqslant \tau$ e $s'_j = -2 - s_j$ para $j \geqslant \tau$.

Observe na Figura 1.5 que s' é um caminho de duração $2n$ e $s'_{2n} = -2$. Por outro lado, todo caminho que termina em -2 passa necessariamente por -1. Além disso, $(s')' = s$, e podemos ver que o mapa $s \mapsto s'$ é uma bijeção entre o conjunto dos caminhos de duração $2n$ que visitam -1 e terminam na origem e o conjunto dos caminhos de duração $2n$ que terminam em -2. Assim, $\#\{s : s_{2n} = 0, \tau(s) \leqslant 2n\} = \#\{s : s_{2n} = -2\} = \binom{2n}{n+1}$. Obtemos, assim, a expressão

$$C_n = \binom{2n}{n} - \binom{2n}{n+1}$$

para o n-ésimo número de Catalan. $\triangle$

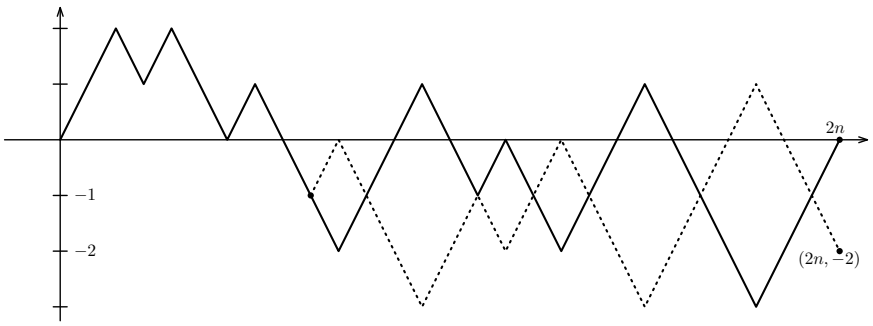


Figura 1.5. Um caminho que passa por -1 e termina na origem e o respectivo caminho refletido, que termina em -2 (passando obrigatoriamente por -1).

1.3 Formulação axiomática de Kolmogorov

Para fazer as operações mais básicas com eventos aleatórios, vamos pedir que nosso espaço de eventos tenha a seguinte estrutura.

Definição 1.33. Dizemos que uma classe $\mathcal{F}$ de subconjuntos de Ω é uma *σ-álgebra* se $\mathcal{F}$ satisfaz às seguintes propriedades:

(1) $\Omega \in \mathcal{F}$;
(2) Para todo $A \in \mathcal{F}$, tem-se que $A^c \in \mathcal{F}$;
(3) Para toda sequência $\{A_n\}_{n=1}^{\infty} \subseteq \mathcal{F}$, vale $(\cup_{n=1}^{\infty} A_n) \in \mathcal{F}$.

Utilizaremos o termo *classe* para denotar um conjunto de subconjuntos de algum espaço amostral. As três propriedades dizem que o evento certo é um dos elementos de $\mathcal{F}$, e que a classe $\mathcal{F}$ é fechada pelas operações de tomar o complementar e uniões enumeráveis. Estas propriedades nos garantem que $\emptyset \in \mathcal{F}$, logo $\mathcal{F}$ é fechada também por uniões finitas. Além disso, $\mathcal{F}$ é fechada com respeito a interseções enumeráveis, pois da lei de De Morgan segue que $\cap_{n=1}^{\infty} A_n = (\cup_{n=1}^{\infty} A_n^c)^c$, o que implica que $\mathcal{F}$ também é fechada pela operação de diferença entre conjuntos, pois $A \setminus B = A \cap B^c$. Ou seja, trabalhar com uma σ-álgebra é algo robusto o suficiente que nos permite fazer as operações elementares de conjuntos, uma quantidade enumerável de vezes, sem sair de $\mathcal{F}$.[5] Mais propriedades de σ-álgebras serão discutidas na Seção 1.4.

Definição 1.34. Seja Ω um espaço amostral e $\mathcal{F}$ uma σ-álgebra em Ω. Uma *medida de probabilidade* $\mathbb{P}$, ou simplesmente *probabilidade*, é uma função $\mathbb{P} : \mathcal{F} \to \mathbb{R}$ satisfazendo às seguintes propriedades:

[5] Daqui surge uma pergunta mais que legítima: por que precisamos da definição de σ-álgebra? Quer dizer, não poderíamos simplesmente tomar $\mathcal{P}(\Omega)$ como espaço de eventos e evitar tudo isso? De fato, é sim possível tomar $\mathcal{F} = \mathcal{P}(\Omega)$ para modelos discretos, mas recordemo-nos de alguns exemplos vistos até aqui, mais precisamente os Exemplos 1.2 e 1.3, e o da agulha de Buffon. O problema nesses exemplos é que existem subconjuntos da reta ou do plano que são tão complicados que não é possível atribuir-lhes comprimento ou área. De forma mais geral, com $\Omega = \mathbb{R}^n$, não existe uma medida de probabilidade $\mathbb{P}$, definida em todos os subconjuntos de $\mathbb{R}^n$, com a propriedade de que $\mathbb{P}(\{\boldsymbol{x}\}) = 0$ para cada ponto $\boldsymbol{x} \in \mathbb{R}^n$.

(1) $\mathbb{P}(A) \geqslant 0$ para todo $A \in \mathcal{F}$.
(2) $\mathbb{P}(\Omega) = 1$.
(3) Se $(A_n)_n$ são eventos disjuntos (isto é, $A_j \cap A_k = \emptyset$ para todos $j \neq k$), então $\mathbb{P}(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mathbb{P}(A_n)$.

Esta última propriedade é chamada *σ-aditividade.*

A partir das propriedades acima, podem-se demonstrar inúmeras outras. Listamos abaixo as mais comuns.

Teorema 1.35. *Sejam $\mathbb{P}$ uma medida de probabilidade, $A, B, A_1, A_2, \cdots \in \mathcal{F}$. Então:*

(4) $\mathbb{P}(\emptyset) = 0$.
(5) $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.
(6) *Se $A \subseteq B$, então $\mathbb{P}(A) \leqslant \mathbb{P}(B)$ e $\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A)$.*
(7) $0 \leqslant \mathbb{P}(A) \leqslant 1$.
(8) $\mathbb{P}(\cup_{n=1}^{\infty} A_n) \leqslant \sum_{n=1}^{\infty} \mathbb{P}(A_n)$.
(9) $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.

Demonstração. Para provar (4), tome $A_1 = \Omega$ e $A_n = \emptyset$ para $n = 2, 3, 4, \ldots$. Se $\mathbb{P}(\emptyset)$ fosse estritamente positivo, a equação que define a Propriedade (3) daria 1 no lado esquerdo e $+\infty$ no lado direito. Para provar (5), basta tomar $A_1 = A$, $A_2 = A^c$ e $A_j = \emptyset$ para $j = 3, 4, \ldots$, então pelas Propriedades (3) e (4) segue que $1 = \mathbb{P}(\Omega) = \mathbb{P}(A \cup A^c \cup \emptyset \cup \cdots \cup \emptyset) = \mathbb{P}(A) + \mathbb{P}(A^c)$.

O item (6) é provado escrevendo $B = A \cup (B \setminus A)$. Com efeito, segue de (3) que $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A)$, logo $\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A)$, e de (1) concluímos que $\mathbb{P}(B) \geqslant \mathbb{P}(A)$. Usando essa última propriedade e observando que $\emptyset \subseteq A \subseteq \Omega$, obtemos $0 = \mathbb{P}(\emptyset) \leqslant \mathbb{P}(A) \leqslant \mathbb{P}(\Omega) = 1$, o que prova (7).

Para provar (8), defina $B_1 = A_1$ e $B_n = A_n \setminus (\cup_{k=1}^{n-1} A_k)$ para $n \geqslant 2$. Observe que os conjuntos $B_1, B_2, \ldots$ são disjuntos, $\cup_{n=1}^{\infty} A_n = \cup_{n=1}^{\infty} B_n$, e que $B_n \subseteq A_n$ para todo n. Usando (3) e (6), obtemos $\mathbb{P}(\cup_{n=1}^{\infty} A_n) = \mathbb{P}(\cup_{n=1}^{\infty} B_n) = \sum_{n=1}^{\infty} \mathbb{P}(B_n) \leqslant \sum_{n=1}^{\infty} \mathbb{P}(A_n)$. Para provar (9), usamos (3) para escrever $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A)$ e $\mathbb{P}(B) = \mathbb{P}(A \cap B) + \mathbb{P}(B \setminus A)$, donde $\mathbb{P}(A \cup B) =$

$\mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$. $\square$

Uma medida de probabilidade $\mathbb{P}$ também tem a propriedade de ser contínua. Dada uma sequência $(A_n)_n$ de eventos, denotamos por $A_n \uparrow A$ a propriedade de que $A_1 \subseteq A_2 \subseteq A_3 \subseteq \cdots$ e $\cup_{n=1}^{\infty} A_n = A$. Analogamente, escrevemos $A_n \downarrow A$ para denotar que $A_1 \supseteq A_2 \supseteq A_3 \supseteq \cdots$ e $\cap_{n=1}^{\infty} A_n = A$.

Teorema 1.36 (Continuidade)**.** *Se $A_n \uparrow A$ ou $A_n \downarrow A$, então $\mathbb{P}(A_n) \to \mathbb{P}(A)$.*

Demonstração. Suponha que $A_n \uparrow A$. Fixe $A_0 = \emptyset$ e defina $B_n = A_n \setminus A_{n-1}$ para $n \in \mathbb{N}$, de modo que $\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} B_n$ e a última união é disjunta. Assim,

$$\begin{aligned}
\mathbb{P}\left(\bigcup_{k=1}^{\infty} A_k\right) &= \mathbb{P}\left(\bigcup_{k=1}^{\infty} B_k\right) = \sum_{k=1}^{\infty} \mathbb{P}(B_k) \\
&= \sum_{k=1}^{\infty} \mathbb{P}(A_k \setminus A_{k-1}) = \lim_{n\to\infty} \sum_{k=1}^{n} \mathbb{P}(A_k \setminus A_{k-1}) \\
&= \lim_{n\to\infty} \sum_{k=1}^{n} \Big((\mathbb{P}(A_k) - \mathbb{P}(A_{k-1})\Big) = \lim_{n\to\infty} \mathbb{P}(A_n),
\end{aligned}$$

provando o primeiro caso. Suponha agora que $A_n \downarrow A = \cap_{n=1}^{\infty} A_n$. Observando que $A_n^c \uparrow \cup_{n=1}^{\infty} A_n^c$, pela parte já demonstrada, $\mathbb{P}(A_n) = 1 - \mathbb{P}(A_n^c) \to 1 - \mathbb{P}(A^c) = \mathbb{P}(A)$. $\square$

Finalmente introduzimos o conceito de espaço de probabilidade, que nada mais é que a justaposição das noções de espaço amostral, eventos aleatórios e medida de probabilidade.

Definição 1.37 (Espaço de probabilidade)**.** Um *espaço de probabilidade* é um trio $(\Omega, \mathcal{F}, \mathbb{P})$, onde

(1) Ω é um conjunto não-vazio;
(2) $\mathcal{F}$ é uma σ-álgebra de subconjuntos de Ω;
(3) $\mathbb{P}$ é uma probabilidade definida em $\mathcal{F}$.

Exemplo 1.38. Lançamento de uma moeda. Este espaço é pequeno o suficiente para que possamos construí-lo explicitamente. Como fizemos anteriormente, as duas faces da moeda serão representadas em $\Omega = \{0, 1\}$. A σ-álgebra $\mathcal{F}$ é dada por $\mathcal{F} = \mathcal{P}(\Omega) = \{\emptyset, \{0\}, \{1\}, \{0, 1\}\}$. A medida de probabilidade $\mathbb{P} : \mathcal{F} \to \mathbb{R}$ é dada por $\mathbb{P}(\emptyset) = 0$, $\mathbb{P}(\{0\}) = \mathbb{P}(\{1\}) = \frac{1}{2}$, $\mathbb{P}(\{0, 1\}) = 1$. $\triangle$

1.4 Espaços de medida

Esta seção pode ser omitida em uma primeira leitura. É pré-requisito para as Seções 5.5, 11.4, e portanto para os Capítulos 12–15. A leitura desta seção pressupõe familiaridade com os tópicos do Apêndice A.3 e usa conceitos desenvolvidos no Apêndice C.

1.4.1 σ-álgebras e conjuntos borelianos

As ferramentas que veremos nas partes mais avançadas deste livro funcionam com uma σ-álgebra $\mathcal{F}$ de eventos. Queremos que essa σ-álgebra seja suficientemente grande de forma a conter os conjuntos que queremos estudar. Em inúmeras situações, para que a teoria funcione bem ou represente adequadamente os objetos estudados, vamos trabalhar com a menor σ-álgebra que contém uma dada classe $\mathcal{E}$, cuja existência e unicidade enunciamos agora.

Proposição 1.39 (σ-álgebra gerada por uma classe de subconjuntos)**.** *Seja Ω um espaço amostral e $\mathcal{E}$ uma classe de subconjuntos de Ω. Então existe uma única σ-álgebra, chamada* a σ-álgebra gerada por $\mathcal{E}$*, denotada por $\sigma(\mathcal{E})$, que satisfaz às seguintes propriedades:*

(1) *$\sigma(\mathcal{E})$ é uma σ-álgebra,*
(2) *$\sigma(\mathcal{E}) \supseteq \mathcal{E}$,*
(3) *se $\mathcal{F} \supseteq \mathcal{E}$ e $\mathcal{F}$ é uma σ-álgebra, então $\mathcal{F} \supseteq \sigma(\mathcal{E})$.*

Daremos a prova no final deste capítulo, em forma de exercício guiado.

Exemplo 1.40. Sejam Ω um espaço amostral, $A \subseteq \Omega$ e $\mathcal{E} = \{A\}$. Afirmamos que a σ-álgebra gerada por $\mathcal{E}$ é $\sigma(\mathcal{E}) = \{\emptyset, \Omega, A, A^c\}$. Com efeito, essa classe é uma σ-álgebra, ela contém $\mathcal{E}$, e qualquer σ-álgebra que contenha $\mathcal{E}$ tem que contê-la. $\triangle$

Exemplo 1.41. Seja $\mathcal{E} = \{A_1, \dots, A_n\}$ uma coleção de subconjuntos disjuntos do espaço amostral Ω, tais que $\cup_{k=1}^n A_k = \Omega$. Então a σ-álgebra gerada por $\mathcal{E}$ é $\mathcal{A} = \{\emptyset\} \cup \{A_{j_1} \cup \cdots \cup A_{j_m} : j_1, \dots, j_m \in \{1, \dots, n\}\}$. Com efeito, $\mathcal{A}$ é uma σ-álgebra, $\mathcal{A} \supseteq \mathcal{E}$, e qualquer σ-álgebra que contenha $\mathcal{E}$ deve contê-la. $\triangle$

A σ-álgebra $\sigma(\mathcal{E})$ contém todos os conjuntos que podem ser obtidos como complementos e uniões enumeráveis dos conjuntos em $\mathcal{E}$, e também os que podem ser obtidos como complemento e união enumerável destes últimos, e assim por diante. Isso é útil porque nos diz que todos os conjuntos que podemos construir a partir de $\mathcal{E}$ com uma sequência enumerável de operações estarão em $\sigma(\mathcal{E})$. Entretanto, essa tentativa de construção quase explícita de $\sigma(\mathcal{E})$ é difícil de formalizar. Quando queremos descrever $\sigma(\mathcal{E})$, é melhor trabalhar diretamente com as três propriedades que a caracterizam, como fizemos nos dois exemplos acima.

Quando o espaço amostral é $\mathbb{R}$, a σ-álgebra mais importante é a seguinte.

Definição 1.42 (Conjuntos borelianos na reta)**.** Para o espaço $\Omega = \mathbb{R}$, a *σ-álgebra de Borel* $\mathcal{B}(\mathbb{R})$ é a σ-álgebra gerada por conjuntos abertos de $\mathbb{R}$. Quando estiver claro no contexto que $\Omega = \mathbb{R}$, podemos escrever $\mathcal{B}$ ao invés de $\mathcal{B}(\mathbb{R})$. Os conjuntos em $\mathcal{B}$ são chamados *borelianos*.

Qualquer conjunto "razoável" é um boreliano. Por exemplo, um conjunto pontual $\{x\}$ pode ser obtido como interseção de intervalos abertos $(x - \frac{1}{n}, x + \frac{1}{n})$, e um conjunto enumerável $\{x_1, x_1, \dots\}$ pode ser obtido união de conjuntos $\{x_n\}$. Logo, qualquer conjunto enumerável, ou cujo complemento é enumerável, é boreliano. Intervalos $(a, b]$ também são borelianos pois $(a, b] = \cap_n (a, b + \frac{1}{n})$. De fato, é bem trabalhoso construir um subconjunto de $\mathbb{R}$ que não seja boreliano. Mesmo assim, devemos ter em mente que a Teoria da Probabilidade está baseada em σ-álgebras, e $\mathcal{B}$ é a σ-álgebra mais

importante em $\mathbb{R}$.

Proposição 1.43. *A classe $\mathcal{B}(\mathbb{R})$ também é a σ-álgebra gerada pela classe dos intervalos da reta, e também é a σ-álgebra gerada pelos intervalos semi-infinitos à esquerda e fechados à direita.*

Demonstração. Denotemos por $\mathcal{I}$ a classe de todos os intervalos da reta, por $\mathcal{O}$ a classe de todos os subconjuntos abertos da reta, e por $\mathcal{D}$ a classe de todos os intervalos semi-infinitos à esquerda e fechados à direita.

Vamos mostrar que $\sigma(\mathcal{D}) \subseteq \sigma(\mathcal{O}) \subseteq \sigma(\mathcal{I}) \subseteq \sigma(\mathcal{D})$.

Seja $D \in \mathcal{D}$. Então $D = (-\infty, b]$ para algum $b \in \mathbb{R}$. Como $(b, +\infty) \in \mathcal{O}$ e $\sigma(\mathcal{O})$ é fechada por complementos, segue que $D \in \sigma(\mathcal{O})$. Logo, $\mathcal{D} \subseteq \sigma(\mathcal{O})$ e, portanto, $\sigma(\mathcal{D}) \subseteq \sigma(\mathcal{O})$. Seja $A \in \mathcal{O}$. Pelo Teorema A.9, $A = \cup_n I_n$, onde $\{I_n\}_n \subseteq \mathcal{I}$. Como $\sigma(\mathcal{I})$ é fechada por uniões enumeráveis, segue que $A \in \sigma(\mathcal{I})$. Logo, $\mathcal{O} \subseteq \sigma(\mathcal{I})$ e, portanto, $\sigma(\mathcal{O}) \subseteq \sigma(\mathcal{I})$. Seja $I \in \mathcal{I}$. O intervalo I pode ser da forma $[a, b)$, $(a, b]$, $[a, b]$, $[a, +\infty)$, etc. Vamos considerar $I = [a, b)$ e deixar os outros sete casos como exercício. Observe que

$$[a, b) = (-\infty, b) \setminus (-\infty, a) = (\cup_n(-\infty, b - n^{-1}]) \setminus (\cup_k(-\infty, a - k^{-1}]).$$

Como todos esses intervalos $(-\infty, x]$ estão em $\mathcal{D}$, e $\sigma(\mathcal{D})$ é fechada por uniões enumeráveis e diferenças, segue que $I \in \sigma(\mathcal{D})$. Logo, $\mathcal{I} \subseteq \sigma(\mathcal{D})$ e, portanto, $\sigma(\mathcal{I}) \subseteq \sigma(\mathcal{D})$. Isso conclui a cadeia de inclusões e a prova da proposição. $\square$

Dado $J \in \mathcal{B}(\mathbb{R})$ não-vazio, por exemplo $J = [0, 1]$, definimos $\mathcal{B}(J) = \{A \in \mathcal{B}(\mathbb{R}) : A \subseteq J\}$. Observe que $\mathcal{B}(J)$ é uma σ-álgebra no espaço amostral J.

Dado $n \in \mathbb{N}$, denotamos o conjunto das sequências de n números reais por

$$\mathbb{R}^n = \{(x_1, \ldots, x_n) : x_k \in \mathbb{R} \text{ para todo } k = 1, \ldots, n\}.$$

Há uma σ-álgebra natural associada ao espaço $\mathbb{R}^n$.

Definição 1.44. Definimos a *σ-álgebra de Borel* $\mathcal{B}(\mathbb{R}^n)$ como a σ-álgebra

em $\mathbb{R}^n$ gerada pela classe dos subconjuntos abertos de $\mathbb{R}^n$. Os conjuntos em $\mathcal{B}(\mathbb{R}^n)$ são chamados *borelianos em* $\mathbb{R}^n$.

1.4.2 Medidas

Um *espaço mensurável* é um par $(\Omega, \mathcal{F})$, onde Ω é um conjunto não-vazio e $\mathcal{F}$ é uma σ-álgebra em Ω.

Definição 1.45 (Medidas e espaços de medida)**.** Seja $(\Omega, \mathcal{F})$ um dado espaço mensurável. Uma função $\mu : \mathcal{F} \to [0, +\infty]$ é chamada *medida em* $(\Omega, \mathcal{F})$ se

(1) $\mu(\emptyset) = 0$,
(2) $\mu(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mu(A_n)$ para toda sequência $(A_n)_n$ de conjuntos em $\mathcal{F}$ disjuntos. Esta propriedade chama-se *σ-aditividade.*

O trio $(\Omega, \mathcal{F}, \mu)$ é chamado *espaço de medida.* Dizemos que μ é uma *medida finita* se $\mu(\Omega) < \infty$. Dizemos que μ é uma *medida σ-finita* se existem conjuntos $(A_n)_{n\in\mathbb{N}}$ em $\mathcal{F}$ tais que $\cup_{n\in\mathbb{N}} A_n = \Omega$ e $\mu(A_n) < \infty$ para todo n.

Quando $\mu(\Omega) = 1$, a medida μ é uma medida de probabilidade, como definida na seção anterior. Vejamos alguns exemplos simples de medidas.

Exemplo 1.46 (Massa puntual de Dirac). Sejam $(\Omega, \mathcal{F})$ um espaço mensurável qualquer e $x \in \Omega$. A função $\delta_x : \mathcal{F} \to \{0, 1\}$ definida por

$$\delta_x(A) = \begin{cases} 1, & \text{se } x \in A \\ 0, & \text{se } x \notin A \end{cases}.$$

é uma medida de probabilidade em $(\Omega, \mathcal{F})$. $\triangle$

Exemplo 1.47 (Medida de contagem). Seja $(\Omega, \mathcal{F})$ um espaço mensurável qualquer. A *medida de contagem em* Ω é definida como $\sum_{x\in\Omega} \delta_x$. Observe que a medida de contagem é finita se, e somente se, Ω for um conjunto finito, e é σ-finita se, e somente se, Ω for um conjunto enumerável. A medida de contagem em $\mathbb{R}$ não é σ-finita. $\triangle$

Proposição 1.48. *Se $(\Omega, \mathcal{F}, \mu)$ é um espaço de medida e $A \in \mathcal{F}$, então a função $\mu_{|_A}$ dada por $\mu_{|_A}(B) = \mu(A \cap B)$ também é uma medida em $(\Omega, \mathcal{F})$.*

Deixamos a prova como exercício.

Proposição 1.49 (Propriedades de uma medida)**.** *Sejam $(\Omega, \mathcal{F}, \mu)$ um espaço de medida e $A_1, A_2, \dots \in \mathcal{F}$. Então:*

(1) $\mu(\cup_{j=1}^{\infty} A_j) \leqslant \sum_{j=1}^{\infty} \mu(A_j)$,
(2) *Se $A_n \uparrow A$, então $\mu(A_n) \to \mu(A)$.*

Essas propriedades chamam-se subaditividade *e* continuidade por baixo*.*

A prova é idêntica àquela já dada para medidas de probabilidade.

Observação 1.50. A "continuidade por cima" pode ser falsa caso $\mu(A_j)$ seja infinita para todo j. Por exemplo, se μ é a medida de contagem em $\mathbb{N}$ e $A_n = \{n, n+1, n+2, \dots\}$, temos que $A_n \downarrow \emptyset$ porém $\mu(A_n) \not\to \mu(\emptyset)$, pois $\mu(\emptyset) = 0$ e $\mu(A_n) = \infty$ para todo n. △

O teorema a seguir, cuja prova será dada no Apêndice D.2, nos garante a existência de uma importante medida em $(\mathbb{R}, \mathcal{B})$ que generaliza a noção de comprimento de intervalos.

Teorema 1.51. *Existe uma única medida m em $(\mathbb{R}, \mathcal{B})$ tal que*

$$m\big((a, b]\big) = b - a$$

para todos $a < b \in \mathbb{R}$. Essa medida m é chamada medida de Lebesgue *em $\mathbb{R}$.*

A medida de Lebesgue é σ-finita, pois tomando $A_n = [-n, n]$, temos que $m(A_n) = 2n < \infty$ para todo n e $A_n \uparrow \mathbb{R}$.

1.5 Exercícios

§1.1

1. Considere o experimento resultante do lançamento de dois dados onde se observa o mínimo entre suas faces. Construa um modelo probabilístico associado.

§1.2

2.

(a) De um baralho comum (52 cartas) são retiradas, sem reposição, uma amostra de 5 cartas. Qual evento é mais provável, sair uma quadra (4 cartas com o mesmo número) ou um *flush* (5 cartas do mesmo naipe, sem formar uma sequência de 5 números consecutivos)?

(b) Repita o exercício anterior, supondo agora um baralho com apenas as cartas $7, 8, 9, 10, J, Q, K$ e A (32 cartas).

3. Para um jantar de gala, o cozinheiro preparou $2n$ pratos, sendo n deles com carne mas sem glúten e os outros n pratos vegetarianos mas com glúten. Dos convidados, $a < n$ são vegetarianos, $b < n$ não comem glúten e $2n - a - b$ convidados comem qualquer coisa. Os pratos são servidos aleatoriamente. Calcule a probabilidade de todas as restrições alimentares serem respeitadas.

4. Sem utilizar a fórmula $\binom{n}{k} = \frac{n!}{k!(n-k)!}$, mostre combinatorialmente a identidade $k\binom{n}{k} = n\binom{n-1}{k-1}$.

5. Prove que $\binom{a+b}{n} = \sum_{k=0}^{n} \binom{a}{k}\binom{b}{n-k}$.

6. Sejam f e g funções n vezes deriváveis. Prove a Fórmula de Leibniz para a n-ésima derivada de $f \cdot g$.

$$\frac{\mathrm{d}^n(f \cdot g)}{\mathrm{d}x^n} = \sum_{k=0}^{n} \binom{n}{k} \frac{\mathrm{d}^k f}{\mathrm{d}x^k} \cdot \frac{\mathrm{d}^{n-k} g}{\mathrm{d}x^{n-k}}.$$

7. Mostre o Teorema das Colunas do Triângulo de Pascal: para todos n e k inteiros não-negativos, $\sum_{j=0}^{k} \binom{n+j}{n} = \binom{k+n+1}{n+1}$.

Dica: Use como hipótese de indução em k que a identidade vale para todo n.

8. Mostre, sem utilizar indução, o Teorema das Diagonais do Triângulo de Pascal: para todos n e k inteiros não-negativos, $\sum_{j=0}^{k} \binom{n+j}{j} = \binom{k+n+1}{k}$.

9. A soma da n-ésima diagonal inversa do Triângulo de Pascal é dada pela expressão $F_n = \sum_{k=0}^{n} \binom{n-k}{k}$. Calcule F_0, F_1, F_2, F_3 e F_4. Conjecture, quem é a sequência $(F_n)_n$? Prove sua conjectura.

10. Prove por indução o Teorema das Linhas e o Teorema Binomial.

11. No final da Seção 1.1.3, observamos que há 21 resultados possíveis no lançamento simultâneo de dois dados idênticos, correspondendo aos 21 pares não-ordenados de números entre 1 e 6. Um jogo de dominó para crianças, em que cada metade de peça pode ser pintada com uma de 7 cores do arco-íris, possui 28 peças, correspondendo aos 28 pares não-ordenados de cores do arco-íris. De forma mais geral, o número de pares não-ordenados de um conjunto com n elementos é $\binom{n+1}{2}$. Prove este fato de quatro formas diferentes:

(a) Por indução em n e usando propriedades do Triângulo de Pascal.
(b) Decompondo a coleção de pares não-ordenados em pares de elementos distintos e pares de elementos iguais.
(c) Descrevendo uma correspondência direta entre pares não-ordenados de $\{1, \ldots, n\}$ e subconjuntos de $\{\star, 1, \ldots, n\}$ com dois elementos, de forma que o novo elemento $\star$ cumpra um papel especial.
(d) Descrevendo uma correspondência direta entre pares não-ordenados de $\{1, \ldots, n\}$ e palavras compostas por $n-1$ símbolos "|" e dois símbolos "·", como por exemplo "|·||||·".

12. Inspirando-se pelo último item do exercício anterior, mostre que o número de soluções inteiras não-negativas da equação $x_1 + \cdots + x_n = k$ é dado por $\binom{n-1+k}{k}$.

13. Oito clientes formam uma fila ao acaso, quatro deles possuem uma única nota de \$10 e os outros quatro com uma única nota de \$20 cada um. O ingresso custa \$10 e o bilheteiro começa a atender os clientes sem troco.

(a) Qual a probabilidade de o bilheteiro não ter problema de troco?

(b) Refaça o item (a) supondo que sejam $2n$ clientes, n clientes com notas de \$10 e n com notas de \$20.

14. Sejam Ω o espaço dos caminhos simples de duração $2n$, conforme definido no Exemplo 1.32, e $\mathbb{P}$ a medida de probabilidade que atribui peso 2^{-2n} a cada caminho. O caminho resultante desse experimento é comumente chamado de *passeio aleatório*. Mostre que

$$\mathbb{P}(\{s : \max\{s_0, \ldots, s_{2n}\} \geqslant 5\}) = 2\,\mathbb{P}(\{s : s_{2n} \geqslant 5\}).$$

15. Vamos mostrar que a probabilidade de um passeio aleatório ainda não ter retornado à origem depois de $2n$ passos é igual à probabilidade de estar na origem no tempo $2n$. Ou seja, vamos mostrar que

$$\mathbb{P}(\{s : \tau_0^+ > 2n\}) = \mathbb{P}(\{s : s_{2n} = 0\}),$$

onde $\tau_0^+ = \tau_0^+(s) = \min\{k \geqslant 1 : s_k = 0\}$ denota o primeiro instante de retorno à origem, com $\min(\emptyset) = +\infty$. Para $k \in \mathbb{N}$, mostre que:

(a) $\mathbb{P}(s_1 = +1, \tau_0^+ \leqslant 2n, s_{2n} = 2k) = \mathbb{P}(s_{2n-1} = -2k-1, s_{2n} = -2k)$.
(b) $\mathbb{P}(s_1 = +1, s_{2n} = 2k) = \mathbb{P}(s_{2n-1} = 2k-1, s_{2n} = 2k-2)$.
(c) $\mathbb{P}(s_1 = +1, \tau_0^+ > 2n) = \mathbb{P}(s_{2n-1} = +1, s_{2n} = 0)$.
(d) $\mathbb{P}(\tau_0^+ > 2n) = \mathbb{P}(s_{2n} = 0)$.

§1.3

16. Sejam A e B eventos quaisquer, mostre que:

(a) $\mathbb{P}(A \cap B) \geqslant 1 - \mathbb{P}(A^c) - \mathbb{P}(B^c)$.
(b) $\mathbb{P}(A \cup B)\mathbb{P}(A \cap B) \leqslant \mathbb{P}(A)\mathbb{P}(B)$. Dê condições necessárias e suficientes para que valha a igualdade.

17. Sejam $(\Omega, \mathcal{F}, \mathbb{P})$ um espaço de probabilidade, A e B eventos quaisquer. Encontre uma expressão para a probabilidade de ocorrer exatamente um dos eventos A ou B.

18. Sejam $\mathbb{P}_1$ e $\mathbb{P}_2$ medidas de probabilidade definidas no mesmo espaço

$(\Omega, \mathcal{F})$. A função $\tilde{\mathbb{P}}(A) = \max\{\mathbb{P}_1(A), \mathbb{P}_2(A)\}$ é medida de probabilidade? Prove ou dê contra-exemplo.

19. Sejam $A_1, A_2, A_3, \ldots$ eventos aleatórios. Prove que:

(a) $\mathbb{P}(A_1 \cap \cdots \cap A_n) \geqslant 1 - (\mathbb{P}(A_1^c) + \cdots + \mathbb{P}(A_n^c))$.
(b) $\mathbb{P}(\cap_{n=1}^{\infty} A_n) \geqslant 1 - \sum_{n=1}^{\infty} \mathbb{P}(A_n^c)$.

20. Suponha que $\mathbb{P}(A_n) = 1$ para todo $n \in \mathbb{N}$. Prove que $\mathbb{P}(\cap_n A_n) = 1$.

21. Dada uma sequência de eventos $(A_n)_n$, definimos os eventos

$$B = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k \qquad \text{e} \qquad C = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k.$$

Mostre que, se $B = C$, então $\mathbb{P}(A_n) \to \mathbb{P}(B)$.

22. Sejam $(\Omega, \mathcal{F}, \mathbb{P})$ um espaço de probabilidade e $(A_n)_n$ e $(B_n)_n$ sequências de eventos tais que $\mathbb{P}(A_n) \to 1$ e $\mathbb{P}(B_n) \to \alpha$. É sempre verdadeiro que $\mathbb{P}(A_n \setminus B_n) \to 1 - \alpha$? Prove ou dê contra-exemplo.

§1.4

23. Sejam Ω um conjunto e $\mathcal{A}$ a classe de todos os subconjuntos que são enumeráveis ou cujo complementar é enumerável. Prove que $\mathcal{A}$ é uma σ-álgebra.

24. Dê um exemplo de um espaço Ω e σ-álgebras $\mathcal{F}_1$ e $\mathcal{F}_2$ em Ω tais que $\mathcal{F}_1 \cup \mathcal{F}_2$ não é uma σ-álgebra em Ω.

25. Neste exercício provaremos a Proposição 1.39.

(a) Sejam $J \neq \emptyset$ um conjunto qualquer de índices e $(\mathcal{F}_j)_{j \in J}$ uma família de σ-álgebras em um mesmo espaço amostral Ω. Mostre que $\cap_{j \in J} \mathcal{F}_j$ é uma σ-álgebra em Ω.
(b) Sejam $\mathcal{E}$ uma classe qualquer de subconjuntos de Ω e J o conjunto de todas as σ-álgebras em Ω que contêm $\mathcal{E}$. Justifique que $J \neq \emptyset$.

(c) Utilize os itens anteriores para provar a existência de uma σ-álgebra satisfazendo aos três itens da proposição.

(d) Conclua a prova da proposição mostrando a unicidade de $\sigma(\mathcal{E})$. Ou seja, mostre que, se $\mathcal{G}$ e $\mathcal{H}$ satisfazem aos três itens da proposição, então $\mathcal{G} = \mathcal{H}$.

26. Sejam $\Omega = \{1, 2, 3, 4\}$ e $\mathcal{E} = \{\{1\}, \{2\}\}$. Quem é $\sigma(\mathcal{E})$? Prove.

27. Seja $\mathcal{E} = \{\{x\} : x \in \Omega\}$ a classe dos subconjuntos unitários de Ω. Quem é $\sigma(\mathcal{E})$? Prove.

28. Seja $(\Omega, \mathcal{F}, \mu)$ um espaço de medida. Prove que são equivalentes:

(i) A medida μ é σ-finita;

(ii) Existem $(A_n)_n$ disjuntos tais que $\cup_n A_n = \Omega$ e $\mu(A_n) < \infty$ para todo n;

(iii) Existem $(A_n)_n$ tal que $A_n \uparrow \Omega$ e $\mu(A_n) < \infty$ para todo n.

29. Seja $(\Omega, \mathcal{F}, \mathbb{P})$, onde Ω é um conjunto qualquer não-enumerável, $\mathcal{F} = \{A \subseteq \Omega : A$ é enumerável ou A^c é enumerável$\}$ e $\mathbb{P}(A) = 0$, se A for enumerável, ou $\mathbb{P}(A) = 1$, se A for não-enumerável. Mostre que $(\Omega, \mathcal{F}, \mathbb{P})$ é um espaço de probabilidade.

30. Sejam $\mathbb{P}_1, \mathbb{P}_2, \ldots$ medidas de probabilidade em $(\Omega, \mathcal{F})$, e sejam $p_1, p_2, \cdots \in [0, 1]$ tais que $\sum_n p_n = 1$. Prove que

$$\mathbb{P} = \sum_n p_n \mathbb{P}_n$$

é uma medida de probabilidade em $(\Omega, \mathcal{F})$.

Capítulo 2

Probabilidade Condicional e Independência

A ocorrência de um certo evento pode alterar a chance de outros eventos também ocorrerem. Suponha que um campeonato de futebol está prestes a começar, e temos uma estimativa da probabilidade de determinado time vencer na quarta rodada. Nossa estimativa certamente deverá ser atualizada após observarmos os resultados das três primeiras rodadas.

Caso a ocorrência de um dado evento não modifique as chances de um outro evento também ocorrer, diremos que esses dois eventos são independentes. Por exemplo, imagine que temos dois maços de baralho e retiramos uma carta de cada, neste caso a probabilidade de que as cartas sejam idênticas não é afetada se soubermos que a primeira carta é um 7♡.

2.1 Probabilidade condicional

A probabilidade condicional é uma nova medida de probabilidade, de forma a representar melhor as chances de eventos aleatórios a partir da observação da ocorrência ou não de um dado evento.

Definição 2.1 (Probabilidade Condicional)**.** Sejam $(\Omega, \mathcal{F}, \mathbb{P})$ espaço de probabilidade e $A, B \in \mathcal{F}$ eventos, com $\mathbb{P}(B) > 0$, definimos a *probabilidade condicional de A dado B* por

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

Quando $\mathbb{P}(B) = 0$, definimos $\mathbb{P}(A|B) = \mathbb{P}(A)$.

Observe que quando $\mathbb{P}(B) = 0$, a razão $\frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$ seria da forma "$\frac{0}{0}$". Alguns livros preferem não definir a probabilidade condicional $\mathbb{P}(A \mid B)$ quando $\mathbb{P}(B) = 0$, já outros optam por definir $\mathbb{P}(A \mid B)$ como 0 ou $\mathbb{P}(A)$. Esta última definição faz com que a maioria dos teoremas estabelecidos neste capítulo permaneçam válidos sem que tenhamos que supor que $\mathbb{P}(B) > 0$ ou considerar o caso $\mathbb{P}(B) = 0$ separadamente.

Proposição 2.2. *Dado $B \in \mathcal{F}$, a função que leva A em $\mathbb{P}(A|B)$ é uma medida de probabilidade em $(\Omega, \mathcal{F})$, isto é, satisfaz às condições da Definição 1.34.*

Demonstração. Se $\mathbb{P}(B) = 0$, não há nada a provar pois temos $\mathbb{P}(A|B) = \mathbb{P}(A)$ e $\mathbb{P}$ é uma medida de probabilidade. Considerando que $\mathbb{P}(B) > 0$, pela definição de probabilidade condicional, $\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \geqslant 0$, pois o denominador é positivo e o numerador é não-negativo;

$$\mathbb{P}(\Omega|B) = \frac{\mathbb{P}(\Omega \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B)}{\mathbb{P}(B)} = 1;$$

se $(A_n)_n$ é uma sequência de eventos disjuntos,

$$\begin{aligned}\mathbb{P}(\cup_n A_n|B) &= \frac{\mathbb{P}((\cup_n A_n) \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(\cup_n (A_n \cap B))}{\mathbb{P}(B)} \\ &= \frac{\sum_n \mathbb{P}(A_n \cap B)}{\mathbb{P}(B)} = \sum_n \mathbb{P}(A_n|B),\end{aligned}$$

onde na segunda igualdade utilizamos a distributividade da união com respeito à interseção e na terceira o fato dos eventos $(A_n \cap B)_n$ também serem disjuntos.

Isto conclui que a função que leva A em $\mathbb{P}(A|B)$ é de fato uma medida de probabilidade. $\square$

Exemplo 2.3. Um dado é lançado. Sabendo-se que o resultado é maior que 3, qual a probabilidade de que seja par? Denotamos o primeiro evento por $B = \{4, 5, 6\}$ e o segundo por $A = \{2, 4, 6\}$. Aplicando a definição, a probabilidade de ser par sabendo-se que o resultado é maior que 3 é dada por

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(\{4, 6\})}{\mathbb{P}(\{4, 5, 6\})} = \frac{2/6}{3/6} = \frac{2}{3}. \qquad \triangle$$

Exemplo 2.4. Suponha que, numa dada população, aproximadamente 30% dos genes de cor de olhos seja para olho claro c, e aproximadamente 70% seja para olhos escuros C. Suponha também que os casais se formam e decidem ter filhos aleatoriamente (sem nenhuma influência da cor dos olhos), de forma que os genes C e c estejam uniformemente espalhados por toda uma população. Assim, um indivíduo escolhido aleatoriamente nesta população terá os genes cc com probabilidade 0,090, ou Cc com probabilidade 0,42, ou CC com probabilidade 0,49. Como ter olhos claros é uma característica genética recessiva, apenas o primeiro grupo terá olhos claros, enquanto os dois últimos terão o mesmo fenótipo, de olhos escuros. Uma pessoa de olhos escuros é selecionada ao acaso. Qual a probabilidade de que essa pessoa tenha o gene recessivo para olhos claros? Denotando $B =$ "olhos escuros"$= \{Cc, CC\}$ e $A =$ "tem um gene de olhos claros"$= \{cc, Cc\}$, temos

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(\{Cc\})}{\mathbb{P}(\{Cc, CC\})} \approx \frac{0{,}42}{0{,}91} \approx 0{,}46,$$

com dois algarismos significativos. $\triangle$

2.1.1 Regra do Produto

A Regra do Produto permite expressar a probabilidade da ocorrência simultânea de diversos eventos a partir do valor de cada probabilidade

condicional de cada um deles dados os eventos anteriores.

Teorema 2.5 (Regra do Produto)**.** *Dados os eventos* $A_1, A_2, \ldots, A_n$*,*

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

Demonstração. Vamos provar por indução em n. Para $n = 1$, a igualdade vale trivialmente. Para $n = 2$, se $\mathbb{P}(A_1) > 0$, então

$$\mathbb{P}(A_2|A_1) = \frac{\mathbb{P}(A_2 \cap A_1)}{\mathbb{P}(A_1)},$$

donde

$$\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1)\mathbb{P}(A_2|A_1);$$

se $\mathbb{P}(A_1) = 0$, então $\mathbb{P}(A_1 \cap A_2) = 0$, logo a igualdade acima continua válida. Suponhamos, por hipótese de indução, que o teorema seja válido para $n = m$. Neste caso, para $n = m + 1$ podemos desenvolver

$$\begin{aligned}
\mathbb{P}(A_1 \cap \cdots \cap A_{m+1}) &= \mathbb{P}(A_1 \cap \cdots \cap A_m)\mathbb{P}(A_{m+1}|A_1 \cap \cdots \cap A_m) \\
&= \mathbb{P}(A_1)\mathbb{P}(A_2|A_1)\mathbb{P}(A_3|A_1 \cap A_2) \cdots \\
&\quad \cdots \mathbb{P}(A_m|A_1 \cap \cdots \cap A_{m-1})\mathbb{P}(A_{m+1}|A_1 \cap \cdots \cap A_m).
\end{aligned}$$

Com efeito, a primeira igualdade corresponde ao caso $n = 2$ já demonstrado, e a segunda corresponde ao caso $n = m$, o que completa a prova. □

Exemplo 2.6. Um móvel tem duas gavetas, a primeira gaveta contém três bolsas e a segunda contém quatro bolsas. A primeira bolsa da primeira gaveta contém duas bolas vermelhas e uma bola azul, e todas as demais bolsas contêm duas bolas azuis. Abre-se uma gaveta, escolhe-se uma bolsa e retira-se uma bola de dentro da bolsa, tudo ao acaso. Qual a probabilidade de que a bola retirada seja vermelha? Denotando $A =$ "abre-se a primeira gaveta", $B =$ "escolhe-se a primeira bolsa" e $C =$ "retira-se a bola vermelha",

pela Regra do Produto obtemos

$$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B|A)\mathbb{P}(C|B \cap A) = \frac{1}{2} \times \frac{1}{3} \times \frac{2}{3} = \frac{1}{9}. \qquad \triangle$$

Exemplo 2.7. Selecionar 3 cartas de um baralho de 52 cartas, ao acaso e sem reposição. Qual a probabilidade de tirar 3 reis? Seja $A_k =$ "tirar rei na k-ésima retirada" e $A =$ "tirar 3 reis" $= A_1 \cap A_2 \cap A_3$. Temos

$$\mathbb{P}(A) = \mathbb{P}(A_1)\mathbb{P}(A_2|A_1)\mathbb{P}(A_3|A_1 \cap A_2) = \frac{4}{52}\frac{3}{51}\frac{2}{50} = \frac{1}{5525}. \qquad \triangle$$

Exemplo 2.8. Continuando o Exemplo 2.4, quando um casal tem um filho, cada um dos pais transmite um dos seus dois genes com mesma probabilidade. Seleciona-se uma criança ao acaso. Qual a probabilidade de a criança e o pai ambos terem olhos claros? Definindo $A =$ "pai tem olhos claros", $B =$ "filho tem olhos claros", $D =$ "o gene transmitido pela mãe é c", temos

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B|A) = \mathbb{P}(A)\mathbb{P}(D) \approx 0{,}090 \times 0{,}30 = 0{,}027,$$

com dois algarismos significativos. $\triangle$

2.1.2 Lei da Probabilidade Total

Dizemos que $B_1, B_2, B_3, \dots \in \mathcal{F}$ formam uma *partição de* Ω se são disjuntos e $\cup_k B_k = \Omega$. Partições são particularmente úteis em contextos onde determinado aspecto divide os resultados possíveis em casos, e é possível expressar determinadas relações em cada um desses casos separadamente.

Teorema 2.9 (Lei da Probabilidade Total)**.** *Sejam $A, B_1, B_2, B_3, \dots$ eventos aleatórios em $(\Omega, \mathcal{F}, \mathbb{P})$ tais que $B_1, B_2, B_3, \dots$ formam uma partição de Ω. Então*

$$\mathbb{P}(A) = \sum_{j=1}^{\infty} \mathbb{P}(B_j)\mathbb{P}(A|B_j).$$

Demonstração. Usando a Regra do Produto,

$$\mathbb{P}(A) = \mathbb{P}\big(\cup_{j=1}^{\infty}(A \cap B_j)\big) = \sum_{j=1}^{\infty} \mathbb{P}(A \cap B_j) = \sum_{j=1}^{\infty} \mathbb{P}(B_j)\mathbb{P}(A|B_j).$$

A primeira igualdade vale porque $A = A \cap (\cup_{j=1}^{\infty} B_j) = \cup_{j=1}^{\infty}(A \cap B_j)$. Na segunda igualdade usamos que esses eventos são disjuntos. Na última igualdade usamos a Regra do Produto. □

A Lei da Probabilidade Total é particularmente útil quando um experimento tem duas etapas, e é possível expressar as probabilidades condicionais de determinado aspecto da etapa final dados os possíveis resultados da etapa inicial.

Exemplo 2.10. Um armário tem duas gavetas, A e B. A gaveta A tem 2 meias azuis e 3 meias pretas, e a gaveta B tem 3 meias azuis e 3 meias vermelhas. Abre-se uma gaveta ao acaso e retira-se uma meia ao acaso da gaveta escolhida. Qual a probabilidade de a meia escolhida ser azul? Para a solução, comecemos observando os valores conhecidos de probabilidade: $\mathbb{P}(A) = \mathbb{P}(B) = \frac{1}{2}$, $\mathbb{P}(\text{azul}|A) = \frac{2}{5}$ e $\mathbb{P}(\text{azul}|B) = \frac{3}{6}$. Assim,

$$\mathbb{P}(\text{azul}) = \mathbb{P}(A)\mathbb{P}(\text{azul}|A) + \mathbb{P}(B)\mathbb{P}(\text{azul}|B) = \frac{1}{2}\frac{2}{5} + \frac{1}{2}\frac{3}{6} = \frac{9}{20}. \qquad \triangle$$

2.1.3 Fórmula de Bayes

A *Fórmula de Bayes*

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

cuja demonstração é imediata, na verdade reflete uma forma de raciocinar extremamente profunda. Estamos interessados na chance de ocorrência do evento B, quando na verdade temos acesso à ocorrência ou não do evento A. Antes de se observar a ocorrência de A ou A^c, a chance de ocorrência de B,

chamada de probabilidade *a priori*, é simplesmente $\mathbb{P}(B)$. Uma vez observada a ocorrência de A ou A^c, procedemos a atualizar a chance do evento B, multiplicando-a pelo *fator de Bayes* dado pela razão entre a probabilidade de se observar A (ou A^c) no cenário em que B ocorre e a probabilidade de se observar A (ou A^c) em geral. O produto entre a probabilidade *a priori* e o fator de Bayes resulta no que chamamos de probabilidade *a posteriori*.

Claro que, usando essa fórmula, precisamos saber $\mathbb{P}(A|B)$ para calcular $\mathbb{P}(B|A)$, mas em muitas situações aquela é mais fácil de calcular do que esta. Entre os exemplos mais simples, estão os experimentos em várias etapas em que observamos apenas o resultado final e queremos estimar as chances de distintas possibilidades nas etapas iniciais. Ou seja, quando queremos determinar a probabilidade condicional de eventos que precedem aquele efetivamente observado.

Exemplo 2.11. Continuando o Exemplo 2.8, se selecionamos uma criança de olhos escuros ao acaso, qual a probabilidade de que o pai tenha olhos claros? Tomando $A =$ "a criança tem olhos claros", $B =$ "o pai tem olhos claros", e $D =$ "o pai transmite o gene de olhos claros", temos

$$\mathbb{P}(A|B) = \mathbb{P}(A|D \cap B) = \mathbb{P}(A|D) \approx 0{,}30$$

que representam a chance de a mãe também transmitir o gene de olhos claros. Pela Fórmula de Bayes,

$$\mathbb{P}(B|A^c) = \frac{\mathbb{P}(A^c|B)}{\mathbb{P}(A^c)}\mathbb{P}(B) \approx \frac{0{,}70}{0{,}910} \cdot 0{,}090 \approx 0{,}77 \cdot 0{,}090 \approx 0{,}069,$$

com dois algarismos significativos. O valor 0,090 é a probabilidade *a priori* de que o pai de uma criança selecionada ao acaso tenha olhos claros. O fator de Bayes resultante da observação de que a criança tem olhos escuros é 0,77, que por ser menor que 1 reduz a probabilidade *a posteriori* para 0,069. $\triangle$

Um uso particular da Fórmula de Bayes é quando conhecemos as probabilidades de uma sequência dos eventos B_j que particionam Ω e as probabilidades condicionais de um evento A dados os eventos dessa partição. Neste caso,

podemos calcular as probabilidades condicionais de ocorrência de cada B_j sabendo-se da ocorrência ou não do evento A, pela fórmula

$$\mathbb{P}(B_j|A) = \frac{\mathbb{P}(A|B_j)}{\mathbb{P}(A)} \cdot \mathbb{P}(B_j).$$

Os valores originais $\mathbb{P}(B_j)$ são as probabilidades *a priori* dos eventos B_j, e os valores $\mathbb{P}(B_j|A)$ são as probabilidades *a posteriori* desses eventos. Como estamos supondo que é possível calcular $\mathbb{P}(B_k)$ e $\mathbb{P}(A|B_k)$ para todo k, muitas vezes o denominador na Fórmula de Bayes será calculado pela Lei da Probabilidade Total, dando origem à fórmula

$$\mathbb{P}(B_j|A) = \frac{\mathbb{P}(B_j)\mathbb{P}(A|B_j)}{\sum_k \mathbb{P}(B_k)\mathbb{P}(A|B_k)}.$$

Exemplo 2.12. No Exemplo 2.10, sabendo-se que uma meia azul foi retirada, qual a probabilidade de ter sido aberta a gaveta A? Pela Fórmula de Bayes,

$$\mathbb{P}(A|\text{azul}) = \frac{\mathbb{P}(A)\mathbb{P}(\text{azul}|A)}{\mathbb{P}(A)\mathbb{P}(\text{azul}|A) + \mathbb{P}(B)\mathbb{P}(\text{azul}|B)} = \frac{\frac{1}{5}}{\frac{1}{2}\,\frac{2}{5} + \frac{1}{2}\,\frac{3}{6}} = \frac{4}{9}. \qquad \triangle$$

2.2 Independência

Dois eventos aleatórios são independentes quando a ocorrência de um deles não aumenta nem diminui a chance relativa de que ocorra o outro.

Definição 2.13 (Eventos independentes)**.** Os eventos aleatórios A e B são ditos *independentes* se

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

Neste caso, dizemos que A é independente de B.

Dois eventos A e B são independentes se, e somente se, $\mathbb{P}(A|B) = \mathbb{P}(A)$. Deixamos a demonstração como exercício (separe em casos!).

Exemplo 2.14. Uma moeda é lançada duas vezes. Sejam $A =$ "a primeira moeda sai cara" e $B =$ "a segunda moeda sai cara". Então A e B são independentes, pois

$$\begin{aligned}
\mathbb{P}(A) &= \mathbb{P}(\{1\} \times \{0,1\}) = \frac{2}{4} = \frac{1}{2} \\
\mathbb{P}(B) &= \mathbb{P}(\{0,1\} \times \{1\}) = \frac{2}{4} = \frac{1}{2} \\
\mathbb{P}(A \cap B) &= \mathbb{P}(\{1\} \times \{1\}) = \frac{1}{4} = \mathbb{P}(A)\mathbb{P}(B).
\end{aligned}$$

△

Veja que no exemplo acima, já sabíamos de antemão que os eventos deveriam ser independentes, pois cada lançamento da moeda não tem interferência alguma sobre o outro. Entretanto, independência não significa necessariamente que os eventos não possuam nenhuma relação entre si.

Exemplo 2.15. Dois dados são lançados. Consideramos os eventos $A =$ "o primeiro dado é par" e $C =$ "a soma dos valores dos dados é par". Então os eventos A e C são independentes, pois

$$\begin{aligned}
\mathbb{P}(A) &= \mathbb{P}(\{2,4,6\} \times \{1,2,3,4,5,6\}) = \frac{18}{36} = \frac{1}{2}, \\
\mathbb{P}(C) &= \mathbb{P}(\{2,4,6\}^2 \cup \{1,3,5\}^2) = \frac{18}{36} = \frac{1}{2}, \\
\mathbb{P}(A \cap C) &= \mathbb{P}(\{2,4,6\}^2) = \frac{9}{36} = \frac{1}{4} = \mathbb{P}(A)\mathbb{P}(C).
\end{aligned}$$

△

Proposição 2.16. *Um evento A é independente de si próprio se, e somente se, $\mathbb{P}(A) = 0$ ou 1.*

Novamente, a demonstração fica como exercício.

Definição 2.17 (Eventos independentes dois a dois)**.** Os eventos aleatórios $(A_j)_{j \in J}$, onde J é um conjunto qualquer de índices, são ditos *independentes dois a dois* se A_k e A_j são independentes para todos $k, j \in J$ com $k \neq j$.

Exemplo 2.18. Dois dados são lançados. Consideramos os eventos $A =$ "o primeiro dado é par", $B =$ "o segundo dado é par" $C =$ "a soma dos valores

dos dados é par". Então

$$\begin{aligned}
\mathbb{P}(A) &= \mathbb{P}(\{2,4,6\} \times \{1,2,3,4,5,6\}) = \frac{18}{36} = \frac{1}{2}, \\
\mathbb{P}(B) &= \mathbb{P}(\{1,2,3,4,5,6\} \times \{2,4,6\}) = \frac{18}{36} = \frac{1}{2}, \\
\mathbb{P}(C) &= \mathbb{P}(\{2,4,6\}^2 \cup \{1,3,5\}^2) = \frac{18}{36} = \frac{1}{2}, \\
\mathbb{P}(A \cap B) &= \mathbb{P}(\{2,4,6\}^2) = \frac{9}{36} = \frac{1}{4} = \mathbb{P}(A)\mathbb{P}(B), \\
\mathbb{P}(A \cap C) &= \mathbb{P}(\{2,4,6\}^2) = \frac{9}{36} = \frac{1}{4} = \mathbb{P}(A)\mathbb{P}(C), \\
\mathbb{P}(B \cap C) &= \mathbb{P}(\{2,4,6\}^2) = \frac{9}{36} = \frac{1}{4} = \mathbb{P}(B)\mathbb{P}(C).
\end{aligned} \qquad \triangle$$

Exemplo 2.19. Lançamento de um dado de 4 faces. Considere $A =$ "par", $B =$ "menor que 3", $C =$ "1 ou 4", ou seja, $A = \{2,4\}$, $B = \{1,2\}$, $C = \{1,4\}$. Então A, B e C são independentes dois a dois. Com efeito,

$$\begin{aligned}
\mathbb{P}(A \cap B) &= \mathbb{P}(\{2\}) = \frac{1}{4} = \mathbb{P}(A)\mathbb{P}(B), \\
\mathbb{P}(A \cap C) &= \mathbb{P}(\{4\}) = \frac{1}{4} = \mathbb{P}(A)\mathbb{P}(C), \\
\mathbb{P}(B \cap C) &= \mathbb{P}(\{1\}) = \frac{1}{4} = \mathbb{P}(B)\mathbb{P}(C).
\end{aligned} \qquad \triangle$$

Definição 2.20 (Eventos independentes)**.** Os eventos aleatórios $(A_j)_{j \in J}$ são ditos *independentes* se, dados quaisquer $n \in \mathbb{N}$ e $j_1, j_2, \ldots, j_n \in J$ distintos, vale

$$\mathbb{P}(A_{j_1} \cap A_{j_2} \cap \cdots \cap A_{j_n}) = \mathbb{P}(A_{j_1})\mathbb{P}(A_{j_2}) \cdots \mathbb{P}(A_{j_n}).$$

Exemplo 2.21. No lançamento de um dado dodecaédrico (12 faces), considere os eventos $A =$ "múltiplo de 3", $B =$ "menor ou igual a 6" e $C =$ "par", isto é, $A = \{3,6,9,12\}$, $B = \{1,2,3,4,5,6\}$ e $C = \{2,4,6,8,10,12\}$. Então A, B e C são independentes, pois

$$\mathbb{P}(A \cap B) = \mathbb{P}(\{3,6\}) = \frac{1}{6} = \mathbb{P}(A)\mathbb{P}(B),$$

$$\begin{aligned}
\mathbb{P}(B \cap C) = \mathbb{P}(\{2,4,6\}) &= \frac{1}{4} = \mathbb{P}(B)\mathbb{P}(C),\\
\mathbb{P}(A \cap C) = \mathbb{P}(\{6,12\}) &= \frac{1}{6} = \mathbb{P}(A)\mathbb{P}(C),\\
\mathbb{P}(A \cap B \cap C) = \mathbb{P}(\{6\}) &= \frac{1}{12} = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C).
\end{aligned} \qquad \triangle$$

Tomando $n = 2$ na Definição 2.20, podemos observar que toda família de eventos independentes é também 2 a 2 independente. Todavia, a recíproca é falsa.

Contra-exemplo 2.22. No Exemplo 2.19, os eventos A, B e C *não são* independentes. Com efeito,

$$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(\emptyset) = 0 \neq \frac{1}{8} = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C). \qquad \triangle$$

Contra-exemplo 2.23. No Exemplo 2.18, os eventos A, B e C *não são* independentes. Com efeito,

$$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(\{2,4,6\}^2) = \frac{1}{4} \neq \frac{1}{8} = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C). \qquad \triangle$$

2.3 Exercícios

§2.1

1. São dadas duas urnas, A e B. A urna A contém 1 bola vermelha e 1 azul. A urna B contém 2 bolas vermelhas e 3 azuis. Uma bola é extraída ao acaso de A e colocada em B. Uma bola então é extraída ao acaso de B. Qual a probabilidade de se retirar uma bola vermelha de B? Qual a probabilidade de ambas as bolas retiradas serem da mesma cor?

2. Paula fez uma prova na semana passada. A probabilidade de aprovação na prova é de $\frac{4}{5}$ caso ela tenha se preparado adequadamente, $\frac{1}{2}$ caso ela tenha estudado apenas na véspera, e $\frac{1}{10}$ caso ela não tenha estudado nada. Sabe-se que as probabilidades de ela ter estudado adequadamente, ter estudado

apenas na véspera e não ter estudado nada são $\frac{3}{5}$, $\frac{1}{5}$ e $\frac{1}{5}$, respectivamente. Qual a probabilidade de Paula ter estudado adequadamente dado que ela foi aprovada?

3. Seja p a proporção de indivíduos na população que estão infectados por um determinado vírus. O exame laboratorial que detecta esse vírus apresenta um resultado *falso negativo* (indica que um indivíduo infectado está saudável) com probabilidade $\frac{1}{1.000}$; enquanto resultados do tipo *falso positivo* (indica que um indivíduo saudável está infectado) ocorrem com probabilidade $\frac{1}{50}$.

(a) Um indivíduo foi escolhido ao acaso e o resultado do exame foi positivo. Calcule $f(p)$, a probabilidade de esse indivíduo estar infectado.

(b) Quanto valem $\lim_{p\to 0^+} f(p)$ e o menor valor de p para o qual $f(p) \geqslant \frac{1}{2}$. Interprete estes resultados.

Nesse problema, o parâmetro p é chamado *prevalência*, $\frac{999}{1.000}$ é a chamada *sensibilidade* e $\frac{49}{50}$ é a chamada *especificidade*.

4. Considere uma questão de múltipla escolha com n opções, e seja p a probabilidade de determinado aluno saber resolver a questão. Se ele sabe resolver, acerta com certeza, e se não sabe, acerta com probabilidade $\frac{1}{n}$. Dado que ele acertou a resposta, qual a probabilidade de que ele sabia resolver? Calcule os limites dessa probabilidade quando $n \to \infty$ e quando $p \to 0^+$.

5. Sejam A, B e C eventos quaisquer. Mostre que

$$\mathbb{P}(A|B) = \mathbb{P}(A|B \cap C)\mathbb{P}(C|B) + \mathbb{P}(A|B \cap C^c)\mathbb{P}(C^c|B).$$

6. Zacarias é um sujeito bastante atrapalhado, ele perdeu sua carteira em uma de suas n gavetas, cada uma delas em diferentes níveis de desordem. Seja p_k a probabilidade de a carteira estar na k-ésima gaveta. Seja q_k a probabilidade condicional de ele encontrar a carteira na k-ésima gaveta dado que ela lá está e a gaveta foi examinada. Calcule a probabilidade da carteira estar na primeira gaveta, dado que Zacarias abriu apenas a k-ésima gaveta e não encontrou sua carteira.

7. Sandro tem em seus bolsos 5 dados, cada um deles com a forma de um dos Poliedros de Platão e com as faces numeradas de um até o número de faces de cada dado. No bolso esquerdo estão o tetraedro, o cubo e o octaedro, enquanto dodecaedro e icosaedro estão no bolso direito. Sandro deve retirar aleatoriamente uma carta de um baralho comum, se a carta for de copas ele escolhe aleatoriamente um dos dados do bolso esquerdo, caso contrário, ele sorteia o dado do bolso direito. Sandro lançará o dado e ganhará um prêmio se a face sorteada for o número 1. Calcule:

(a) A probabilidade de sair o número 1.
(b) A probabilidade condicional de o dado retirado ter sido o cubo dado que saiu o número 1.

8 (Urna de Pólya)**.** Uma urna contém inicialmente p bolas pretas e b bolas brancas. É sorteada uma bola da urna, sua cor é observada, em seguida esta bola é devolvida à urna e também outra bola da mesma cor é adicionada à urna. Repetimos este procedimento indefinidamente. Seja $B_{n,k}$ o evento em que são retiradas exatamente k bolas brancas nas n primeiras extrações. Mostre que

$$\mathbb{P}(B_{n,k}) = \frac{\binom{b+k-1}{b-1}\binom{p+n-k-1}{p-1}}{\binom{p+b+n-1}{n}}.$$

Dica: Mostre por indução em n, para todo p e b.

9. Um casal tem dois filhos que não são gêmeos. Calcule a probabilidade condicional de esse casal ter dois filhos homens, sabendo-se que:

(a) O casal tem um filho homem.
(b) O filho mais velho do casal é homem.
(c) O casal tem um filho homem que nasceu num sábado.
(d) O casal tem um filho homem que não nasceu num sábado.

Respostas aproximadas: 0,33, 0,50, 0,48, 0,36. Comente o porquê de o resultado do item (d) ser próximo ao do item (a) e o do item (c) ser próximo ao do item (b).

10. Sejam $B_1, B_2, \ldots$ eventos disjuntos, A e C eventos aleatórios, e $\varepsilon \in [0, 1]$.

Mostre que:

(a) Se $\mathbb{P}(A|B_n) \geqslant \varepsilon$ para todo n, então $\mathbb{P}(A| \cup_n B_n) \geqslant \varepsilon$.
(b) Se $\mathbb{P}(A|B_n) = \varepsilon$ para todo n, então $\mathbb{P}(A| \cup_n B_n) = \varepsilon$.
(c) Se $\mathbb{P}(A|B_n) = \mathbb{P}(C|B_n)$ para todo n, então $\mathbb{P}(A| \cup_n B_n) = \mathbb{P}(C| \cup_n B_n)$.

11. Mostre que, se $A_n \downarrow A$ e $\mathbb{P}(A_{n+1}|A_n) \leqslant \frac{2}{3}$ para todo n então $\mathbb{P}(A) = 0$.

12. Leonardo prometeu que iria enviar um e-mail para Bernardo até o final do dia. A probabilidade de ele realmente enviar o e-mail conforme prometido é de $\frac{8}{9}$. Cada e-mail enviado tem probabilidade $\frac{3}{4}$ se ser entregue, e probabilidade $\frac{1}{4}$ de ser perdido pelo caminho ou classificado como spam. Dado que Bernardo não recebeu o tal e-mail, calcule a probabilidade de que Leonardo não o tenha enviado.

§2.2

13. Prove que dois eventos A e B são independentes se, e somente se, $\mathbb{P}(A|B) = \mathbb{P}(A)$.

14. Prove que, se A e B são independentes, então B é independente de A^c, e A e A^c são ambos independentes de B^c.

15. Sejam A e B eventos tais que $\mathbb{P}(A|B) = \frac{1}{4}$ e $\mathbb{P}(B|A) = \frac{1}{2}$.

(a) A e B podem ser independentes?
(b) A e B são necessariamente independentes?
(c) A e B podem ser disjuntos?

16. Prove ou dê contra-exemplo:
Para todo espaço de probabilidade $(\Omega, \mathcal{F}, \mathbb{P})$ e todos $A, B, C \in \mathcal{F}$, se $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$, então A, B e C são independentes.

17. Sejam $A_1, \ldots, A_n$ eventos independentes. Sejam $B_1, \ldots, B_n$ eventos tais que, para todo $j = 1, \ldots, n$, ou $B_j = A_j$ ou $B_j = A_j^c$.

(a) Mostre que $A_1, \ldots, A_{n-1}, B_n$ são independentes.
(b) Mostre que $B_1, \ldots, B_{n-1}, B_n$ são independentes.

18. Sejam $A_1, A_2, \ldots, A_n$ eventos independentes e suponha que $\#\Omega < 2^n$. Mostre que $\mathbb{P}(A_k) = 0$ ou 1 para algum $k = 1, \ldots, n$.

19. Em uma gaveta existem 2 maços de baralho fechados. Um deles é um baralho comum de 52 cartas, $\{A, 2, 3, \ldots, 9, 10, J, Q, K\} \times \{\clubsuit, \heartsuit, \spadesuit, \diamondsuit\}$, e outro é um baralho de truco com 40 cartas (não possui as cartas de números '8', '9' e '10').

Um dos maços é retirado da gaveta ao acaso e depois uma carta é sorteada ao acaso do baralho retirado.

(a) Calcule a probabilidade de a carta sorteada ser uma das três figuras reais (J, Q ou K).
(b) Sabendo-se que foi sorteada uma figura real, calcule a probabilidade de o baralho retirado ter sido o baralho comum.
(c) Calcule a probabilidade de a carta sorteada ser de espadas.
(d) Sabendo-se que foi sorteada uma carta de espadas, calcule a probabilidade de o baralho retirado ter sido o baralho de truco.
(e) Sejam $A =$ "Foi retirado o baralho comum", $B =$ "Foi sorteada uma figura real" e $C =$ "Foi sorteada uma carta de espadas". A e B são independentes? A e C são independentes? A, B e C são independentes?
(f) Qual a probabilidade de se sortear uma carta de número '5'?
(g) Sabendo-se que foi sorteado um número (isto é, não foi sorteado A, J, Q nem K), qual a probabilidade de o baralho retirado ter sido o baralho de truco?

20. Dois dados comuns são lançados. Para $n = 1, \ldots, 6$, seja A_n o evento de sair n no primeiro dado. Para $k = 2, \ldots, 12$, seja B_k o evento de a soma dos dois dados ser igual a k. Mostre que A_n e B_k são independentes se, e somente se, $k = 7$.

21. Sejam A e $(B_n)_n$ eventos tais que A e B_n são independentes para todo n. Mostre que, se os eventos $(B_n)_n$ são disjuntos, então A e $\cup_n B_n$ são independentes. Exiba um exemplo ilustrando que a hipótese de que os $(B_n)_n$ são disjuntos não pode ser omitida.

Capítulo 3

Variáveis Aleatórias

Na realização de um experimento aleatório, muitas vezes estamos interessados em uma ou mais quantidades que podem ser observadas durante a realização do experimento. Por exemplo, sortear 11 cartas do baralho e contar quantas dessas cartas são de espadas, ou sortear dois números reais entre 0 e 1 e considerar o menor deles. A essas quantidades damos o nome de *variáveis aleatórias*. Uma variável aleatória é um observável numérico resultante de um experimento.

3.1 Variáveis aleatórias

Uma quantidade numérica que pode ser observada num certo experimento aleatório é representada por uma função $X : \Omega \to \mathbb{R}$.

Exemplo 3.1. Lança-se um dado e observa-se a face superior. Neste caso, podemos tomar $\Omega = \{1, 2, 3, 4, 5, 6\}$ e $X(\omega) = \omega$. $\triangle$

Exemplo 3.2. Lançam-se dois dados e considera-se o maior dos valores. Neste caso, podemos tomar $\Omega = \{1, 2, 3, 4, 5, 6\}^2$ e $X(\omega_1, \omega_2) = \max\{\omega_1, \omega_2\}$. $\triangle$

Vamos impor uma restrição sobre a função X que permitirá atribuir

probabilidade a eventos como "o valor observado de X é no máximo 5".

Definição 3.3 (Variável Aleatória)**.** Uma *variável aleatória* X em um espaço de probabilidade $(\Omega, \mathcal{F}, \mathbb{P})$ é uma função real definida no espaço Ω tal que o conjunto $\{\omega \in \Omega : X(\omega) \leqslant x\}$ é evento aleatório para todo $x \in \mathbb{R}$, isto é, $\{\omega \in \Omega : X(\omega) \leqslant x\} \in \mathcal{F}$ para todo $x \in \mathbb{R}$.

Daqui para frente, denotaremos o evento $\{\omega \in \Omega : X(\omega) \leqslant x\}$ simplesmente por $\{X \leqslant x\}$.

Exemplo 3.4 (Variável aleatória constante)*.* Se $X(\omega) = 5$ para todo $\omega \in \Omega$, então

$$\{X \leqslant a\} = \{\omega : X(\omega) \leqslant a\} = \begin{cases} \Omega, & \text{se } a \geqslant 5, \\ \emptyset, & \text{se } a < 5. \end{cases}$$

Portanto, X é uma variável aleatória. $\triangle$

Dizemos que uma variável aleatória X é *degenerada* se existe $c \in \mathbb{R}$ tal que $\mathbb{P}(X = c) = 1$. No exemplo acima, X é degenerada com $c = 5$.

Exemplo 3.5 (Função indicadora)*.* Dado $A \subseteq \Omega$, definimos

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

Se $A \in \mathcal{F}$ e $X = \mathbb{1}_A$, então

$$\{X \leqslant a\} = \{\omega : X(\omega) \leqslant a\} = \begin{cases} \Omega, & \text{se } a \geqslant 1, \\ A^c, & \text{se } 0 \leqslant a < 1, \\ \emptyset, & \text{se } a < 0. \end{cases}$$

Portanto, X é uma variável aleatória. $\triangle$

Contra-exemplo 3.6. Sejam $\Omega = \{1, 2, 3, 4\}$ e $\mathcal{F} = \{\emptyset, \{1, 2\}, \{3, 4\}, \Omega\}$ e considere os conjuntos $A = \{1, 2\}$ e $B = \{1, 3\}$. Então $\mathbb{1}_A$ é variável aleatória em $(\Omega, \mathcal{F})$, mas $\mathbb{1}_B$ não é. $\triangle$

3.1.1 Espaço induzido e lei de uma variável aleatória

A *σ-álgebra de Borel* na reta $\mathbb{R}$, denotada por $\mathcal{B}$, é a menor σ-álgebra que contém todos os intervalos da reta (veja a Seção 1.4.1 para mais detalhes). Os conjuntos $B \subseteq \mathbb{R}$ tais que $B \in \mathcal{B}$ são chamados *borelianos*. Na prática, essa classe contém todos os subconjuntos de $\mathbb{R}$ que nos interessam.

Dado um espaço de probabilidade $(\Omega, \mathcal{F}, \mathbb{P})$ e uma variável aleatória X, definimos o *espaço de probabilidade induzido por* X como $(\mathbb{R}, \mathcal{B}, \mathbb{P}_X)$, onde

$$\mathbb{P}_X(B) = \mathbb{P}(\{\omega : X(\omega) \in B\}), \quad B \in \mathcal{B}. \tag{3.7}$$

Ou seja, o espaço amostral é o conjunto dos números reais, os eventos aleatórios são os conjuntos borelianos, e a medida de probabilidade é aquela induzida por X. A medida de probabilidade $\mathbb{P}_X$ em $\mathbb{R}$ induzida por X é chamada de *lei da variável aleatória* X ou *distribuição de* X. Na Seção 3.7, mostraremos que $\mathbb{P}_X$ está realmente definida para todo $B \in \mathcal{B}$.

A importância teórica e conceitual do espaço de probabilidade induzido por uma variável aleatória X, bem como sua distribuição $\mathbb{P}_X$, é que ele permite descrever o comportamento estatístico de X abstraindo-se todos os detalhes do espaço de probabilidade original. Mais precisamente, toda pergunta formulada apenas em termos de X pode ser respondida com $\mathbb{P}_X$ ao invés de $\mathbb{P}$.

Exemplo 3.8. Um dado é lançado três vezes. Seja X o valor obtido no primeiro lançamento. Esse experimento pode ser modelado por $\Omega = \{1, 2, 3, 4, 5, 6\}^3$, $\mathcal{F} = \mathcal{P}(\Omega)$ e $\mathbb{P}(A) = \frac{\#A}{216}$ para todo $A \in \mathcal{F}$, nesse caso $X : \Omega \to \mathbb{R}$ é dado por $X(\omega) = \omega_1$, onde cada $\omega \in \Omega$ é identificado como uma tripla $(\omega_1, \omega_2, \omega_3)$. O espaço induzido por X é dado por $(\mathbb{R}, \mathcal{B}, \mathbb{P}_X)$, com $\mathbb{P}_X$ dado por

$$\mathbb{P}_X(B) = \frac{\#(B \cap \{1, 2, 3, 4, 5, 6\})}{6}, \quad B \in \mathcal{B}.$$

Para calcular $\mathbb{P}(1,5 \leqslant X \leqslant 3,4)$, podemos fazer

$$\mathbb{P}(\{\omega : 1{,}5 \leqslant X(\omega) \leqslant 3{,}4\}) = \frac{\#(\{2,3\} \times \{1,2,3,4,5,6\}^2)}{216} = \frac{72}{216} = \frac{1}{3}$$

ou

$$\mathbb{P}_X([1{,}5; 3{,}4]) = \frac{\#\{2,3\}}{6} = \frac{1}{3}. \qquad \triangle$$

3.1.2 Função de distribuição

Assim como a ideia de lei de uma variável, o conceito que definiremos agora é uma outra maneira equivalente de descrever a estrutura probabilística da variável aleatória.

Definição 3.9 (Função de distribuição)**.** A *função de distribuição* da variável aleatória X, é a função $F_X : \mathbb{R} \to [0,1]$, definida como

$$F_X(x) = \mathbb{P}(X \leqslant x), \quad x \in \mathbb{R}.$$

A função de distribuição, também conhecida como função de distribuição acumulada, determina a lei da variável aleatória. Veremos a demostração deste fato na Seção 3.6, mas gostaríamos de enunciá-lo precisamente agora.

Teorema 3.10. *Sejam X e Y variáveis aleatórias. Se $F_X(t) = F_Y(t)$ para todo $t \in \mathbb{R}$, então $\mathbb{P}_X(B) = \mathbb{P}_Y(B)$ para todo $B \in \mathcal{B}$.*

Por isso a função de distribuição é uma característica fundamental da variável aleatória. Caso $F_X = F_Y$ (e portanto $\mathbb{P}_X = \mathbb{P}_Y$), escrevemos $X \sim Y$.

Exemplo 3.11. Duas moedas honestas são lançadas. Seja a variável X que

conta o número de caras observadas. Temos que

$$F_X(t) = \mathbb{P}(X \leqslant t) = \begin{cases} \mathbb{P}(\emptyset) = 0, & t < 0; \\ \mathbb{P}(\{(0,0)\}) = \frac{1}{4}, & 0 \leqslant t < 1; \\ \mathbb{P}(\{(0,0),(0,1),(1,0)\}) = \frac{3}{4}, & 1 \leqslant t < 2; \\ \mathbb{P}(\Omega) = 1, & t \geqslant 2. \end{cases}$$

Observe que o salto da função de distribuição corresponde à probabilidade de a variável aleatória assumir aquele valor, como se vê na Figura 3.1. △

Exemplo 3.12. Selecionamos um ponto ao acaso do intervalo $[a, b]$ com $a < b$. Seja X a variável aleatória que representa a coordenada do ponto. Primeiro observamos que, ao selecionar um ponto ao acaso em um intervalo, estamos dizendo implicitamente que quaisquer subintervalos de mesmo tamanho têm a mesma probabilidade de conter o ponto escolhido. Isso implica que $\mathbb{P}(X \in [c, d]) = \frac{d-c}{b-a}$ para todo $[c, d] \subseteq [a, b]$. Para $t \in [a, b]$, tomando $c = a$ obtemos $\mathbb{P}(X \leqslant t) = \frac{t-a}{b-a}$. Para $t < a$ temos que $\mathbb{P}(X \leqslant t) = 0$, e para $t \geqslant b$ obtemos $\mathbb{P}(X \leqslant t) = 1$. Portanto,

$$F_X(t) = \mathbb{P}(X \leqslant t) = \begin{cases} 0, & t \leqslant a; \\ \dfrac{t-a}{b-a}, & a \leqslant t \leqslant b; \\ 1, & t \geqslant b; \end{cases}$$

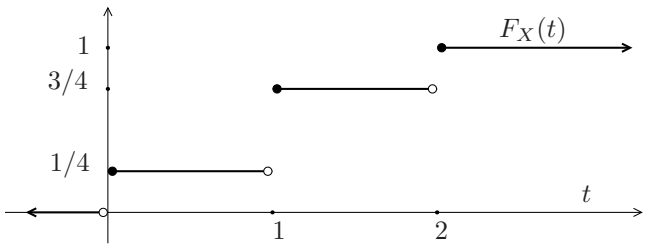


Figura 3.1. Gráfico de uma função de distribuição discreta.

cujo gráfico está ilustrado na Figura 3.2. △

Variáveis aleatórias diferentes podem ter a mesma função de distribuição. Por exemplo, se X a variável aleatória definida no Exemplo 3.1 e definimos $Y = 7 - X$, então, $X \neq Y$ enquanto $F_X(t) = F_Y(t)$ para todo $t \in \mathbb{R}$.

Dada uma sequência $(x_n)_n$ de números reais, denotamos por $x_n \uparrow x$ a propriedade de que $x_1 \leqslant x_2 \leqslant x_3 \leqslant \cdots$ e $x_n \to x$. Analogamente, escrevemos $x_n \downarrow x$ para denotar que $x_1 \geqslant x_2 \geqslant x_3 \geqslant \cdots$ e $x_n \to x$.

Proposição 3.13 (Propriedades da função de distribuição)**.** *Se X é uma variável aleatória, sua função de distribuição F_X satisfaz às seguintes propriedades:*

(1) *F_X é não-decrescente;*
(2) *F_X é contínua à direita;*
(3) $\lim_{x\to-\infty} F_X(x) = 0$ *e* $\lim_{x\to+\infty} F_X(x) = 1$.

Demonstração. Para o item (1), basta ver que $x \leqslant y$ implica $\{X \leqslant x\} \subseteq \{X \leqslant y\}$, que, pelo item (6) do Teorema 1.35, implica

$$F_X(x) = \mathbb{P}(X \leqslant x) \leqslant \mathbb{P}(X \leqslant y) = F_X(y).$$

Para (2), observe que $x_n \downarrow x$ implica $\{X \leqslant x_n\} \downarrow \{X \leqslant x\}$, que, pela

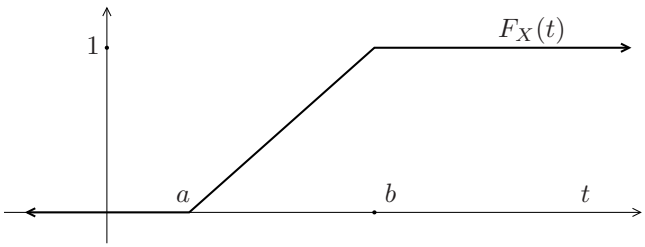


Figura 3.2. Gráfico de uma função de distribuição contínua.

continuidade da probabilidade, implica

$$F_X(x_n) = \mathbb{P}(X \leqslant x_n) \to \mathbb{P}(X \leqslant x) = F_X(x).$$

Pela Observação A.2, isso implica que $\lim_{z \to x^+} F_X(z) = F_X(x)$, ou seja, F_X é contínua à direita. Para (3), observe que $x_n \downarrow -\infty$ implica $\{X \leqslant x_n\} \downarrow \{X \leqslant -\infty\} = \emptyset$, que, pela continuidade da probabilidade, implica

$$F_X(x_n) = \mathbb{P}(X \leqslant x_n) \to \mathbb{P}(\emptyset) = 0.$$

Analogamente, $x_n \uparrow +\infty$ implica $\{X \leqslant x_n\} \uparrow \{X < +\infty\} = \Omega$, que implica $F_X(x_n) = \mathbb{P}(X \leqslant x_n) \to \mathbb{P}(\Omega) = 1$. □

Ao final desta seção, mostraremos que, dada uma função F com as três propriedades dadas pela proposição acima, sempre existe uma variável aleatória cuja função de distribuição é F. Diremos que uma dada F com essas propriedades é *uma função de distribuição*.

De forma geral, uma *função de distribuição* é qualquer função $F(\cdot)$ satisfazendo às três propriedades acima. Ao final desta seção, mostraremos que dada uma função de distribuição F, sempre existe uma variável aleatória cuja função de distribuição é F.

Segue da definição de função de distribuição que $\mathbb{P}(X > a) = 1 - F_X(a)$ e $\mathbb{P}(a < X \leqslant b) = F_X(b) - F_X(a)$, para todos $a < b \in \mathbb{R}$. A proposição abaixo nos diz como obter a distribuição em outros tipos de intervalos. Denotamos $F_X(a-) = \lim_{z \to a^-} F_X(z)$.

Proposição 3.14. *A função de distribuição de uma variável aleatória X satisfaz: $\mathbb{P}(a < X < b) = F_X(b-) - F_X(a)$, $\mathbb{P}(a \leqslant X < b) = F_X(b-) - F_X(a-)$, $\mathbb{P}(a \leqslant X \leqslant b) = F_X(b) - F_X(a-)$ e $\mathbb{P}(X = a) = F_X(a) - F_X(a-)$ para todos números $a < b$. Em particular, $\mathbb{P}(X = a) = 0$ se, e somente se, F_X é contínua em a.*

Demonstração. Provaremos apenas a primeira igualdade, as outras são

análogas. Observe que $(a, b - \frac{1}{n}] \uparrow (a, b)$, donde

$$\{a < X \leqslant b - \tfrac{1}{n}\} \uparrow \{a < X < b\}.$$

Logo, $F_X(b-) - F_X(a) = \lim_n F_X(b - \frac{1}{n}) - F_X(a) = \lim_n \mathbb{P}(a < X \leqslant b - \frac{1}{n}) = \mathbb{P}(a < X < b)$, pela continuidade da probabilidade. $\square$

3.1.3 Função quantil

Imagine que estamos implementando um algoritmo usando uma certa linguagem de programação e queremos simular uma variável aleatória X com uma dada função de distribuição F. Em geral, as linguagens de programação são capazes de gerar uma variável aleatória U distribuída "uniformemente" no intervalo $(0, 1)$. Veremos na Seção 3.3 o que quer dizer estar distribuída uniformemente em $(0, 1)$, mas por enquanto basta saber que $\mathbb{P}(U \leqslant u) = u$ para $u \in [0, 1]$. A partir de uma variável U com essa distribuição, é possível obter uma variável X com função de distribuição F?

Caso o valor de U seja, por exemplo, 0,272106341..., queremos atribuir a X o valor x que ocupa, dentre os possíveis valores assumidos por X, a "posição" correspondente a 0,272106341... numa escala entre 0 e 1. Mais precisamente, buscamos o valor x tal que $\mathbb{P}(X < x) \leqslant$ 0,272106341... e $\mathbb{P}(X > x) \leqslant$ 0,727893658.... Observe que, ao buscar tal valor de x, estamos tentando inverter a função de distribuição de X.

A *função quantil* de uma variável aleatória é a inversa da sua função de distribuição, no seguinte sentido. Dada uma função de distribuição F, definimos

$$F^{-1}(u) = \min\{x \in \mathbb{R} : F(x) \geqslant u\},$$

ou seja, $F^{-1}(u)$ é o ponto x mais à esquerda a partir do qual o gráfico de F fica acima da altura u. Caso F seja como o da Figura 3.2, $F^{-1}(u)$ nada mais é do que a pré-imagem de u. Porém, no exemplo da função F da Figura 3.1, os valores $u = \frac{1}{4}$ e $u = \frac{3}{4}$ têm todo um intervalo como pré-imagem, e qualquer

outro $0 < u < 1$ tem pré-imagem vazia. Para $u = \frac{1}{4}$ e $u = \frac{3}{4}$, tomamos o ponto mais à esquerda do intervalo e, para os demais pontos, completamos o gráfico de F com segmentos verticais para definir a inversa. A definição de F^{-1} é ilustrada na Figura 3.3.

Mencionamos aqui algumas propriedades da função quantil que podem ser observadas visualmente, e cuja prova será omitida. Para todo $x \in \mathbb{R}$, temos $F^{-1}(F(x)) \leqslant x$, valendo a igualdade se F é estritamente crescente num intervalo aberto contendo x. Para todo $0 < u < 1$, temos $F(F^{-1}(u)) \geqslant u$, valendo a igualdade se F é contínua num intervalo aberto contendo $F^{-1}(u)$.

A propriedade que vamos usar é que, dados $x \in \mathbb{R}$ e $0 < u < 1$,

$$F^{-1}(u) \leqslant x \qquad \Longleftrightarrow \qquad u \leqslant F(x). \tag{3.15}$$

Essa propriedade segue da definição de F^{-1} e do fato de F ser não-decrescente e contínua à direita.

A discussão acima culmina com a demonstração formal da existência de variáveis aleatórias com quaisquer distribuições. No mesmo espírito do parágrafo inicial acima, o ponto de partida é a existência de uma variável aleatória uniforme, o que enunciaremos agora.

Teorema 3.16 (Variável aleatória uniforme)**.** *Existe uma variável aleatória U em um espaço de probabilidade $(\Omega, \mathcal{F}, \mathbb{P})$ tal que $F_U(u) = u$ para $u \in [0, 1]$.*

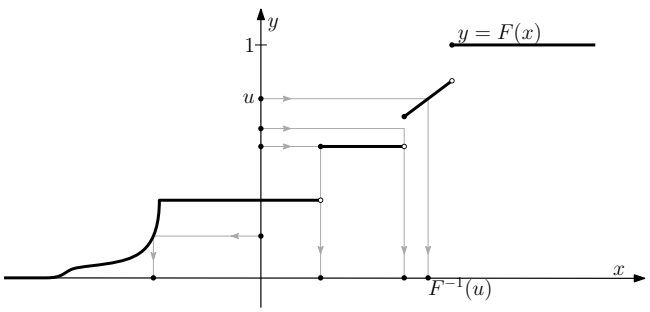


Figura 3.3. Ilustração de como é obtida a função quantil.

A demonstração exige ferramentas avançadas e será dada na Seção 3.6.

Teorema 3.17. *Se uma função F satisfaz às três propriedades listadas na Proposição 3.13, então existe um espaço de probabilidade e uma variável aleatória cuja função de distribuição é F.*

Demonstração. Seja U uma variável aleatória com distribuição uniforme em $(0,1)$ e defina $X = F^{-1}(U)$. Para cada $x \in \mathbb{R}$, por (3.15) temos

$$\mathbb{P}(X \leqslant x) = \mathbb{P}\big(F^{-1}(U) \leqslant x\big) = \mathbb{P}\big(U \leqslant F(x)\big) = F(x),$$

portanto F é a função de distribuição de X, como queríamos demonstrar. □

3.2 Variáveis aleatórias discretas

Dizemos que uma variável aleatória X, sua função de distribuição F_X e sua lei $\mathbb{P}_X$ são *discretas* se existe um conjunto enumerável $A = \{x_1, x_2, x_3, \dots\} \subseteq \mathbb{R}$ tal que $\mathbb{P}(X \in A) = 1$.

Definimos a *função de probabilidade* de uma variável aleatória X como

$$p_X(x) = \mathbb{P}(X = x).$$

O tratamento de variáveis aleatórias discretas é feito em termos de somatórios com a função de probabilidade. Por exemplo, a lei de uma variável aleatória discreta é dada por

$$\mathbb{P}_X(B) = \sum_{x \in B} p_X(x) \quad \forall B \in \mathcal{B}$$

e sua função de distribuição é dada por

$$F_X(t) = \sum_{x \leqslant t} p_X(x) \quad \forall t \in \mathbb{R}.$$

Com efeito, esta última equação é um caso particular da anterior tomando-se $B = (-\infty, t]$, e para justificar aquela usamos σ-aditividade da medida de probabilidade $\mathbb{P}$:

$$\mathbb{P}_X(B) = \mathbb{P}_X(B \cap A) = \mathbb{P}(\bigcup_{x \in B \cap A} \{X = x_k\}) = \sum_{x \in B \cap A} \mathbb{P}(X = x) = \sum_{x \in B} p_X(x).$$

A primeira e última igualdades valem pois $\mathbb{P}_X(A^c) = 0$ e $p_X(x) = 0$ para todo $x \notin A$.

A função de distribuição de uma variável aleatória discreta tipicamente se parece à da Figura 3.1, sendo constante por partes e dando um saltos de tamanho $p_X(t)$ em cada ponto $t \in \mathbb{R}$ com $p_X(t) > 0$. Como

$$p_X(t) = F_X(t) - F_X(t-),$$

a função de distribuição determina a função de probabilidade de uma variável aleatória, e pela equação anterior vale a recíproca. Portanto, para determinar o comportamento estatístico de uma variável aleatória discreta, é equivalente especificar p_X, F_X, ou $\mathbb{P}_X$. A primeira normalmente é mais simples.

De forma geral, uma *função de probabilidade* é qualquer função $p(\cdot)$ satisfazendo

$$p(x) \geqslant 0 \ \ \forall x \in \mathbb{R}, \quad \sum_{x \in \mathbb{R}} p(x) = 1.$$

Dada uma função de probabilidade $p(\cdot)$, existe um espaço de probabilidade onde está definida uma variável aleatória discreta cuja função de probabilidade é $p(\cdot)$. Por certo, basta tomar $\Omega = \mathbb{R}$, $\mathcal{F} = \mathcal{B}$ e $\mathbb{P}(A) = \sum_{\omega \in A} p(\omega)$ e $X(\omega) = \omega$.

Finalizamos esta seção com alguns exemplos conhecidos de variáveis aleatórias discretas.

Distribuição uniforme discreta

Seja $I = \{x_1, x_2, \ldots, x_k\}$ um conjunto finito, dizemos que X tem distribuição uniforme discreta (ou equiprovável) em I, quando sua função de probabilidade for constante em I, isto é

$$p_X(x_j) = \frac{1}{k}, \quad \text{para todo } j = 1, 2, \ldots, k.$$

Distribuição de Bernoulli

Dado $p \in [0, 1]$, dizemos que $X : \Omega \to \mathbb{R}$ é uma *variável de Bernoulli com parâmetro* p, o que denotamos $X \sim \text{Bernoulli}(p)$, se $p_X(1) = p$ e $p_X(0) = 1-p$.

Chamamos de *ensaios de Bernoulli* à realização sucessiva e independente de um mesmo experimento, sendo que cada realização resulta em "sucesso" com probabilidade p e "fracasso" com probabilidade $1 - p$, e dizemos que p é o parâmetro desses ensaios. O caso mais simples é quando realizamos apenas um ensaio, e tomamos X como sendo a indicadora do evento de sucesso nessa única tentativa, e nesse caso temos $X \sim \text{Bernoulli}(p)$.

Distribuição geométrica

Considere uma sequência de ensaios de Bernoulli, com o mesmo parâmetro p, e seja X o número de ensaios realizados até que se obtenha o primeiro sucesso. Se denotamos por A_n o evento de sucesso na n-ésima tentativa, podemos encontrar a função de probabilidade de X como segue:

$$p_X(n) = \mathbb{P}(X = n) = \mathbb{P}(A_1^c \cap \cdots \cap A_{n-1}^c \cap A_n) = p(1-p)^{n-1}, \quad n \in \mathbb{N}.$$

Na última igualdade, utilizamos a independência dos eventos $(A_n)_n$. Dizemos que X tem *distribuição geométrica de parâmetro* p, o que denotaremos por $X \sim \text{Geom}(p)$.

Exemplo 3.18. Lançar um par de dados até obter números iguais. Se X

denota o número de lançamentos necessários, então $X \sim \text{Geom}(\frac{1}{6})$. $\triangle$

Observação 3.19. Alguns textos definem a distribuição geométrica contando o número de falhas anteriores ao primeiro sucesso, ao invés de contar o número total de ensaios. Isso faz com que as distribuições sejam defasadas de uma unidade, pois a função de probabilidade passa a ser dada por $p(1-p)^n$ para $n \in \mathbb{N}_0$. $\triangle$

Distribuição binomial

Agora, consideremos uma repetição de n ensaios de Bernoulli com parâmetro p e seja X o número de sucessos obtidos nesses n ensaios. Para calcularmos sua função de probabilidade, observemos que se há exatamente k sucessos, então há $n-k$ fracassos. Devido à independência dos ensaios, cada uma das sequências com n sucessos e $n-k$ fracassos ocorre com probabilidade $p^k(1-p)^{n-k}$. Logo, a probabilidade de haver exatamente k sucessos é $p^k(1-p)^{n-k}$ vezes o número de sequências binárias de comprimento n onde o 1 aparece k vezes e o 0 aparece $n-k$ vezes. Isto é, a função de probabilidade de X é dada por

$$p_X(k) = \tbinom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, 2, \ldots, n.$$

Neste caso, dizemos que X tem *distribuição binomial com parâmetros n e p*, o que denotamos por $X \sim \text{Binom}(n, p)$. Pelo Teorema Binomial,

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

ou seja, p_X definida acima é de fato uma função de probabilidade.

Exemplo 3.20. Lançar um dado 4 vezes e contar o número de vezes que se obtém um número múltiplo de 3. Neste caso, $X \sim \text{Binom}(4, \frac{1}{3})$. A

probabilidade de se obter um múltiplo de 3 duas vezes é dada por

$$\mathbb{P}(X=2) = p_X(2) = \binom{4}{2}(\tfrac{1}{3})^2(\tfrac{2}{3})^{4-2} = \frac{4!}{2!(4-2)!}\frac{2^2}{3^4} = \frac{8}{27}. \qquad \triangle$$

Distribuição de Poisson

Consideremos agora a repetição de n ensaios de variáveis de Bernoulli, todas elas com o mesmo parâmetro p. Suponha que n é um número muito grande, e $p = \frac{\lambda}{n}$, isto é, a probabilidade de sucesso é inversamente proporcional ao número de repetições. Se X é a variável que conta o número de sucessos, então $X \sim \text{Binom}(n, \frac{\lambda}{n})$. O que podemos afirmar sobre a função de probabilidade de X quando $n \to \infty$? Para todo k fixo, vale

$$\begin{aligned} p_X(k) = \mathbb{P}(X=k) &= \binom{n}{k}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k} \\ &= \frac{\lambda^k}{k!}\left(\frac{n}{n}\frac{n-1}{n}\cdots\frac{n-k+1}{n}\right)\left(1-\frac{\lambda}{n}\right)^{n-k} \to \frac{e^{-\lambda}\lambda^k}{k!} \end{aligned}$$

quando $n \to \infty$. Ademais, podemos observar que $\sum_{k=0}^{\infty}\frac{e^{-\lambda}\lambda^k}{k!} = 1$, de modo que o limite acima define uma função de probabilidade.

Dizemos que uma variável aleatória Z tem *distribuição de Poisson com parâmetro* λ, o que denotaremos por $Z \sim \text{Poisson}(\lambda)$, se sua função de probabilidade é dada por

$$p_Z(k) = \frac{e^{-\lambda}\lambda^k}{k!}, \qquad \text{para todo } k = 0, 1, 2, 3, \ldots.$$

Exemplo 3.21. A padaria da esquina produz deliciosos panetones, cuja receita leva 50 uvas-passas por panetone, e a produção é feita em larga escala. Patrícia acaba de comprar um desses panetones. Qual a distribuição do número de uvas-passas do panetone adquirido por Patrícia? Se o panetone comprado por Patrícia foi produzido dentro de um lote de n panetones, temos que o número de passas de seu panetone tem distribuição $\text{Binom}(50n, \frac{1}{n})$,

pois $50n$ é o número de uvas-passas utilizadas e $\frac{1}{n}$ é a probabilidade de que cada uma dessas uvas-passas esteja justamente ne porção da massa que foi para o panetone de Patrícia. Porém, sabendo que a produção é feita em larga escala, o valor n se torna irrelevante e tal distribuição é bem aproximada por uma distribuição de Poisson(50). $\triangle$

Distribuições hipergeométrica e binomial negativa

Considere uma urna contendo N bolas, sendo K bolas pretas e $N-K$ bolas brancas. Retiramos, sem reposição, uma amostra de n bolas. Seja X a variável aleatória que denota o número de bolas pretas retiradas nas n extrações. Observe que X satisfaz $0 \leqslant X \leqslant K$ e $0 \leqslant n - X \leqslant N - K$. A função de probabilidade de X é dada por:

$$p_X(k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$$

para k tal que $0 \leqslant k \leqslant K$ e $0 \leqslant n - k \leqslant N - K$. Dizemos que uma variável aleatória com essa função de probabilidade tem *distribuição hipergeométrica de parâmetros N, K e n*, o que denotamos por $X \sim \mathrm{Hgeo}(N, K, n)$. Observe que, como as extrações são sem reposição, a independência é perdida. Caso contrário, o número de bolas pretas extraídas teria uma distribuição binomial de parâmetros n e $\frac{K}{N}$.

Exemplo 3.22. De um baralho comum extraímos, sem reposição, uma amostra de 13 cartas. Seja X a variável que denota o quantidade de cartas retiradas com um número par. Então $X \sim \mathrm{Hgeo}(52, 20, 13)$. $\triangle$

Considere uma sequência de ensaios de Bernoulli, com o mesmo parâmetro p. Seja X a variável aleatória que denota o número de ensaios realizados até que se obtenha k sucessos, onde $k \in \mathbb{N}$ é outro parâmetro do experimento. Podemos observar que X assume valores no conjunto $\{k, k+1, \dots\}$. A função

de probabilidade de X é dada por:

$$\begin{aligned} p_X(n) &= \mathbb{P}(X = n) = \mathbb{P}(B_{n-1,k-1} \cap A_n) \\ &= \tbinom{n-1}{k-1} p^k (1-p)^{n-k}, \quad n \in \{k, k+1, \dots\}. \end{aligned}$$

Neste caso, dizemos que X tem *distribuição binomial negativa de parâmetros k e p*.

Exemplo 3.23. Lançar um par de dados até obter o oitavo par de números iguais. Se X denota o número de lançamentos necessários, então X tem distribuição binomial negativa de parâmetros 8 e $\frac{1}{6}$. △

Observação 3.24. Alguns textos definem a distribuição binomial negativa contando o número de falhas anteriores ao k-ésimo sucesso, ao invés de contar o número total de ensaios. △

3.3 Variáveis aleatórias absolutamente contínuas

Seja $f_X : \mathbb{R} \to \mathbb{R}$ uma função não-negativa e contínua por partes. Dizemos que uma variável aleatória X, sua função de distribuição F_X e sua lei $\mathbb{P}_X$ são *absolutamente contínuas com densidade f_X* se

$$\mathbb{P}_X(B) = \mathbb{P}(X \in B) = \int_B f_X(x)\,\mathrm{d}x$$

para todo intervalo B. Neste caso, dizemos que f_X é *uma função de densidade de probabilidade de X*, ou simplesmente *uma função de densidade de X*.

No tratamento de variáveis aleatórias que possuem densidade, tudo pode ser feito em termos de integrais. A função de distribuição de uma variável aleatória absolutamente contínua com densidade f_X é dada por

$$F_X(t) = \int_{-\infty}^{t} f_X(s)\,\mathrm{d}s. \tag{3.25}$$

Exemplo 3.26. Sortear um número em $[0,1]$. Definimos

$$f_X(x) = \begin{cases} 1, & x \in [0,1] \\ 0, & \text{caso contrário,} \end{cases}$$

e assim

$$F_X(t) = \int_{-\infty}^{t} f_X(x)\,\mathrm{d}x = \begin{cases} 0, & t \leqslant 0, \\ t, & 0 \leqslant t \leqslant 1, \\ 1, & t \geqslant 1. \end{cases} \qquad \triangle$$

Se suspeitamos que uma variável aleatória possui densidade, podemos obter um candidato a densidade calculando

$$f_X(x) = \frac{\mathrm{d}}{\mathrm{d}x} F_X(x).$$

Depois podemos confirmar se f_X é de fato uma função densidade de X verificando (3.25).

Note que a função F_X pode não ser diferenciável em todos os pontos, como a do exemplo acima que não é diferenciável em $x = 0$ ou $x = 1$. Porém, o valor de f_X nestes dois pontos não importa pois não afetam o valor da integral. Por certo, nos cálculos referentes à variável aleatória X, a função f_X sempre aparece dentro de uma integral, e se modificamos f_X em um conjunto finito de pontos, obtemos outra função de densidade para X sem que isso cause nenhum tipo de problema.

Se $f : \mathbb{R} \to \mathbb{R}$ é uma função satisfazendo

$$f(x) \geqslant 0 \quad \forall x \in \mathbb{R}, \qquad\qquad \int_{-\infty}^{+\infty} f(x)\,\mathrm{d}x = 1,$$

diremos que f é *uma função de densidade*. Dada uma função de densidade f, sempre podemos definir uma função de distribuição F pela fórmula (3.25), logo, pelo Teorema 3.17, existe uma variável aleatória definida em algum espaço de probabilidade e que tem f como sua função de densidade.

A seguir descreveremos alguns dos exemplos mais importantes de variáveis aleatórias absolutamente contínuas. Obviamente há outras, não descritas aqui e que também são muito importantes, sobretudo em Estatística.

Distribuição uniforme Dizemos que a variável aleatória X tem distribuição uniforme no intervalo $[a, b]$, denotado por $X \sim \mathcal{U}[a, b]$, se

$$f_X(x) = \frac{1}{b-a}\,\mathbb{1}_{[a,b]}(x) = \begin{cases} \frac{1}{b-a}, & x \in [a, b], \\ 0, & x \notin [a, b]. \end{cases}$$

A distribuição uniforme é a distribuição absolutamente contínua mais simples. Segundo esta distribuição, a probabilidade de X estar em um dado subintervalo de $[a, b]$ depende apenas do comprimento desse subintervalo.

A distribuição uniforme pode ser pensada como o limite de uma distribuição uniforme discreta em $\{a, a+\frac{b-a}{n}, a+2\frac{b-a}{n}, \ldots, a+(n-2)\frac{b-a}{n}, a+(n-1)\frac{b-a}{n}, b\}$, quando n é muito grande.

Exemplo 3.27. O ponto de ruptura X de algum cabo numa rede elétrica de 5 km pode ser modelado por uma variável aleatória com distribuição uniforme em $[0, 5]$. Neste caso, $f_X = \frac{1}{5}\mathbb{1}_{[0,5]}$. A probabilidade de um determinado cabo se romper nos primeiros 800m da rede é igual a $\int_0^{0,8} \frac{1}{5}\mathrm{d}x = \frac{4}{25}$. $\triangle$

Distribuição exponencial Dizemos que X tem distribuição exponencial com parâmetro $\lambda > 0$, denotado por $X \sim \mathrm{Exp}(\lambda)$, se sua função de distribuição for dada por

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

A distribuição exponencial se caracteriza pela propriedade conhecida por perda de memória,

$$\mathbb{P}(X > t + s | X > s) = \frac{\mathbb{P}(X > t + s)}{\mathbb{P}(X > s)} = \frac{e^{-\lambda(t+s)}}{e^{-\lambda s}} = e^{-\lambda t} = \mathbb{P}(X > t)$$

para todos $t, s > 0$, ou seja, a variável X esquece aquilo que já "viveu".

Essa é a distribuição adequada para modelar o decaimento radioativo de átomos instáveis, bem como a vida útil de inúmeros materiais e objetos inanimados, como lâmpadas incandescentes ou óleos isolantes. Esses objetos deixam de funcionar não por envelhecimento ou deterioração ao longo do tempo, mas sim por um evento raro que ocorre independentemente da idade acumulada. A distribuição exponencial pode ser pensada como o limite de distribuições geométricas com pequenos intervalos de tempo. Isto é, se $X \sim \frac{1}{n}\,\mathrm{Geom}(\frac{\lambda}{n})$ com n muito grande, então a distribuição de X se aproxima da distribuição exponencial com parâmetro λ.

Distribuição normal Dizemos que a variável aleatória X tem distribuição normal com parâmetros $\mu \in \mathbb{R}$ e $\sigma^2 > 0$, denotado por $X \sim \mathcal{N}(\mu, \sigma^2)$, se X tem como densidade

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R}.$$

O parâmetro μ é o valor central de X, em torno do qual a distribuição é simétrica, e σ é relacionado à dispersão de X em torno de sua média, dado pela distância entre o centro e os pontos de inflexão. Nos Capítulos 5 e 6 veremos que μ e σ são a "média" e o "desvio-padrão" de X. Quando $\mu = 0$ e $\sigma^2 = 1$, a distribuição $\mathcal{N} = \mathcal{N}(0, 1)$ é chamada *normal padrão*.

Exercício 3.28. Mostre que se $X \sim \mathcal{N}(\mu, \sigma^2)$ então a variável aleatória $\frac{X-\mu}{\sigma}$ tem distribuição normal padrão. △

Sendo assim, podemos determinar a função de distribuição de uma variável normal qualquer se conhecermos a função de distribuição de uma normal

padrão.

Como veremos ao longo deste livro, a distribuição normal tem um papel fundamental em probabilidade, de modo que com muita frequência precisamos determinar sua função de distribuição, ou seja determinar o valor da integral

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{x^2}{2}} \mathrm{d}x.$$

A integral acima não pode ser expressa como composição de funções elementares. Sendo assim, devemos calcular a integral acima por outros métodos, sem apelar ao Teorema Fundamental do Cálculo. Valores aproximados para a função $\Phi(z)$ podem ser encontrados na tabela na página 484.

Exemplo 3.29. Seja $X \sim \mathcal{N}(2,5)$, podemos determinar $F_X(3)$ observando que $\frac{X-2}{\sqrt{5}}$ tem distribuição normal padrão, logo

$$F_X(3) = \mathbb{P}(X \leqslant 3) = \mathbb{P}(\tfrac{X-2}{\sqrt{5}} \leqslant \tfrac{3-2}{\sqrt{5}}) \approx \Phi(0{,}447) \approx 0{,}67. \qquad \triangle$$

Apesar de não ser possível expressar a integral indefinida de $e^{-\frac{x^2}{2}}$ como uma função elementar, podemos mostrar que a função f_X é de fato uma densidade. Supondo que $X \sim \mathcal{N}$ é uma normal padrão,

$$\begin{aligned}\left(\int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} \mathrm{d}x\right)^2 &= \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} \mathrm{d}x \int_{-\infty}^{+\infty} e^{-\frac{y^2}{2}} \mathrm{d}y = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} e^{-\frac{x^2+y^2}{2}} \mathrm{d}x\mathrm{d}y \\ &= \int_0^{2\pi}\int_0^{+\infty} r e^{-\frac{r^2}{2}} \mathrm{d}r\mathrm{d}\theta = 2\pi\left[-e^{-\frac{r^2}{2}}\right]_0^{\infty} = 2\pi,\end{aligned}$$

onde realizamos uma mudança de coordenadas cartesianas para polares. Para o caso $X \sim \mathcal{N}(\mu, \sigma^2)$, basta fazer a mudança de variáveis $x - \mu = \sigma u$, para reduzirmos ao caso da normal padrão.

Outras distribuições absolutamente contínuas Veremos agora outras distribuições absolutamente contínuas de grande importância em aplicações,

e que aparecerão em alguns exemplos e exercícios ao longo do livro.

Dizemos que X tem *distribuição Gama com parâmetros* $\alpha > 0$ *e* $\beta > 0$, notação $X \sim \text{Gama}(\alpha, \beta)$, se admite função densidade dada por

$$f_X(x) = c\,\beta^\alpha x^{\alpha-1} e^{-\beta x}\,\mathbb{1}_{(0,+\infty)}(x)$$

onde $c = c(\alpha) = 1/\int_0^{+\infty} x^{\alpha-1} e^{-x} \mathrm{d}x$.

Dizemos que X tem *distribuição de Cauchy com parâmetros* $a \in \mathbb{R}$ *e* $b > 0$, notação $X \sim \text{Cauchy}(a, b)$, se X possui densidade dada por

$$f_X(x) = \frac{b}{\pi[b^2 + (x-a)^2]}.$$

Dizemos que X tem *distribuição de Laplace com parâmetros* $a \in \mathbb{R}$ *e* $b > 0$, notação $X \sim \text{Laplace}(a, b)$, se X possui densidade dada por

$$f_X(x) = \frac{1}{2b} e^{-\frac{|x-a|}{b}}.$$

Dizemos que uma X tem *distribuição de Gumbel* se sua função de distribuição é dada por

$$F_X(x) = e^{-e^{-x}}.$$

Dizemos que X tem *distribuição Beta com parâmetros* $a, b > 0$, notação $X \sim \text{Beta}(a, b)$, se X possui densidade dada por

$$f_X(x) = c\,x^{a-1}(1-x)^{b-1}\,\mathbb{1}_{(0,1)}(x),$$

onde $c = c(a, b) = 1/\int_0^1 x^{a-1}(1-x)^{b-1} \mathrm{d}x$.

3.4 Distribuição condicional dado um evento

Esta seção será necessária para o estudo da esperança condicional a ser desenvolvido no Capítulo 11 e, exceto por isso, pode ser omitida. Dado um evento A com $\mathbb{P}(A) > 0$, definimos a *função de distribuição condicional de X dado A* como

$$F_{X|A}(t) = \mathbb{P}(X \leqslant t \,|\, A), \quad t \in \mathbb{R}.$$

Exemplo 3.30. Considere dois lançamentos de uma moeda honesta e seja X o número de "caras" obtidas. Neste caso,

$$F_X(t) = \begin{cases} 0, & t < 0, \\ \frac{1}{4}, & 0 \leqslant t < 1, \\ \frac{3}{4}, & 1 \leqslant t < 2, \\ 1, & t \geqslant 2. \end{cases}$$

Seja A o evento "pelo menos uma moeda deu cara". Temos

$$F_{X|A}(t) = \begin{cases} 0, & t < 1, \\ \frac{2}{3}, & 1 \leqslant t < 2, \\ 1, & t \geqslant 2. \end{cases} \qquad \triangle$$

Exemplo 3.31. Se X é uma variável aleatória com distribuição $X \sim \mathcal{U}[0,1]$ e $A = \{X \leqslant \frac{1}{2}\}$, então a função de distribuição condicional de X dado A é

$$F_{X|A}(x) = \mathbb{P}(X \leqslant x|A) = \begin{cases} 0, & x \leqslant 0, \\ 2x, & 0 \leqslant x \leqslant \frac{1}{2}, \\ 1, & x \geqslant \frac{1}{2}. \end{cases} \qquad \triangle$$

Se X é discreta, definimos a *função de probabilidade condicional de X dado A*, denotada por $p_{X|A}(\,\cdot\,)$, como a função de probabilidade associada à função de distribuição $F_{X|A}(\,\cdot\,)$.

Exemplo 3.32. No Exemplo 3.30,

$$p_{X|A}(x) = \begin{cases} \frac{2}{3}, & x = 1, \\ \frac{1}{3}, & x = 2, \\ 0, & \text{caso contrário.} \end{cases} \qquad \triangle$$

Se X tem densidade f_X, podemos tomar

$$f_{X|A}(x) = \frac{\mathrm{d}}{\mathrm{d}x} F_{X|A}(x)$$

e, caso $f_{X|A}$ seja integrável e satisfaça

$$F_{X|A}(x) = \int_{-\infty}^{x} f_{X|A}(s)\,\mathrm{d}s \qquad \text{para todo } x \in \mathbb{R},$$

dizemos que $f_{X|A}$ é uma *função de densidade condicional de X dado A*. Uma função de densidade condicional sempre existe, porém a prova deste fato exige ferramentas mais avançadas (ver Seção 11.5), e a função $f_{X|A}$ pode não ser Riemann-integrável (ver Seção 5.5.4).

Exemplo 3.33. No Exemplo 3.31, uma densidade condicional é dada por

$$f_{X|A}(x) = \frac{\mathrm{d}}{\mathrm{d}x} F_{X|A}(x) = 2\,\mathbb{1}_{[0,\frac{1}{2}]},$$

cuja integral de fato coincide com $F_{X|A}$. $\triangle$

3.5 Distribuições mistas e singulares

Uma variável aleatória discreta X vive em um conjunto enumerável de pontos, cada um dos quais tem probabilidade de ocorrência positiva. Nesse contexto, tudo se expressa em termos de somatórios ponderados pela função p_X.

Uma variável aleatória X com densidade f_X vive em $\mathbb{R}$, sua distribuição em cada intervalo $(x, x + \Delta x]$ é similar à de uma distribuição uniforme

em $(x, x+\Delta x]$, exceto que seu peso é ponderado por um fator $f_X(x)\cdot\Delta x$. Nesse contexto, tudo se expressa em termos de integrais do elemento infinitesimal de probabilidade dado por $f_X(x)\,\mathrm{d}x$.

Existem variáveis aleatórias que são misturas dos tipos discreto e absolutamente contínuo. Mais precisamente, dizemos que X é uma *variável aleatória mista com componentes discreta com função de probabilidade* p_X *e absolutamente contínua com densidade* f_X se

$$\mathbb{P}(a < X \leqslant b) = \sum_{a<x\leqslant b} p_X(x) + \int_a^b f_X(x)\,\mathrm{d}x, \quad a < b.$$

Neste caso, suas propriedades serão determinadas por combinações de somatórios e integrais. Formalmente, se X é absolutamente contínua com densidade f_X, podemos dizer que X é mista tomando $p_X = 0$; analogamente se X é discreta com função p_X, X é um caso particular de variável mista com $f_X = 0$.

Exemplo 3.34. Sejam $X \sim \mathrm{Exp}(\lambda)$ e $Y = \min(X, 1)$. Neste caso, Y é uma variável aleatória mista com função de distribuição dada por

$$F_Y(y) = \begin{cases} 0, & y < 0, \\ 1 - e^{-\lambda y}, & 0 \leqslant y < 1, \\ 1. & 1 \leqslant y. \end{cases}$$

Logo, as componentes discreta e absolutamente contínua de Y são as funções

$$p_Y(y) = \begin{cases} e^{-\lambda}, & y = 1, \\ 0, & y \neq 1, \end{cases} \qquad f_Y(y) = \begin{cases} \lambda e^{-\lambda y}, & y \in (0,1), \\ 0, & y \notin (0,1). \end{cases} \qquad \triangle$$

Há uma outra categoria bastante curiosa de variáveis aleatórias. São variáveis aleatórias cuja função de distribuição é contínua, logo não há uma componente discreta, mas tampouco possuem função de densidade, ou seja não há componente absolutamente contínua. Tais variáveis são ditas *singulares*.

Exemplo 3.35 (Distribuição de Cantor). Começamos definindo uma função $\Psi : (0,1] \to (0,1]$ com propriedades diabólicas (que será uma "inversa" da chamada "função de Cantor"). Dado $x \in (0,1]$, escreva $x = (0{,}x_1x_2x_3x_4\dots)_2$ na base 2, ou seja, cada dígito x_j é um número em $\{0,1\}$. Caso x tenha mais de uma expansão binária, escolhemos a que termina com infinitos dígitos "1" ao invés da expansão finita. O número $y = \Psi(x)$ é definido tomando-se $y_j = 2x_j \in \{0,2\}$, e escrevendo-se $y = (0{,}y_1y_2y_3y_4\dots)_3$ na base 3.

Afirmamos que Ψ é injetiva. Com efeito, suponha que $0 < x < x' \leqslant 1$. Então existe k tal que $x_k = 0$, $x'_k = 1$, e $x_j = x'_j$ para todo $j < k$. Ademais, existe $l > k$ tal que $x'_l = 1$, pois não temos expansões finitas. Vamos mostrar que $y < y'$, onde $y = \Psi(x)$ e $y' = \Psi(x')$. Como justificado acima, $y_j = y'_j$ para $j < k$, $y_k = 0$, $y'_k = 2$, e $y'_l = 2$. Daí segue que $y' > y$, provando a injetividade. Afirmamos também que Ψ não toma valores em $(\frac{1}{3}, \frac{2}{3}]$. Com efeito, os números $y \in (\frac{1}{3}, \frac{2}{3}]$ são exatamente os números com primeiro dígito $y_1 = 1$ (note que $\frac{1}{3} = (0.0222\dots)_3$ e $\frac{2}{3} = (0.1222\dots)_3$). Logo, a imagem de Ψ está contida em um conjunto de comprimento $\frac{2}{3}$. Da mesma forma, Ψ não toma valores em $(\frac{1}{9}, \frac{2}{9}]$ ou $(\frac{7}{9}, \frac{8}{9}]$, pois estes são os números cujos dois primeiros dígitos são 01 e 21. Logo, a imagem de Ψ, que denotamos $C \subseteq (0,1]$, está contida em um conjunto de comprimento $(\frac{2}{3})^2$. Procedendo desta maneira, vemos que a C está contido em um conjunto de comprimento $(\frac{2}{3})^k$ para qualquer k e, portanto, C tem "comprimento" zero.

Agora seja X uma variável aleatória com distribuição uniforme em $(0,1]$, e tome $Y = \Psi(X)$. Como F_X é contínua e Ψ é injetiva, $\mathbb{P}(Y = y) = 0$ para qualquer y, ou seja, F_Y é contínua. Entretanto, $\mathbb{P}(Y \in C) = 1$, e C tem "comprimento" zero, logo Y não pode ter densidade. $\triangle$

Vemos que a variável aleatória Y do exemplo acima ocupa um lugar estranho entre as discretas e as absolutamente contínuas. Por um lado, nenhum ponto em particular tem probabilidade positiva de ocorrer, isto é $p_Y = 0$, o que afasta operações com somatórios como no caso discreto. Por outro lado, essa variável aleatória vive em um conjunto pequeno da reta (mais precisamente em C, que tem "comprimento" zero), não sendo aplicável tampouco o uso de

integrais em $f(x)\,\mathrm{d}x$ para nenhuma f.

3.6 Existência e unicidade de distribuições

Para o leitor que passou pela Seção 1.4, continuamos colecionando ferramentas que serão necessárias para estudar os capítulos mais avançados.

Aqui provamos os Teoremas 3.10 e 3.16. No caminho para a prova daquele, vamos definir π-sistemas e enunciar um teorema sobre unicidade de medidas, que será utilizado em outros capítulos.

Definição 3.36 (π-sistemas)**.** Uma classe $\mathcal{C}$ de conjuntos é chamado um *π-sistema* se é fechada por interseções, isto é, $A \cap B \in \mathcal{C}$ para todos $A, B \in \mathcal{C}$.

Teorema 3.37 (Unicidade de medidas)**.** *Sejam μ e ν medidas num espaço mensurável $(\Omega, \mathcal{F})$, e seja $\mathcal{C} \subseteq \mathcal{F}$ um π-sistema. Suponha que $\mu(A_n) = \nu(A_n) < \infty$ para alguma sequência $A_n \uparrow \Omega$ de conjuntos em $\mathcal{C}$. Se $\mu = \nu$ em $\mathcal{C}$, então $\mu = \nu$ em $\sigma(\mathcal{C})$.*

A demonstração será dada no Apêndice D.1.

Demonstração do Teorema 3.10. Seja $\mathcal{C} = \{(-\infty, b] : b \in \mathbb{R}\}$ a classe de todos os intervalos fechados e semi-infinitos à esquerda. Observe que $\mathcal{C}$ é um π-sistema e, pela Proposição 1.43, $\sigma(\mathcal{C}) = \mathcal{B}$. Por hipótese,

$$\mathbb{P}_X((-\infty, b]) = F_X(b) = F_Y(b) = \mathbb{P}_Y((-\infty, b]),$$

portanto $\mathbb{P}_X(A) = \mathbb{P}_Y(A)$, para todo $A \in \mathcal{C}$. Além disso, a sequência $(A_n)_n$ dada por $A_n = (-\infty, n]$, satisfaz $A_n \in \mathcal{C}$ e $\mathbb{P}_X(A_n) = \mathbb{P}_Y(A_n) < \infty$ para todo $n \in \mathbb{N}$ e $A_n \uparrow \mathbb{R}$. Pelo Teorema 3.37, segue que $\mathbb{P}_X(C) = \mathbb{P}_Y(C)$, para todo $C \in \sigma(\mathcal{C})$, o que conclui a prova. $\square$

Concluímos esta breve seção com uma demonstração da existência de uma variável aleatória com distribuição $\mathcal{U}[0, 1]$.

Demonstração do Teorema 3.16. Recordemos que existe uma medida m em $(\mathbb{R}, \mathcal{B})$ tal que $m((a,b]) = b - a$ para todos $a < b \in \mathbb{R}$, como enunciado no Teorema 1.51. Tome $\Omega = \mathbb{R}$, $\mathcal{F} = \mathcal{B}$, $\mathbb{P} = m_{|_{[0,1]}}$ e $U(\omega) = \omega$. Então $F_U(u) = m([0,1] \cap (-\infty, u]) = u$ para todo $u \in [0,1]$. □

3.7 Funções mensuráveis

Esta seção é uma continuação da anterior. Aqui vamos introduzir o conceito de funções mensuráveis e estudar suas principais propriedades.

Definição 3.38 (Função mensurável)**.** Sejam $(\Omega_1, \mathcal{F}_1)$ e $(\Omega_2, \mathcal{F}_2)$ dois espaços mensuráveis. Dizemos que uma função $f : \Omega_1 \to \Omega_2$ é $\mathcal{F}_1/\mathcal{F}_2$*-mensurável* se, para todo $A \in \mathcal{F}_2$, o conjunto $f^{-1}(A) = \{\omega \in \Omega_1 : f(\omega) \in A\}$ está em $\mathcal{F}_1$.

Quando as σ-álgebras $\mathcal{F}_1$ e $\mathcal{F}_2$ estiverem claras no contexto, podemos reduzir '$\mathcal{F}_1/\mathcal{F}_2$-mensurável' para apenas 'mensurável' ou '$\mathcal{F}_1$-mensurável'. Quando o espaço for $\mathbb{R}$ ou $\mathbb{R}^n$, vamos sempre considerar a σ-álgebra de Borel.

Como é de se esperar, a composição de funções mensuráveis é mensurável.

Lema 3.39 (Composição de funções mensuráveis)**.** *Sejam* $(\Omega_1, \mathcal{F}_1)$*,* $(\Omega_2, \mathcal{F}_2)$ *e* $(\Omega_3, \mathcal{F}_3)$ *espaços mensuráveis e* $f : \Omega_1 \to \Omega_2$*,* $g : \Omega_2 \to \Omega_3$ *funções mensuráveis. Então, a composição* $g \circ f$ *é mensurável.*

Demonstração. Seja $A \in \mathcal{F}_3$. Como g é mensurável, $g^{-1}(A) \in \mathcal{F}_2$. Como f é mensurável, $(g \circ f)^{-1}(A) = f^{-1}(g^{-1}(A)) \in \mathcal{F}_1$. Como isso vale para todo $A \in \mathcal{F}_3$, concluímos que $g \circ f$ é mensurável. □

Observe que pré-imagem comuta com operações de conjuntos: $f^{-1}(A \setminus B) = f^{-1}(A) \setminus f^{-1}(B)$, $f^{-1}(A \cap B) = f^{-1}(A) \cap f^{-1}(B)$, etc. Dado um espaço amostral Ω_1, um espaço mensurável $(\Omega_2, \mathcal{F}_2)$ e uma função $f : \Omega_1 \to \Omega_2$, podemos construir uma σ-álgebra em Ω_1 induzida por f.

Definição 3.40 (σ-álgebra gerada por uma função f)**.** Sejam Ω_1 um conjunto não-vazio, $(\Omega_2, \mathcal{F}_2)$ um espaço mensurável e $f : \Omega_1 \to \Omega_2$. A *$\sigma$-álgebra gerada por f* em Ω_1 é definida como $\sigma(f) = \{f^{-1}(A) : A \in \mathcal{F}_2\}$.

Observe que $\sigma(f)$ é de fato uma σ-álgebra (exercício!). A notação $\sigma(f)$ não deixa explícito o fato de que $\sigma(f)$ também depende de $\mathcal{F}_2$. Uma função qualquer $f : \Omega_1 \to \Omega_2$ é mensurável se, e somente se, $\sigma(f) \subseteq \mathcal{F}_1$.

Definição 3.41. Sejam $f : \Omega_1 \to \Omega_2$ uma função mensurável e μ uma medida em Ω_1. Definimos a *medida push-forward* $f_*\mu$ em Ω_2 dada por

$$(f_*\mu)(A) = \mu(f^{-1}(A))$$

Deixamos como exercício a prova de que $f_*\mu$ é uma medida.

Muitas vezes, é inconveniente verificar que $f^{-1}(A) \in \mathcal{F}_1$ para todo $A \in \mathcal{F}_2$, pois a classe $\mathcal{F}_2$ pode ser muito grande. O lema abaixo simplifica essa tarefa.

Lema 3.42 (Critério de mensurabilidade)**.** *Sejam $(\Omega_1, \mathcal{F}_1)$ e $(\Omega_2, \mathcal{F}_2)$ espaços mensuráveis, $\mathcal{E} \subseteq \mathcal{F}_2$ uma classe de conjuntos que gera $\mathcal{F}_2$, e $f : \Omega_1 \to \Omega_2$. Se $f^{-1}(A) \in \mathcal{F}_1$ para todo $A \in \mathcal{E}$, então f é mensurável.*

Demonstração. É suficiente mostrar que a coleção $\mathcal{G}$ dada por

$$\mathcal{G} = \{A \subseteq \Omega_2 : f^{-1}(A) \in \mathcal{F}_1\}$$

é uma σ-álgebra em Ω_2, pois por hipótese $\mathcal{E} \subseteq \mathcal{G}$, de forma que $\mathcal{F}_2 = \sigma(\mathcal{E}) \subseteq \mathcal{G}$. Primeiro, $\Omega_2 \in \mathcal{G}$, pois $f^{-1}(\Omega_2) = \Omega_1 \in \mathcal{F}_1$. Segundo, dado $A \in \mathcal{G}$, como $f^{-1}(\Omega_2 \setminus A) = \Omega_1 \setminus f^{-1}(A) \in \mathcal{F}_1$, segue que $A^c \in \mathcal{G}$. Terceiro, dados $A_1, A_2, \dots \in \mathcal{G}$ como $f^{-1}(\cup_{j=1}^{\infty} A_j) = \cup_{j=1}^{\infty} f^{-1}(A_j) \in \mathcal{F}_1$, segue que $\bigcup_{j=1}^{\infty} A_j \in \mathcal{G}$. Portanto, $\mathcal{G}$ é σ-álgebra sobre Ω_2. $\square$

Como espaços de probabilidade são espaços de medida, podemos reformular conceitos envolvendo variáveis aleatórias na linguagem desta seção. Pelo lema acima, uma função $X : \Omega \to \mathbb{R}$ é uma variável aleatória se, e somente

se, é uma função mensurável. Ademais, a distribuição de X, que denotamos $\mathbb{P}_X$, é o push-forward da medida $\mathbb{P}$ pela função X, ou seja, $\mathbb{P}_X = X_*\mathbb{P}$. Em particular, a fórmula (3.7) está definida para todo $B \in \mathcal{B}$. A σ-álgebra gerada por X é denotada por $\sigma(X)$.

Lema 3.43 (Funções contínuas)**.** *Toda $f : \mathbb{R}^n \to \mathbb{R}$ contínua é mensurável.*

Demonstração. Tomamos $\mathcal{E}$ como a classe dos conjuntos abertos em $\mathbb{R}$. Seja $B \in \mathcal{E}$. Pelo Teorema A.10, $f^{-1}(B)$ é aberto, logo $f^{-1}(B) \in \mathcal{B}(\mathbb{R}^n)$. Como $\sigma(\mathcal{E}) = \mathcal{B}(\mathbb{R})$, pelo Lema 3.42, f é mensurável. □

Lema 3.44 (Funções monótonas)**.** *Toda $f : \mathbb{R} \to \mathbb{R}$ monótona é mensurável.*

Demonstração. Tomamos $\mathcal{E}$ como a classe dos intervalos (possivelmente vazios) de $\mathbb{R}$. Como a pré-imagem de um intervalo por uma função monótona é sempre um intervalo, $f^{-1}(B) \in \mathcal{E} \subseteq \mathcal{B}$ para todo $B \in \mathcal{E}$. Pela Proposição 1.43, $\sigma(\mathcal{E}) = \mathcal{B}$. Logo, pelo Lema 3.42, f é mensurável. □

Lema 3.45 (Justaposição de funções reais)**.** *Seja $(\Omega, \mathcal{F})$ um espaço mensurável e $f_1, \ldots, f_n : \Omega \to \mathbb{R}$ funções reais. Considere a função $\boldsymbol{f} : \Omega \to \mathbb{R}^n$ dada por $\boldsymbol{f} = (f_1, \ldots, f_n)$. Então $\boldsymbol{f}$ é $\mathcal{F}/\mathcal{B}(\mathbb{R}^n)$-mensurável se, e somente se, todas as $f_1, \ldots, f_n$ são $\mathcal{F}/\mathcal{B}(\mathbb{R})$-mensuráveis.*

Demonstração. Suponha que $\boldsymbol{f}$ seja mensurável. Como a projeção $\pi_k : \mathbb{R}^n \to \mathbb{R}$ dada por $\pi_k(x_1, \ldots, x_n) = x_k$ é uma função contínua, é mensurável. Como $f_k = \pi_k \circ \boldsymbol{f}$, segue que f_k é mensurável. Reciprocamente, suponha que $f_1, \ldots, f_n$ são mensuráveis. Seja $A \subseteq \mathbb{R}^n$ aberto. Pelo Teorema A.11, existe uma sequência de paralelepípedos abertos $(I_j)_{j\in\mathbb{N}}$ tais que $A = \cup_j I_j$. Cada I_j é da forma $(a_1, b_1) \times \cdots \times (a_n, b_n)$. Assim, $\{\boldsymbol{f} \in I_j\} = \cap_{k=1}^n \{f_k \in (a_k, b_k)\} \in \mathcal{F}$, logo $\{\boldsymbol{f} \in A\} = \cup_j \{\boldsymbol{f} \in I_j\} \in \mathcal{F}$. Pelo Lema 3.42, $\boldsymbol{f}$ é mensurável. □

Veremos mais abaixo que algumas operações usuais com funções reais mensuráveis resultam em funções também mensuráveis. Para tratar de forma mais robusta supremos e limites de funções, sem ter que ficar separando

em casos, é mais conveniente considerar funções que tomam valores na reta estendida $[-\infty, +\infty]$.

Definimos a σ-álgebra de Borel em $[-\infty, +\infty]$ como

$$\mathcal{B}([-\infty, +\infty]) = \{A \subseteq [-\infty, +\infty] : A \cap \mathbb{R} \in \mathcal{B}(\mathbb{R})\}.$$

Em outras palavras, essa σ-álgebra é formada por conjuntos da forma $B \cup D$, com $B \in \mathcal{B}(\mathbb{R})$ e $D \subseteq \{-\infty, +\infty\}$.

Proposição 3.46. *Uma função $f : \Omega \to [-\infty, +\infty]$ é mensurável se, e somente se, $\{f \leqslant a\} \in \mathcal{F}$ para todo $a \in \mathbb{R}$.*

Demonstração. Basta mostrar que $\sigma(\mathcal{E}) = \mathcal{B}([-\infty, +\infty])$, onde $\mathcal{E}$ é a classe $\{[-\infty, a] : a \in \mathbb{R}\}$. Observe que $\{-\infty\} = \cap_{a \in \mathbb{Z}}[-\infty, a] \in \sigma(\mathcal{E})$, pois $\sigma(\mathcal{E})$ é fechada por interseções enumeráveis. Da mesma forma, $[-\infty, +\infty) = \cup_{a \in \mathbb{Z}}[-\infty, a] \in \sigma(\mathcal{E})$, logo $\{+\infty\}$ e $\mathbb{R}$ também estão em $\sigma(\mathcal{E})$. Defina $\mathcal{G} = \{A \cap \mathbb{R} : A \in \sigma(\mathcal{E})\}$. Observe que $\mathcal{G}$ é uma σ-álgebra em $\mathbb{R}$ e, como $(-\infty, a] = [-\infty, a] \cap \mathbb{R} \in \mathcal{G}$ para todo $a \in \mathbb{R}$, temos que $\mathcal{B}(\mathbb{R}) \subseteq \mathcal{G}$. Assim, $\sigma(\mathcal{E})$ contém $\mathcal{B}(\mathbb{R})$, $\{-\infty\}$ e $\{+\infty\}$. Portanto, $\sigma(\mathcal{E})$ contém $\mathcal{B}([-\infty, +\infty])$, que contém $\mathcal{E}$, provando que $\sigma(\mathcal{E}) = \mathcal{B}([-\infty, +\infty])$. $\square$

Uma função $f : \Omega \to \mathbb{R}$ ou $f : \Omega \to [0, +\infty]$ é um caso particular de uma função $f : \Omega \to [-\infty, +\infty]$, e o critério de mensurabilidade é o mesmo. Assim sendo, propriedades de funções mensuráveis tomando valores em $[-\infty, +\infty]$ também valem para funções mensuráveis tomando valores em $\mathbb{R}$.

Definição 3.47 (Variável aleatória estendida)**.** Seja $(\Omega, \mathcal{F}, \mathbb{P})$ um um espaço de probabilidade. Uma *variável aleatória estendida* é uma função mensurável $X : \Omega \to [-\infty, +\infty]$.

Uma propriedade que vamos usar inúmeras vezes, quase sempre de forma implícita, é que a mensurabilidade é preservada por vários tipos de operações básicas com funções. As propriedades enunciadas abaixo serão usadas em capítulos subsequentes sem que se lhes faça referência explícita.

As provas dos lemas abaixo encontram-se no Apêndice D.3.

Lema 3.48 (Comparação de funções)**.** *Sejam* $(\Omega, \mathcal{F})$ *um espaço mensurável e* $f, g : \Omega \to [-\infty, +\infty]$ *funções mensuráveis. Então, os subconjuntos de* Ω *dados por* $\{f < g\}$, $\{f \leqslant g\}$, $\{f = g\}$, *e* $\{f \neq g\}$ *pertencem à* σ*-álgebra* $\mathcal{F}$.

Somas e produtos de funções mensuráveis também são mensuráveis.

Lema 3.49 (Somas e produtos)**.** *Seja* $(\Omega, \mathcal{F})$ *um espaço mensurável e* $f, g : \Omega \to [-\infty, +\infty]$ *funções mensuráveis. Então* fg *é uma função mensurável. Se* $f + g$ *está bem definida para todo* ω, *então é uma função mensurável.*

O $\limsup$ e correlatos também preservam mensurabilidade.

Lema 3.50 (Supremos e limites de funções)**.** *Sejam* $(\Omega, \mathcal{F})$ *um espaço mensurável e* $(f_n)_{n \geqslant 1}$ *uma sequência de funções mensuráveis de* Ω *em* $[-\infty, +\infty]$. *Então, as funções* $\sup_n f_n$, $\inf_n f_n$, $\limsup_n f_n$, $\liminf_n f_n$ *são mensuráveis. Em particular, se definimos a função*

$$f(\omega) = \begin{cases} \lim_n f_n(\omega), & \textit{caso o limite exista}, \\ 0, & \textit{caso contrário}, \end{cases}$$

então f *é uma função mensurável. Se as funções* $(f_n)_n$ *são não-negativas, então* $\sum_n f_n$ *é uma função mensurável.*

3.8 Exercícios

§3.1

1. Seja $F(x)$ a função

$$F(x) = \begin{cases} 0, & x < 0, \\ x + \frac{1}{2}, & 0 \leqslant x \leqslant \frac{1}{2} \\ 1, & x > \frac{1}{2}. \end{cases}$$

Mostre que F é de fato uma função de distribuição e calcule:

(a) $\mathbb{P}(X > \frac{1}{8})$
(b) $\mathbb{P}(\frac{1}{8} < X < \frac{2}{5})$
(c) $\mathbb{P}(X < \frac{2}{5} \mid X > \frac{1}{8})$

2. Encontre os valores das constantes reais α e β de modo que a função F abaixo seja função de distribuição de alguma variável aleatória definida em algum espaço de probabilidade:

$$F(x) = \begin{cases} 0, & x \leqslant 0, \\ \alpha + \beta e^{-x^2/2}, & x > 0. \end{cases}$$

3. Sejam X, Y, Z e W variáveis aleatórias onde X é uma variável qualquer, $Y = \max\{0, X\}$, $Z = \max\{0, -X\}$ e $W = |X|$. Escreva as funções de distribuição F_Y, F_Z e F_W em termos da função de distribuição F_X.

4. Mostre que, se $\mathbb{P}(X = Y) = 1$, então $F_X = F_Y$.

5. Seja X uma variável aleatória. Prove que $\mathbb{P}_X$ é uma medida de probabilidade.

6. Seja X uma variável aleatória contínua e F_X sua função de distribuição. Considere $Y = F_X(X)$. Mostre que Y tem distribuição uniforme em $(0, 1)$,

(a) supondo que F_X é contínua e estritamente crescente,
(b) supondo apenas que F_X é contínua.

§3.2

7. Seja X o número de caras obtidas em 4 lançamentos de uma moeda honesta. Escreva a função de probabilidade e a função de distribuição de X e esboce os seus gráficos.

8. Seja X uma variável aleatória com função de probabilidade $\mathbb{P}(X = x) = cx^2$, onde c é uma constante e $x = 1, 2, 3, 4, 5$.

(a) Encontre $p_X(x)$ e $F_X(x)$.

(b) Calcule $\mathbb{P}(X \text{ é ímpar})$.

9. Temos duas urnas, em cada uma delas há cinco bolas numeradas de 1 a 5. Retiramos, sem reposição, 3 pares de bolas, onde cada par é sempre formado por uma bola de cada urna. Seja X o quantidade de pares de bolas com o mesmo número. Determine a função de probabilidade de X e desenhe o gráfico de sua função de distribuição.

10. Seja $X \sim \text{Geom}(p)$, mostre que $\mathbb{P}(X \geqslant m+n | X > n) = \mathbb{P}(X \geqslant m)$ para todos m e n inteiros não-negativos.

11. Suponha que, para cada $N \in \mathbb{N}$, temos uma variável aleatória X_N com distribuição $\text{Hgeo}(N, K, n)$, onde $K = K(N)$ é tal que $\frac{K}{N} \to p \in [0,1]$ quando $N \to \infty$. Mostre que

$$\lim_{N\to\infty} \mathbb{P}(X_N = k) = \mathbb{P}(Z = k)$$

para todo $k = 0, \ldots, n$, onde $Z \sim \text{Binom}(n, p)$.

12 (Fósforos de Banach)**.** Um indivíduo tem duas caixas de fósforos, uma em cada bolso e cada uma delas com n palitos. Toda vez que ele precisa de um fósforo, ele escolhe aleatoriamente uma das caixas, retira um palito e a devolve para o mesmo bolso. Ele percebe que uma caixa está vazia somente quando vai pegar um palito e não há nenhum nesta caixa. Seja X o número de palitos na outra caixa quando ele percebe que a primeira está vazia. Mostre que

$$p_X(k) = \tfrac{1}{2^{2n-k}}\tbinom{2n-k}{n-k}.$$

13. Considere um jogo de dominó para crianças que contém 28 peças distintas, sendo cada metade de peça é pintada com uma das 7 cores do arco-íris. Júlia recebe 2 peças escolhidas ao acaso. Seja X o número de cores distintas que Júlia observa entre as peças recebidas. Calcule p_X e faça o gráfico de F_X.

14. Sejam $X : \Omega \to \mathbb{R}$ uma variável aleatória que assume finitos valores, denotados $x_1, \ldots, x_n$. Para cada $k = 1, \ldots, n$ seja $p_k = p_X(x_k)$, onde p_X é a

função de probabilidade de X. Definimos a *entropia de* X, $H(X)$, como

$$H(X) = -\sum_{k=1}^{n} p_k \log p_k,$$

com a convenção que $0 \cdot \log 0 = 0$.

(a) Calcule $H(X)$, se X é a variável degenerada em x_j, ou seja, $p_X(x_k) = \delta_j(k)$ para todo $k = 1, \ldots, n$.
(b) Seja $X \sim \text{Binom}(3, \frac{1}{2})$. Calcule $H(X)$.
(c) Determine o vetor $p^* = (p_1^*, \ldots, p_n^*)$ que resulta e máxima entropia. Para isso pode-se usar a técnica do multiplicador de Lagrange: sujeito à restrição $p_1 + \cdots + p_n = 1$, o ponto p^* que maximiza H é o ponto em que $\nabla H(p)$ é ortogonal à superfície $\{p : p_1 + \cdots + p_n = 1\}$.
(d) Observando que $H(X) \geqslant 0$ para toda variável X e levando em consideração os itens anteriores, tente interpretar o significado de entropia de uma variável aleatória.

§3.3

15. Seja X uma variável aleatória absolutamente contínua com função densidade f_X par, isto é, $f_X(x) = f_X(-x)$ para todo $x \in \mathbb{R}$. Mostre que

(a) $F_X(x) = 1 - F_X(-x)$.
(b) $F_X(0) = \frac{1}{2}$.
(c) $\mathbb{P}(-x < X < x) = 2F_X(x) - 1$ para todo $x > 0$.
(d) $\mathbb{P}(X > x) = \frac{1}{2} - \int_0^x f_X(t)\mathrm{d}t$ para todo $x > 0$.

16. Se $f(t) = c\ 3t^2 e^{-t} \mathbb{1}_{[0,2]}(t)$ é uma função de densidade, ache c.

17. Esboce os gráficos das funções de densidade e distribuição de uma variável aleatória com distribuição normal de parâmetros μ e σ^2.

18. Seja $X \sim \mathcal{U}[1, 2]$. Qual a distribuição de $Y = 3 - 2X$?

19. Seja $X \sim \text{Exp}(\lambda)$ e $N = \lceil X \rceil$. Encontre a distribuição de N.

20.

(a) Se $U \sim \mathcal{U}[0,1]$, encontre uma densidade de $\sqrt{5}\tan(\pi U - \frac{\pi}{2})$.
(b) Se $X \sim \text{Laplace}(0,1)$, encontre uma densidade de $|X|$.
(c) Se $Z \sim \mathcal{N}(0,1)$, encontre uma densidade de Z^2.

21. Sejam X uma variável aleatória absolutamente contínua com densidade $f_X(x)$, e $Y = aX + b$ com $a > 0$ e $b \in \mathbb{R}$. Mostre que Y é absolutamente contínua com densidade $f_Y(y) = \frac{1}{a} f_X(\frac{y-b}{a})$.

22. Mostre que, se $X \sim \text{Cauchy}(0,1)$, então $a + bX \sim \text{Cauchy}(a,b)$.

§3.4

23. Se $X \sim \text{Exp}(\lambda)$, mostre que, para todo $s > 0$, a distribuição condicional de X dado que $X > s$ é igual à distribuição de $X + s$.

24. Se Z tem distribuição geométrica, mostre que a distribuição condicional de Z dado que $Z > n$ é igual à distribuição de $Z + n$ para todo $n \in \mathbb{N}$.

25. Seja $X \sim \mathcal{U}[a,b]$. Considere o evento $A = \{c \leqslant X \leqslant d\}$, onde $c < d \in \mathbb{R}$. Determine a função de distribuição condicional $F_{X|A}$.

26. Sejam $X \sim \text{Exp}(\lambda)$ e $Y = 5X$, ache a função de distribuição de Y. Ache a função de distribuição condicional e uma densidade condicional de Y dado que $\{X > 3\}$.

§3.5

27. Sejam $X \sim \text{Exp}(1)$ e $Y = \max\{X, 1\}$. Encontre as funções de distribuição, de probabilidade e de densidade de Y.

28. Seja $X \sim \mathcal{U}[0,3]$. Esboce o gráfico da função de distribuição de $Y = \min(2X, X^2)$. A variável Y tem função de densidade?

29. Sejam $X \sim \mathcal{U}[0,2]$ e $Y = \max(\lceil X \rceil, \frac{3}{2}X, X^2)$, onde $\lceil z \rceil$ é o menor inteiro maior que ou igual a z.

(a) Esboce o gráfico da função de distribuição de Y.

(b) Encontre as funções de probabilidade e de densidade de Y.

30. Seja Y a variável aleatória definida no Exemplo 3.35.

(a) Quanto vale $\mathbb{P}(Y \in (\frac{1}{8}, \frac{6}{7}))$?
(b) Esboce o gráfico de sua função de distribuição F_Y.
(c) Determine a função de distribuição condicional $F_{2Y|\{Y \leqslant \frac{1}{3}\}}$.
(d) Determine a função de distribuição $F_{2^n Y|\{Y \leqslant \frac{1}{3^n}\}}$ para todo $n \in \mathbb{N}$. Tente interpretar este exercício geometricamente, em termos do gráfico da função de distribuição.

§3.6

31. Dê pelo menos sete exemplos de π-sistemas em $\mathbb{R}$ que geram $\mathcal{B}(\mathbb{R})$.

32. Em $\Omega = \{1, 2, 3, 4\}$, dê um exemplo de medidas de probabilidade $\mathbb{P}$ e $\mathbf{P}$ tais que a classe $\mathcal{D} = \{A \in \mathcal{P}(\Omega) : \mathbb{P}(A) = \mathbf{P}(A)\}$ gera $\mathcal{P}(\Omega)$ mas não é um π-sistema.

§3.7

33. Sejam $f : \Omega_1 \to \Omega_2$ uma função qualquer e $\mathcal{F}_2$ é uma σ-álgebra em Ω_2. Prove que $f^{-1}(\mathcal{F}_2) = \{f^{-1}(A) : A \in \mathcal{F}_2\}$ é uma σ-álgebra em Ω_1.

34. Sejam $(\Omega, \mathcal{F})$ e $(\Omega', \mathcal{F}')$ espaços mensuráveis. Mostre que

(a) A função identidade $I : \Omega \to \Omega$, dada por $I(\omega) = \omega$ é $\mathcal{F}/\mathcal{F}$-mensurável.
(b) Se $\mathcal{F} = \mathcal{P}(\Omega)$, então toda função $f : \Omega \to \Omega'$ é $\mathcal{F}/\mathcal{F}'$-mensurável.
(c) Se $\mathcal{F}' = \{\emptyset, \Omega'\}$, então toda função $f : \Omega \to \Omega'$ é $\mathcal{F}/\mathcal{F}'$-mensurável.
(d) Se $\Omega' = \{0, 1\}$, $\mathcal{F}' = \mathcal{P}(\Omega')$ e $A \subseteq \Omega$, então a função indicadora $\mathbb{1}_A : \Omega \to \Omega'$ é $\mathcal{F}/\mathcal{F}'$-mensurável se, e somente se $A \in \mathcal{F}$.

35. Sejam X uma variável aleatória e $g : \mathbb{R} \to \mathbb{R}$ uma função mensurável. Defina $Y(\omega) = g(X(\omega))$. Mostre que Y é uma variável aleatória. Mostre que $\sigma(Y) \subseteq \sigma(X)$.

36. Sejam A e B subconjuntos de Ω e $f : \Omega \to \mathbb{R}$ a função $f = a\mathbb{1}_A + b\mathbb{1}_B$ com $a, b > 0$. Descreva $\sigma(f)$ explicitamente.

37. Sejam μ a medida de contagem em $\mathbb{Z}$ e $f(x) = |x|$. Descreva a medida $f_*\mu$ explicitamente.

38. Seja $(X_n)_n$ uma sequência de variáveis aleatórias estendidas. Defina

$$X = \begin{cases} \lim_n X_n, & \text{se o limite existe,} \\ 0, & \text{caso contrário.} \end{cases}$$

Mostre que X é uma variável aleatória estendida.

39. Considere o espaço mensurável $([0,1], \mathcal{B}([0,1]))$, e defina as funções $f, g : [0,1] \to \{0,1,2,3\}$ por

$$f(\omega) = \begin{cases} 1, & \omega \in [0, \frac{1}{4}] \\ 2, & \omega \in (\frac{1}{4}, \frac{1}{2}] \\ 3, & \omega \in (\frac{1}{2}, 1] \end{cases} \quad , \quad g(\omega) = \begin{cases} 0, & \omega \in [0, \frac{1}{2}] \\ 1, & \omega \in (\frac{1}{2}, 1] \end{cases} \quad .$$

Verifique explicitamente que $\sigma(g) \subseteq \sigma(f)$.

Capítulo 4

Vetores Aleatórios

Imagine que queiramos produzir duas variáveis aleatórias com distribuição Bernoulli($\frac{1}{2}$). Uma forma seria lançar uma moeda duas vezes e considerar o par $\boldsymbol{X} = (Z, W)$, onde Z é o resultado do primeiro lançamento e W do segundo. Outras formas seriam lançar a moeda apenas uma vez e copiar o resultado ou trocar o resultado, obtendo os pares $\boldsymbol{Y} = (Z, Z)$ e $\tilde{\boldsymbol{Y}} = (Z, 1 - Z)$, respectivamente. Em todos os casos, produziu-se um par de variáveis aleatórias distribuídas como Bernoulli($\frac{1}{2}$). Entretanto, o comportamento *conjunto* dessas variáveis aleatórias é bem diferente nos três casos.

Neste capítulo vamos estudar *vetores aleatórios*, isto é, coleções de variáveis aleatórias em que se considera seu comportamento estatístico conjunto. Vamos estender os conceitos função de probabilidade, função de densidade e função de distribuição, e posteriormente explorar conceitos de independência, entre outros.

4.1 Vetores aleatórios

Começamos introduzindo o conceito que intitula este capítulo.

Definição 4.1 (Vetor aleatório)**.** Dizemos que $\boldsymbol{X} = (X_1, \ldots, X_n)$ é um *vetor aleatório* se X_j for variável aleatória para todo $j = 1, \ldots, n$.

Dado um espaço de probabilidade $(\Omega, \mathcal{F}, \mathbb{P})$ e um vetor aleatório $\boldsymbol{X}$, definimos o *espaço de probabilidade induzido por* $\boldsymbol{X}$ como $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mathbb{P}_{\boldsymbol{X}})$, onde

$$\mathbb{P}_{\boldsymbol{X}}(B) = \mathbb{P}(\{\omega : \boldsymbol{X}(\omega) \in B\}), \quad B \in \mathcal{B}(\mathbb{R}^n), \tag{4.2}$$

e $\mathcal{B}(\mathbb{R}^n)$ é a σ-álgebra de Borel em $\mathbb{R}^n$, definida na Seção 1.4.1. Ou seja, o espaço amostral é o conjunto $\mathbb{R}^n$ dos vetores n-dimensionais, os eventos aleatórios são os conjuntos borelianos em $\mathbb{R}^n$, e a medida de probabilidade é aquela induzida por $\boldsymbol{X}$. Chamaremos de *distribuição conjunta do vetor aleatório* $\boldsymbol{X}$ a medida de probabilidade $\mathbb{P}_{\boldsymbol{X}}$ em $\mathbb{R}^n$ induzida por $\boldsymbol{X}$. Segue do Lema 3.45 que $\mathbb{P}_{\boldsymbol{X}}$ está definida para todo $B \in \mathcal{B}(\mathbb{R}^n)$.

Dada $\boldsymbol{x} \in \mathbb{R}^n$ uma n-upla de números reais, $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$, denotamos por $\boldsymbol{x} \leqslant \boldsymbol{y}$ o conjunto de desigualdades $x_j \leqslant y_j, j = 1, \ldots, n$, isto é, $\boldsymbol{x} \leqslant \boldsymbol{y}$ se, e somente se, as desigualdades valem para todas as coordenadas simultaneamente.

Definição 4.3 (Função de distribuição conjunta)**.** A *função de distribuição conjunta* de um vetor aleatório $\boldsymbol{X}$, denotada por $F_{\boldsymbol{X}}$, é a função $F_{\boldsymbol{X}} : \mathbb{R}^n \to \mathbb{R}$ dada por

$$F_{\boldsymbol{X}}(\boldsymbol{x}) = \mathbb{P}(\boldsymbol{X} \leqslant \boldsymbol{x}).$$

Exemplo 4.4. Lançamos duas moedas honestas e consideramos X_1 = quantidade de caras, X_2 = quantidade de coroas, e $\boldsymbol{X} = (X_1, X_2)$. Os valores de $F_{\boldsymbol{X}}$ são mostrados na Figura 4.1. △

A prova do teorema abaixo é análoga à do Teorema 3.10, e será dada no Apêndice D.4.

Teorema 4.5. *Se $\boldsymbol{X}$ e $\boldsymbol{Y}$ são vetores aleatórios tais que $F_{\boldsymbol{X}}(\boldsymbol{t}) = F_{\boldsymbol{Y}}(\boldsymbol{t})$ para todo $\boldsymbol{t} \in \mathbb{R}^n$, então $\mathbb{P}_{\boldsymbol{X}}(B) = \mathbb{P}_{\boldsymbol{Y}}(B)$ para todo $B \in \mathcal{B}(\mathbb{R}^n)$.*

Ou seja, como no caso de variáveis aleatórias, a função de distribuição conjunta determina a distribuição conjunta de um vetor aleatório. Entretanto, não é

usual ou prático especificar $\mathbb{P}_{\boldsymbol{X}}$ a partir de uma descrição explícita de $F_{\boldsymbol{X}}$. Tampouco existe uma forma natural de simular um vetor aleatório a partir de uma função de distribuição conjunta qualquer usando variáveis uniformes, e não existe uma noção natural de quantil (mediana, quartil, percentil, etc.) de vetores aleatórios. De fato, uma generalização do Teorema 3.17 é complicada e de pouco uso prático. Ainda assim, a função de distribuição conjunta tem propriedades muito úteis, como veremos ao longo desta seção.

A partir da função de distribuição conjunta, pode-se obter o comportamento de cada variável isoladamente. A distribuição de uma das coordenadas obtida a partir da distribuição conjunta é chamada de *distribuição marginal*.

Proposição 4.6. *Seja F_{X_k} a função de distribuição da k-ésima coordenada do vetor $\boldsymbol{X} = (X_1, \ldots, X_n)$. Então*

$$F_{X_k}(x_k) = \lim_{\substack{x_j \to +\infty \\ j \neq k}} F_{\boldsymbol{X}}(x_1, \ldots, x_n),$$

em que o limite é tomado em todas as coordenadas, exceto a k-ésima.

Demonstração. Reordenando as coordenadas, podemos supor sem perda de

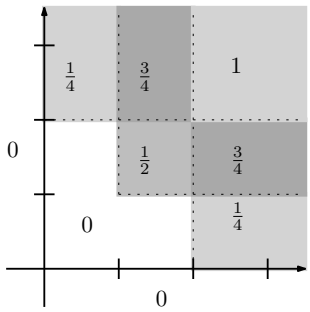


Figura 4.1. Valores de $F_{\boldsymbol{X}}(t_1, t_2)$ para cada $(t_1, t_2) \in \mathbb{R}^2$ no Exemplo 4.4.

generalidade que $k = 1$, e temos que mostrar que

$$F_{X_1}(x_1) = \lim_{x_2 \to +\infty} \cdots \lim_{x_n \to +\infty} F_{\boldsymbol{X}}(x_1, \ldots, x_n). \tag{4.7}$$

No primeiro limite, dada $(x_n^j)_j$ tal que $x_n^j \uparrow +\infty$, e fixados $(x_1, \ldots, x_{n-1})$, temos $\{\boldsymbol{X} \leqslant (x_1, \ldots x_{n-1}, x_n^j)\} \uparrow \{X_1 \leqslant x_1, \ldots X_{n-1} \leqslant x_{n-1}\}$ quando $j \to \infty$, logo $F_{\boldsymbol{X}}(x_1, \ldots x_{n-1}, x_n^j) \to F_{X_1,\ldots,X_{n-1}}(x_1, \ldots, x_{n-1})$, ou seja,

$$F_{X_1,\ldots,X_{n-1}}(x_1, \ldots, x_{n-1}) = \lim_{x_n \to +\infty} F_{\boldsymbol{X}}(x_1, \ldots, x_n).$$

Tomando $(x_{n-1}^j)_j$ tal que $x_{n-1}^j \uparrow +\infty$, de forma análoga chegamos a

$$F_{X_1,\ldots,X_{n-2}}(x_1, \ldots, x_{n-2}) = \lim_{x_{n-1} \to +\infty} \lim_{x_n \to +\infty} F_{\boldsymbol{X}}(x_1, \ldots, x_n).$$

Tomando $(x_{n-2}^j)_j$ tal que $x_{n-2}^j \uparrow +\infty$, e assim sucessivamente, chegamos a (4.7), o que conclui a demonstração. □

Exemplo 4.8. No Exemplo 4.4, tomando o limite em cada variável temos

$$F_{X_1}(x_1) = \lim_{x_2 \to +\infty} F_{X_1,X_2}(x_1, x_2) = \begin{cases} 0, & x_1 < 0, \\ \frac{1}{4}, & 0 \leqslant x_1 < 1, \\ \frac{3}{4}, & 1 \leqslant x_1 < 2, \\ 1, & x_1 \geqslant 2, \end{cases}$$

e, similarmente, encontramos a mesma expressão para F_{X_2}. △

Definição 4.9 (Variáveis aleatórias independentes)**.** Dizemos que as variáveis aleatórias $X_1, X_2, \ldots, X_n$ são *independentes* se

$$\mathbb{P}(X_1 \in B_1, \ldots, X_n \in B_n) = \mathbb{P}(X_1 \in B_1) \cdots \mathbb{P}(X_n \in B_n)$$

para todos $B_1, \ldots, B_n \in \mathcal{B}$. Dito de forma mais sofisticada,

$$\mathbb{P}_{\boldsymbol{X}}(B_1 \times \cdots \times B_n) = \mathbb{P}_{X_1}(B_1) \cdots \mathbb{P}_{X_n}(B_n).$$

Se J é uma coleção qualquer de índices, dizemos que a família de variáveis $(X_j)_{j\in J}$ são *independentes* se $X_{j_1}, \ldots, X_{j_k}$ são independentes para todo $k \in \mathbb{N}$ e toda escolha $j_1, \ldots, j_k \in J$ distintos.

Dada uma família de variáveis aleatórias independentes, qualquer subfamília é também formada por variáveis aleatórias independentes.

Muitas vezes vamos considerar famílias de variáveis aleatórias que, além de serem independentes, têm a mesma distribuição, o que chamamos de *independentes e identicamente distribuídas*, ou simplesmente i.i.d.

Proposição 4.10 (Critério de independência)**.** *Seja* $\boldsymbol{X} = (X_1, \ldots, X_n)$ *um vetor aleatório. Então* $X_1, \ldots, X_n$ *são independentes se, e somente se,* $F_{\boldsymbol{X}}(\boldsymbol{x}) = F_{X_1}(x_1) \cdots F_{X_n}(x_n)$ *para todo* $\boldsymbol{x} \in \mathbb{R}^n$.

Demonstração. A ideia da prova é a seguinte: se as $X_1, \ldots, X_n$ são independentes, então $F_{\boldsymbol{X}}$ é dada pelo produto $F_{X_1} \cdots F_{X_n}$ e, por outro lado, o Teorema 4.5 nos diz que $F_{\boldsymbol{X}}$ determina $\mathbb{P}_{\boldsymbol{X}}$, logo $\mathbb{P}_{\boldsymbol{X}}(B_1 \times \cdots \times B_n) = \mathbb{P}_{X_1}(B_1) \cdots \mathbb{P}_{X_n}(B_n)$. A prova completa será dada será dada no Apêndice D.4. □

Assim como fizemos no caso de independência de eventos, podemos definir independência duas a duas para famílias de variáveis aleatórias.

Definição 4.11. Seja J uma coleção qualquer de índices e $(X_j)_{j\in J}$ uma família de variáveis aleatórias. Dizemos que as variáveis $(X_j)_{j\in J}$ são *independentes duas a duas* se X_k e X_j são independentes para quaisquer $k \neq j \in J$.

Observe que uma família de variáveis aleatórias independentes também é independente duas a duas. Entretanto não vale a recíproca.

Contra-exemplo 4.12. Sejam X e Y independentes assumindo os valores -1 ou $+1$ com probabilidade $\frac{1}{2}$ cada, e tome $Z = XY$. Então X, Y e Z não são

independentes, pois

$$\mathbb{P}(X = 1, Y = 1, Z = 1) = \frac{1}{4} \neq \frac{1}{8} = \mathbb{P}(X = 1)\mathbb{P}(Y = 1)\mathbb{P}(Z = 1).$$

Entretanto, X, Y e Z são independentes duas a duas. $\triangle$

Concluímos esta seção com a observação seguinte, cuja prova será dada na Seção 13.1.

Observação 4.13. Funções mensuráveis de famílias disjuntas de variáveis aleatórias independentes resultam em variáveis independentes. Por exemplo, se X, Y e Z são variáveis aleatórias independentes, então $X + Y$ e Z também são independentes. $\triangle$

Apesar de que propriedades análogas àquelas listadas na Proposição 3.13 são satisfeitas pela função de distribuição conjunta, como vemos na Figura 4.1, um análogo do Teorema 3.17 pode falhar, dependendo de como for feita a generalização. Por exemplo, a função $F : \mathbb{R}^2 \to [0, 1]$ dada por

$$F(x, y) = \begin{cases} 1, & x \geqslant 0, y \geqslant 0, x + y \geqslant 1, \\ 0, & \text{caso contrário}, \end{cases}$$

não pode ser função de distribuição conjunta de nenhum vetor aleatório, apesar de ser monótona e contínua à direita em cada variável, tender a 1 quando todas as coordenadas tendem a $+\infty$, e tender a 0 quando uma delas tende a $-\infty$. Com efeito, se tal F fosse função de distribuição conjunta de algum par (X, Y), teríamos

$$\mathbb{P}\big(a < X \leqslant b, c < Y \leqslant d\big) = F(b, d) - F(b, c) - [F(a, d) - F(a, c)]$$

para todos $a < b$ e $c < d$ e, tomando $a = c = 0$ e $b = d = 1$, teríamos $\mathbb{P}\big(0 < X \leqslant 1, 0 < Y \leqslant 1\big) = -1$, o que seria absurdo.

4.2 Vetores aleatórios discretos e contínuos

Dizemos que um vetor aleatório $\boldsymbol{X}$ é *discreto* se existe um subconjunto enumerável $A \subseteq \mathbb{R}^n$ tal que $\mathbb{P}(\boldsymbol{X} \in A) = 1$. Neste caso, a *função de probabilidade conjunta* de $\boldsymbol{X}$ é dada por

$$p_{\boldsymbol{X}}(\boldsymbol{x}) = \mathbb{P}(\boldsymbol{X} = \boldsymbol{x}).$$

Um vetor aleatório $\boldsymbol{X}$ é discreto se, e somente se, suas coordenadas $X_1, \ldots, X_n$ são discretas.

A *função de probabilidade marginal* de uma componente X_k é obtida somando-se nas demais variáveis:

$$\begin{aligned} p_{X_k}(x_k) &= \mathbb{P}(X_k = x_k) \\ &= \sum_{x_1} \cdots \sum_{x_{k-1}} \sum_{x_{k+1}} \cdots \sum_{x_n} p_{\boldsymbol{X}}(x_1, \ldots, x_{k-1}, x_k, x_{k+1}, \ldots, x_n). \end{aligned}$$

Para a última igualdade acima basta observar que o evento $\{X_k = x_k\}$ pode ser escrito como a união disjunta de

$$\{X_k = x_k, \boldsymbol{X} \in A\} = \cup_{x_1} \cdots \cup_{x_{k-1}} \cup_{x_{k+1}} \cdots \cup_{x_n} \{(X_1, \ldots, X_n) = (x_1, \ldots, x_n)\}$$

com o evento $\{X_k = x_k, \boldsymbol{X} \notin A\}$, que tem probabilidade zero.

Exemplo 4.14. No Exemplo 4.4, vamos obter a função de probabilidade conjunta de $\boldsymbol{X}$, e as funções de probabilidade marginais de X_1 e X_2.

Primeiramente, pela descrição do vetor (X_1, X_2), sabemos que $p_{X_1,X_2}(0,2) = p_{X_1,X_2}(2,0) = \frac{1}{4}$, $p_{X_1,X_2}(1,1) = \frac{1}{2}$, e p_{X_1,X_2} vale zero fora desses três pontos. Somando em sobre x_2, encontramos a função de probabilidade marginal, que é dada por $p_{X_1}(0) = p_{X_1}(2) = \frac{1}{4}$ e $p_{X_1}(1) = \frac{1}{2}$.

Observe que os mesmos valores podem ser lidos a partir dos saltos da função de distribuição marginal obtida no Exemplo 4.8. △

Proposição 4.15 (Critério de independência, caso discreto)**.** *Seja* $\boldsymbol{X} =$

$(X_1, \ldots, X_n)$ um vetor aleatório discreto. Então $X_1, \ldots, X_n$ são independentes se, e somente se, $p_{\boldsymbol{X}}(\boldsymbol{x}) = p_{X_1}(x_1) \cdots p_{X_n}(x_n)$ para todo $\boldsymbol{x} \in \mathbb{R}^n$.

Demonstração. A implicação direta é trivial. Para a implicação inversa, suponha que $p_{\boldsymbol{X}}(\boldsymbol{x}) = p_{X_1}(x_1) \cdots p_{X_n}(x_n)$ para todo $\boldsymbol{x} \in \mathbb{R}^n$. Neste caso,

$$\begin{aligned}\mathbb{P}(X_1 \in B_1, \ldots, X_n \in B_n) &= \sum_{x_1 \in B_1} \cdots \sum_{x_n \in B_n} p_{\boldsymbol{X}}(\boldsymbol{x}) = \\ &= \left[\sum_{x_1 \in B_1} p_{X_1}(x_1)\right] \cdots \left[\sum_{x_n \in B_b} p_{X_n}(x_n)\right] = \mathbb{P}(X_1 \in B_1) \cdots \mathbb{P}(X_b \in B_n),\end{aligned}$$

para todos $B_1, \ldots, B_n \in \mathcal{B}$, e portanto as $X_1, \ldots, X_n$ são independentes. $\square$

Dados um vetor aleatório $\boldsymbol{X}$ e uma função mensurável[6] $f_{\boldsymbol{X}} : \mathbb{R}^n \to \mathbb{R}$, dizemos que $\boldsymbol{X}$ é *absolutamente contínuo com função de densidade conjunta* $f_{\boldsymbol{X}}$ se

$$\mathbb{P}(a_j < X_j \leqslant b_j,\ j = 1, \ldots, n) = \int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} f_{\boldsymbol{X}}(x_1, \ldots, x_n)\, \mathrm{d}x_n \cdots \mathrm{d}x_1$$

para todas as escolhas de $a_j < b_j$ para $j = 1, \ldots, n$.

A função de distribuição conjunta de um vetor aleatório absolutamente contínuo com densidade conjunta $f_{\boldsymbol{X}}$ pode ser calculada por:

$$F_{\boldsymbol{X}}(\boldsymbol{x}) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_{\boldsymbol{X}}(\boldsymbol{s})\, \mathrm{d}s_n \cdots \mathrm{d}s_1.$$

Se suspeitamos que um vetor aleatório possui densidade conjunta, podemos obter um candidato a densidade calculando

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{\partial^n}{\partial x_1 \cdots \partial x_n} F_{\boldsymbol{X}}(x_1, \ldots, x_n).$$

[6] Qualquer função que é contínua ou que pode ser aproximada por funções contínuas é uma função mensurável. De fato é muito difícil construir uma função que não seja mensurável. Veja a Seção 3.7 para mais detalhes.

Uma vez calculada $f_{\boldsymbol{X}}$ por essa fórmula, podemos verificar se é de fato uma densidade conjunta mostrando que sua integral coincide com $F_{\boldsymbol{X}}$.

A *função de densidade marginal* de uma componente X_k é obtida integrando-se nas demais variáveis:

$$f_{X_k}(x_k) = \underbrace{\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty}}_{n-1 \text{ vezes}} f_{\boldsymbol{X}}(x_1, \ldots, x_k, \ldots, x_n) \underbrace{\mathrm{d}x_1 \cdots \mathrm{d}x_n}_{\text{exceto } x_k}.$$

Vejamos que a densidade marginal f_{X_k} é uma densidade de X_k. Como $f_{\boldsymbol{X}}$ é não-negativa, segue que f_{X_k} será não-negativa para cada k. Além disso, para todo intervalo B, definindo a região $\boldsymbol{B} = \{\boldsymbol{x} \in \mathbb{R}^n : x_k \in B\}$, obtemos

$$\begin{aligned}
\mathbb{P}_{X_k}(B) = \mathbb{P}(X_k \in B) = \mathbb{P}(\boldsymbol{X} \in \boldsymbol{B}) &= \int_{\boldsymbol{B}} f_{\boldsymbol{X}}(\boldsymbol{x}) \, \mathrm{d}^n \boldsymbol{x} \\
&= \int_B \underbrace{\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty}}_{n-1 \text{ vezes}} f_{\boldsymbol{X}}(x_1, \ldots, x_k, \ldots, x_n) \underbrace{\mathrm{d}x_1 \cdots \mathrm{d}x_n}_{\text{exceto } x_k} \mathrm{d}x_k \\
&= \int_B f_{X_k}(x_k) \mathrm{d}x_k.
\end{aligned}$$

Logo, $f_{X_k}(x_k)$ é uma densidade de X_k.

Exemplo 4.16. Considere o quadrado $Q = \{(x, y) \in \mathbb{R}^2 : |x| + |y| < 1\}$, e suponha que X e Y tenham densidade conjunta dada por $f_{X,Y}(x, y) = \frac{1}{2}\mathbb{1}_Q(x, y)$. Podemos determinar a densidade marginal de X integrando:

$$f_X(x) = \int_{-\infty}^{+\infty} \tfrac{1}{2}\mathbb{1}_Q(x, y) \, \mathrm{d}y = (1 - |x|)\mathbb{1}_{[-1,1]}(x).$$

Analogamente, se integramos em x obtemos $f_Y(y) = (1 - |y|)\mathbb{1}_{[-1,1]}(y)$. $\triangle$

Proposição 4.17 (Critério de independência, caso contínuo)**.** *Seja* $\boldsymbol{X} = (X_1, \ldots, X_n)$ *um vetor aleatório cujas coordenadas* $X_1, \ldots, X_n$ *tenham função de densidade* $f_{X_1}, \ldots, f_{X_n}$. *Então* $X_1, \ldots, X_n$ *são independentes se, e somente se, a função dada por* $f_{\boldsymbol{X}}(\boldsymbol{x}) = f_{X_1}(x_1) \cdots f_{X_n}(x_n)$ *é uma densidade*

conjunta de $\boldsymbol{X}$.

Demonstração. Suponha que $f_{\boldsymbol{X}}(\boldsymbol{x}) = f_{X_1}(x_1)\cdots f_{X_n}(x_n)$ seja uma densidade conjunta de $\boldsymbol{X}$. Integrando,

$$F_{\boldsymbol{X}}(\boldsymbol{x}) = \int_{-\infty}^{x_1}\cdots\int_{-\infty}^{x_n} f_{X_1}(s_1)\cdots f_{X_n}(s_n)\,\mathrm{d}s_n\cdots\mathrm{d}s_1 = F_{X_1}(x_1)\cdots F_{X_n}(x_n)$$

e, pela Proposição 4.10, $X_1,\ldots,X_n$ são independentes.

Reciprocamente, suponha que $X_1,\ldots,X_n$ são independentes. Neste caso,

$$\begin{aligned} F_{\boldsymbol{X}}(\boldsymbol{x}) &= F_{X_1}(x_1)\cdots F_{X_n}(x_n) \\ &= \int_{-\infty}^{x_1} f_{X_1}(t_1)\,\mathrm{d}t\cdots\int_{-\infty}^{x_n} f_{X_n}(t_n)\,\mathrm{d}t \\ &= \int_{-\infty}^{x_1}\cdots\int_{-\infty}^{x_n} \left(f_{X_1}(t_1)\cdots f_{X_n}(t_n)\right)\mathrm{d}t_n\cdots\mathrm{d}t_1 \end{aligned}$$

e, portanto, $f_{X_1}(x_1)\cdots f_{X_n}(x_n)$ é uma densidade conjunta de $\boldsymbol{X}$. □

Observemos que a proposição acima *não* fornece um critério imediato para concluir que $X_1,\ldots,X_n$ *não* são independentes, pois densidades conjuntas não são únicas. Para isso, devemos verificar que existe toda uma região não-degenerada $(a_1,b_1]\times\cdots\times(a_n,b_n]$ onde $f_{\boldsymbol{X}}(\boldsymbol{x}) \neq f_{X_1}(x_1)\cdots f_{X_n}(x_n)$, para justificar que a Definição 4.9 não é satisfeita, ou mostrar a desigualdade em um ponto onde todas as funções envolvidas são contínuas.

Exemplo 4.18. No Exemplo 4.16, verifiquemos formalmente que X e Y não são independentes. Para isto, basta observar que $f_{X,Y}(x,y) = \frac{1}{2}\mathbb{1}_Q(x,y) = 0$ para todo $(x,y)\in(\frac{2}{3},\frac{3}{4}]\times(\frac{2}{3},\frac{3}{4}]$, enquanto $f_X(x)f_Y(y) = (1-x)(1-y) > 0$. △

Se um vetor aleatório $\boldsymbol{X}$ possui densidade conjunta, então cada uma das suas componentes $X_1,\ldots,X_n$ são variáveis aleatórias que também possuem densidade (a marginal), mas não vale a recíproca! Abaixo temos um exemplo simples de vetor aleatório contínuo que não tem densidade.

Contra-exemplo 4.19 (Inexistência de densidade conjunta). Seja $X \sim \mathcal{U}[0,1]$, $Y = 1 - X$. Apesar de X e Y serem absolutamente contínuas com densidade $f_X = f_Y = \mathbb{1}_{[0,1]}$, o vetor (X,Y) é singular. Mais precisamente, $\mathbb{P}((X,Y) = (x,y)) = 0$ para todo ponto $(x,y) \in \mathbb{R}^2$, mas (X,Y) não tem densidade conjunta. Para ver que ele não pode ter densidade, observe que (X,Y) assume valores em um segmento de reta, e a integral de qualquer função em um segmento de reta sempre dá zero. É interessante também analisar a função de distribuição conjunta, dada por

$$F_{X,Y}(x,y) = \begin{cases} 0, & x \leqslant 0 \text{ ou } y \leqslant 0 \text{ ou } x+y \leqslant 1, \\ y, & 0 \leqslant y \leqslant 1, x \geqslant 1, \\ x, & 0 \leqslant x \leqslant 1, y \geqslant 1, \\ x+y-1, & x \leqslant 1, y \leqslant 1, x+y \geqslant 1, \\ 1, & x \geqslant 1, y \geqslant 1. \end{cases}$$

Calculando a derivada cruzada, vemos que $\frac{\partial^2}{\partial y \partial x} F_{\boldsymbol{X}}(x,y) = 0$ para todo par (x,y) no plano $\mathbb{R}^2$, exceto em duas retas e um segmento de reta, onde a derivada cruzada não existe. $\triangle$

4.3 Soma de variáveis independentes

Trataremos nesta seção de um caso especial de função de vetor aleatório que permeará todo o livro. Somas de variáveis aleatórias independentes aparecem naturalmente em Probabilidade, Estatística e nas Ciências Naturais. Dadas X e Y variáveis aleatórias independentes, qual a distribuição de $X+Y$?

No caso em que o vetor (X,Y) é discreto, $X+Y$ é variável aleatória discreta e sua função de probabilidade p_{X+Y} pode ser calculada via Lei da Probabilidade Total:

$$p_{X+Y}(z) = \sum_x \mathbb{P}(X = x, Y = z - x) = \sum_x \mathbb{P}(X = x)\mathbb{P}(Y = z - x)$$

$$= \sum_x p_X(x) p_Y(z-x).$$

Observe que, como o vetor (X, Y) é discreto, as parcelas da soma acima são não-nulas para no máximo uma quantidade enumerável valores de x.

Exemplo 4.20. Sejam $X \sim \text{Binom}(n, p)$ e $Y \sim \text{Binom}(m, p)$. A função de probabilidade de $X + Y$ pode ser calculada por

$$\begin{aligned}
p_{X+Y}(k) &= \sum_{j=0}^{\infty} \mathbb{P}(X = j)\mathbb{P}(Y = k - j) \\
&= \sum_{j=0}^{k} \tbinom{n}{j} p^j (1-p)^{n-j} \tbinom{m}{k-j} p^{k-j} (1-p)^{m-k+j} \\
&= p^k (1-p)^{m+n-k} \sum_{j=0}^{k} \tbinom{n}{j} \tbinom{m}{k-j} \\
&= \tbinom{n+m}{k} p^k (1-p)^{m+n-k}.
\end{aligned}$$

Portanto, $X + Y \sim \text{Binom}(n + m, p)$, o que não chega a ser uma surpresa, dada a definição de variável binomial. Indutivamente, podemos generalizar este fato e mostrar que se $X_k \sim \text{Binom}(n_k, p)$ para todo $k = 1, \ldots, l$ são independentes, então $\sum_{k=1}^{l} X_k \sim \text{Binom}(\sum_{k=1}^{l} n_k, p)$. $\triangle$

Quando as variáveis X e Y forem independentes e absolutamente contínuas com densidades f_X e f_Y, respectivamente, temos a relação análoga

$$f_{X+Y}(z) = \int_{-\infty}^{+\infty} f_X(x) f_Y(z-x)\, \mathrm{d}x. \tag{4.21}$$

Para justificá-la, tomamos $B = \{(x, y) : x + y \leqslant z\}$ e desenvolvemos

$$\begin{aligned}
F_{X+Y}(z) &= \iint_B f_X(x) f_Y(y)\, \mathrm{d}x\mathrm{d}y \\
&= \int_{-\infty}^{+\infty} \int_{-\infty}^{z-x} f_X(x) f_Y(y)\, \mathrm{d}y\mathrm{d}x
\end{aligned}$$

$$
\begin{aligned}
&= \int_{-\infty}^{+\infty} \int_{-\infty}^{z} f_X(x) f_Y(u-x)\,\mathrm{d}u\mathrm{d}x \\
&= \int_{-\infty}^{z} \Big(\int_{-\infty}^{+\infty} f_X(x) f_Y(u-x)\,\mathrm{d}x \Big) \mathrm{d}u,
\end{aligned}
$$

para todo $z \in \mathbb{R}$, o que mostra que a função definida em (4.21) é de fato uma densidade de $X+Y$.

Soma de normais independentes

O caso quando X e Y são distribuições normais é tão importante que preferimos enunciá-lo como a proposição seguinte.

Proposição 4.22. *Sejam $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ e $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$ independentes. Então $X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.*

Demonstração. Como $X_1 - \mu_1 \sim \mathcal{N}(0, \sigma_1^2)$ e $X_2 - \mu_2 \sim \mathcal{N}(0, \sigma_2^2)$, podemos supor sem perda de generalidade que $\mu_1 = \mu_2 = 0$. A densidade de $X+Y$ é então dada por

$$
f_{X+Y}(z) = \frac{1}{2\pi\sigma_1\sigma_2} \int_{-\infty}^{+\infty} e^{-\frac{(z-x)^2}{2\sigma_2^2}}\, e^{-\frac{x^2}{2\sigma_1^2}}\,\mathrm{d}x.
$$

Para simplificar, escrevemos $a = 2\sigma_2^2$ e $b = 2\sigma_1^2$. Utilizando a identidade

$$
\frac{1}{a}(z-x)^2 + \frac{1}{b}x^2 = \Big(\frac{a+b}{ab}\Big)\Big(x - \frac{bz}{a+b}\Big)^2 + \frac{1}{a+b}z^2,
$$

válida para todos $a > 0$, $b > 0$, $z \in \mathbb{R}$ e $x \in \mathbb{R}$, obtemos

$$
\begin{aligned}
f_{X+Y}(z) &= \frac{1}{2\pi\sigma_1\sigma_2} \int_{-\infty}^{+\infty} e^{-\frac{(z-x)^2}{a}} e^{-\frac{x^2}{b}}\,\mathrm{d}x \\
&= \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{z^2}{a+b}} \int_{-\infty}^{+\infty} e^{-(\frac{a+b}{ab})(x-\frac{bz}{a+b})^2}\,\mathrm{d}x \\
&= \frac{1}{2\pi\sigma_1\sigma_2} \cdot \sqrt{\pi\tfrac{ab}{a+b}} \cdot e^{-\frac{z^2}{a+b}}
\end{aligned}
$$

$$= \frac{1}{\sqrt{2\pi(\sigma_1^2+\sigma_2^2)}} \cdot e^{-\frac{z^2}{2(\sigma_1^2+\sigma_2^2)}}.$$

Portanto f_{X+Y} corresponde à densidade de uma distribuição $\mathcal{N}(0, \sigma_1^2 + \sigma_2^2)$, como queríamos demonstrar. $\square$

Usando a Observação 4.13, pode-se verificar indutivamente que se $X_1, \ldots, X_n$ são variáveis aleatórias independentes com $X_k \sim \mathcal{N}(\mu_k, \sigma_k^2)$, então $\sum_{k=1}^n X_k \sim \mathcal{N}(\sum_{k=1}^n \mu_k, \sum_{k=1}^n \sigma_k^2)$. Em palavras, soma de normais independentes é uma variável normal cujos parâmetros são as somas dos respectivos parâmetros.

Exemplo 4.23. Uma fábrica produz parafusos cujos comprimentos são independentes e têm distribuição $\mathcal{N}(\mu, \sigma^2)$. É retirada uma amostra aleatória de n parafusos e calculada a média aritmética dos comprimentos da amostra. Qual a probabilidade da média observada diferir do valor médio teórico, μ, mais que uma tolerância δ? Sejam $X_1, \ldots, X_n$ os comprimentos dos parafusos da amostra e $S_n = \sum_{k=1}^n X_k$, estamos interessados em determinar quanto vale

$$\mathbb{P}(\left|\tfrac{S_n}{n} - \mu\right| < \delta).$$

Como $X_1, \ldots, X_n$ são independentes e com distribuição $\mathcal{N}(\mu, \sigma^2)$, segue que $S_n \sim \mathcal{N}(n\mu, n\sigma^2)$, ou seja, $\frac{S_n - n\mu}{\sigma\sqrt{n}}$ tem distribuição $\mathcal{N}(0, 1)$. Desse modo

$$\begin{aligned}\mathbb{P}(\left|\tfrac{S_n}{n} - \mu\right| < \delta) &= \mathbb{P}(\left|\tfrac{S_n - n\mu}{\sigma\sqrt{n}}\right| < \tfrac{\delta\sqrt{n}}{\sigma}) \\ &= \Phi(\tfrac{\delta\sqrt{n}}{\sigma}) - \Phi(-\tfrac{\delta\sqrt{n}}{\sigma}) = 2\Phi(\tfrac{\delta\sqrt{n}}{\sigma}) - 1.\end{aligned}$$

Por exemplo, se $n = 25$, $\delta = 0{,}200$ e $\sigma = 0{,}600$, podemos calcular o valor acima consultando a tabela na página 484 e obter

$$2\Phi(\tfrac{\delta\sqrt{n}}{\sigma}) - 1 \approx 2\Phi(1{,}67) - 1 \approx 0{,}904,$$

com três algarismos significativos.

Outra pergunta que poderíamos fazer é: qual o tamanho mínimo da amostra de modo que a média observada difira de μ por no máximo δ unidades com

probabilidade superior a $1-\alpha$ (nível de confiabilidade). Ou seja, devemos encontrar o menor z_α tal que

$$\mathbb{P}(|\mathcal{N}(0,1)| \leqslant z_\alpha) = 2\Phi(z_\alpha) - 1 \geqslant 1-\alpha.$$

Novamente com auxílio da tabela da distribuição normal, podemos encontrar z_α de modo que $\Phi(z_\alpha) = 1 - \frac{\alpha}{2}$ e posteriormente escolher n de modo que $\frac{\delta\sqrt{n}}{\sigma} \geqslant z_\alpha$. Por exemplo, se $\alpha = 0{,}060$, $\delta = 0{,}200$ e $\sigma = 0{,}600$, então $z_\alpha \approx 1{,}88$ e $n = \lceil z_\alpha^2 \sigma^2 \delta^{-2} \rceil = 32$. Ou seja, um aumento de 25 para 32 no tamanho da amostra aumenta de 0,904 para 0,940 nosso grau de confiança de a média observada diferir de μ menos que $\delta = 0{,}200$.

Há três variáveis que se relacionam neste problema: n, δ e α. Dadas quaisquer duas delas podemos determinar a terceira. Como exemplo do caso faltante, suponha que uma amostra de tamanho $n = 25$ é retirada, gostaríamos de afirmar que com probabilidade superior a 0,990 nossa média observada tem um grau de precisão δ. Qual o menor valor de δ? Como no caso anterior, para $\alpha = 0{,}010$, resolvendo $\Phi(z_\alpha) = 1 - \frac{\alpha}{2}$, obtemos $z_{0,\alpha} \approx 2{,}58$. Assim, $\delta = z\sigma n^{-1/2} \approx 0{,}310$. Ou seja, para a amostra de tamanho $n = 25$, a margem de erro aumentou de 0,200 para 0,310 se quisermos aumentar a confiança de 0,904 para 0,990. $\triangle$

4.4 Método do jacobiano

Seja $\boldsymbol{X}$ um vetor aleatório com densidade conjunta $f_{\boldsymbol{X}}$ e que assume valores em um domínio $G_0 \subseteq \mathbb{R}^n$. Suponha que estamos interessados em estudar o vetor aleatório $\boldsymbol{Y}$ dado por uma transformação $\boldsymbol{Y} = \boldsymbol{g}(\boldsymbol{X})$. Vamos considerar o caso em que $\boldsymbol{g} : G_0 \to G$, $G \subseteq \mathbb{R}^n$, é bijetiva e diferenciável, com inversa $\boldsymbol{g}^{-1} = \boldsymbol{h} : G \to G_0$ também diferenciável. Escrevemos a transformação

inversa como $\boldsymbol{X} = \boldsymbol{h}(\boldsymbol{Y})$ e definimos os jacobianos:

$$J_{\boldsymbol{h}}(\boldsymbol{y}) = \det\left(\frac{\partial \boldsymbol{x}}{\partial \boldsymbol{y}}\right) = \det\begin{bmatrix} \frac{\partial x_1}{\partial y_1} & \cdots & \frac{\partial x_1}{\partial y_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial x_n}{\partial y_1} & \cdots & \frac{\partial x_n}{\partial y_n} \end{bmatrix}$$

e

$$J_{\boldsymbol{g}}(\boldsymbol{x}) = \det\left(\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}}\right) = \det\begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_n}{\partial x_1} & \cdots & \frac{\partial y_n}{\partial x_n} \end{bmatrix}.$$

Do Cálculo em várias variáveis, sabemos que o jacobiano satisfaz à seguinte identidade:

$$J_{\boldsymbol{h}}(\boldsymbol{y}) = \frac{1}{J_{\boldsymbol{g}}(\boldsymbol{x})}.$$

Proposição 4.24 (Método do jacobiano)**.** *Sejam $G_0, G \subseteq \mathbb{R}^n$, $\boldsymbol{g} : G_0 \to G$ uma bijeção e $\boldsymbol{h} = \boldsymbol{g}^{-1}$, e suponha que $\boldsymbol{g}$ e $\boldsymbol{h}$ sejam diferenciáveis. Se $\boldsymbol{X}$ é um vetor aleatório absolutamente contínuo com densidade $f_{\boldsymbol{X}}$ assumindo valores em G_0, e $\boldsymbol{Y} = \boldsymbol{g}(\boldsymbol{X})$, então uma densidade $f_{\boldsymbol{Y}}$ pode ser obtida a partir de $f_{\boldsymbol{X}}$ pela relação*

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = \left|J_{\boldsymbol{h}}(\boldsymbol{y})\right| \cdot f_{\boldsymbol{X}}\big(\boldsymbol{h}(\boldsymbol{y})\big) = \frac{1}{\left|J_{\boldsymbol{g}}(\boldsymbol{h}(\boldsymbol{y}))\right|}\, f_{\boldsymbol{X}}\big(\boldsymbol{h}(\boldsymbol{y})\big).$$

Ideia da prova. Do cálculo de várias variáveis, sabemos que se o jacobiano for não-nulo para todo $y \in G$, vale a fórmula de mudança de variáveis

$$\int_A f(\boldsymbol{x})\,\mathrm{d}^n\boldsymbol{x} = \int_{g(A)} f(\boldsymbol{h}(\boldsymbol{y}))\,\left|J_{\boldsymbol{h}}(\boldsymbol{y})\right|\,\mathrm{d}^n\boldsymbol{y}$$

para qualquer f integrável em A, onde $A \subseteq G_0$. Como $\mathbb{P}(\boldsymbol{Y} \in \boldsymbol{g}(A))$ é dada por $\mathbb{P}(\boldsymbol{X} \in A)$, e esta última é dada pelo lado esquerdo da expressão acima com $f = f_{\boldsymbol{X}}$, o integrando do lado direito é uma densidade de $\boldsymbol{Y}$. $\square$

Exemplo 4.25. Considere o vetor aleatório $\boldsymbol{X} = (X_1, X_2)$ com densidade

$$f_{\boldsymbol{X}}(x_1, x_2) = \begin{cases} 4x_1x_2, & x_1, x_2 \in [0,1], \\ 0, & \text{caso contrário}, \end{cases}$$

e o vetor $\boldsymbol{Y}$ dado por $Y_1 = X_1/X_2$ e $Y_2 = X_1X_2$. Escrevendo $\boldsymbol{y} = \boldsymbol{g}(\boldsymbol{x}) = (x_1/x_2, x_1x_2)$,

$$\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}} = \begin{bmatrix} 1/x_2 & -x_1/x_2^2 \\ x_2 & x_1 \end{bmatrix}$$

e $J_{\boldsymbol{g}}(\boldsymbol{x}) = 2x_1/x_2$. Obtendo $\boldsymbol{x}$ em função de $\boldsymbol{y}$, temos que $x_1 = \sqrt{y_1y_2}$, $x_2 = \sqrt{y_2/y_1}$ e os valores possíveis de $\boldsymbol{y}$ são

$$G = \left\{(y_1, y_2) : 0 < y_2 < y_1, 0 < y_2 < \tfrac{1}{y_1}\right\}.$$

Agora,

$$J_{\boldsymbol{g}}(\boldsymbol{h}(\boldsymbol{y})) = \frac{2\sqrt{y_1y_2}}{\sqrt{y_2/y_1}} = 2y_1$$

e

$$f_{\boldsymbol{X}}(\boldsymbol{h}(\boldsymbol{y})) = 4\sqrt{y_1y_2}\sqrt{y_2/y_1} = 4y_2.$$

Portanto,

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = \frac{1}{|J_{\boldsymbol{g}}(\boldsymbol{x})|}\, f_{\boldsymbol{X}}(\boldsymbol{h}(\boldsymbol{y})) = \begin{cases} 2y_2/y_1, & 0 < y_2 < 1, y_2 < y_1 < 1/y_2, \\ 0, & \text{caso contrário}. \end{cases} \qquad \triangle$$

Exemplo 4.26. Se X e Y são independentes e distribuídas como $\mathcal{N}(0,1)$, então $X+Y$ e $X-Y$ são independentes e ambas distribuídas como $\mathcal{N}(0,2)$.

Com efeito, definindo $\boldsymbol{Z} = (X, Y)$ e $\boldsymbol{W} = (X+Y, X-Y)$, temos que $\boldsymbol{W} = \boldsymbol{g}(\boldsymbol{Z})$, onde $\boldsymbol{g}(x,y) = (x+y, x-y)$. Logo,

$$\frac{\partial \boldsymbol{w}}{\partial \boldsymbol{z}} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix},$$

donde $J_{\boldsymbol{g}}(\boldsymbol{z}) = -2$. Escrevendo $\boldsymbol{z}$ como função de $\boldsymbol{w}$, obtemos $x = \frac{w_1+w_2}{2}$ e $y = \frac{w_1-w_2}{2}$. Ainda,

$$f_{\boldsymbol{Z}}(\boldsymbol{z}) = f_{X,Y}(x,y) = f_X(x)f_Y(y) = \frac{1}{\sqrt{2\pi}}\, e^{\frac{-x^2}{2}} \cdot \frac{1}{\sqrt{2\pi}}\, e^{\frac{-y^2}{2}},$$

logo

$$\begin{aligned} f_{\boldsymbol{Z}}(\boldsymbol{h}(\boldsymbol{w})) =& \frac{1}{2\pi}\, e^{\frac{-\left(\frac{w_1+w_2}{2}\right)^2}{2}}\, e^{\frac{-\left(\frac{w_1-w_2}{2}\right)^2}{2}} \\ =& \frac{e^{-\frac{w_1^2+w_2^2+2w_1w_2+w_1^2+w_2^2-2w_1w_2}{8}}}{2\pi} = \frac{1}{2\pi}\, e^{\frac{-w_1^2}{4}}\, e^{\frac{-w_2^2}{4}} \end{aligned}$$

e, substituindo,

$$f_{\boldsymbol{W}}(\boldsymbol{w}) = \frac{1}{|J_{\boldsymbol{g}}(\boldsymbol{h}(\boldsymbol{w}))|}\, f_{\boldsymbol{Z}}(\boldsymbol{h}(\boldsymbol{w})) = \frac{1}{4\pi}\, e^{\frac{-w_1^2}{4}}\, e^{\frac{-w_2^2}{4}} = f_{\mathcal{N}(0,2)}(w_1) f_{\mathcal{N}(0,2)}(w_2).$$

Portanto, W_1 e W_2 são independentes e com distribuição $\mathcal{N}(0,2)$. $\triangle$

4.5 Sequência de variáveis independentes

No capítulo anterior, mostramos a existência de uma distribuição uniforme (Teorema 3.16) e que a partir desta distribuição uniforme podemos mostrar a existência de uma variável aleatória com qualquer função de distribuição, F, pré-especificada (Teorema 3.17).

Aqui daremos um passo adiante nessa direção. Mostraremos que a partir de uma variável com distribuição uniforme, podemos construir uma sequência de variáveis independentes e cada uma com uma distribuição especificada, que pode ou não ser a mesma para cada variável desta sequência.

Lema 4.27. *O conjunto* $\mathbb{N} \times \mathbb{N}$ *é enumerável.*

Demonstração. A função $f : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ dada por $f(j,k) = 2^{j-1} \cdot (2k-1)$

é uma bijeção pois todo número natural pode ser fatorado de forma única como uma potência de dois vezes um número ímpar. □

Dado $x \in [0,1)$, consideremos a sequência de dígitos $(x_n)_n$ de sua expansão binária $x = \sum_{n=1}^{\infty} \frac{x_n}{2^n}$, com $x_n \in \{0,1\}$ para todo $n \in \mathbb{N}$. Caso haja duas expansões distintas, uma terminada com infinitos zeros e outra com infinitos uns, consideremos apenas a primeira.

Lema 4.28. *Seja U uma variável aleatória com distribuição uniforme em $[0,1)$ e defina $(X_n)_n$, tomando valores em $\{0,1\}$, a sequência de dígitos na expansão binária de U. Então as variáveis $(X_n)_n$ são i.i.d. com distribuição uniforme no conjunto $\{0,1\}$.*

Demonstração. Com efeito, basta observar que $X_1 = x_1, \ldots, X_n = x_n$ se, e somente se, $0 \leqslant U - (\frac{x_1}{2} + \frac{x_2}{2^2} + \cdots + \frac{x_n}{2^n}) < 2^{-n}$, donde $\mathbb{P}(X_1 = x_1, \ldots, X_n = x_n) = \frac{1}{2^n}$, para todos $n \in \mathbb{N}$ e $x_1, \ldots, x_n \in \{0,1\}$. Logo, a independência segue da Proposição 4.15. □

O próximo lema é uma espécie de recíproca do lema anterior.

Lema 4.29. *Seja $(X_n)_n$ uma sequência de variáveis aleatórias i.i.d. com distribuição uniforme no conjunto $\{0,1\}$ e defina $U = \sum_{n=1}^{\infty} \frac{X_n}{2^n}$. Então U é uma variável aleatória com distribuição uniforme no intervalo $[0,1)$.*

Demonstração. Que U é uma variável aleatória segue direto do Lema 3.50. Sejam $n \in \mathbb{N}$ e $k \in \{0, \ldots, 2^n - 1\}$. Escreva $k = \sum_{j=1}^{n} 2^{n-j} x_j$ com $x_j \in \{0,1\}$ para todo j, isto é, $(x_1, \ldots, x_n)$ é a expansão binária de k. Observe que

$$\mathbb{P}(\tfrac{k}{2^n} \leqslant U < \tfrac{k+1}{2^n}) = \mathbb{P}((X_1, \ldots, X_n) = (x_1, \ldots, x_n)) = \tfrac{1}{2^n},$$

pois os eventos acima diferem apenas nos casos $\{X_j = 0 \text{ para todo } j > n\}$ ou $\{X_j = 1 \text{ para todo } j > n\}$, que têm probabilidade zero. Somando sobre $k = 0, \ldots, m-1$, obtemos

$$\mathbb{P}(U < \tfrac{m}{2^n}) = \tfrac{m}{2^n}$$

para todo $m = 0, \ldots, 2^n$. Finalmente, dado $u \in [0,1)$ e $n \in \mathbb{N}$, existe m tal que $\frac{m}{2^n} \leqslant u < \frac{m+1}{2^n}$, de forma que

$$u - \tfrac{1}{2^n} \leqslant \tfrac{m}{2^n} = \mathbb{P}(U < \tfrac{m}{2^n}) \leqslant \mathbb{P}(U \leqslant u) \leqslant \mathbb{P}(U < \tfrac{m+1}{2^n}) = \tfrac{m+1}{2^n} \leqslant u + \tfrac{1}{2^n}.$$

Como isso vale para todo n, temos $\mathbb{P}(U \leqslant u) = u$, portanto $U \sim \mathcal{U}[0,1]$. □

Agora, estamos aptos a provar o principal objetivo desta seção.

Teorema 4.30 (Existência de uma sequência independente)**.** *Seja $(F_n)_n$ uma sequência de funções de distribuição. Então, existem um espaço de probabilidade e uma sequência de variáveis aleatórias independentes, $(X_n)_n$, de modo que X_n tenha função de distribuição F_n para todo $n \in \mathbb{N}$.*

Demonstração. Pelo Teorema 3.16, existe um espaço de probabilidade onde podemos definir uma variável aleatória U com distribuição uniforme em $[0,1)$. Definimos $(Y_n)_n$ como sendo a sequência de dígitos na expansão binária de U. Pelo Lema 4.28, as variáveis $(Y_n)_n$ são i.i.d. com distribuição uniforme no conjunto $\{0,1\}$. Seja $f : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ uma bijeção, o que é garantido pelo Lema 4.27. Definimos a sequência de sequências $((Z_{j,k})_{k\in\mathbb{N}})_{j\in\mathbb{N}}$ por $Z_{j,k} = Y_{f(j,k)}$. Observe que $((Z_{j,k})_{k\in\mathbb{N}})_{j\in\mathbb{N}}$ continua sendo formada por variáveis i.i.d. com distribuição uniforme no conjunto $\{0,1\}$.

Para cada $j \in \mathbb{N}$ fixo, usamos a sequência $(Z_{j,k})_k$ para definir $U_j = \sum_{k=1}^{\infty} 2^{-k} Z_{j,k}$. Pelo Lema 4.29, U_j tem distribuição uniforme em $[0,1)$. Além disso, como cada variável da sequência $(U_j)_j$ depende de subconjuntos disjuntos de variáveis da sequência $(Y_n)_n$, concluímos que as variáveis $(U_j)_j$ são independentes,[7] logo, i.i.d. com distribuição uniforme em $[0,1)$.

Finalmente, tomamos $X_j = F_j^{-1}(U_j)$ para todo j. Como a família $(U_j)_j$ é independente, a família $(X_j)_j$ também será independente. Ademais, $X_j \sim F_j$ para todo j, veja a demonstração do Teorema 3.17. Isso conclui a prova. □

[7] Esta conclusão bastante intuitiva se deve a uma generalização da Observação 4.13, que será justificada de forma mais rigorosa pelo Corolário 13.12.

4.6 Exercícios

§4.1

1. Seja (X, Y) vetor aleatório tal que $(X, Y) \sim (Y, X)$. Mostre que $\mathbb{P}(X \geqslant Y) = \frac{1+\mathbb{P}(X=Y)}{2}$.

2. Sejam $X \sim \text{Binom}(2, \frac{1}{2})$ e $Y \sim \mathcal{U}[0, 2]$ variáveis aleatórias independentes. Esboce o gráfico da função de distribuição de $\min(X, Y)$.

3. Seja X uma variável aleatória qualquer. Mostre que X é independente de si mesma se, e somente se, existe $c \in \mathbb{R}$ tal que $\mathbb{P}(X = c) = 1$.

4. Sejam $X_1, \ldots, X_n$ variáveis aleatórias independentes, com $X_k \sim \text{Exp}(\lambda)$ para todo $k = 1, \ldots n$. Determine a função de distribuição de $Y = \min\{X_1, \ldots, X_n\}$. Trata-se de alguma distribuição conhecida? Se sim, qual?

5. Sejam Y e U duas variáveis aleatórias independentes e com leis $Y \sim \mathcal{N}(0, 1)$ e $\mathbb{P}(U = -1) = \mathbb{P}(U = +1) = \frac{1}{2}$. Ache a distribuição de $Z = UY$.

6. Um ponto P é escolhido com distribuição uniforme no círculo unitário $C = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 < 1\}$. Informalmente, isso quer dizer que a probabilidade de (X, Y) tomar valor em algum conjunto $A \subseteq C$ é dada pela razão entre as áreas de A e de C.

(a) Sejam (X, Y) as coordenadas cartesianas de P. As variáveis X e Y são independentes? Justifique.

(b) Sejam (R, θ) as coordenadas polares de P. Determine as funções de distribuição marginal F_R e F_θ bem como a conjunta $F_{R,\theta}$.

Dica: Utilize apenas argumentos geométricos.

(c) As variáveis R e θ são independentes? Justifique.

7. Dentre as funções abaixo, quais poderiam ser funções de distribuição conjunta de algum vetor aleatório?

(a)

$$F(x, y) = \begin{cases} 1 - \frac{1}{xy}, & x, y \geqslant 1, \\ 0, & x < 1 \text{ ou } y < 1. \end{cases}$$

(b)

$$F(x,y) = \begin{cases} 1 - e^{-x} - e^{-y} + e^{-x-y}, & x, y \geqslant 0, \\ 0, & x < 0 \text{ ou } y < 0. \end{cases}$$

(c)

$$F(x,y) = \begin{cases} 1 - e^{-x-y}, & x, y \geqslant 0, \\ 0, & x < 0 \text{ ou } y < 0. \end{cases}$$

§4.2

8. Suponha que $p_{X,Y}(0,1) = p_{X,Y}(1,0) = p_{X,Y}(1,2) = p_{X,Y}(2,1) = \frac{1}{4}$.

(a) Encontre as distribuições marginais de X e Y.
(b) Determine se X e Y são independentes.

9. Sejam X e Y variáveis aleatórias independentes, discretas e com distribuições Poisson(λ_1) e Poisson(λ_2), respectivamente. Mostre que, dada a ocorrência do evento $\{X + Y = n\}$, a probabilidade condicional de $X = k$ é

$$\mathbb{P}(X = k | X + Y = n) = \binom{n}{k} \left(\frac{\lambda_1}{\lambda_1 + \lambda_2}\right)^k \left(\frac{\lambda_2}{\lambda_1 + \lambda_2}\right)^{n-k}.$$

Interprete essa identidade.

10. Sejam $n \in \mathbb{N}$ e (X, Y) com função de probabilidade conjunta

$$p(j,k) = \begin{cases} c, & \text{se } 1 \leqslant j = k \leqslant n \text{ ou } 1 \leqslant j = n + 1 - k \leqslant n, \\ 0, & \text{caso contrário.} \end{cases}$$

(a) Determine o valor da constante c.
(b) Determine as distribuições marginais de X e Y.
(c) X e Y são independentes?

11. Seja (X, Y, Z) um vetor aleatório com função de densidade conjunta dada

por

$$f(x,y,z) = \begin{cases} cxy^2z, & \text{se } 0 < x \leqslant 1,\ 0 < y \leqslant 1 \text{ e } 0 < z \leqslant \sqrt{2}, \\ 0, & \text{caso contrário.} \end{cases}$$

Encontre o valor de c e a função de densidade marginal de X.

12. Considere um vetor aleatório (Z, W) absolutamente contínuo com densidade

$$f_{Z,W}(z,w) = \begin{cases} c, & 0 < z < 1, 0 < w < z, \\ 0, & \text{caso contrário.} \end{cases}$$

Encontre c e $F_{Z,W}$.

13. Uma densidade conjunta de X e Y é dada por

$$f(x,y) = \begin{cases} \frac{ce^{-y}}{x^3}, & x > 1, y > 0 \\ 0, & \text{caso contrário.} \end{cases}$$

Encontre a constante c. Diga se X e Y são independentes e por quê.

14. Seja (X, Y) um vetor aleatório absolutamente contínuo com função de distribuição conjunta dada por $F_{X,Y}(x,y) = 0$ se $x < 0$ ou $y < 0$ e $F_{X,Y}(x,y) = 1 - e^{-x} + e^{-x-y} - xe^{-x} - e^{-y} + xe^{-x-y}$ caso contrário.

(a) Encontre uma densidade conjunta e diga se X e Y são independentes.
(b) Encontre a distribuição marginal F_Y.

15. Seja $\boldsymbol{X} = (X_1, X_2)$ um vetor com densidade conjunta dada por

$$f_{\boldsymbol{X}}(x,y) = \frac{\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]\right\}}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}.$$

Tal vetor é dito ter *distribuição normal bivariada* com parâmetros $\mu_1, \mu_2 \in \mathbb{R}$, $\sigma_1, \sigma_2 > 0$ e $\rho \in [0, 1)$. Mostre que:

(a) $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ e $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$.

(b) As variáveis X_1 e X_2 são independentes se, e somente se, $\rho = 0$.

16. Sejam $Z_1, Z_2, \ldots, Z_n$ variáveis aleatórias i.i.d. assumindo valores no conjunto $\{z_1, \ldots, z_r\}$ tais que $\mathbb{P}(Z_1 = z_j) = p_j$ para todo $j = 1, \ldots, r$ (onde $p_1 + \cdots + p_r = 1$). Defina o vetor aleatório $\boldsymbol{X} = (X_1, \ldots, X_r)$, onde para cada $k = 1, \ldots, r$, $X_k = \#\{j : Z_j = k\}$. Determine a distribuição do vetor $\boldsymbol{X}$. Dizemos que o vetor $\boldsymbol{X}$ tem *distribuição multinomial de parâmeros* $n, p_1, \ldots, p_r$.

§4.3

17. Sejam $X \sim \text{Poisson}(\lambda)$ e $Y \sim \text{Poisson}(\mu)$ independentes. Qual a distribuição de $X + Y$?

18. Sejam $X_1, X_2, \ldots, X_k$ variáveis aleatórias i.i.d. com distribuição $\text{Geom}(p)$. Qual a distribuição de $X_1 + \cdots + X_k$? É alguma distribuição conhecida?

19. Sejam X e Y variáveis aleatórias independentes ambas com distribuição geométrica de parâmetro p. Determine a função de probabilidade de $X - Y$.

20. Sejam $X_1, \ldots, X_n$ variáveis aleatórias independentes, com $X_k \sim \mathcal{U}[-k, +k]$ para todo $k = 1, \ldots n$. Calcule $\mathbb{P}(X_1 + \cdots + X_n \geqslant 0)$.

21. Sejam $X_1, \ldots, X_n$ variáveis independentes e identicamente distribuídas com distribuição $\text{Exp}(\lambda)$, mostre que $X_1 + \cdots + X_n \sim \text{Gama}(n, \lambda)$.

22. Uma caixa contém 10 parafusos, cujos tamanhos são normais independentes, com $\mu = 21{,}40\,\text{mm}$ e $\sigma = 0{,}50\,\text{mm}$. Calcule a probabilidade de que nenhum dos parafusos tenha mais de $22{,}0\,\text{mm}$.

23. Um elevador pode suportar uma carga de 10 pessoas ou um peso total de $750\,\text{kg}$. Assumindo que apenas homens tomam o elevador e que seus pesos são normalmente distribuídos com $\mu = 74{,}0\,\text{kg}$ e $\sigma = 2{,}00\,\text{kg}$, qual a probabilidade de que o peso limite seja excedido para um grupo de 10 homens escolhidos aleatoriamente?

24. A distribuição dos comprimentos dos elos da corrente de bicicleta é normal, com parâmetros $\mu = 2{,}00\,\text{cm}$ e $\sigma^2 = 0{,}01\,\text{cm}^2$. Para que uma corrente se ajuste à bicicleta, deve ter comprimento total entre 58,00 e 61,00 cm. Qual é a probabilidade de uma corrente com 30 elos não se ajustar à bicicleta?

25. O peso de uma determinada fruta é uma variável aleatória com distribuição normal $\mu = 200{,}0\,\text{g}$ e $\sigma = 50{,}0\,\text{g}$. Determine a probabilidade de um lote contendo 100 unidades dessa fruta pesar mais que 21,00 kg e diga com quantos algarismos significativos é possível determinar essa probabilidade.

§4.4

26. Sejam X e Y variáveis aleatórias independentes, ambas com distribuição Exp(1).

(a) Use o método do jacobiano para determinar a distribuição conjunta de $X + Y$ e $\frac{X}{X+Y}$.
(b) Diga se $X + Y$ e $\frac{X}{X+Y}$ são independentes.
(c) Encontre a distribuição de $\frac{X}{X+Y}$.

27. Sejam X e Y independentes com densidades f_X e f_Y. Usando o método do jacobiano, dê uma prova alternativa para a fórmula

$$f_{X+Y}(t) = \int_{-\infty}^{\infty} f_X(t-s) f_Y(s)\,\mathrm{d}s \qquad \forall t \in \mathbb{R}.$$

Sugestão: Encontre a densidade conjunta de $X + Y$ e Y.

28. Se X e Y são independentes e distribuídas como $\mathcal{N}(0,1)$, mostre que $4X + 3Y$ e $3X - 4Y$ são independentes e ambas distribuídas como $\mathcal{N}(0,25)$.

29. Sejam X e Y independentes com distribuição $\mathcal{U}[0,1]$, e defina $W = X+Y$ e $Z = X - Y$. Calcule a densidade conjunta de W e Z, e determine se elas são independentes.

Capítulo 5

Esperança Matemática

A esperança $\mathbb{E}X$ de uma variável aleatória X é a média dos valores assumidos por X, ponderada pela probabilidade de X assumir esses valores. Podemos pensar em $\mathbb{E}X$ como sendo o "centro de massa" de X. A esperança de X é, em vários sentidos, a melhor aproximação determinística para a variável aleatória X.

Uma das justificativas mais importantes, que veremos mais adiante, é a lei dos grandes números: se $X_1, \ldots, X_n$ são independentes e têm a mesma distribuição, então a média observada $\frac{1}{n}\sum_{j=1}^{n} X_j$ se aproxima de $\mathbb{E}X_1$ quando tomamos n grande.

Começaremos definindo a esperança e estudando suas propriedades básicas, o que cumpriremos nas duas primeiras seções. Depois veremos propriedades de convergência da esperança, que serão usadas em algumas passagens específicas dos próximos capítulos, e a definição de esperança condicional dado um evento, que será usada no capítulo de esperança condicional.

Na última seção, que constitui a metade deste capítulo, vamos construir a integral de Lebesgue e estudar algumas das suas propriedades mais importantes. Essa seção será usada nos tópicos mais avançados deste livro.

5.1 Variáveis aleatórias simples

Uma variável aleatória X é dita *simples* se assume apenas finitos valores.

Definição 5.1. Dada uma variável aleatória simples X, definimos a *esperança de X*, ou *média de X*, denotada por $\mathbb{E}X$, por

$$\mathbb{E}X = \sum_x x \cdot \mathbb{P}(X = x).$$

Note que na soma acima há apenas finitos termos não-nulos.[8]

A esperança de X pode ser pensada como o "centro de massa" da variável aleatória X, como ilustrado na Figura 5.1.

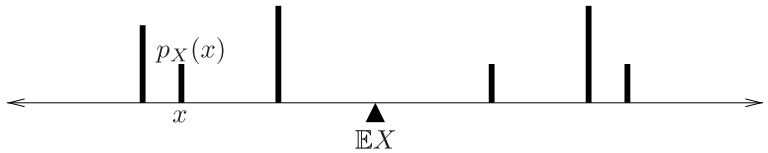


Figura 5.1. A esperança de X como o centro de massa de p_X.

Outra interpretação de $\mathbb{E}X$ é dada pelo valor médio após muitas realizações de um mesmo experimento. Sejam $a_1, \ldots, a_k$ os possíveis valores recebidos em uma rodada de um jogo e X o resultado efetivamente ganho após a rodada. Suponhamos também que jogaremos esse jogo n vezes, e denotamos o resultado de cada jogada por $X_1, \ldots, X_n$, independentes e com a mesma distribuição de X. A noção intuitiva de probabilidade como frequência relativa diz que a proporção dentre essas n repetições em que o resultado é a_r se aproxima de $\mathbb{P}(X = a_r)$ para n grande, ou seja,

$$\frac{1}{n} \sum_{j=1}^{n} \mathbb{1}_{\{X_j = a_r\}} \approx \mathbb{P}(X = a_r).$$

[8] Sempre que aparecer "$\sum_x$", a soma é realizada sobre todos os valores de x para os quais o termo da soma é não-nulo.

Dessa forma, para o ganho total dividido pelo número de jogadas, obtemos

$$\frac{1}{n}\sum_{j=1}^{n} X_j = \frac{1}{n}\sum_{j=1}^{n}\sum_{r=1}^{k} a_r \mathbb{1}_{\{X_j=a_r\}} =$$
$$= \sum_{r=1}^{k} a_r \left(\frac{1}{n}\sum_{j=1}^{n} \mathbb{1}_{\{X_j=a_r\}}\right) \approx \sum_{r=1}^{k} a_r \cdot \mathbb{P}(X=a_r) = \mathbb{E}X.$$

Vejamos alguns exemplos simples.

Exemplo 5.2. Ao lançar um dado, seja X o valor observado em sua face superior. Neste caso,

$$\mathbb{E}X = 1 \times \mathbb{P}(X=1) + \cdots + 6 \times \mathbb{P}(X=6) = \frac{1}{6} + \frac{2}{6} + \cdots + \frac{6}{6} = \frac{21}{6} = \frac{7}{2}. \quad \triangle$$

Exemplo 5.3. Lançamos uma moeda 4 vezes, seja X a variável aleatória que conta quantas vezes saem cara. Neste caso,

$$\mathbb{E}X = 0 \times \frac{1}{16} + 1 \times \frac{4}{16} + 2 \times \frac{6}{16} + 3 \times \frac{4}{16} + 4 \times \frac{1}{16} = \frac{32}{16} = 2. \quad \triangle$$

Exemplo 5.4. Seja X dada por $X = \mathbb{1}_A$ para algum $A \in \mathcal{F}$. Neste caso, $\mathbb{E}X = 0 \times \mathbb{P}(A^c) + 1 \times \mathbb{P}(A) = \mathbb{P}(A)$. Ou seja,

$$\mathbb{E}[\mathbb{1}_A] = \mathbb{P}(A).$$

Reciprocamente, se $X \sim \text{Bernoulli}(p)$ então podemos considerar o evento A dado por $A = \{\omega : X(\omega) = 1\}$, de forma que $X = \mathbb{1}_A$ e $\mathbb{E}X = p$. $\triangle$

Exemplo 5.5. Ao lançar um dado duas vezes, seja X a soma dos valores observados. Neste caso,

$$\mathbb{E}X = 2 \times \frac{1}{36} + 3 \times \frac{2}{36} + 4 \times \frac{3}{36} + 5 \times \frac{4}{36} + 6 \times \frac{5}{36} + 7 \times \frac{6}{36} +$$
$$+ 8 \times \frac{5}{36} + 9 \times \frac{4}{36} + 10 \times \frac{3}{36} + 11 \times \frac{2}{36} + 12 \times \frac{1}{36} = 7. \quad \triangle$$

Exemplo 5.6. Retiramos 3 cartas de um baralho comum, seja X o número de reis retirados. Neste caso,

$$\begin{aligned}\mathbb{E}X = 0 \times \frac{48 \cdot 47 \cdot 46}{52 \cdot 51 \cdot 50} + 1 \times \frac{3 \cdot 48 \cdot 47 \cdot 4}{52 \cdot 51 \cdot 50} + \\ + 2 \times \frac{3 \cdot 48 \cdot 4 \cdot 3}{52 \cdot 51 \cdot 50} + 3 \times \frac{4 \cdot 3 \cdot 2}{52 \cdot 51 \cdot 50} = \frac{3}{13}.\end{aligned}$$ △

Exemplo 5.7. Se $X \sim \text{Binom}(n, p)$, então

$$\begin{aligned}\mathbb{E}X &= \sum_{k=0}^{n} k\tbinom{n}{k}p^k(1-p)^{n-k} = \sum_{k=1}^{n} n\tbinom{n-1}{k-1}p^k(1-p)^{n-k} \\ &= n\sum_{j=0}^{n-1}\tbinom{n-1}{j}p^{j+1}(1-p)^{n-j-1} = np\sum_{j=0}^{n-1}\tbinom{n-1}{j}p^j(1-p)^{n-1-j} \\ &= np[p+(1-p)]^{n-1} = np.\end{aligned}$$ △

Exemplo 5.8. Seja X a variável que denota o produto dos valores observados ao lançarmos um dado duas vezes.

$$\begin{aligned}\mathbb{E}X = \frac{1}{36}\big(&1 \cdot 1 + 2 \cdot 2 + 3 \cdot 2 + 4 \cdot 3 + 5 \cdot 2 + 6 \cdot 4 + \\ &+ 8 \cdot 2 + 9 \cdot 1 + 10 \cdot 2 + 12 \cdot 4 + 15 \cdot 2 + 16 \cdot 1 + \\ &+ 18 \cdot 2 + 20 \cdot 2 + 24 \cdot 2 + 25 \cdot 1 + 30 \cdot 2 + 36 \cdot 1\big) = \frac{49}{4}.\end{aligned}$$ △

Uma propriedade fundamental da esperança é a seguinte.

Teorema 5.9 (Linearidade)**.** *Sejam X e Y variáveis aleatórias simples e $a, b \in \mathbb{R}$. Então $\mathbb{E}[aX + bY] = a\,\mathbb{E}X + b\,\mathbb{E}Y$.*

Antes da demonstração, vejamos alguns exemplos do uso da linearidade.

Exemplo 5.10. No Exemplo 5.3, podemos escrever $X = X_1 + X_2 + X_3 + X_4$, onde X_k denota a função indicadora de que saiu cara no k-ésimo lançamento. Logo, $\mathbb{E}X = \mathbb{E}X_1 + \mathbb{E}X_2 + \mathbb{E}X_3 + \mathbb{E}X_4 = 4 \cdot \frac{1}{2} = 2$. △

Exemplo 5.11. No Exemplo 5.5, observamos que $X = Y + Z$, onde Y e Z

representam os valores observados no primeiro e no segundo lançamentos do dado. Logo,

$$\mathbb{E}X = \mathbb{E}Y + \mathbb{E}Z = \frac{7}{2} + \frac{7}{2} = 7. \qquad \triangle$$

Exemplo 5.12. No Exemplo 5.6, observamos que $X = X_1 + X_2 + X_3$, onde X_k é a indicadora de que a k-ésima carta retirada é rei. Ao contrário dos exemplos anteriores, aqui X_1, X_2 e X_3 não são independentes. Ainda assim, cada uma individualmente satisfaz $\mathbb{E}X_k = \frac{1}{13}$, e podemos calcular

$$\mathbb{E}X = \mathbb{E}X_1 + \mathbb{E}X_2 + \mathbb{E}X_3 = \frac{3}{13}. \qquad \triangle$$

Exemplo 5.13. No Exemplo 5.7, observamos que X tem a mesma distribuição de $X_1 + \cdots + X_n$, com X_k i.i.d. Bernoulli(p), e portanto

$$\mathbb{E}X = \mathbb{E}X_1 + \cdots + \mathbb{E}X_n = np. \qquad \triangle$$

Nos exemplos anteriores, é mais curto calcular a esperança usando linearidade do que usando a distribuição de X diretamente. Existem muitos outros casos em que descrever a distribuição de X é muito difícil ou até mesmo impossível, mas ainda assim é possível calcular a esperança usando linearidade.

Exemplo 5.14. Uma gaveta contém 10 pares de meias, todos diferentes. Abre-se a gaveta no escuro e retiram-se 6 meias. Qual é a esperança do número de pares X formados pelas meias retiradas? Se numeramos as meias retiradas, e contamos a quantidade N de meias cujo par também foi retirado, estaremos contando cada par duas vezes, portanto $N = 2X$. A probabilidade de que o par da primeira meia retirada também seja retirado é $\frac{5}{19}$. Somando sobre as 6 meias, obtemos $\mathbb{E}[2X] = \mathbb{E}N = 6 \times \frac{5}{19}$, e portanto $\mathbb{E}X = \frac{15}{19}$. $\triangle$

Uma prova curta do Teorema 5.9 utiliza a seguinte propriedade.

Teorema 5.15. *Sejam X e Y variáveis aleatórias simples. Então,*

$$\mathbb{E}[g(X,Y)] = \sum_x \sum_y g(x,y) \cdot \mathbb{P}(X = x, Y = y)$$

para qualquer função $g : \mathbb{R}^2 \to \mathbb{R}$. *Em particular,*

$$\mathbb{E}[g(X)] = \sum_x g(x) \cdot \mathbb{P}(X = x)$$

para qualquer função $g : \mathbb{R} \to \mathbb{R}$.

Demonstração. Particionando Ω segundo os valores de X e Y,

$$\begin{aligned}
\mathbb{E}[g(X,Y)] &= \sum_z z \cdot \mathbb{P}\big(g(X,Y) = z\big) \\
&= \sum_z z \cdot \sum_x \sum_y \mathbb{1}_{g(x,y)=z} \cdot \mathbb{P}\big(X = x, Y = y\big) \\
&= \sum_x \sum_y \Big(\sum_z z \cdot \mathbb{1}_{g(x,y)=z} \Big) \cdot \mathbb{P}\big(X = x, Y = y\big) \\
&= \sum_x \sum_y g(x,y) \cdot \mathbb{P}\big(X = x, Y = y\big).
\end{aligned}$$

A segunda parte é um caso particular tomando-se Y constante. □

Exemplo 5.16. Se $X \sim \text{Binom}(n,p)$, então

$$\begin{aligned}
\mathbb{E}[2^X + X^2 - X] &= \sum_{k=0}^{n} (2^k + k(k-1)) \cdot \tbinom{n}{k} p^k (1-p)^{n-k} \\
&= \sum_{k=0}^{n} \Big(\tbinom{n}{k} (2p)^k (1-p)^{n-k} + n(k-1) \tbinom{n-1}{k-1} p^k (1-p)^{n-k} \Big) \\
&= (2p + 1 - p)^n + \sum_{k=0}^{n} n(n-1) \tbinom{n-2}{k-2} p^k (1-p)^{n-k} \\
&= (1+p)^n + n(n-1)p^2 \sum_{j=0}^{n-2} \tbinom{n-2}{j} p^j (1-p)^{n-2-j} \\
&= (1+p)^n + (np)^2 - np^2.
\end{aligned}$$

△

Demonstração do Teorema 5.9. Usando o Teorema 5.15 três vezes,

$$\begin{aligned}\mathbb{E}[aX+bY] &= \sum_x \sum_y (ax+by)\cdot \mathbb{P}(X=x, Y=y)\\ &= a\sum_x \sum_y x\cdot \mathbb{P}(X=x,Y=y) + b\sum_x\sum_y y\cdot \mathbb{P}(X=x,Y=y)\\ &= a\,\mathbb{E}X + b\,\mathbb{E}Y,\end{aligned}$$

o que conclui a prova. □

Teorema 5.17. *Se X e Y são simples e independentes, então*

$$\mathbb{E}[XY] = \mathbb{E}X\cdot\mathbb{E}Y.$$

Demonstração. Usando o Teorema 5.15 e a independência,

$$\begin{aligned}\mathbb{E}[XY] &= \sum_x\sum_y xy\cdot\mathbb{P}(X=x,Y=y)\\ &= \sum_x\sum_y xy\cdot\mathbb{P}(X=x)\mathbb{P}(Y=y)\\ &= \Big(\sum_x x\cdot\mathbb{P}(X=x)\Big)\Big(\sum_y y\cdot\mathbb{P}(Y=y)\Big) = \mathbb{E}X\cdot\mathbb{E}Y,\end{aligned}$$

o que conclui a prova. □

Exemplo 5.18. No Exemplo 5.8, observemos que $X = YZ$, onde Y e Z denotam o primeiro e segundo valores observados nos lançamentos do dado. Logo,

$$\mathbb{E}X = \mathbb{E}Y\cdot\mathbb{E}Z = \frac{7}{2}\cdot\frac{7}{2} = \frac{49}{4}.$$

Repare como o cálculo foi simplificado. △

5.2 Esperança matemática

Nesta seção vamos definir a esperança de uma variável aleatória qualquer, e estudar suas principais propriedades. Vimos na seção anterior que $\mathbb{E}X$ representa a média dos valores que X pode tomar, ponderada por sua probabilidade. A generalização dessa mesma ideia seria, por analogia,

$$\mathbb{E}X = \sum_x x \cdot p_X(x)$$

no caso de X ser discreta e

$$\mathbb{E}X = \int_{-\infty}^{+\infty} x \cdot f_X(x)\,\mathrm{d}x$$

no caso de X ter densidade.

As fórmulas acima são a forma correta de pensar na esperança como valor médio, mas elas se restringem ao caso de variáveis discretas ou que tenham densidade. Além disso, dificilmente permitiriam comparar uma variável discreta com uma contínua, e não cobrem o caso de variáveis que não são nem de um tipo nem de outro (como será feito na prova da Desigualdade de Tchebyshev). Em particular, uma das propriedades fundamentais da esperança, a linearidade, não pode ser demonstrada a partir das fórmulas acima (de fato, a soma de duas variáveis com densidade pode não ser de

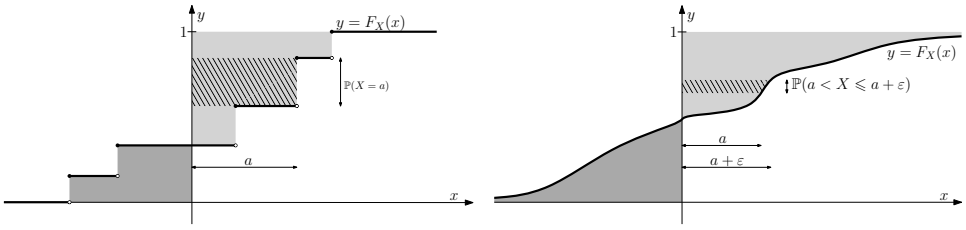


Figura 5.2. No lado esquerdo, a esperança de uma variável aleatória simples, vista como uma área determinada pela função de distribuição. No lado direito, a generalização do mesmo conceito para uma variável aleatória qualquer.

nenhum desses dois tipos). Por isso, daremos uma definição unificada de $\mathbb{E}X$, que não usa f_X nem p_X.

Para motivar a definição, reparemos que a esperança de uma variável aleatória simples é dada pela diferença entre as áreas cinza-clara e cinza-escura no lado esquerdo da Figura 5.2. A esperança de uma variável aleatória qualquer é definida por analogia, como mostrado no lado direito da mesma figura.

Definição 5.19. Definimos a *esperança* de uma variável aleatória X por

$$\mathbb{E}X \;=\; \int_0^{+\infty} \mathbb{P}(X > x)\,\mathrm{d}x \;-\; \int_{-\infty}^{0} \mathbb{P}(X < x)\,\mathrm{d}x$$

se pelo menos uma das duas integrais impróprias acima for finita. Caso contrário, dizemos que $\mathbb{E}X$ *não está definida*. Quando ambas as integrais forem finitas, dizemos que a variável aleatória X é *integrável*. Observe que essas integrais impróprias de Riemann sempre estão definidas pois o integrando é uma função monótona e não-negativa (Teorema A.4).

Pela definição acima, X é integrável quando $\mathbb{E}X$ é um número real. Entretanto, essa definição permite que $\mathbb{E}X$ assuma os valores $-\infty$ ou $+\infty$ quando apenas uma das integrais acima for finita, e neste caso X não será integrável.

Exemplo 5.20 (Exponencial). Se $X \sim \mathrm{Exp}(\lambda)$, então

$$\mathbb{E}X = \int_0^{+\infty} \mathbb{P}(X > x)\,\mathrm{d}x = \int_0^{+\infty} e^{-\lambda x}\mathrm{d}x = \frac{1}{\lambda}.$$

Ou seja, X é integrável e sua esperança é $\frac{1}{\lambda}$. △

Exemplo 5.21 (Normal). Seja $X \sim \mathcal{N}(0,1)$. Então

$$\begin{aligned}\int_0^{+\infty} \mathbb{P}(X > x)\,\mathrm{d}x &= \int_0^{2} \mathbb{P}(X > x)\,\mathrm{d}x + \int_2^{+\infty}\int_x^{+\infty} f_X(t)\,\mathrm{d}t\,\mathrm{d}x \\ &\leqslant 2 + \int_2^{+\infty}\int_x^{+\infty} f_X(t)\,\mathrm{d}t\,\mathrm{d}x\end{aligned}$$

$$\leqslant 2 + \int_2^{+\infty} \int_x^{+\infty} e^{-t} \, \mathrm{d}t \, \mathrm{d}x = 2 + e^{-2} < \infty.$$

Qualquer que seja o valor dessa integral, é um número finito e, por simetria, será igual a $\int_{-\infty}^0 \mathbb{P}(X < x) \, \mathrm{d}x$. Portanto, X é integrável e $\mathbb{E}X = 0$. △

Exemplo 5.22. Seja X com densidade $f_X(x) = x^{-2} \mathbb{1}_{(-\infty,-2]\cup[2,+\infty)}(x)$. Este exemplo tem simetria como o anterior, porém não podemos chegar à mesma conclusão pois as integrais não são finitas. Com efeito,

$$\int_0^{+\infty} \mathbb{P}(X > x) \, \mathrm{d}x \geqslant \int_2^{+\infty} \int_x^{+\infty} f_X(t) \, \mathrm{d}t \, \mathrm{d}x = \int_2^{+\infty} x^{-1} \, \mathrm{d}x = +\infty.$$

Uma estimativa idêntica mostra que $\int_{-\infty}^0 \mathbb{P}(X < x) \, \mathrm{d}x = +\infty$. Assim sendo, a esperança de X não está definida, apesar da simetria. △

Exemplo 5.23. Seja X com densidade $f_X(x) = x^{-2} \mathbb{1}_{(-\infty,-2]}(x) + x^{-3} \mathbb{1}_{[1,+\infty)}(x)$. Como no exemplo acima, para a parte negativa temos $\int_{-\infty}^0 \mathbb{P}(X < x) \, \mathrm{d}x = +\infty$. Por outro lado, para a parte positiva,

$$\int_0^{+\infty} \mathbb{P}(X > x) \, \mathrm{d}x = \frac{1}{2} + \int_1^{+\infty} \int_x^{+\infty} f_X(t) \, \mathrm{d}t \, \mathrm{d}x = \frac{1}{2} + \int_1^{+\infty} \frac{x^{-2}}{2} \, \mathrm{d}x = 1$$

e, portanto, $\mathbb{E}X = 1 - (+\infty) = -\infty$. Neste exemplo, X não é integrável. △

Observação 5.24. A esperança de uma variável aleatória depende apenas da sua distribuição. Mais precisamente, se X e Y têm a mesma distribuição e $\mathbb{E}X$ está definida, então $\mathbb{E}Y = \mathbb{E}X$. Em particular, se $\mathbb{P}(X = Y) = 1$ e $\mathbb{E}X$ está definida, então $\mathbb{E}Y = \mathbb{E}X$. △

Observação 5.25. Uma variável aleatória X é integrável se, e somente se, $\mathbb{E}|X| < \infty$, pois

$$\mathbb{E}|X| = \int_0^{+\infty} \mathbb{P}(|X| > x) \, \mathrm{d}x = \int_0^{+\infty} \mathbb{P}(X > x) \, \mathrm{d}x \; + \int_{-\infty}^0 \mathbb{P}(X < x) \, \mathrm{d}x.$$

A esperança de uma variável aleatória não-negativa sempre está definida, podendo assumir um valor finito ou $+\infty$. △

A seguir veremos as principais propriedades da esperança, fórmulas para calcular $\mathbb{E}X$ no caso de X ser discreta ou absolutamente contínua, bem como a esperança de funções de variáveis aleatórias desses tipos.

Teorema 5.26 (Variáveis aleatórias discretas)**.** *Se X é discreta, então*

$$\mathbb{E}X = \sum_{x<0} x \cdot p_X(x) + \sum_{x>0} x \cdot p_X(x)$$

caso uma das duas séries convirja. Caso contrário, $\mathbb{E}X$ não está definida.

Demonstração. Já sabemos que a identidade vale se X assume finitos valores. Podemos mostrar que também vale para X discreta aproximando-a por variáveis que assumem finitos valores. Uma prova completa será dada na Seção 5.3. □

Exemplo 5.27 (Poisson). Se $X \sim \text{Poisson}(\lambda)$, então

$$\mathbb{E}X = \sum_{n=0}^{\infty} n \frac{\lambda^n e^{-\lambda}}{n!} = \sum_{n=1}^{\infty} \frac{\lambda^n e^{-\lambda}}{(n-1)!} =$$
$$= \lambda e^{-\lambda} \sum_{n=1}^{\infty} \frac{\lambda^{n-1}}{(n-1)!} = \lambda e^{-\lambda} \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} = \lambda e^{-\lambda} e^{\lambda} = \lambda.$$

Portanto, a esperança de uma variável aleatória Poisson com parâmetro λ é o próprio λ. △

Vejamos agora como calcular a esperança de uma variável com densidade.

Teorema 5.28 (Variáveis aleatórias absolutamente contínuas)**.** *Seja X uma variável aleatória absolutamente contínua com densidade f_X. Então*

$$\mathbb{E}X = \int_0^{+\infty} x \cdot f_X(x)\,\mathrm{d}x \;+\; \int_{-\infty}^{0} x \cdot f_X(x)\,\mathrm{d}x$$

caso uma das duas integrais seja finita. Caso contrário, $\mathbb{E}X$ não está definida.

Demonstração. A prova completa será dada na Seção 5.5.5, mas aqui damos dois argumentos que ajudam a entender de onde vem a fórmula. O primeiro supõe que podemos inverter as integrais:

$$\begin{aligned}\int_0^{+\infty} \mathbb{P}(X > s)\,\mathrm{d}s &= \int_0^{+\infty}\int_s^{+\infty} f_X(x)\,\mathrm{d}x\,\mathrm{d}s \\ &= \int_0^{+\infty}\int_0^{x} f_X(x)\,\mathrm{d}s\,\mathrm{d}x = \int_0^{+\infty} x\cdot f_X(x)\,\mathrm{d}x.\end{aligned}$$

Vejamos uma prova mais gráfica supondo que f_X é contínua e vale zero fora de $[0, M]$ para algum M. Integrando por partes, obtemos

$$\begin{aligned}\int_0^{+\infty} \mathbb{P}(X > x)\,\mathrm{d}x &= \int_0^{+\infty} [1 - F_X(x)]\,\mathrm{d}x = \int_0^{+\infty} x\cdot F_X'(x)\,\mathrm{d}x \\ &= \int_0^{+\infty} x\cdot f_X(x)\,\mathrm{d}x.\end{aligned}$$

A integral $\int_{-\infty}^0 \mathbb{P}(X < x)\,\mathrm{d}x$ é tratada de forma análoga. □

Exemplo 5.29 (Uniforme). Se $X \sim \mathcal{U}[a, b]$, então

$$\mathbb{E}X = \int_a^b x\,\tfrac{1}{b-a}\,\mathrm{d}x = \frac{a+b}{2}.$$

Isto é, a esperança de uma variável aleatória com distribuição uniforme em um intervalo é o ponto médio deste intervalo. △

Teorema 5.30. *Sejam X e Y variáveis aleatórias e $a, b \in \mathbb{R}$. Então valem as seguintes propriedades:*

(1) *Unitariedade: se $X = \mathbb{1}_A$ para algum $A \in \mathcal{F}$, então $\mathbb{E}X = \mathbb{P}(A)$.*
(2) *Monotonicidade: se $0 \leqslant X \leqslant Y$ para todo ω, então $0 \leqslant \mathbb{E}X \leqslant \mathbb{E}Y$.*
(3) *Linearidade: vale $\mathbb{E}[aX + bY] = a\,\mathbb{E}X + b\,\mathbb{E}Y$, supondo que Y seja integrável e $\mathbb{E}X$ esteja definida.*

Demonstração. Para provar (1), observamos que se $X = \mathbb{1}_A$, então X é uma variável aleatória simples, donde $\mathbb{E}X = 1\cdot\mathbb{P}(X = 1) = \mathbb{P}(A)$. Para (2), basta

desenvolver

$$\mathbb{E}X = \int_0^{+\infty} \mathbb{P}(X > x)\,\mathrm{d}x \leqslant \int_0^{+\infty} \mathbb{P}(Y > x)\,\mathrm{d}x = \mathbb{E}Y.$$

Já sabemos que (3) vale se X e Y são variáveis aleatórias simples, podemos mostrar que continua válido para X e Y quaisquer aproximando-as por variáveis desse tipo. Uma prova completa será dada na Seção 5.3. □

Diversas propriedades da esperança decorrem do teorema acima.

Proposição 5.31. *Dada uma variável aleatória X e $a, b, c \in \mathbb{R}$, valem:*

(1) *Se $X = c$ para todo ω, então $\mathbb{E}X = c$.*
(2) *Se $a \leqslant X \leqslant b$ para todo ω, então $a \leqslant \mathbb{E}X \leqslant b$.*
(3) *Se $X \geqslant 0$ para todo ω, então $\mathbb{E}X \geqslant 0$.*
(4) *Se $X \geqslant 0$ para todo ω e $\mathbb{E}X = 0$, então $\mathbb{P}(X = 0) = 1$.*

Demonstração. Vamos mostrar apenas a última, deixando as demais como exercício. Para todo $\varepsilon > 0$, temos $X \geqslant \varepsilon \cdot \mathbb{1}_{\{X \geqslant \varepsilon\}}$. Por monotonicidade, $\mathbb{E}X \geqslant \mathbb{E}\big[\varepsilon \mathbb{1}_{\{X \geqslant \varepsilon\}}\big] = \varepsilon \mathbb{P}(X \geqslant \varepsilon)$. Logo, $\mathbb{P}(X \geqslant \varepsilon) = 0$ para todo $\varepsilon > 0$, donde $\mathbb{P}(X > 0) = \lim_k \mathbb{P}(X \geqslant k^{-1}) = 0$, pois $\{X \geqslant k^{-1}\} \uparrow \{X > 0\}$. □

Há situações em que somente podemos afirmar que algo ocorre com probabilidade 1, mesmo que não ocorra para todo ω. Um exemplo é o item (4) da proposição acima. Daqui para frente, diremos que um evento A ocorre *quase certamente*, abreviado por q.c., quando $\mathbb{P}(A) = 1$. Usando a Observação 5.24, podemos obter uma versão mais forte da proposição acima, trocando "para todo ω" por "quase certamente" em cada item.

Proposição 5.32 (Esperança de variáveis aleatórias independentes)**.** *Se X e Y são independentes e não-negativas, então*

$$\mathbb{E}[XY] = \mathbb{E}X \cdot \mathbb{E}Y.$$

A mesma identidade vale supondo X e Y independentes e integráveis.

Demonstração. Já sabemos que a identidade vale se X e Y assumem finitos valores. Podemos mostrar o caso geral aproximando-as por variáveis desse tipo. Uma prova completa será dada na Seção 5.3. □

Proposição 5.33. *Se X assume valores em $\mathbb{N}_0$, então*

$$\mathbb{E}X = \sum_{n=1}^{\infty} \mathbb{P}(X \geqslant n).$$

Demonstração. Basta ver que

$$\int_0^n \mathbb{P}(X > x)\,\mathrm{d}x = \sum_{k=1}^{n} \int_{k-1}^{k} \mathbb{P}(X > x)\,\mathrm{d}x = \sum_{k=1}^{n} \mathbb{P}(X \geqslant k)$$

e tomar o limite com $n \to \infty$. □

Exemplo 5.34 (Geométrica). Se $X \sim \mathrm{Geom}(p)$ então

$$\mathbb{E}X = \sum_{n=1}^{\infty} \mathbb{P}(X \geqslant n) = \sum_{n=1}^{\infty} (1-p)^{n-1} = \sum_{j=0}^{\infty} (1-p)^j = \frac{1}{1-(1-p)} = \frac{1}{p}. \quad \triangle$$

Proposição 5.35 (Critério de integrabilidade)**.** *Uma variável aleatória X é integrável se, e somente se, $\sum_{n=1}^{\infty} \mathbb{P}(|X| \geqslant n) < \infty$.*

Demonstração. Sabemos que X é integrável se, e somente se, $\mathbb{E}|X| < \infty$. Seja $Y = \lfloor |X| \rfloor$. Como $0 \leqslant Y \leqslant |X| \leqslant Y+1$, segue que $\mathbb{E}Y \leqslant \mathbb{E}|X| \leqslant 1+\mathbb{E}Y$. E como Y assume valores em $\mathbb{N}_0$, segue da Proposição 5.33 que

$$\sum_{n=1}^{\infty} \mathbb{P}(Y \geqslant n) \leqslant \mathbb{E}|X| \leqslant 1 + \sum_{n=1}^{\infty} \mathbb{P}(Y \geqslant n).$$

Observando que $\mathbb{P}(Y \geqslant n) = \mathbb{P}(|X| \geqslant n)$, concluímos a demonstração. □

Suponha que queiramos calcular a esperança de uma variável aleatória descrita explicitamente como uma função de outra. Mais precisamente, suponha

que $Y = g(X)$, para alguma função g. Certamente, uma forma de obter $\mathbb{E}Y$ é calcular $F_Y(y)$ para todo $y \in \mathbb{R}$, a partir da distribuição de X, e depois calcular a esperança usando os Teoremas 5.26 e 5.28 ou a própria Definição 5.19, aplicados a Y. Entretanto, existe outra maneira, que pode ser mais conveniente.

Teorema 5.36 (Funções de variáveis aleatórias discretas)**.** *Sejam X uma variável aleatória discreta e $g : \mathbb{R} \to \mathbb{R}$ uma função não-negativa. Então*

$$\mathbb{E}[g(X)] = \sum_x g(x) \cdot p_X(x).$$

Se X toma apenas finitos valores, a identidade acima é a segunda parte do Teorema 5.15. O caso geral será tratado na Seção 5.3.

Exemplo 5.37. Se $X \sim \text{Poisson}(\lambda)$, então

$$\begin{aligned}\mathbb{E}X^2 &= \sum_{n=0}^{\infty} n^2 \frac{\lambda^n e^{-\lambda}}{n!} = \sum_{n=1}^{\infty} n \frac{\lambda^n e^{-\lambda}}{(n-1)!} = \sum_{n=1}^{\infty} \frac{\lambda^n e^{-\lambda}}{(n-1)!} + \sum_{n=2}^{\infty} \frac{\lambda^n e^{-\lambda}}{(n-2)!} \\ &= \lambda e^{-\lambda} \sum_{n=1}^{\infty} \frac{\lambda^{n-1}}{(n-1)!} + \lambda^2 e^{-\lambda} \sum_{n=2}^{\infty} \frac{\lambda^{n-2}}{(n-2)!} = \lambda + \lambda^2.\end{aligned}$$

△

Exemplo 5.38. Seja $X \sim \text{Geom}(p)$. Vamos calcular

$$\mathbb{E}X^2 = \sum_{n=1}^{\infty} n^2 (1-p)^{n-1} p$$

derivando séries de potência. Para isso, escrevemos $q = 1 - p$ e

$$\begin{aligned}\mathbb{E}X^2 &= p \sum_{n=1}^{\infty} n q^{n-1} + pq \sum_{n=1}^{\infty} n(n-1) q^{n-2} \\ &= p \frac{\mathrm{d}}{\mathrm{d}q} \sum_{n=0}^{\infty} q^n + pq \frac{\mathrm{d}^2}{\mathrm{d}q^2} \sum_{n=0}^{\infty} q^n \\ &= p \frac{1}{(1-q)^2} + pq \frac{2}{(1-q)^3}\end{aligned}$$

$$= \frac{1}{p} + 2 \cdot \frac{1-p}{p^2} = \frac{2-p}{p^2}.$$

Admitimos a propriedade de que séries de potência podem ser derivadas termo a termo, dentro do seu raio de convergência, neste caso $|q| < 1$. $\triangle$

Teorema 5.39 (Funções de variáveis absolutamente contínuas)**.** *Seja X uma variável aleatória absolutamente contínua com densidade f_X. Seja g uma função não-negativa e contínua por partes. Então*

$$\mathbb{E}[g(X)] = \int_{-\infty}^{+\infty} g(x)\, f_X(x)\, \mathrm{d}x.$$

Observe que, pelo Teorema A.5, a função $g \cdot f_X$ é integrável.

Demonstração. Provaremos aqui supondo que g é uma função-degrau que assume finitos valores. A prova do caso geral será dada na Seção 5.5.5. Sejam $y_1, \ldots, y_n$ os valores assumidos por g, e $A_k = \{x \in \mathbb{R} : g(x) = y_k\}$. Assim, podemos escrever $g(x) = \sum_k y_k \mathbb{1}_{A_k}(x)$, logo $Y = \sum_{k=1}^n y_k \mathbb{1}_{\{X \in A_k\}}$. Portanto,

$$\mathbb{E}Y = \sum_{k=1}^n y_k \mathbb{P}(X \in A_k) = \sum_{k=1}^n \int_{A_k} y_k f_X(x)\, \mathrm{d}x = \int_{-\infty}^{+\infty} g(x)\, f_X(x)\, \mathrm{d}x,$$

pois $g(x) = y_k$ para todo $x \in A_k$. $\square$

Exemplo 5.40. Se $X \sim \mathrm{Exp}(\lambda)$, então

$$\mathbb{E}X^2 = \int_0^{+\infty} x^2 \lambda e^{-\lambda x}\, \mathrm{d}x = \frac{2}{\lambda^2}. \qquad \triangle$$

Exemplo 5.41. Se $X \sim \mathcal{N}(0,1)$, então

$$\mathbb{E}X^2 = \int_{-\infty}^{+\infty} x^2 \frac{e^{-x^2/2}}{\sqrt{2\pi}} \mathrm{d}x = 1,$$

onde usamos o fato de que $\int_{-\infty}^{+\infty} e^{-x^2/2}\mathrm{d}x = \sqrt{2\pi}$. Ademais,

$$\mathbb{E}|X| = \int_{-\infty}^{+\infty} |x| \frac{e^{-x^2/2}}{\sqrt{2\pi}} \mathrm{d}x = \sqrt{\frac{2}{\pi}}. \qquad \triangle$$

Observação 5.42. O teorema anterior tem seu análogo multidimensional. Se um vetor aleatório $\boldsymbol{X} = (X_1, \ldots, X_n)$ possui densidade conjunta $f_{\boldsymbol{X}}$ e se $g : \mathbb{R}^n \to \mathbb{R}$ é uma função não-negativa e contínua, então

$$\mathbb{E}[g(\boldsymbol{X})] = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} g(x_1, \ldots, x_n)\, f_{\boldsymbol{X}}(x_1, \ldots, x_n)\, \mathrm{d}x_1 \cdots \mathrm{d}x_n. \tag{5.43}$$

Esta fórmula será usada em alguns exemplos e exercícios e será provada no final deste capítulo. $\triangle$

Os Teoremas 5.36 e 5.39 valem supondo que $\mathbb{E}[g(X)]$ está definida, ao invés de supor g não-negativa. Para justificar essa extensão, basta separar os pontos onde g é positiva e negativa. A *parte positiva* e a *parte negativa* de um número $z \in [-\infty, +\infty]$ são denotadas por

$$z^+ = \begin{cases} z, & z \geqslant 0, \\ 0, & z \leqslant 0, \end{cases} \qquad z^- = \begin{cases} 0, & z \geqslant 0, \\ -z, & z \leqslant 0. \end{cases}$$

Com essa definição, $z = z^+ - z^-$ e $|z| = z^+ + z^-$. Da mesma forma, podemos definir $g^+(x) = [g(x)]^+$ e $g^-(x) = [g(x)]^-$ de forma que $g = g^+ - g^-$. Observe que $\mathbb{E}[g^+(X)]$ e $\mathbb{E}[g^-(X)]$ sempre estão definidas. Ademais,

$$\mathbb{E}[g(X)] = \mathbb{E}[g^+(X)] - \mathbb{E}[g^-(X)]$$

se uma das duas for finita, e $\mathbb{E}[g(X)]$ não está definida caso contrário. Em geral, não sabemos de antemão se uma variável aleatória é integrável; assim sendo, a abordagem mais simples é de fato separar as partes positiva e negativa e usar esses dois teoremas tal como foram enunciados.

Existem outras formas de se definir a esperança, todas elas equivalentes.

Isso também se reflete em distintas notações, que aparecem em diferentes bibliografias:

$$\mathbb{E}X = \int_\Omega X \, \mathrm{d}\mathbb{P}, \qquad \mathbb{E}X = \int_\mathbb{R} x \, \mathrm{d}\mathbb{P}_X, \qquad \mathbb{E}X = \int_\mathbb{R} x \, \mathrm{d}F_X(x).$$

A definição que usamos aqui corresponde à primeira, que se refere à integral de Lebesgue no espaço de probabilidade Ω visto como um espaço de medida. A segunda integral também é uma integral de Lebesgue, porém no espaço de probabilidade $(\mathbb{R}, \mathcal{B}, \mathbb{P}_X)$ onde $\mathbb{P}_X$ é a lei de X. A terceira integral é conhecida como integral de Stieltjes com função integradora dada por F_X.

A fórmula que usamos na Definição 5.19 é um atalho pouco utilizado porém mais curto e acessível para definir a integral de Lebesgue da função X com respeito à medida $\mathbb{P}$, sem usar esse nome. Essa forma é equivalente à definição mais comum de integral de Lebesgue a partir de funções simples, que será vista na Seção 5.5. A terceira fórmula é equivalente à segunda, porém definida de uma forma mais explícita no caso de integrandos contínuos, o que torna alguns cálculos mais transparentes. A igualdade entre a primeira e a segunda fórmulas acima é assegurada pelo Teorema 5.65.

5.3 Aproximação e convergência da esperança

Em inúmeras situações estaremos interessados no limite de $\mathbb{E}X_n$ para uma sequência $X_1, X_2, X_3, \ldots$ de variáveis aleatórias. Mais precisamente, gostaríamos de poder comutar a esperança com um limite em n, e por isso queremos estabelecer condições sob as quais $\mathbb{E}[\lim_n X_n] = \lim_n \mathbb{E}X_n$. Vale lembrar que derivadas e integrais também são limites.

Para ler o livro em um nível mais básico, esta seção pode ser omitida. Para a leitura em um nível intermediário, as ferramentas desta seção serão usadas principalmente nas Seções 7.4, 8.3 e 10.3. A leitura em nível mais avançado requer o conteúdo da Seção 5.5, que inclui essas ferramentas.

5.3.1 Teorema da Convergência Monótona

Começamos pelo Teorema da Convergência Monótona, que tem menos pré-requisitos técnicos, e o usaremos para demonstrar várias propriedades da esperança enunciadas na seção anterior, usando aproximação por variáveis aleatórias simples e separando as partes positiva e negativa. Este teorema também será usado na Seção 8.3.

Teorema 5.44 (Teorema da Convergência Monótona)**.** *Sejam X e $(X_n)_n$ variáveis aleatórias não-negativas, tais que, para todo ω, $(X_n(\omega))_n$ é não-decrescente e converge para $X(\omega)$. Então $\mathbb{E}X_n \to \mathbb{E}X$ quando $n \to \infty$.*

Demonstração. Como $0 \leqslant X_n \leqslant X_{n+1} \leqslant X$, temos que $\lim_n \mathbb{E}X_n$ existe e $\lim_n \mathbb{E}X_n \leqslant \mathbb{E}X$, restando mostrar a desigualdade oposta. Seja $M < \mathbb{E}X$ qualquer, e tome K tal que $\int_0^K \mathbb{P}(X > x)\,\mathrm{d}x > M$. Seja $N \in \mathbb{N}$. Escrevemos

$$\begin{aligned}\mathbb{E}X_n &= \int_0^{+\infty} \mathbb{P}(X_n > x)\,\mathrm{d}x \geqslant \int_0^K \mathbb{P}(X_n > x)\,\mathrm{d}x = \\ &= \sum_{j=1}^N \int_{(j-1)\frac{K}{N}}^{j\frac{K}{N}} \mathbb{P}(X_n > x)\,\mathrm{d}x \geqslant \sum_{j=1}^N \tfrac{K}{N}\mathbb{P}(X_n > j\tfrac{K}{N}) \to \sum_{j=1}^N \tfrac{K}{N}\mathbb{P}(X > j\tfrac{K}{N}),\end{aligned}$$

pois $\{X_n > a\} \uparrow \{X > a\}$. De forma análoga,

$$M \leqslant \int_0^K \mathbb{P}(X > x)\,\mathrm{d}x \leqslant \sum_{j=0}^{N-1} \tfrac{K}{N}\mathbb{P}(X > j\tfrac{K}{N}) \leqslant \tfrac{K}{N} + \sum_{j=1}^N \tfrac{K}{N}\mathbb{P}(X > j\tfrac{K}{N}).$$

Combinando as desigualdades acima, $\lim_n \mathbb{E}X_n \geqslant M - \frac{K}{N}$. Como isso vale para todo $N \in \mathbb{N}$, segue que $\lim_n \mathbb{E}X_n \geqslant M$. Como isso vale para todo $M < \mathbb{E}X$, segue que $\lim_n \mathbb{E}X_n \geqslant \mathbb{E}X$, concluindo a prova. □

Uma primeira aplicação deste teorema é a fórmula para esperança de funções de variáveis aleatórias discretas.

Demonstração do Teorema 5.36. Seja $B = \{x_1, x_2, x_3, \dots\} \subseteq \mathbb{R}$ tal que

$\mathbb{P}(X \notin B) = 0$. Podemos supor que $X(\omega) \in B$ para todo $\omega \in \Omega$, pois caso contrário bastaria considerar $\tilde{X} = X\mathbb{1}_{\{X \in B\}}$. Defina $g_n(x) = \sum_{k=1}^{n} g(x_k)\mathbb{1}_{\{x_k\}}(x)$ e observe que $g_n(x) \uparrow g(x)$ para todo $x \in B$, donde $g_n(X) \uparrow g(X)$. Usando o Teorema 5.15 e o Teorema da Convergência Monótona,

$$\mathbb{E}[g(X)] = \lim_n \mathbb{E}[g_n(X)] = \lim_n \sum_{k=1}^{n} g(x_k)\mathbb{P}(X = x_k) = \sum_x g(x)\mathbb{P}(X = x),$$

provando o Teorema 5.36. □

Demonstração do Teorema 5.26. Como $\mathbb{E}X = \mathbb{E}X^+ - \mathbb{E}X^-$, basta aplicar o Teorema 5.36 com $g(x) = x^+$ e $g(x) = x^-$, respectivamente. □

Para provar a propriedade de linearidade do Teorema 5.30, usaremos linearidade para variáveis aleatórias simples combinada com o fato de que sempre é possível aproximar variáveis não-negativas por variáveis simples.

Proposição 5.45. *Existe uma sequência de funções não-decrescentes $(\psi_n)_n : \mathbb{R}_+ \to \mathbb{R}_+$, cada uma das quais assumindo finitos valores, e tais que $\psi_n(x) \uparrow x$ para todo $x \geqslant 0$.*

Demonstração. Para $n \in \mathbb{N}$, seja ψ_n definida por

$$\psi_n(x) = 2^{-n} \cdot \max\{j \in \{0, 1, \ldots, 2^n n\} \,:\, 2^{-n} j \leqslant x\}, \tag{5.46}$$

ilustrada na Figura 5.3. Observe que cada ψ_n assume finitos valores. Ademais, dado $x \geqslant 0$, temos que $x \geqslant \psi_{n+1}(x) \geqslant \psi_n(x)$ para todo $n \in \mathbb{N}$, e para $n \geqslant x$ temos também $\psi_n(x) \geqslant x - 2^{-n}$. Portanto, esta sequência de funções satisfaz às propriedades enunciadas. □

Demonstração do Teorema 5.30. Já provamos unitariedade e monotonicidade. Para a linearidade, primeiro observamos que da definição de esperança segue que $\mathbb{E}[aX] = a\,\mathbb{E}X$, restando apenas mostrar que $\mathbb{E}[X+Y] = \mathbb{E}X + \mathbb{E}Y$.

Suponha inicialmente que X e Y sejam não-negativas. Seja $(\psi_n)_n$ a sequência de funções monótonas dadas pela Proposição 5.45. Usando o Teorema 5.9 e o Teorema da Convergência Monótona três vezes,

$$\mathbb{E}[X+Y] = \lim_n \mathbb{E}[\psi_n(X)+\psi_n(Y)] = \lim_n (\mathbb{E}[\psi_n(X)]+\mathbb{E}[\psi_n(Y)]) = \mathbb{E}X+\mathbb{E}Y.$$

Finalmente, sejam X e Y duas variáveis aleatórias quaisquer. Expandindo,

$$(X+Y)^+ - (X+Y)^- = X+Y = X^+ - X^- + Y^+ - Y^-,$$

donde

$$(X+Y)^+ + X^- + Y^- = (X+Y)^- + X^+ + Y^+.$$

Como todas as variáveis aleatórias acima são não-negativas, pelo caso anterior temos

$$\mathbb{E}(X+Y)^+ + \mathbb{E}X^- + \mathbb{E}Y^- = \mathbb{E}(X+Y)^- + \mathbb{E}X^+ + \mathbb{E}Y^+.$$

Como estamos supondo que $\mathbb{E}X$ está definida, temos $\mathbb{E}X^- < \infty$ ou $\mathbb{E}X^+ < \infty$. Suponhamos que valha o primeiro caso (se for o segundo, o argumento será análogo). Como Y é integrável, $\mathbb{E}Y^- < \infty$ e $\mathbb{E}(X+Y)^- \leqslant \mathbb{E}X^- + \mathbb{E}Y^- < \infty$.

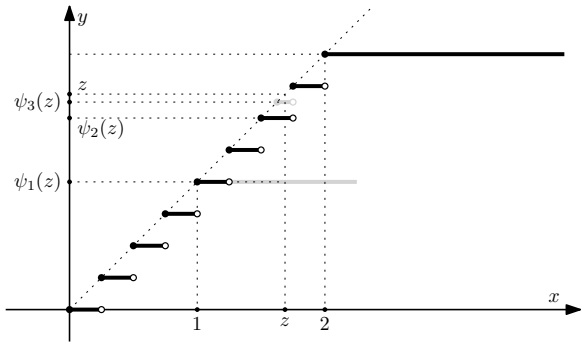


Figura 5.3. Gráfico de ψ_2 e a aproximação $\psi_n(z) \uparrow z$ para um z fixo.

Assim, podemos subtrair esses três termos, obtendo

$$\mathbb{E}(X+Y)^+ - \mathbb{E}(X+Y)^- = (\mathbb{E}X^+ - \mathbb{E}X^-) + (\mathbb{E}Y^+ - \mathbb{E}Y^-),$$

o que conclui a prova da linearidade. □

Demonstração da Proposição 5.32. Suponha que X e Y são não-negativas. Seja $(\psi_n)_n$ a sequência de funções monótonas dadas pela Proposição 5.45. Usando o Teorema 5.17 e o Teorema da Convergência Monótona três vezes,

$$\mathbb{E}[XY] = \lim_n \mathbb{E}[\psi_n(X)\psi_n(Y)] = \lim_n (\mathbb{E}[\psi_n(X)] \cdot \mathbb{E}[\psi_n(Y)]) = \mathbb{E}X \cdot \mathbb{E}Y.$$

Suponha agora que X e Y são integráveis. Usando o caso anterior,

$$\begin{aligned}\mathbb{E}[XY] &= \mathbb{E}[X^+Y^+ - X^+Y^- - X^-Y^+ + X^-Y^-] \\ &= \mathbb{E}X^+ \cdot \mathbb{E}Y^+ - \mathbb{E}X^+ \cdot \mathbb{E}Y^- - \mathbb{E}X^- \cdot \mathbb{E}Y^+ + \mathbb{E}X^- \cdot \mathbb{E}Y^- \\ &= (\mathbb{E}X^+ - \mathbb{E}X^-)(\mathbb{E}Y^+ - \mathbb{E}Y^-) = \mathbb{E}X \cdot \mathbb{E}Y.\end{aligned}$$ □

5.3.2 Teorema da Convergência Dominada

Veremos agora o Teorema da Convergência Dominada, cuja demonstração exige ferramentas um pouco mais avançadas. Isto nos permitirá, por exemplo, dar condições para tomar derivadas dentro da esperança, o que será usado nas Seções 10.1 e 10.2.

Teorema 5.47 (Teorema da Convergência Dominada)**.** *Sejam $(X_n)_n, X$ e Y variáveis aleatórias. Se $\mathbb{P}(X_n \to X) = 1$ e $\mathbb{P}(|X_n| \leqslant Y) = 1$ para todo $n \in \mathbb{N}$, com Y integrável, então $\mathbb{E}X_n \to \mathbb{E}X$ quando $n \to \infty$.*

Demonstração. A prova usa o seguinte fato. Dada uma sequência $(Z_n)_n$ de variáveis aleatórias não-negativas tais que $\liminf_n Z_n(\omega)$ é finito para todo $\omega \in \Omega$, vale

$$\mathbb{E}[\liminf_n Z_n] \leqslant \liminf_n \mathbb{E}Z_n. \tag{5.48}$$

Com efeito, tomando $Y_n = \inf_{k \geqslant n} Z_k$ e definindo $Y = \liminf_n Z_n$, obtemos $0 \leqslant Y_n \uparrow Y$. Pelo Teorema da Convergência Monótona, $\lim_n \mathbb{E}Y_n = \mathbb{E}Y$. Como $Y_n \leqslant Z_n$, temos $\liminf_n \mathbb{E}Z_n \geqslant \liminf_n \mathbb{E}Y_n = \mathbb{E}Y = \mathbb{E}[\liminf_n Z_n]$, provando a desigualdade acima.

Passemos à prova do teorema. Sem perda de generalidade, podemos supor que $X_n \to X$ e $\sup_n |X_n| \leqslant Y$ para todo $\omega \in \Omega$ (exercício!). Como $X_n \leqslant Y$, temos $Y - X_n \geqslant 0$, e portanto podemos aplicar (5.48), obtendo

$$\liminf_n \mathbb{E}[Y - X_n] \geqslant \mathbb{E}[\liminf_n (Y - X_n)] = \mathbb{E}[Y - X]$$

Como Y é integrável, vale $\mathbb{E}[Y - X_n] = \mathbb{E}Y - \mathbb{E}X_n$, $\mathbb{E}[Y - X] = \mathbb{E}Y - \mathbb{E}X$ e

$$\limsup_n \mathbb{E}X_n \leqslant \mathbb{E}X.$$

Aplicando o mesmo argumento com $-X_n$ no lugar de X_n, obtemos a desigualdade oposta $\liminf_n \mathbb{E}X_n \geqslant \mathbb{E}X$, e portanto $\mathbb{E}X_n \to \mathbb{E}X$. □

Uma das aplicações desse teorema é quando estamos considerando uma grandeza expressa em termos de uma variável aleatória e um parâmetro real. Sob certas condições, podemos derivar dentro da esperança.

Teorema 5.49. *Seja $f : [a, b] \times \mathbb{R} \to \mathbb{R}$ contínua, com $\frac{\partial}{\partial t} f(t, x)$ também contínua, e seja X uma variável aleatória. Suponha que $f(t, X)$ é integrável para todo $t \in [a, b]$, e que existe uma variável aleatória integrável Y tal que*

$$\left|\tfrac{\partial}{\partial t} f(t, X)\right| \leqslant Y \text{ q.c., para todo } t \in [a, b].$$

Então $\mathbb{E}[f(t, X)]$ é diferenciável em $[a, b]$ e

$$\tfrac{\mathrm{d}}{\mathrm{d}t}\, \mathbb{E}[f(t, X)] = \mathbb{E}\Big[\tfrac{\partial}{\partial t} f(t, X)\Big].$$

Em $t = a$ e $t = b$, as derivadas acima são derivadas laterais. O mesmo vale se consideramos o intervalo aberto (a, b).

Demonstração. Pelo Teorema do Valor Médio, para todos $x \in \mathbb{R}$ e todos $s, t \in [a, b]$ tais que $s \neq t$,

$$\frac{f(s,x) - f(t,x)}{s-t} = \tfrac{\partial}{\partial t} f(\theta, x),$$

onde $\theta = \theta(s,t,x)$ está entre t e s. Portanto,

$$\Big|\frac{f(s,X) - f(t,X)}{s-t}\Big| = \Big|\tfrac{\partial}{\partial t} f(\theta, X)\Big| \leqslant Y \text{ q.c.}$$

Tomando uma sequência $s_n \to t$ qualquer, podemos usar o Teorema da Convergência Dominada, obtendo

$$\lim_{s\to t} \mathbb{E}\Big[\frac{f(s,X) - f(t,X)}{s-t}\Big] = \mathbb{E}\Big[\lim_{s\to t} \frac{f(s,X) - f(t,X)}{s-t}\Big].$$

O termo da esquerda é igual a $\frac{\mathrm{d}}{\mathrm{d}t}\,\mathbb{E}[f(t,X)]$ e o termo da direita é $\mathbb{E}\big[\frac{\partial}{\partial t} f(t,X)\big]$. Caso $t = a$ ou $t = b$, tomamos $s \to t^{\pm}$. Isso conclui a prova. $\square$

5.4 Esperança condicional dado um evento

Assim como a Seção 3.4, esta seção será necessária para o estudo da esperança condicional a ser desenvolvido no Capítulo 11 e, exceto por isso, pode ser omitida.

A informação sobre a ocorrência de um certo evento $A \in \mathcal{F}$ com $\mathbb{P}(A) > 0$ leva à definição de uma nova medida $\mathbf{P}$ em $(\Omega, \mathcal{F})$, dada pela relação $\mathbf{P}(B) = \mathbb{P}(B|A)$, $B \in \mathcal{F}$. A distribuição de qualquer variável aleatória X também é afetada neste caso. Como vimos na Seção 3.4, X passa a ter uma nova função de distribuição $F_{X|A}(\cdot)$, uma nova lei $\mathbb{P}_{X|A}(\cdot)$.

Nesta situação, X também terá uma nova média $\mathbb{E}[X|A]$, dada por

$$\mathbb{E}[X|A] = \int_0^{+\infty} \mathbb{P}(X > x|A)\ \mathrm{d}x - \int_{-\infty}^0 \mathbb{P}(X < x|A)\ \mathrm{d}x,$$

que pode ser calculada a partir de $F_{X|A}$, $p_{X|A}$ ou $f_{X|A}$ conforme o caso.

Exemplo 5.50. Seja X a variável aleatória que representa o resultado do lançamento de um dado, isto é, $X \sim U_d\{1,2,3,4,5,6\}$. Vamos calcular $\mathbb{E}[X|A]$, onde A é dado pelo evento "X é par". Primeiro encontramos a função de probabilidade condicional:

$$p_{X|A}(x) = \mathbb{P}(X = x|A) = \frac{1}{3}\mathbb{1}_{\{2,4,6\}}(x)$$

e em seguida a esperança

$$\mathbb{E}[X|A] = \sum_x x \cdot p_{X|A}(x) = 4. \qquad \triangle$$

Exemplo 5.51. Sejam X a variável aleatória e A o evento definidos no Exemplo 3.30. Podemos calcular a esperança condicional de X dado A por

$$\mathbb{E}[X|A] = \sum_x x \cdot p_{X|A}(x) = \frac{2}{3} \cdot 1 + \frac{1}{3} \cdot 2 = \frac{4}{3}. \qquad \triangle$$

Exemplo 5.52. Assim como fizemos no exemplo anterior, considerando agora X e A como definidos no Exemplo 3.31, podemos calcular esperança condicional como

$$\mathbb{E}[X|A] = \int_0^{+\infty} x\, f_{X|A}(x)\, \mathrm{d}x = \frac{1}{4}. \qquad \triangle$$

A proposição seguinte nos fornece uma definição equivalente de esperança condicional de uma variável aleatória dado um evento. Tal definição será muito útil conforme veremos no Capítulo 11.

Proposição 5.53. *Sejam X uma variável aleatória e A um evento tal que $\mathbb{P}(A) > 0$. Então podemos escrever*

$$\mathbb{E}[X|A] = \frac{\mathbb{E}[X \cdot \mathbb{1}_A]}{\mathbb{P}(A)}.$$

Demonstração. Basta observar que

$$\begin{aligned}
\mathbb{E}[X|A] &= \int_0^{+\infty} \mathbb{P}(X > x|A)\ \mathrm{d}x - \int_{-\infty}^0 \mathbb{P}(X < x|A)\ \mathrm{d}x \\
&= \int_0^{+\infty} \frac{\mathbb{P}(\{X > x\} \cap A)}{\mathbb{P}(A)} \mathrm{d}x - \int_{-\infty}^0 \frac{\mathbb{P}(\{X < x\} \cap A)}{\mathbb{P}(A)} \mathrm{d}x \\
&= \frac{\int_0^{+\infty} \mathbb{P}(X\mathbb{1}_A > x)\ \mathrm{d}x\ -\ \int_{-\infty}^0 \mathbb{P}(X\mathbb{1}_A < x)\ \mathrm{d}x}{\mathbb{P}(A)} \\
&= \frac{\mathbb{E}[X \cdot \mathbb{1}_A]}{\mathbb{P}(A)}.
\end{aligned}$$

□

Em muitas situações, mesmo teóricas, é conveniente separar a distribuição de uma variável segundo a ocorrência ou não de um certo evento (ou, de forma mais geral, de uma sequência de eventos que particione o espaço amostral). Esse abaixo é um análogo da Lei da Probabilidade Total para a esperança condicional.

Proposição 5.54. *Sejam X uma variável aleatória e $A_1, A_2, A_3, \ldots$ uma partição do espaço amostral. Então*

$$\mathbb{E}X = \sum_{n=1}^{\infty} \mathbb{P}(A_n)\, \mathbb{E}[X|A_n].$$

Demonstração. Basta escrever $X = \sum_{n=1}^{\infty} X\mathbb{1}_{A_n}$ e desenvolver

$$\mathbb{E}X = \mathbb{E}[\textstyle\sum_{n=1}^{\infty} X\mathbb{1}_{A_n}] = \sum_{n=1}^{\infty} \mathbb{E}[X\mathbb{1}_{A_n}] = \sum_{n=1}^{\infty} \mathbb{P}(A_n)\, \mathbb{E}[X|A_n],$$

usando o Teorema da Convergência Monótona e a proposição anterior. □

Exemplo 5.55. No Exemplo 5.50, podemos calcular

$$\mathbb{E}X = \frac{1}{2} \cdot 4 + \frac{1}{2} \cdot 3 = \frac{7}{2}.$$

△

5.5 Integral de Lebesgue

Em Análise Real, primeiro vemos a integral de Riemann, que aproxima a área abaixo de uma curva por retângulos verticais. Entretanto, este não é o procedimento mais adequado quando analisamos sequências de funções. Por exemplo, suponha que queremos mostrar que

$$\lim_{n\to\infty} \int_0^{+\infty} \frac{e^{-nx^2}}{\sqrt{x}}\,\mathrm{d}x = 0$$

ou

$$\frac{\mathrm{d}}{\mathrm{d}t} \int_0^{+\infty} e^{tx} f(x)\,\mathrm{d}x = \int_0^{+\infty} x e^{tx} f(x)\,\mathrm{d}x$$

para alguma função contínua e limitada f. Em princípio, poderíamos usar um emaranhado de teoremas do Cálculo Avançado. A integral de Lebesgue, por outro lado, aproxima a área abaixo de uma curva por retângulos horizontais, e essa pequena diferença a torna muitíssimo mais flexível. Por exemplo, as identidades acima ficam muito mais simples se usamos o Teorema da Convergência Dominada, que apresentaremos nesta seção.

O principal motivo para usar a integral de Lebesgue é a forma robusta com que ela comuta com limites, derivadas, séries e outras integrais.

Outra razão para introduzir a integral de Lebesgue é que a maioria dos

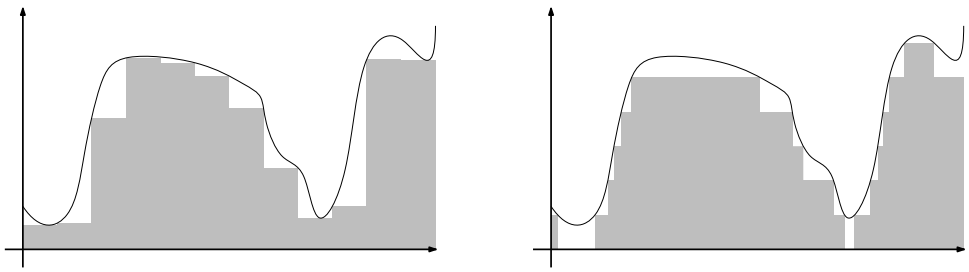


Figura 5.4. Comparação entre a integral de Riemann em $\mathbb{R}$ e a integral de Lebesgue em um espaço de medida.

espaços de medida (incluindo espaços de probabilidade!) não podem ser facilmente particionados em pequenos intervalos ou cubos contíguos como em $\mathbb{R}^n$, mas ainda assim podemos medir suas partes. Para definir a integral neste caso, ao invés de particionar o domínio e medir a altura do gráfico no contradomínio, particionamos o contradomínio e medimos pedaços do domínio, como mostrado na Figura 5.4.

5.5.1 Construção

Vamos construir a integral de uma função f com respeito a uma medida μ, denotada

$$\int_\Omega f \,\mathrm{d}\mu \quad \text{ou} \quad \int_\Omega f(\omega)\,\mu(\mathrm{d}\omega).$$

Seja $(\Omega, \mathcal{F}, \mu)$ um espaço de medida. Dizemos que $f : \Omega \to \mathbb{R}$ é uma *função simples* se é mensurável e toma apenas finitos valores.

Se f é uma função simples não-negativa, definimos sua *integral* por

$$\int_\Omega f \,\mathrm{d}\mu = \sum_x x \cdot \mu(\{\omega \in \Omega : f(\omega) = x\}).$$

Teorema 5.56. *Sejam f e g funções simples não-negativas e $a \in [0, +\infty)$. Então valem as seguintes propriedades:*

(1) *Unitariedade: $\int_\Omega \mathbb{1}_A \,\mathrm{d}\mu = \mu(A)$ para todo $A \in \mathcal{F}$.*
(2) *Monotonicidade: se $f \geqslant g \geqslant 0$, então $\int_\Omega f \,\mathrm{d}\mu \geqslant \int_\Omega g \,\mathrm{d}\mu \geqslant 0$.*
(3) *Linearidade: $\int_\Omega (af + g) \,\mathrm{d}\mu = a \int_\Omega f \,\mathrm{d}\mu + \int_\Omega g \,\mathrm{d}\mu$.*

Demonstração. O operador é unitário por construção. Linearidade é provada de forma idêntica ao Teorema 5.9. Para monotonicidade, observamos que $\int_\Omega f \,\mathrm{d}\mu = \int_\Omega (g + f - g)\mathrm{d}\mu = \int_\Omega g \,\mathrm{d}\mu + \int_\Omega (f - g)\mathrm{d}\mu \geqslant \int_\Omega g \,\mathrm{d}\mu \geqslant 0$. □

O próximo passo é definir a integral de uma função $f : \Omega \to [0, +\infty]$

mensurável, o que fazemos aproximando-a por baixo por funções simples:

$$\int_\Omega f \,\mathrm{d}\mu = \sup\Big\{ \textstyle\int_\Omega g \,\mathrm{d}\mu : 0 \leqslant g \leqslant f \text{ e } g \text{ é simples}\Big\}. \tag{5.57}$$

Note que para funções não-negativas simples, essa definição coincide com a anterior tomando-se $g = f$.

Teorema 5.58. *Vale o Teorema 5.56 supondo funções $f, g : \Omega \to [0, +\infty]$ mensuráveis e constante $a \in [0, +\infty]$.*

O operador é unitário por construção e é monótono por inclusão: quanto maior f, mais funções simples entram no supremo em (5.57). Falta provar a linearidade. Para isso, usaremos o Teorema da Convergência Monótona, um dos teoremas fulcrais da Teoria da Medida.

Teorema 5.59 (Teorema da Convergência Monótona)**.** *Seja $(f_n)_n$ uma sequência de funções mensuráveis de Ω em $[0, +\infty]$ tais que $f_{n+1}(\omega) \geqslant f_n(\omega)$ para todo $n \in \mathbb{N}$ e $\omega \in \Omega$. Então*

$$\lim_{n\to\infty} \int_\Omega f_n(\omega)\mu(\mathrm{d}\omega) = \int_\Omega (\lim_{n\to\infty} f_n(\omega))\mu(\mathrm{d}\omega).$$

Demonstração. Denotemos $\lim_{n\to\infty} f_n = f$. Pela monotonicidade da integral, a sequência $\int_\Omega f_n \,\mathrm{d}\mu$ é não-decrescente e limitada por $\int_\Omega f \,\mathrm{d}\mu$. Logo, ela tem limite, e

$$\lim_{n\to\infty} \int_\Omega f_n \,\mathrm{d}\mu \leqslant \int_\Omega f \,\mathrm{d}\mu.$$

Resta mostrar a desigualdade oposta. Seja $0 \leqslant g \leqslant f$ simples, tomando valores $a_1, \ldots, a_k$. Seja $0 < \alpha < 1$ e $A_n = \{\omega : f_n(\omega) \geqslant \alpha g(\omega)\}$. Como $f \geqslant g$ e $0 \leqslant f_n \uparrow f$, para cada ω existe n tal que $f_n(\omega) \geqslant \alpha g(\omega)$, logo $A_n \uparrow \Omega$ e

$$\begin{aligned}\int_\Omega f_n \,\mathrm{d}\mu &\geqslant \int_\Omega (f_n \mathbb{1}_{A_n}) \,\mathrm{d}\mu \geqslant \int_\Omega (\alpha g \mathbb{1}_{A_n}) \,\mathrm{d}\mu \\ &= \alpha \sum_{j=1}^k a_j \mu(\{\omega \in A_n : g(\omega) = a_j\}) \to \alpha \sum_{j=1}^k a_j \mu(g = a_j)\end{aligned}$$

$$= \alpha \int_\Omega g \, \mathrm{d}\mu.$$

Logo,

$$\lim_{n\to\infty} \int_\Omega f_n \, \mathrm{d}\mu \geqslant \alpha \int_\Omega g \, \mathrm{d}\mu.$$

Como isso vale para todo $0 < \alpha < 1$, concluímos que

$$\lim_{n\to\infty} \int_\Omega f_n \, \mathrm{d}\mu \geqslant \int_\Omega g \, \mathrm{d}\mu.$$

Como isso vale para toda função simples g tal que $0 \leqslant g \leqslant f$,

$$\lim_{n\to\infty} \int_\Omega f_n \, \mathrm{d}\mu \geqslant \int_\Omega f \, \mathrm{d}\mu. \qquad \square$$

Como fizemos para a esperança, inúmeras propriedades da integral de Lebesgue serão demonstradas combinando-se o Teorema da Convergência Monótona com a aproximação por funções simples.

Proposição 5.60. *Dada uma função mensurável $f : \Omega \to [0, +\infty]$, existe uma sequência de funções simples f_n tais que $0 \leqslant f_n \uparrow f$ para todo ω.*

Demonstração. Seja $\psi_n : [0, +\infty] \to [0, +\infty)$ a função definida em (5.46) e observamos que $0 \leqslant \psi_n(x) \uparrow x$ para todo $x \in [0, +\infty]$. Tomando $f_n = \psi_n \circ f$ obtemos a sequência desejada. $\square$

Agora estamos aptos a terminar a prova do Teorema 5.58.

Demonstração do Teorema 5.58. Falta provar a linearidade. Sejam $f, g : \Omega \to [0, +\infty]$ funções mensuráveis $a \in [0, +\infty]$. Se $a \in [0, +\infty)$, temos $\int_\Omega (af) \, \mathrm{d}\mu = a \int_\Omega f \, \mathrm{d}\mu$ diretamente de (5.57). Se $a = +\infty$, demonstra-se a mesma igualdade tomando-se uma sequência $(a_n)_n$, tal que $a_n \uparrow +\infty$ e aplicando-se o Teorema da Convergência Monótona. Finalmente, pela Proposição 5.60, existem sequências de funções simples não-negativas $f_n \uparrow f$ e

$g_n \uparrow g$. Usando o Teorema da Convergência Monótona três vezes e linearidade para a integral de funções simples não-negativas,

$$\int_\Omega (f+g)\,\mathrm{d}\mu = \lim_{n\to\infty}\int_\Omega (f_n+g_n)\,\mathrm{d}\mu = \lim_{n\to\infty}\Big(\int_\Omega f_n\,\mathrm{d}\mu + \int_\Omega g_n\,\mathrm{d}\mu\Big) =$$
$$= \lim_{n\to\infty}\int_\Omega f_n\,\mathrm{d}\mu + \lim_{n\to\infty}\int_\Omega g_n\,\mathrm{d}\mu = \int_\Omega f\,\mathrm{d}\mu + \int_\Omega g\,\mathrm{d}\mu,$$

o que conclui a prova. □

Dizemos que uma propriedade vale para *μ-quase todo* $\omega \in \Omega$, ou em *μ-quase toda parte*, abreviado por μ-q.t.p., se existe um conjunto $A \in \mathcal{F}$ tal que $\mu(A^c) = 0$ e tal que essa propriedade vale para todo $\omega \in A$. Caso μ esteja claro no contexto, podemos dizer simplesmente "q.t.p." omitindo μ.

Exercício 5.61. Seja $f : \Omega \to [0, +\infty]$ uma função mensurável. Mostre que $\int_\Omega f\,\mathrm{d}\mu = 0$ se, e somente se, $f = 0$ q.t.p. △

Dada uma função mensurável $f : \Omega \to [-\infty, +\infty]$, denote sua parte positiva por f^+ e sua parte negativa por f^-. Definimos

$$\int_\Omega f\,\mathrm{d}\mu = \int_\Omega f^+\,\mathrm{d}\mu - \int_\Omega f^-\,\mathrm{d}\mu$$

caso uma das integrais seja finita, caso contrário dizemos que $\int_\Omega f\,\mathrm{d}\mu$ não está definida. Quando ambas as integrais são finitas, dizemos que f *é integrável*. Observe que, se f é integrável, então f é finita q.t.p.

Essa definição também dá origem a um operador linear.

Teorema 5.62. *Sejam $f, g : \Omega \to \mathbb{R}$ funções mensuráveis. Sejam $a, b \in \mathbb{R}$. Suponha que g seja integrável. Então valem as seguintes propriedades.*

(1) *Monotonicidade: se $f \geqslant g$ para todo ω, então $\int_\Omega f\,\mathrm{d}\mu$ está definida e $\int_\Omega f\,\mathrm{d}\mu \geqslant \int_\Omega g\,\mathrm{d}\mu$.*
(2) *Linearidade: $\int_\Omega (af+bg)\,\mathrm{d}\mu = a\int_\Omega f\,\mathrm{d}\mu + b\int_\Omega g\,\mathrm{d}\mu$, desde que $\int_\Omega f\,\mathrm{d}\mu$ esteja definida.*

Demonstração. Começamos pela linearidade. Mostraremos primeiro que $\int_\Omega (af)\mathrm{d}\mu = a\int_\Omega f\,\mathrm{d}\mu$. Suponhamos inicialmente que $a > 0$. Desenvolvendo,

$$\begin{aligned}\int_\Omega (af)\,\mathrm{d}\mu &= \int_\Omega (af^+)\,\mathrm{d}\mu - \int_\Omega (af^-)\,\mathrm{d}\mu \\ &= a\int_\Omega f^+\,\mathrm{d}\mu - a\int_\Omega f^-\,\mathrm{d}\mu \\ &= a\Big(\int_\Omega f^+\,\mathrm{d}\mu - \int_\Omega f^-\,\mathrm{d}\mu\Big) = a\int_\Omega f\,\mathrm{d}\mu.\end{aligned}$$

Essas expressões não contêm "$\infty - \infty$" porque estamos supondo que $\int_\Omega f\,\mathrm{d}\mu$ está definida. Para o caso $a \leqslant 0$ basta considerar $-f$ no lugar de f.

Para terminar a prova da linearidade, resta mostrar que $\int_\Omega (f+g)\,\mathrm{d}\mu = \int_\Omega f\,\mathrm{d}\mu + \int_\Omega g\,\mathrm{d}\mu$. Observamos que

$$(f+g)^+ - (f+g)^- = f + g = f^+ - f^- + g^+ - g^-,$$

donde

$$(f+g)^+ + f^- + g^- = (f+g)^- + f^+ + g^+.$$

Observando que todas as funções acima são não-negativas, pela aditividade dada pelo Teorema 5.58 segue que

$$\int_\Omega (f+g)^+\,\mathrm{d}\mu + \int_\Omega f^-\,\mathrm{d}\mu + \int_\Omega g^-\,\mathrm{d}\mu = \int_\Omega (f+g)^-\,\mathrm{d}\mu + \int_\Omega f^+\,\mathrm{d}\mu + \int_\Omega g^+\,\mathrm{d}\mu.$$

Como estamos supondo que $\int_\Omega f\,\mathrm{d}\mu$ está definida, temos $\int_\Omega f^-\,\mathrm{d}\mu < \infty$ ou $\int_\Omega f^+\,\mathrm{d}\mu < \infty$. Suponhamos que valha o primeiro caso; se for o segundo, o argumento será análogo. Como g é integrável, temos $\int_\Omega g^-\,\mathrm{d}\mu < \infty$ e $\int_\Omega (f+g)^-\,\mathrm{d}\mu \leqslant \int_\Omega f^-\,\mathrm{d}\mu + \int_\Omega g^-\,\mathrm{d}\mu < \infty$. Assim, podemos subtrair esses três termos de ambos os lados, obtendo

$$\int_\Omega (f+g)^+\,\mathrm{d}\mu - \int_\Omega (f+g)^-\,\mathrm{d}\mu = \int_\Omega f^+\,\mathrm{d}\mu - \int_\Omega f^-\,\mathrm{d}\mu + \int_\Omega g^+\,\mathrm{d}\mu - \int_\Omega g^-\,\mathrm{d}\mu,$$

o que conclui a prova da linearidade.

Para monotonicidade, observe que, sendo g integrável e $f \geqslant g$, temos $f^- \leqslant g^-$, logo $\int_\Omega f^- \, \mathrm{d}\mu < \infty$ e, portanto, $\int_\Omega f \, \mathrm{d}\mu$ está definida. Por linearidade,

$$\int_\Omega f \, \mathrm{d}\mu = \int_\Omega g \, \mathrm{d}\mu + \int_\Omega (f-g) \mathrm{d}\mu \geqslant \int_\Omega g \, \mathrm{d}\mu,$$

o que conclui a prova. □

5.5.2 Principais propriedades

Dado um conjunto $B \in \mathcal{F}$, a integral de f em B é definida como

$$\int_B f \, \mathrm{d}\mu = \int_\Omega (f \mathbb{1}_B) \, \mathrm{d}\mu.$$

Exercício 5.63. Seja $f : \Omega \to [-\infty, +\infty]$ uma função mensurável. Mostre que $\int_A f \, \mathrm{d}\mu \geqslant 0$ para todo $A \in \mathcal{F}$ se, e somente se, $f \geqslant 0$ q.t.p. △

A esperança de uma variável aleatória X nada mais é do que a integral de Lebesgue da função mensurável X com respeito à medida $\mathbb{P}$.

Teorema 5.64. *Seja $(\Omega, \mathcal{F}, \mathbb{P})$ um espaço de probabilidade e X uma variável aleatória. Então a esperança de X é dada por*

$$\mathbb{E}X = \int_\Omega X \, \mathrm{d}\mathbb{P},$$

desde que uma das duas esteja definida.

Demonstração. Se X é da forma $X = \sum_{k=1}^n a_k \mathbb{1}_{A_k}$, pela unitariedade e linearidade da esperança e da integral, segue que $\mathbb{E}X = \sum_{k=1}^n a_k \mathbb{P}(A_k) = \int_\Omega X \, \mathrm{d}\mathbb{P}$. Se X é não-negativa, obtemos a identidade usando ambas as versões do Teorema da Convergência Monótona, para a esperança e para a integral. Finalmente, no caso em que X é uma variável aleatória qualquer, basta observar que $\mathbb{E}X = \mathbb{E}X^+ - \mathbb{E}X^- = \int_\Omega X^+ \, \mathrm{d}\mathbb{P} - \int_\Omega X^- \, \mathrm{d}\mathbb{P} = \int_\Omega X \, \mathrm{d}\mathbb{P}$, desde que uma das duas diferenças acima esteja definida. □

A integral de Lebesgue tem a vantagem adicional de nos permitir estender a noção de esperança para variáveis aleatórias estendidas $X : \Omega \to [-\infty, +\infty]$, via a fórmula $\mathbb{E}X = \int_\Omega X \, \mathrm{d}\mathbb{P}$.

Seja f uma função mensurável de $(\Omega_1, \mathcal{F}_1)$ em $(\Omega_2, \mathcal{F}_2)$. Dada uma função mensurável $g : \Omega_2 \to [-\infty, +\infty]$, podemos definir o seu *pull-back* em Ω_1 dado pela função $f^*g = g \circ f$ em Ω_1. Juntamente com $\sigma(f)$ e $f_*\mu$ definidos na Seção 3.7, obtemos o seguinte diagrama:

$$\begin{tikzcd}
\Omega_1 \arrow[r, dotted, no head] \arrow[d, "f"] & \sigma(f) \arrow[r, dotted, no head] & \mu \arrow[r, dotted, no head] \arrow[d, "\circ f^{-1}"] & f^*g \\
\Omega_2 \arrow[r, dotted, no head] & \mathcal{F}_2 \arrow[r, dotted, no head] \arrow[u, "f^{-1}"] & f_*\mu \arrow[r, dotted, no head] & g \arrow[u, "\circ f"]
\end{tikzcd}$$

Em suma, a partir de uma função que leva pontos $\omega_1 \in \Omega_1$ a pontos $\omega_2 \in \Omega_2$, podemos puxar uma σ-álgebra em Ω_2 de volta para uma σ-álgebra em Ω_1, e empurrar uma medida em Ω_1 adiante para uma medida em Ω_2, e puxar um observável definido em Ω_2 de volta para um observável definido em Ω_1.

Teorema 5.65 (Mudança de variável)**.** *Sejam $(\Omega_1, \mathcal{F}_1, \mu)$ um espaço de medida, $(\Omega_2, \mathcal{F}_2)$ um espaço mensurável, $f : \Omega_1 \to \Omega_2$ uma função mensurável, e $g : \Omega_2 \to [0, +\infty]$ também mensurável. Então*

$$\int_{\Omega_2} g \, \mathrm{d}(f_*\mu) = \int_{\Omega_1} (f^*g) \, \mathrm{d}\mu.$$

Demonstração. Supondo que $g = \mathbb{1}_C$ para algum $C \in \mathcal{F}_2$, temos que $f^*g = \mathbb{1}_D$ onde $D = f^{-1}(C) \in \mathcal{F}_1$. Substituindo essas identidades, obtemos

$$\int_{\Omega_2} g \, \mathrm{d}(f_*\mu) = (f_*\mu)(C) = \mu(f^{-1}(C)) = \mu(D) = \int_{\Omega_1} (f^*g) \, \mathrm{d}\mu.$$

Por linearidade, a identidade vale para funções simples não-negativas. Pelo Teorema da Convergência Monótona, mostramos que vale para qualquer $g : \Omega_2 \to [0, +\infty]$ mensurável. $\square$

Quando consideramos uma variável aleatória X em um espaço de probabilidade, aplicando o teorema acima com $g(x) = x^+$ e $g(x) = x^-$, obtemos

$$\mathbb{E}X = \int_{\mathbb{R}} x \, \mathbb{P}_X(\mathrm{d}x),$$

que é válido desde que um dos lados esteja definido.

Proposição 5.66 (Funções iguais q.t.p.)**.** *Sejam $(\Omega, \mathcal{F}, \mu)$ um espaço de medida e $f, g : \Omega \to [-\infty, +\infty]$ funções mensuráveis. Se $f = g$ q.t.p., então*

$$\int_\Omega f \, \mathrm{d}\mu = \int_\Omega g \, \mathrm{d}\mu,$$

desde que uma das duas esteja definida.

Demonstração. Tome $A = \{\omega : f(\omega) = g(\omega)\} \in \mathcal{F}$. Defina $f_1 = f\mathbb{1}_A$ e $f_2 = f_1 + \infty \cdot \mathbb{1}_{A^c}$. Então $0 \leqslant f_1^+ \leqslant f^+ \leqslant f_2^+$ e $0 \leqslant f_1^+ \leqslant g^+ \leqslant f_2^+$. Como $\int_\Omega f_2 \, \mathrm{d}\mu = \int_\Omega f_1 \, \mathrm{d}\mu + \infty \cdot \mu(A^c) = \int_\Omega f_1 \, \mathrm{d}\mu$, por monotonicidade obtemos $\int_\Omega f^+ \, \mathrm{d}\mu = \int_\Omega g^+ \, \mathrm{d}\mu$. Por um argumento similar, $\int_\Omega f^- \, \mathrm{d}\mu = \int_\Omega g^- \, \mathrm{d}\mu$. Assumindo que uma dessas duas seja finita, obtemos $\int_\Omega f \, \mathrm{d}\mu = \int_\Omega g \, \mathrm{d}\mu$. □

Finalmente, mostraremos que o operador integral é linear não apenas no integrando mas também na medida com respeito à qual se integra.

Proposição 5.67. *Sejam μ_1 e μ_2 medidas em $(\Omega, \mathcal{F})$ e $f : \Omega \to [0, +\infty]$ mensurável. Então $\int_\Omega f \, \mathrm{d}(\mu_1 + \mu_2) = \int_\Omega f \, \mathrm{d}\mu_1 + \int_\Omega f \, \mathrm{d}\mu_2$.*

Demonstração. Se $f = \mathbb{1}_A$ para algum $A \in \mathcal{F}$, ambos os lados se reduzem a $\mu_1(A) + \mu_2(A)$ e portanto vale a identidade. Estendemos a identidade para funções simples não-negativas por linearidade, e para funções mensuráveis não-negativas usando o Teorema da Convergência Monótona. □

5.5.3 Convergência

Conforme discutido na Seção 5.3, em inúmeras situações queremos tomar um limite dentro da integral. Começamos discutindo alguns casos para ver o que poderia dar errado.

Pensemos a região abaixo do gráfico de uma função real não-negativa como tendo uma "área", "volume" ou "massa". Se a função é dada por $f(x) = n \cdot \mathbb{1}_{(0,\frac{1}{n}]}(x)$, ou por $g(x) = \mathbb{1}_{(n,n+1]}(x)$, essa massa é sempre igual a 1 e, no entanto, desaparece quando tomamos o limite em n. Podemos dizer que a massa "escapou ao infinito". No primeiro exemplo, escapou verticalmente e, no segundo, horizontalmente. Os três teoremas de convergência explicam o que pode acontecer com a massa no limite.

O Teorema da Convergência Monótona diz que nada de estranho pode acontecer com uma sequência crescente de funções, mais precisamente que não se pode ganhar massa. O próximo teorema nos diz que para uma sequência de funções mensuráveis não-negativas, até podemos perder massa no limite, mas nunca ganhar.

Teorema 5.68 (Lema de Fatou)**.** *Seja* $(\Omega, \mathcal{F}, \mu)$ *um espaço de medida e* $(f_n)_{n\geqslant 1}$ *uma sequência de funções mensuráveis de* Ω *em* $[0, +\infty]$*. Então,*

$$\int_\Omega (\liminf_{n\to\infty} f_n(\omega))\, \mu(\mathrm{d}\omega) \leqslant \liminf_{n\to\infty} \int_\Omega f_n(\omega)\, \mu(\mathrm{d}\omega).$$

Demonstração. Tomando $g_n(\omega) = \inf_{k\geqslant n} f_k(\omega)$, temos que

$$0 \leqslant g_n(\omega) \uparrow \liminf_{n\to\infty} f_n(\omega).$$

Usando o Teorema da Convergência Monótona, como $g_n(\omega) \leqslant f_n(\omega)$,

$$\begin{aligned}\liminf_{n\to\infty} \int_\Omega f_n(\omega)\mu(\mathrm{d}\omega) &\geqslant \lim_{n\to\infty} \int_\Omega g_n(\omega)\mu(\mathrm{d}\omega) = \\ &= \int_\Omega \lim_n g_n(\omega)\mu(\mathrm{d}\omega) = \int_\Omega (\liminf_{n\to\infty} f_n(\omega))\mu(\mathrm{d}\omega),\end{aligned}$$

o que conclui a prova. □

A possibilidade de desigualdade estrita no Lema de Fatou é ilustrada pelos exemplos $n \cdot \mathbb{1}_{(0,\frac{1}{n}]}$ e $\mathbb{1}_{(n,n+1]}$. Um exemplo extremo é $\mathbb{1}_{[n,+\infty)}$. Nesses exemplos, a massa escapou para o infinito. Outra situação que leva à desigualdade estrita é quando a massa fica zanzando, como por exemplo a sequência $(f_n)_n$ dada por $f_n = \mathbb{1}_{(0,1]}$ para n par e $f_n = \mathbb{1}_{(1,2]}$ para n ímpar.

Observe que o Lema de Fatou quase não tem hipóteses. Vimos algo análogo em (5.48), porém optamos por não dar-lhe nome porque, naquele contexto, vinha com a hipótese (desnecessária) de que $\liminf_n f_n$ fosse finito.

O Teorema da Convergência Dominada diz que, se os gráficos das funções da sequência $(f_n)_n$ estão confinados em uma região de massa finita, então não pode haver perda ou ganho de massa no limite. Isso porque o gráfico de f_n divide esta região de massa finita em duas partes e, caso houvesse perda de massa em uma das partes, necessariamente haveria ganho na outra.

Teorema 5.69 (Teorema da Convergência Dominada)**.** *Seja $(\Omega, \mathcal{F}, \mu)$ um espaço de medida e $(f_n)_{n\geqslant 1}$ uma sequência de funções mensuráveis de Ω em $[-\infty, +\infty]$, tais que $f_n(\omega) \to f(\omega)$ quando $n \to \infty$ para μ-quase todo ω. Suponha que existe $g : \Omega \to [0, +\infty]$ integrável, tal que $|f_n| \leqslant g$ q.t.p. para todo $n \geqslant 1$. Então f é integrável e*

$$\lim_{n\to\infty} \int_\Omega f_n(\omega)\mu(\mathrm{d}\omega) = \int_\Omega \big(\lim_{n\to\infty} f_n(\omega)\big)\mu(\mathrm{d}\omega).$$

Demonstração. Modificando f_n, f, g num conjunto de medida zero, podemos supor que $f_n(\omega) \to f(\omega)$ e $|f_n| \leqslant g < \infty$ para todo ω. Como $|f| = \lim_n |f_n| \leqslant |g|$, temos que f é integrável.

Agora observe que $f_n + g \geqslant 0$ para todo n. Aplicando o Lema de Fatou, obtemos $\int_\Omega (f + g)\,\mathrm{d}\mu \leqslant \liminf_{n\to\infty} \int_\Omega (f_n + g)\,\mathrm{d}\mu$ e, como g é integrável, $\int_\Omega f\,\mathrm{d}\mu \leqslant \liminf_{n\to\infty} \int_\Omega f_n\,\mathrm{d}\mu$. De forma análoga, $-f_n + g \geqslant 0$ para todo n, o que nos dá $\int_\Omega (-f+g)\,\mathrm{d}\mu \leqslant \liminf_{n\to\infty} \int_\Omega (-f_n+g)\,\mathrm{d}\mu$ e portanto $\int_\Omega f\,\mathrm{d}\mu \geqslant$

$\limsup_{n\to\infty} \int_\Omega f_n \, \mathrm{d}\mu$. Dessas duas desigualdades, obtemos $\lim_{n\to\infty} \int_\Omega f_n \, \mathrm{d}\mu = \int_\Omega f \, \mathrm{d}\mu$, o que conclui a prova. □

Mencionamos brevemente uma importante aplicação do Teorema da Convergência Dominada, que permite derivar dentro da integral. O enunciado e prova são análogos aos do Teorema 5.49 e serão omitidos.

5.5.4 Integral de Riemann e integral imprópria

Estamos introduzindo uma nova definição de integral. Por certo, isso gera inquietudes mais que legítimas. Nas definições e teoremas envolvendo variáveis aleatórias absolutamente contínuas, podemos substituir a integral de Riemann pela de Lebesgue? Em outras palavras, esta generaliza aquela? Neste caso, teria esta alguma vantagem? E na direção oposta, teria aquela alguma vantagem sobre esta? Para a penúltima pergunta, basta mencionar os Teoremas da Convergência Monótona e da Convergência Dominada.

Teorema 5.70. *Seja $f : [a, b] \to \mathbb{R}$ uma função mensurável. Se f é Riemann-integrável então também é Lebesgue-integrável e os valores das integrais coincidem, isto é,*

$$\int_a^b f(x) \, \mathrm{d}x = \int_{[a,b]} f \, \mathrm{d}m,$$

onde m é a medida de Lebesgue definida na Seção 1.4.2.

Demonstração. Se a $\int_a^b f(x) \, \mathrm{d}x = L \in \mathbb{R}$, então para todo $\varepsilon > 0$, existem funções-degrau g e h tais que $g \leqslant f \leqslant h$ e $L - \varepsilon < \int_a^b g(x) \, \mathrm{d}x \leqslant \int_a^b h(x) \, \mathrm{d}x < L + \varepsilon$. Logo, g e h (e portanto f) são limitadas, e por conseguinte $\int_{[a,b]} f \, \mathrm{d}m$ está definida. Por outro lado, como g e h são funções-degrau, sua integral de Riemann e de Lebesgue coincidem, logo $L - \varepsilon < \int_{[a,b]} f \, \mathrm{d}m < L + \varepsilon$. Como isso vale para todo $\varepsilon > 0$, concluímos que $\int_{[a,b]} f \, \mathrm{d}m = L$. □

A recíproca é falsa. Uma função pode ser Lebesgue-integrável sem ser Riemann-integrável. Caso f não seja limitada, não será Riemann-integrável

mesmo que seja contínua em quase todo ponto.

Exemplo 5.71. Um exemplo simples é $f : [0,1] \to \mathbb{R}$ dada por $f = \mathbb{1}_{\mathbb{Q}}$. Como $m(\mathbb{Q}) = 0$, temos $\int_{[0,1]} f \, \mathrm{d}m = 0$ mas esta função não é Riemann-integrável. Para verificarmos este fato, considere qualquer partição de $[0,1]$ em finitos intervalos não-degenerados, e observe que todos os intervalos conterão números racionais e também irracionais, pois tanto os números racionais quanto os irracionais são densos na reta. Logo, toda função-degrau $g \leqslant f$ deve satisfazer $g \leqslant 0$ q.t.p, e toda função-degrau $h \geqslant f$ deve satisfazer $h \geqslant 1$ q.t.p. Portanto, $\int g \, \mathrm{d}x \leqslant 0$ e $\int h \, \mathrm{d}x \geqslant 1$ e tomando $\varepsilon = \frac{1}{3}$ vemos que não pode haver um número L tal que $L - \varepsilon < \int g \, \mathrm{d}x \leqslant \int h \, \mathrm{d}x < L + \varepsilon$. $\triangle$

Daqui em diante, usaremos a notação

$$\int_a^b f(x) \, \mathrm{d}x \quad \text{e} \quad \int_A f(x) \, \mathrm{d}x$$

para denotar a integral de f com respeito à medida de Lebesgue no intervalo $[a,b] \subseteq \mathbb{R}$ ou no conjunto $A \in \mathcal{B}$, respectivamente.

O Teorema 5.70 considera apenas funções mensuráveis. Tecnicamente, uma função pode ser Riemann-integrável sem ser mensurável (com respeito à σ-álgebra de Borel). Entretanto, se $f : [a,b] \to \mathbb{R}$ é Riemann-integrável, seguindo a prova desse teorema pode-se mostrar que existe uma função mensurável g e um conjunto $N \subseteq [a,b]$ tais que $m(N) = 0$ e $f = g$ em $[a,b] \setminus N$. As integrais de f e g coincidem em qualquer subintervalo de $[a,b]$.

Mencionamos brevemente a integral imprópria. Como a integral de Riemann é definida apenas para intervalos finitos, a teoria baseada nela trata as integrais em intervalos infinitos como um limite. Existem funções f que não são Lebesgue-integráveis e para as quais o limite

$$\lim_{z \to +\infty} \int_0^z f(x) \, \mathrm{d}x$$

existe e é finito. Ou seja, há funções para as quais a integral imprópria está definida mas integral de Lebesgue não está. Porém, a exclusão dessas funções

torna a teoria de integração de Lebesgue muito mais robusta. Quase todos os teoremas deste capítulo deixam de valer se atribuímos significado de integral ao limite acima, incluindo a mudança de variável, regra da cadeia, cálculo da esperança, e teoremas de convergência. Se usássemos esse limite para calcular a esperança de uma variável aleatória, a Lei dos Grandes Números não seria válida. Mas isso não quer dizer que o limite acima não tenha importância. Ao contrário, ele é conhecido como integral de Dirichlet, tem um papel central em Análise Harmônica, e inclusive será usado na Seção 10.4.

Pode-se fazer discussão semelhante com respeito a séries condicionalmente convergentes. Observe que a soma de uma família de números estendidos não-negativos, definida em (C.1), coincide com a integral de Lebesgue da função $f(\alpha) = x_\alpha$ com respeito à medida de contagem em Λ. Dessa forma, a soma de uma família de números estendidos indexados por $\mathbb{N}$ é dada por

$$\sum_{k\in\mathbb{N}} x_k = \sum_{k=1}^{\infty} x_k^+ - \sum_{k=1}^{\infty} x_k^-,$$

desde que uma das duas séries do lado direito convirja, e, neste caso,

$$\sum_{k\in\mathbb{N}} x_k = \sum_{k=1}^{\infty} x_k^+ - \sum_{k=1}^{\infty} x_k^- = \lim_k \sum_{j=1}^{k} x_j^+ - \lim_k \sum_{j=1}^{k} x_j^- = \lim_k \sum_{j=1}^{k} x_j = \sum_{k=1}^{\infty} x_k.$$

Novamente, esta definição coincide com a integral de $f(k) = x_k$ com respeito à medida de contagem em $\mathbb{N}$. Caso ambas as séries divirjam, dizemos que $\sum_{k\in\mathbb{N}} x_k$ *não está definida.* Com essa definição, a soma $\sum_{k\in\mathbb{N}}(-1)^k \frac{1}{k}$ não está definida, da mesma forma que $\int_{\mathbb{R}} x^{-1} \operatorname{sen} x \, \mathrm{d}x$ não está definida. Note que há uma diferença entre "$\sum_{k\in\mathbb{N}}$" definida aqui e "$\sum_{k=1}^{\infty}$" definida no Apêndice A.1, pois a série harmônica alternada $\sum_{k=1}^{\infty}(-1)^k \frac{1}{k}$ converge.

5.5.5 Densidade de medidas

Nesta seção vamos explorar o conceito de densidade e como se aplica a variáveis aleatórias absolutamente contínuas.

Proposição 5.72. *Seja $(\Omega, \mathcal{F}, \mu)$ um espaço de medida e $f : \Omega \to [0, +\infty]$ uma função mensurável. Então,*

$$\nu(A) = \int_A f \, \mathrm{d}\mu, \quad \forall A \in \mathcal{F}, \tag{5.73}$$

define uma outra medida em $(\Omega, \mathcal{F})$.

A demonstração fica como exercício. Quando ν é definida a partir de μ desta maneira, costuma-se denotar $\nu = f\mu$ e $\mathrm{d}\nu = f \, \mathrm{d}\mu$.

Exemplo 5.74. Seja $\Omega = \{1, 2, 3, 4, 5, 6\}$, $\mathcal{F} = \mathcal{P}(\Omega)$ e $\mathbb{P}(A) = \frac{\#A}{6}$. Considere a função $g(n) = \frac{2}{7}n$. Pela Proposição 5.72, a função $\mathbf{P}$ dada por $\mathbf{P}(A) = \int_\Omega (g\mathbb{1}_A)\mathrm{d}\mathbb{P}$ define uma nova medida em $(\Omega, \mathcal{F})$. De fato, isso define uma medida de probabilidade (exercício!). $\triangle$

Quando duas medidas μ e ν são relacionadas por (5.73), chamamos a função f de *derivada de Radon-Nikodým de ν com respeito a μ*, denotada por $\frac{\mathrm{d}\nu}{\mathrm{d}\mu}$, de forma que $\nu(A) = \int_A \frac{\mathrm{d}\nu}{\mathrm{d}\mu}(\omega)\mu(\mathrm{d}\omega)$. Neste caso, dizemos que a derivada $\frac{\mathrm{d}\nu}{\mathrm{d}\mu}$ existe. A proposição abaixo diz que a derivada de Radon-Nikodým é essencialmente única.

Proposição 5.75. *Sejam μ uma medida σ-finita e $f, g : \Omega \to [-\infty, +\infty]$ duas funções mensuráveis. Se $\int_\Omega f \, \mathrm{d}\mu$ e $\int_\Omega g \, \mathrm{d}\mu$ estão definidas, e $\int_A f \, \mathrm{d}\mu = \int_A g \, \mathrm{d}\mu$ para todo $A \in \mathcal{F}$, então $f = g$ q.t.p.*

Demonstração. Supomos inicialmente que f e g são não-negativas. Tome $B_n \uparrow \Omega$ tais que $\mu(B_n) < \infty$ para todo n. Defina $A_n = \{\omega \in B_n : 0 \leqslant g(\omega) \leqslant n, g(\omega) < f(\omega)\}$. Como $\mu(A_n) \leqslant \mu(B_n) < \infty$, temos que $g\mathbb{1}_{A_n}$ é finita e integrável. Logo, $\int_{A_n}(f - g)\mathrm{d}\mu = \int_{A_n} f\mathrm{d}\mu - \int_{A_n} g\mathrm{d}\mu = 0$. Pelo Exercício 5.61, $\mu(A_n) = 0$. Mas $A_n \uparrow \{f > g\}$, logo $\mu(f > g) = 0$. Um argumento idêntico mostra que $\mu(f < g) = 0$, concluindo a prova.

Consideramos agora o caso geral. Como as integrais de f e g estão definidas e coincidem, podemos supor sem perda de generalidade que f^- e g^- são integráveis. Logo, f^- e g^- são finitas para μ-quase todo ω. Modificando f e

g em um conjunto de medida nula, podemos supor também que f^- e g^- são finitas para todo ω. Aplicando o caso anterior às funções $(f + f^- + g^-)$ e $(g + f^- + g^-)$, podemos concluir que essas duas funções coincidem q.t.p., e portanto $f = g$ q.t.p. $\square$

Observação 5.76. A proposição é falsa sem a hipótese de que μ é σ-finita. Tomando $\Omega = \{1\}$, $\mu(\{1\}) = +\infty$, $f(1) = 1$ e $g(1) = 2$, podemos ver que $\int_A f \, \mathrm{d}\mu = \int_A g \, \mathrm{d}\mu$ para todo $A \subseteq \Omega$ apesar de que $f \neq g$ em todo ponto. $\triangle$

Proposição 5.77 (Regra da cadeia)**.** *Sejam ν, μ medidas em um espaço mensurável $(\Omega, \mathcal{F})$ tais que $\frac{\mathrm{d}\nu}{\mathrm{d}\mu}$ existe, e seja $g : \Omega \to [0, +\infty]$ mensurável. Então,*

$$\int_\Omega g \, \mathrm{d}\nu = \int_\Omega g(\omega) \tfrac{\mathrm{d}\nu}{\mathrm{d}\mu}(\omega) \, \mu(\mathrm{d}\omega).$$

Demonstração. Suponha que $g = \mathbb{1}_A$ para algum $A \in \mathcal{F}$. Neste caso, temos

$$\int_\Omega \mathbb{1}_A \, \mathrm{d}\nu = \nu(A) = \int_A \tfrac{\mathrm{d}\nu}{\mathrm{d}\mu} \, \mathrm{d}\mu = \int_\Omega \mathbb{1}_A \cdot \tfrac{\mathrm{d}\nu}{\mathrm{d}\mu} \, \mathrm{d}\mu,$$

provando a identidade desejada. Por linearidade, a proposição também vale quando g é uma função simples não-negativa. Pelo Teorema da Convergência Monótona, vale para o caso geral. $\square$

Definição 5.78 (Variáveis aleatórias absolutamente contínuas)**.** Dizemos que uma variável aleatória X em um espaço de probabilidade $(\Omega, \mathcal{F}, \mathbb{P})$ é uma *variável aleatória absolutamente contínua* se $\frac{\mathrm{d}\mathbb{P}_X}{\mathrm{d}m}$ existe. Neste caso, $f_X = \frac{\mathrm{d}\mathbb{P}_X}{\mathrm{d}m}$ é uma densidade de X. Veja que f_X é determinada por X e $\mathbb{P}$, mas este último é omitido na notação.

Pela Proposição 5.75, a densidade de uma variável aleatória absolutamente contínua é única no sentido de que qualquer outra é igual em quase todo ponto. Repare que na Seção 3.3 não demos uma definição geral de distribuição absolutamente contínua, pois nem sempre existe uma densidade Riemann-integrável. Por outro lado, tomando a integral em (5.73) como de Lebesgue, a definição acima é a mais geral possível, como veremos na Seção 11.5.

Recordemos que uma variável aleatória é discreta se existe uma função não-negativa p_X tal que $\mathbb{P}_X(A) = \sum_x p_X(x)\delta_x(A)$ para todo $A \in \mathcal{B}$. Isso é o mesmo que $\frac{\mathrm{d}\mathbb{P}_X}{\mathrm{d}\nu} = p_X$, onde ν é a medida de contagem em $\mathbb{R}$. Mais geralmente, X é mista com componentes discreta e absolutamente contínua, como definido na Seção 3.5, se $\mathbb{P}_X = \mu_d + \mu_c$, onde $\frac{\mathrm{d}\mu_d}{\mathrm{d}\nu} = p_X$ e $\frac{\mathrm{d}\mu_c}{\mathrm{d}m} = f_X$.

Teorema 5.79 (Esperança de funções de variáveis mistas)**.** *Seja X uma variável aleatória mista com componentes discreta e absolutamente contínua e seja $g : \mathbb{R} \to [0, +\infty]$ mensurável. Então*

$$\mathbb{E}[g(X)] = \sum_x g(x) \cdot p_X(x) + \int_{\mathbb{R}} g(x) f_X(x)\,\mathrm{d}x.$$

Demonstração. Aplicando os Teoremas 5.64 e 5.65, depois a Proposição 5.67 e a regra da cadeia, obtemos

$$\begin{aligned}\mathbb{E}[g(X)] &= \int_\Omega g(X)\,\mathrm{d}\mathbb{P} = \int_{\mathbb{R}} g\,\mathrm{d}\mathbb{P}_X = \int_{\mathbb{R}} g\,\mathrm{d}\mu_d + \int_{\mathbb{R}} g\,\mathrm{d}\mu_c \\ &= \int_{\mathbb{R}} g\,p_X\,\mathrm{d}\nu + \int_{\mathbb{R}} g\,f_X\,\mathrm{d}m = \sum_x g(x) \cdot p_X(x) + \int_{\mathbb{R}} g\,f_X\,\mathrm{d}m,\end{aligned}$$

o que prova o teorema. □

Observe que o Teorema 5.39 corresponde a caso particular em que $p_X = 0$. Já o Teorema 5.28 segue do Teorema 5.39 tomando-se $g(x) = x^+$ e $g(x) = x^-$.

5.5.6 Espaços produto e integrais iteradas

Veremos agora condições sob as quais uma função $f(x, y)$ de duas variáveis x e y pode ser integrada com respeito a ambas variáveis, e se o resultado independe da ordem de integração. Para isso precisamos primeiro formalizar o espaço onde estará definida uma função de duas variáveis.

Definição 5.80 (σ-álgebra produto)**.** Sejam $(\Omega_1, \mathcal{F}_1)$ e $(\Omega_2, \mathcal{F}_2)$ dois espaços mensuráveis. Definimos *a σ-álgebra produto de $\mathcal{F}_1$ e $\mathcal{F}_2$*, denotada por $\mathcal{F}_1 \otimes \mathcal{F}_2$,

como sendo a σ-álgebra gerada pela classe[9]

$$\mathcal{F}_1 \times \mathcal{F}_2 = \{A_1 \times A_2 \subseteq \Omega_1 \times \Omega_2 : A_1 \in \mathcal{F}_1,\ A_2 \in \mathcal{F}_2\}$$

no espaço amostral $\Omega_1 \times \Omega_2$.

Em outras palavras, $\mathcal{F}_1 \otimes \mathcal{F}_2$ é a σ-álgebra gerada pela coleção de todos os "retângulos" cujos "lados" estão em $\mathcal{F}_1$ e $\mathcal{F}_2$. Note que $\mathcal{F}_1 \otimes \mathcal{F}_2$ é muito maior do que o $\mathcal{F}_1 \times \mathcal{F}_2$. Por exemplo, em $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$, observamos que o círculo

$$\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 < 1\},$$

está em $\mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R})$, pois é união enumerável de retângulos, mas não está em $\mathcal{B}(\mathbb{R}) \times \mathcal{B}(\mathbb{R})$. Literalmente, o círculo não é um retângulo!

O seguinte teorema garante a existência de uma medida no espaço produto que fatora para retângulos. A prova será dada no Apêndice D.4.

Teorema 5.81 (Medida produto)**.** *Sejam $(\Omega_1, \mathcal{F}_1, \mu_1)$ e $(\Omega_2, \mathcal{F}_2, \mu_2)$ dois espaços de medida σ-finita. Então, existe uma única medida ν na σ-álgebra $\mathcal{F}_1 \otimes \mathcal{F}_2$ tal que*

$$\nu(A_1 \times A_2) = \mu_1(A_1) \cdot \mu_2(A_2)$$

para todos $A_1 \in \mathcal{F}_1, A_2 \in \mathcal{F}_2$. Essa medida ν, denotada por $\mu_1 \otimes \mu_2$ é chamada medida produto *de μ_1 e μ_2.*

Para que uma integral iterada faça sentido, ao integrar com respeito a uma das variáveis, o resultado deveria ser uma função mensurável da outra variável. O lema abaixo também será provado no Apêndice D.4.

Lema 5.82. *Sejam $(\Omega_1, \mathcal{F}_1, \mu)$ e $(\Omega_2, \mathcal{F}_2, \nu)$ espaços de medida σ-finita. Seja $g : \Omega_1 \times \Omega_2 \to [0, +\infty]$ uma função mensurável com respeito à σ-álgebra produto $\mathcal{F}_1 \otimes \mathcal{F}_2$. Para todo $x \in \Omega_1$ fixo, a função $y \mapsto g(x, y)$ é $\mathcal{F}_2$-mensurável. Ademais, a integral $\int_{\Omega_2} g(x, y)\nu(\mathrm{d}y)$ define uma função $\mathcal{F}_1$-mensurável de x.*

[9] Aqui há um pequeno abuso de notação: $\mathcal{F}_1 \times \mathcal{F}_2$ não é exatamente um produto cartesiano, pois é uma coleção de retângulos $A \times B$ e não uma coleção de pares (A, B).

Analogamente, para todo $y \in \Omega_2$ fixo, a função $x \mapsto g(x,y)$ é $\mathcal{F}_1$-mensurável e a integral $\int_{\Omega_1} g(x,y)\mu(\mathrm{d}x)$ define uma função $\mathcal{F}_2$-mensurável de y.

Sabendo que a integral com respeito a uma das variáveis resulta em uma função mensurável da outra variável, passamos a estudar integrais iteradas.

Teorema 5.83 (Teorema de Tonelli)**.** *Sejam $(\Omega_1, \mathcal{F}_1, \mu)$ e $(\Omega_2, \mathcal{F}_2, \nu)$ espaços de medida σ-finita, e $g : \Omega_1 \times \Omega_2 \to [0, +\infty]$ uma função mensurável. Então*

$$\int_{\Omega_1} \Big(\int_{\Omega_2} g(x,y)\, \nu(\mathrm{d}y) \Big) \mu(\mathrm{d}x) = \int_{\Omega_2} \Big(\int_{\Omega_1} g(x,y)\, \mu(\mathrm{d}x) \Big) \nu(\mathrm{d}y)$$

e essas integrais são iguais a $\int_{\Omega_1 \times \Omega_2} g \, \mathrm{d}(\mu \otimes \nu)$.

A prova será dada no Apêndice D.4. Uma aplicação do Teorema de Tonelli é uma prova alternativa de que a densidade de uma variável aleatória normal é de fato uma função de densidade.

Exemplo 5.84 (Densidade da normal)*.* Vamos calcular

$$\int_0^{+\infty} \Big(\int_0^{+\infty} e^{-u^2} \mathrm{d}u \Big) e^{-y^2} \mathrm{d}y.$$

Com uma mudança de variáveis reescrevemos essa integral como

$$\int_0^{+\infty} \Big(\int_0^{+\infty} y e^{-x^2 y^2} \mathrm{d}x \Big) e^{-y^2} \mathrm{d}y,$$

que, pelo Teorema de Tonelli, é igual a

$$\int_0^{+\infty} \Big(\int_0^{+\infty} y e^{-(1+x^2)y^2} \mathrm{d}y \Big) \mathrm{d}x,$$

pois os integrandos são não-negativos. Como a integral iterada acima é igual a $\frac{\pi}{4}$, concluímos que

$$\int_0^{+\infty} e^{-x^2} \mathrm{d}x = \frac{1}{2}\sqrt{\pi}.$$

Com isso provamos que

$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\,\mathrm{d}x = 1$$

sem ter que usar coordenadas polares. △

A teoria acima trata sempre de funções não-negativas. Se consideramos uma função mensurável qualquer, ainda podemos aplicar o Lema 5.82 e o Teorema de Tonelli às suas partes positiva e negativa. Entretanto, poderá haver valores de x para os quais a integral em y não estará definida e vice-versa. Isso será de fato um problema, a não ser que integral interna esteja definida para quase todo valor da variável externa.

Teorema 5.85 (Teorema de Fubini)**.** *Sejam* $(\Omega_1, \mathcal{F}_1, \mu)$ *e* $(\Omega_2, \mathcal{F}_2, \nu)$ *espaços de medida* σ*-finita e* $f : \Omega_1 \times \Omega_2 \to [-\infty, +\infty]$ *uma função mensurável. Se*

$$\int_{\Omega_1} \Big(\int_{\Omega_2} |f(x,y)|\, \nu(\mathrm{d}y) \Big) \mu(\mathrm{d}x) < \infty,$$

então: $\int_{\Omega_2} f(x,y)\, \nu(\mathrm{d}y)$ *está definida para* μ*-quase todo* x*,* $\int_{\Omega_1} f(x,y)\, \mu(\mathrm{d}x)$ *está definida para* ν*-quase todo* y*, vale a igualdade*

$$\int_{\Omega_1} \Big(\int_{\Omega_2} f(x,y)\, \nu(\mathrm{d}y) \Big) \mu(\mathrm{d}x) = \int_{\Omega_2} \Big(\int_{\Omega_1} f(x,y)\, \mu(\mathrm{d}x) \Big) \nu(\mathrm{d}y),$$

e essas integrais são iguais a $\int_{\Omega_1 \times \Omega_2} f\, \mathrm{d}(\mu \otimes \nu)$*. Nas duas integrais iteradas logo acima, podemos substituir o valor da integral interna por zero no conjunto (de medida nula) de pontos para os quais ela não está definida.*

A prova será dada no Apêndice D.4. Em sua versão mais corrente, o enunciado do Teorema de Fubini tem como hipótese que f seja integrável com respeito a $\mu \otimes \nu$, que no enunciado acima foi substituída por outra mais conveniente graças ao Teorema de Tonelli. Vejamos um exemplo de integrais iteradas que não comutam, e o que nos diz o Teorema de Fubini nesse caso.

Exemplo 5.86. Seja $f : (0,1) \times (0,1) \to \mathbb{R}$ definida por

$$f(x,y) = \frac{x-y}{(x+y)^3}.$$

Fazendo $u = x + y$,

$$\int_0^1 \int_0^1 \frac{x-y}{(x+y)^3}\,\mathrm{d}x\,\mathrm{d}y = \int_0^1 \int_y^{1+y} \frac{u-2y}{u^3}\,\mathrm{d}u\,\mathrm{d}y = -\frac{1}{2}.$$

Por outro lado, também substituindo $u = x + y$,

$$\int_0^1 \int_0^1 \frac{x-y}{(x+y)^3}\,\mathrm{d}y\,\mathrm{d}x = \int_0^1 \int_x^{1+x} \frac{2x-u}{u^3}\mathrm{d}u\,\mathrm{d}x = \frac{1}{2}.$$

Podemos assim concluir que

$$\int_0^1 \int_0^1 |f(x,y)|\,\mathrm{d}x\,\mathrm{d}y = +\infty,$$

pois caso contrário teríamos $\int_0^1 \int_0^1 f(x,y)\,\mathrm{d}x\,\mathrm{d}y = \int_0^1 \int_0^1 f(x,y)\,\mathrm{d}y\,\mathrm{d}x$. △

Como toda soma é uma integral de Lebesgue com respeito à medida de contagem, podemos obter propriedades muito úteis de somas infinitas iteradas. Por exemplo, o Teorema C.2 segue direto do Teorema de Tonelli.

Teorema 5.87 (Teorema de Fubini para somas)**.** *Seja* $(x_{m,n})_{m,n\in\mathbb{N}}$ *uma sequência de números reais estendidos duplamente indexada por* m *e* n*.*

$$\textit{Se} \quad \sum_{m=1}^{\infty}\sum_{n=1}^{\infty} |x_{m,n}| < \infty, \qquad \textit{então} \quad \sum_{m=1}^{\infty}\sum_{n=1}^{\infty} x_{m,n} = \sum_{n=1}^{\infty}\sum_{m=1}^{\infty} x_{m,n}$$

estão definidas e coincidem.

Demonstração. Segue direto do Teorema de Fubini, pois somas infinitas são integrais com respeito à medida de contagem e, quando são somáveis, coincidem com a respectiva série. □

Vejamos um exemplo de somas iteradas que não comutam.

Exemplo 5.88. Considere a sequência duplamente indexada

$$x_{m,n} = \begin{array}{|cccccc} 1 & -1 & 0 & 0 & 0 & \dots \\ 0 & 1 & -1 & 0 & 0 & \dots \\ 0 & 0 & 1 & -1 & 0 & \dots \\ 0 & 0 & 0 & 1 & -1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \ddots \end{array}$$

que não é somável. Veja que somando-se colunas e depois linhas obtém-se 1 mas somando-se linhas e depois colunas obtém-se 0. △

Terminamos este capítulo com uma observação sobre esperança de funções de vetores aleatórios com densidade conjunta.

Observação 5.89 (Vetores aleatórios com densidade conjunta). A medida de Lebesgue em $\mathbb{R}^n$, denotada $\boldsymbol{m}$, é definida recursivamente como $\boldsymbol{m} = m \otimes \cdots \otimes m$. Seja $h : \mathbb{R}^n \to [0, +\infty]$ mensurável. Como $\mathcal{B}(\mathbb{R}^n) = \mathcal{B}(\mathbb{R}) \otimes \cdots \otimes \mathcal{B}(\mathbb{R})$,[10] concluímos que h é mensurável com respeito à σ-álgebra produto. Usando o Teorema de Tonelli recursivamente, $n-1$ vezes, obtemos

$$\int_{\mathbb{R}^n} h \, \mathrm{d}\boldsymbol{m} = \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} h(x_1, \dots, x_n) \, \mathrm{d}x_n \cdots \mathrm{d}x_1. \tag{5.90}$$

Munidos da identidade acima, concluímos que a definição de densidade conjunta dada na Seção 4.2 diz simplesmente que

$$\mathbb{P}_{\boldsymbol{X}}(B) = \int_B f_{\boldsymbol{X}} \, \mathrm{d}\boldsymbol{m}$$

para todo paralelepípedo $B \subseteq \mathbb{R}^n$, o que implica a mesma identidade para todo boreliano $B \in \mathcal{B}(\mathbb{R}^n)$ usando o Teorema 3.37 (unicidade de medidas). Ou seja, $\boldsymbol{X}$ ter densidade conjunta $f_{\boldsymbol{X}}$ é o mesmo que $\frac{\mathrm{d}\mathbb{P}_{\boldsymbol{X}}}{\mathrm{d}\boldsymbol{m}} = f_{\boldsymbol{X}}$.

[10] O produto de n espaços de medida pode ser definido recursivamente, veja o Apêndice D.4 para uma descrição desse argumento e para a prova de que $\mathcal{B}(\mathbb{R}^n) = \mathcal{B}(\mathbb{R}) \otimes \cdots \otimes \mathcal{B}(\mathbb{R})$.

Podemos agora justificar a fórmula (5.43). Para isso, observamos que

$$\begin{aligned}
\mathbb{E}[g(\boldsymbol{X})] &= \int_{\Omega} g(\boldsymbol{X}(\omega))\,\mathbb{P}(\mathrm{d}\omega) \\
&= \int_{\mathbb{R}^n} g(\boldsymbol{x})\,\mathbb{P}_{\boldsymbol{X}}(\mathrm{d}\boldsymbol{x}) \\
&= \int_{\mathbb{R}^n} g(\boldsymbol{x}) f_{\boldsymbol{X}}(\boldsymbol{x})\,\boldsymbol{m}(\mathrm{d}\boldsymbol{x}) \\
&= \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} g(x_1,\ldots,x_n)\,f_{\boldsymbol{X}}(x_1,\ldots,x_n)\,\mathrm{d}x_n \cdots \mathrm{d}x_1
\end{aligned}$$

para qualquer $g : \mathbb{R}^n \to [0,+\infty)$ mensurável. Em particular, pelo Lema 3.43, vale para g contínua e não-negativa. Nas igualdades acima, usamos os Teoremas 5.64 e 5.65, a regra da cadeia, e a fórmula (5.90), nesta ordem. $\triangle$

5.6 Exercícios

§5.1

1. Um método usado por cientistas para estimar a quantidade de animais em determinado local é o de capturar alguns deles ao acaso, marcá-los, soltá-los, voltar a capturar outros tantos, e contar quantos deles estão marcados. Caso sejam poucos os marcados, isso indica que a população é grande. Imagine que há 80 animais de determinada espécie, os cientistas marcam 20 deles ao acaso, e depois capturam 20 deles ao acaso. Calcule a esperança do número de animais que terão a marca.

2. Um baralho tem 52 cartas, sendo 13 de cada naipe. As cartas são embaralhadas e um jogador recebe 10 dessas cartas. Calcule a esperança do número de cartas de espadas recebida pelo jogador.

3. Pedro aprendeu um truque para ganhar dinheiro na roleta, e o vem aplicando com sucesso diário há uma semana. Ele começa com \$31,00 e aposta \$1,00 nos vermelhos contra os pretos. Se ganha, vai embora feliz com \$32,00. Se perde, seu capital baixa a \$30,00, então ele aposta \$2,00 nos

vermelhos, podendo ganhar e sair feliz com \$32,00, ou perder e continuar seu método. Nas próximas rodadas, ele aposta \$4,00, \$8,00 e \$16,00, se necessário, até ganhar. Veja que Pedro sempre termina saindo do jogo com \$32,00, a não ser que ele seja tão azarado que todas as rodadas resultem em números pretos. Supondo que não existe a casa verde, de forma que ambos vermelhos e pretos tenham probabilidade $\frac{1}{2}$, calcule a esperança do lucro de Pedro cada vez que ele entra numa casa de jogos determinado a aplicar essa estratégia. Calcule a esperança do número de rodadas que Pedro apostará antes de ir embora e do capital que Pedro terá ao ir embora.

4. Considere o seguinte jogo de azar. Uma urna contém 18 bolas, sendo 9 azuis e 9 brancas. Retiram-se 3 bolas da urna ao acaso. As bolas retiradas são descartadas e o jogador marca 1 ponto se pelo menos 2 dessas 3 bolas forem azuis. Em seguida retiram-se outras 3 bolas da urna ao acaso, as bolas retiradas são descartadas e o jogador marca 1 ponto se pelo menos 2 dessas 3 bolas forem azuis. Repete-se o procedimento até que a urna esteja vazia. Ao final, o jogador recebe um prêmio X igual ao total de pontos marcados. Calcule $\mathbb{E}X$.

5. Temos duas urnas, a primeira urna contém n bolas brancas numeradas de 1 a n, enquanto a segunda possui n bolas pretas numeradas de 1 a n. Sorteamos uma bola de cada urna e observamos os respectivos números. Dizemos que há uma coincidência se os números sorteados são iguais. Descartamos as bolas sorteadas e repetimos o procedimento até que ambas as urnas fiquem vazias. Calcule a esperança do número total de coincidências.

6. Seja $X \sim \text{Binom}(n, p)$. Calcule $\mathbb{E}X^2$.

7. Temos cinco dados com a forma de cada um dos Poliedros de Platão. O tetraedro tem suas faces numeradas de 1 a 4, o cubo de 1 a 6, o octaedro de 1 a 8, o dodecaedro de 1 a 12 e o icosaedro tem suas faces numeradas de 1 a 20. Lançamos todos os cinco dados simultaneamente. Calcule a esperança do produto do valor exibido pelo icosaedro com a soma dos valores exibidos pelos outros quatro dados.

8. Dois dados são lançados simultaneamente. Calcule a esperança do maior valor exibido.

§5.2

9. Seja $X \sim \mathcal{U}[0,1]$. Calcule $\mathbb{E}X^t$ para todo $t \in \mathbb{R}$.

10. Sejam $X \sim \mathcal{N}(0,1)$ e $Y = \max\{0, X\}$. Calcule $\mathbb{E}Y$ e $\mathbb{E}Y^2$.

11. Seja $X \sim \text{Beta}(a,b)$ com $a, b > 0$. Calcule $\mathbb{E}X$.

12. Seja $X \sim \text{Gama}(n, \beta)$, com $n \in \mathbb{N}$. Calcule $\mathbb{E}X$.

13. Seja $X \sim \text{Cauchy}(a,b)$. Mostre que $\mathbb{E}X$ não está definida.

14. Seja X uma variável aleatória tal que $\mathbb{P}(X \geqslant \mu + x) = \mathbb{P}(X \leqslant \mu - x)$ para todo $x \in \mathbb{R}$. Mostre que, se X é integrável, então $\mathbb{E}X = \mu$.

15. Sejam $X_1, \ldots, X_n$ variáveis aleatórias não-negativas (ou integráveis) independentes. Prove que $\mathbb{E}[\prod_j X_j] = \prod_j \mathbb{E}X_j$.

16. Sejam $X_1, X_2, X_3, \ldots$ uma sequência de variáveis independentes com distribuição $\mathcal{U}[0,1]$ e tome a variável aleatória N como sendo o menor n tal que $X_1 + X_2 + \cdots + X_n \geqslant 1$. Mostre que $\mathbb{E}N = e$.

17. Sejam $X_0, X_1, X_2, X_3, \ldots$ uma sequência de variáveis independentes com distribuição $\mathcal{U}[0,1]$ e tome a variável aleatória N como sendo o menor n tal que $X_n \geqslant X_0$. Mostre que $\mathbb{E}N = +\infty$.

18. Prove que, se $\mathbb{E}X$ está definida e $A \in \mathcal{F}$, então $\mathbb{E}[X\mathbb{1}_A]$ está definida.

19. Prove que, se X é integrável $A \in \mathcal{F}$, então $X\mathbb{1}_A$ é integrável.

20. Seja X uma variável aleatória tal que $\mathbb{E}X$ está definida. Defina

$$Y = \begin{cases} X, & X \leqslant a, \\ a, & \text{caso contrário}, \end{cases}$$

onde $a \in \mathbb{R}$ é constante. Mostre que $\mathbb{E}Y \leqslant \mathbb{E}X$.

21. Seja X uma variável aleatória não-degenerada tal que $\mathbb{P}(a \leqslant X \leqslant b) = 1$. Mostre que $a < \mathbb{E}X < b$.

22. Sejam X e Y variáveis aleatórias. Mostre que:

(a) Se $\mathbb{E}X$ está definida, então $|\mathbb{E}X| \leqslant \mathbb{E}|X|$;
(b) Se X é integrável, então $\mathbb{E}[X - \mathbb{E}X] = 0$;
(c) Se $0 \leqslant |X| \leqslant Y$ q.c. e Y é integrável, então X é integrável.

23. Sejam X uma variável aleatória e $p > 0$ tais que $z^p\mathbb{P}(|X| > z) \leqslant M$ para todo $z > 0$. Mostre que $\mathbb{E}|X|^q < \infty$ para todo $q \in [0, p)$. Dê um exemplo ilustrando que é possível termos $\mathbb{E}|X|^p = +\infty$.

§5.3

24. Seja $U \sim \mathcal{U}[0,1]$ e defina $X_n = n\mathbb{1}_{\{0<U\leqslant\frac{1}{n}\}}$.

(a) A sequência $(X_n)_n$ satisfaz às hipóteses dos Teoremas da Convergência Monótona ou Dominada?
(b) Calcule $\lim_n \mathbb{E}X_n$ e $\mathbb{E}[\lim_n X_n]$.

25. Sejam X uma variável aleatória integrável e $(A_n)_n$ uma sequência de eventos tal que $A_n \downarrow \emptyset$. Mostre que $\lim_n \mathbb{E}[X\mathbb{1}_{A_n}] = 0$.

26. Sejam $(X_n)_n$ e X variáveis aleatórias tais que $X_n \downarrow X \geqslant 0$ q.c. e X_1 é integrável. Mostre que $\mathbb{E}X_n \to \mathbb{E}X$.

27. Sejam $(X_n)_n$ e X variáveis aleatórias tais que $\mathbb{P}(X_n \to X) = 1$ e existe $M \in \mathbb{R}$ tal que $|X_n| \leqslant M$ q.c. para todo n. Mostre que $\mathbb{E}X_n \to \mathbb{E}X$.

28. Dê um exemplo de uma sequência de variáveis aleatórias $(X_n)_n$ tal que

$$\mathbb{E}\Big[\sum_{n=1}^{\infty} X_n\Big] \neq \sum_{n=1}^{\infty} \mathbb{E}X_n$$

e ambos os lados da equação acima estejam bem definidos.

§5.4

29. Seja $X \sim \mathcal{U}[-1,1]$. Considerando os eventos $A_1 = \{X \geqslant 0\}$ e $A_2 = \{X < 0\}$, calcule

(a) A distribuição condicional de X dado A_1.
(b) A distribuição condicional de X dado A_2.
(c) $\mathbb{E}[X|A_1]$.
(d) $\mathbb{E}[X|A_2]$.

30. Seja X uma variável aleatória exponencial com parâmetro λ. Encontre $\mathbb{E}[X \mid X > 2]$.

31. Se $X \sim \mathrm{Geom}(p)$, encontre $\mathbb{E}[X \mid X > 5]$.

§5.5

32. Seja $f : \Omega \to [-\infty, +\infty]$ uma função mensurável. Suponha que $\int_\Omega f \, \mathrm{d}\mu$ esteja definida. Mostre que $|\int_\Omega f \, \mathrm{d}\mu| \leqslant \int_\Omega |f| \, \mathrm{d}\mu$.

33. Sejam $(\Omega, \mathcal{F}, \mu)$ um espaço de medida e $f : \Omega \to [-\infty, +\infty]$ uma função mensurável. Prove que f é integrável se, e somente se, $\int_\Omega |f| \, \mathrm{d}\mu < \infty$.

34. Sejam $(\Omega, \mathcal{F}, \mu)$ um espaço de medida e $f, g : \Omega \to [-\infty, +\infty]$ funções mensuráveis. Suponha que $|f| \leqslant g$ e g é integrável. Mostre que f é integrável.

35. Sejam $(\Omega, \mathcal{F}, \mu)$ um espaço de medida, $f : \Omega \to [-\infty, +\infty]$ uma função mensurável e $A \in \mathcal{F}$.

(a) Prove que, se $\int_\Omega f \, \mathrm{d}\mu$ está definida, então $\int_\Omega f \mathbb{1}_A \, \mathrm{d}\mu$ está definida.
(b) Prove que, se f é integrável, então $f\mathbb{1}_A$ é integrável.

36. Mostre que a coleção de todas as funções mensuráveis e integráveis em $(\Omega, \mathcal{F}, \mu)$ é um espaço vetorial real.

37. Seja $(X_n)_n$ uma sequência de variáveis aleatórias estendidas não-negativas. Mostre que $\mathbb{E}[\sum_n X_n] = \sum_n \mathbb{E}X_n$.

38. Seja $(X_n)_n$ uma sequência i.i.d. de variáveis aleatórias estendidas não-negativas com $\mathbb{E}X_1 < \infty$. Mostre que $\sum_{n=1}^\infty \frac{X_n}{n^2} < \infty$ quase certamente.

39. Seja $(X_n)_n$ uma sequência de variáveis aleatórias estendidas tais que $\sum_n \mathbb{E}|X_n| < \infty$. Mostre que a série $\sum_{n=1}^{\infty} X_n$ converge quase certamente, e $\mathbb{E}[\sum_{n=1}^{\infty} X_n] = \sum_{n=1}^{\infty} \mathbb{E}X_n$.

40. Seja $(X_n)_n$ uma sequência de variáveis aleatórias i.i.d. com $\mathbb{P}(X_1 = -1) = \mathbb{P}(X_1 = 0) = \mathbb{P}(X_1 = +1) = \frac{1}{3}$. Calcule $\mathbb{E}[\liminf_n X_n]$ e $\liminf_n \mathbb{E}X_n$. Existe alguma generalização do Lema de Fatou que poderia aplicar-se aqui?

41. Sejam $(\Omega, \mathcal{F}, \mu)$ espaço de medida onde μ é a medida de contagem e $f : \Omega \to [-\infty, +\infty]$ uma função mensurável. Mostre que, se $\int_\Omega f \mathrm{d}\mu$ é finito, então $\{\omega \in \Omega : f(\omega) \neq 0\}$ é enumerável.

42. Seja Y a variável aleatória definida no Exemplo 3.34. Calcule $\mathbb{E}Y^2$.

43. Sejam μ_1 e μ_2 medidas σ-finitas tais que μ_2 tenha densidade f com respeito a μ_1. Mostre que, se $f > 0$ q.t.p., então $\frac{1}{f}$ é uma densidade de μ_1 com respeito a μ_2.

44. Considere o espaço mensurável $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$ e seja δ_k a medida de Dirac no ponto $k \in \mathbb{N}$, como definida no Exemplo 1.46. Determine uma função $f : \mathbb{N} \to [0, +\infty]$ que corresponda à derivada de Radon-Nikodým $\frac{\mathrm{d}\delta_k}{\mathrm{d}\nu}$, onde ν denota a medida de contagem em $\mathbb{N}$.

45. Mostre que duas variáveis aleatórias X e Y são independentes se, e somente se, $\mathbb{P}_{X,Y} = \mathbb{P}_X \otimes \mathbb{P}_Y$.

46. Sejam X e Y variáveis aleatórias independentes. Prove que

$$\mathbb{P}(X \leqslant Y) = \mathbb{E}[F_X(Y)].$$

47. Sejam $(\Omega_1, \mathcal{F}_1, \mu_1)$ e $(\Omega_2, \mathcal{F}_2, \mu_2)$ espaços de medida, onde $\Omega_1 = \Omega_2 = [0, 1]$, $\mathcal{F}_1 = \mathcal{F}_2 = \mathcal{B}([0, 1])$, μ_1 é a medida de Lebesgue e μ_2 é a medida da contagem. Seja $f : \Omega_1 \times \Omega_2 \to \mathbb{R}$ definida por $f(\omega_1, \omega_2) = \mathbb{1}_{\{\omega_1 = \omega_2\}}$. Calcule

$$\int_{\Omega_1} \int_{\Omega_2} f(\omega_1, \omega_2)\, \mu_2(\mathrm{d}\omega_2)\mu_1(\mathrm{d}\omega_1) \quad \text{e} \quad \int_{\Omega_2} \int_{\Omega_1} f(\omega_1, \omega_2)\, \mu_1(\mathrm{d}\omega_1)\mu_2(\mathrm{d}\omega_2). \tag{5.91}$$

Vale o Teorema de Tonelli? Por quê?

Capítulo 6

Momentos e Desigualdades

Neste capítulo introduzimos as noções de momentos de variáveis aleatórias, de variância e covariância, e estudaremos várias desigualdades envolvendo momentos, que serão fundamentais nos capítulos seguintes.

6.1 Momentos e variância

Sejam X e Y variáveis aleatórias com distribuições uniformes em $[-1, +1]$ e $\{-1, +1\}$, respectivamente. Está claro que X tem uma dispersão de seus valores em torno de sua média menor que a de Y. Um dos objetivos desta seção é quantificar o quanto uma variável aleatória se espalha em torno de determinado valor e, em particular, o quanto ela se espalha em torno de sua média.

Definição 6.1. Dada uma variável aleatória X e $k = 1, 2, 3, \ldots$, definimos o *k-ésimo momento* de X como $\mathbb{E}X^k$, caso X^k seja integrável. Neste caso, dizemos que X tem k-ésimo momento finito, e definimos o *k-ésimo momento central* de X como $\mathbb{E}(X - \mathbb{E}X)^k$.

Observamos que, se X tem k-ésimo momento finito, então X tem j-ésimo

momento finito para $j = 1, \ldots, k$ (pois $|X|^j \leqslant 1 + |X|^k$) e, em particular, o k-ésimo momento central está bem definido.

Exemplo 6.2. Se $X \sim \mathcal{U}[0,1]$, então

$$\mathbb{E}X = \int_0^1 x \,\mathrm{d}x = \frac{1}{2}, \quad \mathbb{E}X^2 = \int_0^1 x^2 \,\mathrm{d}x = \frac{1}{3}, \quad \mathbb{E}X^k = \int_0^1 x^k \,\mathrm{d}x = \frac{1}{k+1},$$

e o segundo momento central é dado por

$$\mathbb{E}(X - \tfrac{1}{2})^2 = \textstyle\int_0^1 (x - \frac{1}{2})^2 \,\mathrm{d}x = \frac{1}{12}. \qquad \triangle$$

Como veremos a seguir, o segundo momento central é um excelente quantificador da dispersão de variável aleatória X em torno de sua média, com propriedades muito especiais.

Definição 6.3 (Variância)**.** Seja X uma variável aleatória integrável. Define-se a *variância* da variável aleatória X, denotada por $\mathbb{V}X$, como

$$\mathbb{V}X = \mathbb{E}(X - \mathbb{E}X)^2.$$

Exemplo 6.4. Se $X \sim \mathcal{U}[0,1]$, então $\mathbb{E}X = \frac{1}{2}$ e $\mathbb{V}X = \frac{1}{12}$, como calculado no Exemplo 6.2. $\triangle$

Proposição 6.5 (Propriedades da variância)**.** *Seja X uma variável aleatória com segundo momento finito. Então:*

(1) $\mathbb{V}X = \mathbb{E}X^2 - (\mathbb{E}X)^2$. *Em particular,* $\mathbb{V}X \leqslant \mathbb{E}X^2$.
(2) $\mathbb{V}X \geqslant 0$. *Além disso,* $\mathbb{V}X = 0$ *se, e somente se,* $X = \mathbb{E}X$ *q.c.*
(3) $\mathbb{V}[aX + b] = a^2\,\mathbb{V}X$.

Demonstração. Para o item (1), basta expandir $\mathbb{V}X = \mathbb{E}(X - \mathbb{E}X)^2 = \mathbb{E}[X^2 - 2X\mathbb{E}X + (\mathbb{E}X)^2] = \mathbb{E}X^2 - (\mathbb{E}X)^2$.

Para o item (2), $\mathbb{V}X \geqslant 0$ pois $(X - \mathbb{E}X)^2 \geqslant 0$. Pelo item (4) da Proposição 5.31, $\mathbb{V}X = 0$ implica que $X = \mathbb{E}X$ quase certamente; a recíproca é imediata.

Para provar o item (3), expandimos $\mathbb{V}[aX+b] = \mathbb{E}(aX+b-\mathbb{E}[aX+b])^2 = a^2\,\mathbb{E}(X-\mathbb{E}X)^2 = a^2\,\mathbb{V}X$. $\square$

Exemplo 6.6. Se $X \sim \text{Bernoulli}(p)$, então

$$\mathbb{E}X = p, \quad \mathbb{E}X^2 = p, \quad \mathbb{V}X = \mathbb{E}X^2 - (\mathbb{E}X)^2 = p - p^2 = p(1-p). \qquad \triangle$$

Observe que a variância é máxima no caso simétrico $p = \frac{1}{2}$.

Exemplo 6.7. Seja $X \sim \mathcal{N}(\mu, \sigma^2)$. Sabemos que $X = \sigma Z + \mu$ com $Z \sim \mathcal{N}(0,1)$ (Exercício 3.28) e que $\mathbb{E}Z^2 = 1$ (Exemplo 5.41). Usando proposição acima, obtemos $\mathbb{V}X = \sigma^2$. $\triangle$

Exemplo 6.8. Se $X \sim \mathcal{U}[a,b]$ então $X \sim a + (b-a)U$, onde U é uma uniforme no intervalo $[0,1]$, cuja variância foi calculada no Exemplo 6.4. Logo, segue do item (3) da proposição anterior que $\mathbb{V}X = \frac{(b-a)^2}{12}$. $\triangle$

Podemos observar que $\mathbb{V}X$ é uma medida da dispersão de X em torno de sua média, mas que dimensionalmente não é expressa nas mesmas unidades de X. Por exemplo, se X for medida em kg então $\mathbb{V}X$ é medida em kg^2. Para que tenhamos uma medida de dispersão na mesma escala da variável aleatória X, somos motivados a introduzir a próxima definição.

Definição 6.9 (Desvio-padrão)**.** O *desvio-padrão* $\sigma(X)$ da variável aleatória X é dado pela raiz quadrada da variância [11]

$$\sigma(X) = \sqrt{\mathbb{V}X}.$$

Exemplo 6.10. Se $X \sim \text{Bernoulli}(\frac{1}{2})$, então

$$\sigma(X) = \sqrt{\mathbb{V}X} = \sqrt{\tfrac{1}{4}} = \tfrac{1}{2}. \qquad \triangle$$

Exemplo 6.11. Se $X \sim \mathcal{N}(\mu, \sigma^2)$, do Exemplo 6.7, segue que $\sigma(X) = \sigma$. $\triangle$

[11] O significado de $\sigma(X)$ neste capítulo, bem como no Capítulo 9, é completamente diferente daquele nas seções mais avançadas dos Capítulos 3 e 5, bem como nos Capítulos 11, 12 e 13. Essa reutilização da letra σ não deve causar confusão.

Gostaríamos de estudar a variância da soma de duas ou mais variáveis aleatórias e, se possível, relacioná-la com a variância das variáveis envolvidas. Expandindo a fórmula da variância, obtemos

$$\mathbb{V}[X+Y] = \mathbb{V}X + \mathbb{V}Y + 2\,\mathbb{E}[(X-\mathbb{E}X)\cdot(Y-\mathbb{E}Y)].$$

Ou seja, a variância de $X+Y$ é igual à soma das variâncias de X e de Y, mais um termo cruzado que envolve ambas as variáveis. Isto nos motiva a introduzir o seguinte conceito.

Definição 6.12 (Covariância)**.** Dadas duas variáveis aleatórias X e Y com segundo momento finito, definimos a *covariância de X e Y* como

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

A expressão acima está definida e é finita, pois $|xy| \leqslant x^2+y^2$.

Observe que $\mathbf{Cov}(X,X) = \mathbb{V}X$, $\mathbf{Cov}(X,Y) = \mathbf{Cov}(Y,X)$ (a covariância é simétrica), e

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

Se $\mathbf{Cov}(X,Y)=0$, dizemos que X e Y são *não-correlacionadas*, e isso vale se, e somente se, $\mathbb{V}[X+Y] = \mathbb{V}X + \mathbb{V}Y$. Se as variáveis aleatórias X e Y são independentes e têm segundo momento finito então X e Y são não-correlacionadas e $\mathbb{V}[X+Y] = \mathbb{V}X + \mathbb{V}Y$. Entretanto, nem sempre vale a recíproca, pois $\mathbb{E}[XY] = \mathbb{E}X\cdot\mathbb{E}Y$ não implica X e Y independentes.

Exemplo 6.13. Sejam X e Y variáveis aleatórias tomando valores $-1,0,1$ com distribuição conjunta dada por $p(-1,-1)=p(-1,1)=p(1,-1)=p(1,1)=p(0,0)=\frac{1}{5}$. Então $\mathbf{Cov}(X,Y)=0$, mas X e Y não são independentes. $\triangle$

Outras propriedades importantes da covariância são dadas na proposição abaixo.

Proposição 6.14 (Propriedades da Covariância)**.** *Sejam $X, X_1, \ldots, X_n$ e $Y, Y_1, \ldots, Y_m$ variáveis aleatórias com segundo momento finito e $a_1, \ldots, a_n$, e $b_1, \ldots, b_m, c$ números reais. Então:*

(1) $\mathbf{Cov}(X, c) = 0$ *para todo* $c \in \mathbb{R}$*;*
(2) $\mathbf{Cov}(\sum_k a_k X_k, \sum_j b_j Y_j) = \sum_k \sum_j a_k b_j \, \mathbf{Cov}(X_k, Y_j)$;
(3) $\mathbb{V}[\sum_{k=1}^n X_k] = \sum_{k=1}^n \mathbb{V} X_k + 2 \sum_{1 \leqslant k < j \leqslant n} \mathbf{Cov}(X_k, X_j)$.

Demonstração. Expandindo $\mathbf{Cov}(X, c) = \mathbb{E}[cX] - \mathbb{E}X \cdot \mathbb{E}[c] = c\,\mathbb{E}X - c\,\mathbb{E}X = 0$, mostramos o item (1). O item (2) segue da expansão

$$\begin{aligned}
&\mathbb{E}[(\sum_k a_k X_k)(\sum_j b_j Y_j)] - \mathbb{E}[\sum_k a_k X_k]\mathbb{E}[\sum_j b_j Y_j] \\
&\quad = \mathbb{E}[\sum_{k,j} a_k b_j X_k Y_j] - (\sum_k a_k \mathbb{E}X_k)(\sum_j b_j \mathbb{E}Y_j) \\
&\quad = \sum_{k,j} a_k b_j \mathbb{E}[X_k Y_j] - (\sum_{k,j} a_k b_j \mathbb{E}X_k \cdot \mathbb{E}Y_j) \\
&\quad = \sum_k \sum_j a_k b_j \, \mathbf{Cov}(X_k, Y_j).
\end{aligned}$$

Para provar o item (3), expandimos

$$\begin{aligned}
\mathbb{V}[\sum_{k=1}^n X_k] &= \mathbf{Cov}\Big(\sum_{k=1}^n X_k, \sum_{j=1}^n X_j\Big) = \sum_{k=1}^n \sum_{j=1}^n \mathbf{Cov}(X_k, X_j) \\
&= \sum_{k=1}^n \mathbf{Cov}(X_k, X_k) + \sum_{k=1}^n \sum_{j=1}^n \mathbb{1}_{k \neq j} \, \mathbf{Cov}(X_k, X_j)
\end{aligned}$$

e observamos que todo par de k e j distintos aparece duas vezes na última soma acima. □

Pelo último item acima, se as variáveis aleatórias são não-correlacionadas, então a variância da soma é a soma das variâncias.

6.2 Correlação

Nesta seção vamos introduzir o coeficiente de correlação, estudar suas propriedades, e relacioná-lo com o método dos mínimos quadrados. Observamos

que esta seção não será usada no restante do livro.

Dada uma variável aleatória X não-degenerada com segundo momento finito, definimos a *padronização de* X como sendo a variável aleatória

$$\tilde{X} = \frac{X - \mathbb{E}X}{\sqrt{\mathbb{V}X}}.$$

Ou seja, a padronização é a transformação afim crescente que leva X em $\tilde{X}$ com as propriedades que $\mathbb{E}\tilde{X} = 0$ e $\mathbb{V}\tilde{X} = 1$. Observe também que a padronização de X é uma variável aleatória adimensional, pois $\sqrt{\mathbb{V}X}$ é medido na mesma unidade de X.

Definição 6.15 (Coeficiente de correlação)**.** Dadas duas variáveis aleatórias X e Y com variâncias finitas e positivas, definimos o *coeficiente de correlação* $\rho(X, Y)$ entre X e Y como:

$$\rho(X, Y) = \mathbf{Cov}\left(\frac{X - \mathbb{E}X}{\sigma(X)}, \frac{Y - \mathbb{E}Y}{\sigma(Y)}\right).$$

O coeficiente de correlação é adimensional, pois ele depende apenas das padronizações de X e Y. Outras propriedades do coeficiente de correlação são dadas na proposição a seguir.

Proposição 6.16 (Propriedades do coeficiente de correlação)**.** *Dadas X e Y variáveis aleatórias com variâncias finitas e positivas, valem:*

(1) $\rho(X, Y) = \rho(Y, X)$;
(2) $\rho(X, Y) = \frac{\mathbf{Cov}(X,Y)}{\sigma(X)\sigma(Y)}$;
(3) $\rho(X, X) = 1$;
(4) $\rho(aX + b, Y) = \frac{a}{|a|}\rho(X, Y)$ *se* $a, b \in \mathbb{R},\ a \neq 0$;
(5) $\rho(aX + b, cY + d) = \frac{ac}{|ac|}\rho(X, Y)$ *se* $a, b, c, d \in \mathbb{R}$ *e* $a, c \neq 0$.

Demonstração. O item (1) é imediato da definição de coeficiente de correlação e da simetria da covariância. Para o item (2) basta utilizar o item (2) da Proposição 6.14. Do item (2) segue que $\rho(X, X) = \frac{\mathbf{Cov}(X,X)}{(\sigma(X))^2} = \frac{\mathbb{V}X}{(\sigma(X)^2)} = 1$,

o que prova o item (3). Para mostrar o item (4), calculamos

$$\begin{aligned}\rho(aX+b,Y) &= \frac{\mathbf{Cov}(aX+b,Y)}{\sigma(aX+b)\sigma(Y)} = \frac{a\,\mathbf{Cov}(X,Y)+b\,\mathbf{Cov}(1,Y)}{|a|\sigma(X)\sigma(Y)}\\ &= \frac{a\,\mathbf{Cov}(X,Y)}{|a|\sigma(X)\sigma(Y)} = \frac{a}{|a|}\rho(X,Y),\end{aligned}$$

onde a segunda e a terceira igualdades acima seguem dos itens (2) e (1) da Proposição 6.14, respectivamente.

Provamos o item (5) utilizando os itens (1) e (4):

$$\begin{aligned}\rho(aX+b,cY+d) &= \tfrac{a}{|a|}\rho(X,cY+d) = \tfrac{a}{|a|}\rho(cY+d,X)\\ &= \frac{ac}{|ac|}\rho(Y,X) = \tfrac{ac}{|ac|}\rho(X,Y).\end{aligned}$$ □

Exemplo 6.17. Sejam (X,Y) vetor aleatório com densidade conjunta dada por $f_{XY}(x,y) = \mathbb{1}_{[0,1]}(x)\mathbb{1}_{[0,1]}(y)$, $Z = \min\{X,Y\}$ e $W = \max\{X,Y\}$, então:

$$\begin{aligned}\mathbb{E}[ZW] = \mathbb{E}[XY] &= \int_0^1\int_0^1 xy\,\mathrm{d}x\mathrm{d}y = \frac{1}{4}\\ \mathbb{E}Z &= \int_0^1\left[\int_0^x y\mathrm{d}y + \int_x^1 x\mathrm{d}y\right]\mathrm{d}x = \int_0^1 (\tfrac{x^2}{2}+x-x^2)\mathrm{d}x = \frac{1}{3}\\ \mathbb{E}W &= \int_0^1\left[\int_0^x x\mathrm{d}y + \int_x^1 y\mathrm{d}y\right]\mathrm{d}x = \int_0^1 (x^2+\tfrac{1}{2}-\tfrac{x^2}{2})\mathrm{d}x = \frac{2}{3}\end{aligned}$$

Logo, $\mathbf{Cov}(Z,W) = \mathbb{E}[ZW] - \mathbb{E}Z\cdot\mathbb{E}W = \frac{1}{36}$.

Continuando,

$$\begin{aligned}\mathbb{E}Z^2 &= \int_0^1\left[\int_0^x y^2\mathrm{d}y + \int_x^1 x^2\mathrm{d}y\right]\mathrm{d}x = \int_0^1 (\tfrac{x^3}{3}+x^2-x^3)\mathrm{d}x = \frac{1}{6}\\ \mathbb{V}Z &= \mathbb{E}Z^2 - (\mathbb{E}Z)^2 = \frac{1}{6}-\frac{1}{9} = \frac{1}{18}\\ \mathbb{V}W &= \cdots\text{exercício}\cdots = \frac{1}{18}\end{aligned}$$

$$\rho(Z,W) = \frac{\mathbf{Cov}(Z,W)}{\sigma(Z)\sigma(W)} = \frac{1/36}{\sqrt{1/18}\sqrt{1/18}} = \frac{1}{2}. \qquad \triangle$$

Já sabíamos que o coeficiente de correlação é invariante pela padronização das variáveis. O último item da proposição acima nos diz algo mais forte. O valor absoluto do coeficiente de correlação $\rho(X,Y)$ é preservado por quaisquer transformações afins não-constantes que façamos nas variáveis X e Y.

O coeficiente de correlação é uma indicação do grau de dependência linear entre as variáveis aleatórias X e Y. A proposição a seguir dá ainda mais sentido a esta afirmação.

Proposição 6.18. *Sejam X e Y variáveis aleatórias não-degeneradas com segundo momento finito. Então $-1 \leqslant \rho(X,Y) \leqslant 1$. Ademais, $\rho(X,Y) = \pm 1$ se, e somente se, $Y = \pm aX + b$ q.c. para algum $a > 0$ e $b \in \mathbb{R}$.*

Veremos a demonstração na próxima seção, como corolário da Desigualdade de Cauchy-Schwarz.

Correlação e o método dos mínimos quadrados

O leitor talvez se lembre dos laboratórios de ciências naturais, em que escolhiam-se distintos valores $x_1, \ldots, x_n$ de uma determinada grandeza, observavam-se valores correspondentes $y_1, \ldots, y_n$ de uma outra grandeza que supostamente depende da primeira, e tentava-se traçar a reta $y = ax + b$ que melhor se aproximasse dos n pontos $(x_1, y_y), \ldots, (x_n, y_n)$ no plano, como ilustrado na Figura 6.1. O critério mais comum para dizer que uma reta $y = ax + b$ se aproxime desses pontos mais que outras é o de minimizar o erro quadrático médio, dado por $\frac{1}{n}\sum_j (ax_j + b - y_j)^2$.

Usando notação $\overline{z_j} = \frac{1}{n}\sum_j z_j$, queremos minimizar $g(a,b) = \overline{(ax_j + b - y_j)^2}$. O ponto que minimiza g satisfaz $\vec{\nabla} g = \vec{0}$, ou seja, $\frac{\partial g}{\partial a} = \frac{\partial g}{\partial b} = 0$.

Calculando as derivadas parciais, obtemos

$$\tfrac{\partial}{\partial a} g(a,b) = \overline{2x_j(ax_j + b - y_j)} = 2a\,\overline{x_j^2} + 2b\,\overline{x_j} - 2\,\overline{x_j y_j},$$

$$\frac{\partial}{\partial b} g(a,b) = 2\,\overline{ax_j + b - y_j} = 2a\,\overline{x_j} + 2b - 2\,\overline{y_j}.$$

Resolvendo o sistema linear $\frac{\partial g}{\partial a} = \frac{\partial g}{\partial b} = 0$, obtemos

$$a = \frac{\overline{x_j y_j} - \overline{x_j} \cdot \overline{y_j}}{\overline{x_j^2} - \overline{x_j}^2} \quad \text{e} \quad b = \overline{y_j} - a\,\overline{x_j}.$$

Agora considere o experimento aleatório que consiste em selecionar um desses n pontos ao acaso, ou seja, assuma que (X, Y) é um vetor aleatório que assume os valores $(x_1, y_1), \ldots, (x_n, y_n)$ com probabilidade $\frac{1}{n}$ cada. Neste caso, a solução acima pode ser escrita como

$$a = \frac{\mathbf{Cov}(X,Y)}{\mathbb{V}X} = \rho(X,Y)\frac{\sigma(Y)}{\sigma(X)} \quad \text{e} \quad b = \mathbb{E}Y - a\,\mathbb{E}X.$$

Ou seja, se padronizamos tanto X quanto Y, a reta passará pela origem e sua inclinação será justamente o coeficiente de correlação.

Ademais, o problema original pode ser reescrito como: encontre a e b tal que

$$\mathbb{E}(aX + b - Y)^2$$

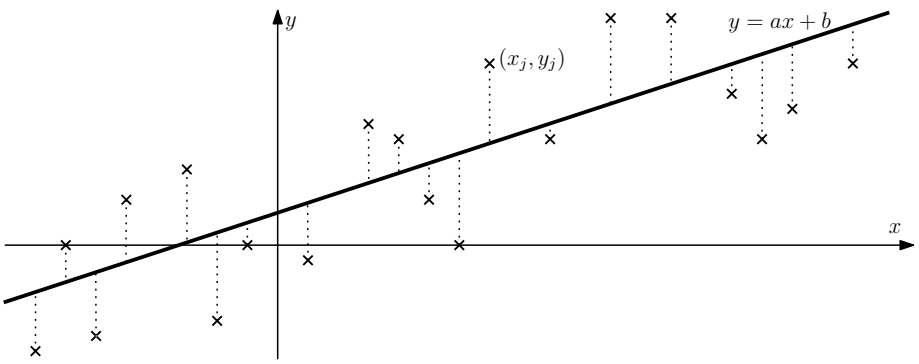


Figura 6.1. Uma coleção de 20 pontos e a reta que minimiza a soma dos quadrados dos comprimentos dos segmentos tracejados.

seja o menor possível. Ou seja, de todas as variáveis aleatórias $\hat{Y}$ que podem ser expressas como $\hat{Y} = aX + b$ para algum a e algum b, encontramos aquela que minimiza $\mathbb{E}(\hat{Y} - Y)^2$.

6.3 Desigualdades básicas

Nesta seção, provaremos uma série de desigualdades de fundamental importância, que fornecem estimativas para probabilidades de eventos ou momentos de variáveis aleatórias.

O primeiro teorema é emblemático, a simplicidade de sua prova contrasta com sua enorme aplicabilidade, como veremos ao longo deste livro. Repare que começamos a utilizar esperança e momentos para estimar probabilidades.

Teorema 6.19 (Desigualdade de Markov)**.** *Sejam X uma variável aleatória, $\lambda > 0$ e $t > 0$. Então*

$$\mathbb{P}(|X| \geqslant \lambda) \leqslant \frac{\mathbb{E}|X|^t}{\lambda^t}.$$

Demonstração. Como $|X|^t \geqslant \lambda^t \cdot \mathbb{1}_{\{|X|^t \geqslant \lambda^t\}}$, segue que $\mathbb{E}|X|^t \geqslant \lambda^t \cdot \mathbb{P}(|X|^t \geqslant \lambda^t)$. Portanto, $\mathbb{P}(|X| \geqslant \lambda) = \mathbb{P}(|X|^t \geqslant \lambda^t) \leqslant \frac{\mathbb{E}|X|^t}{\lambda^t}$. □

Teorema 6.20 (Desigualdade de Tchebyshev)**.** *Seja X uma variável aleatória integrável e seja $\lambda > 0$ uma constante. Então*

$$\mathbb{P}(|X - \mathbb{E}X| \geqslant \lambda) \leqslant \frac{\mathbb{V}X}{\lambda^2}.$$

Demonstração. Aplicamos a Desigualdade de Markov a $X - \mathbb{E}X$ com $t = 2$:

$$\mathbb{P}(|X - \mathbb{E}X| \geqslant \lambda) \leqslant \frac{\mathbb{E}[(X - \mathbb{E}X)^2]}{\lambda^2} = \frac{\mathbb{V}X}{\lambda^2}.$$ □

Exemplo 6.21. Estimar a probabilidade de uma variável aleatória X não diferir de sua média μ por mais que duas vezes o valor do seu desvio-padrão

σ. Usando a Desigualdade de Tchebyshev,

$$\begin{aligned}\mathbb{P}(\mu - 2\sigma < X < \mu + 2\sigma) &= 1 - \mathbb{P}(|X - \mathbb{E}X| \geqslant 2\sigma) \\ &\geqslant 1 - \frac{\mathbb{V}X}{(2\sigma)^2} = 1 - \frac{\sigma^2}{4\sigma^2} = \frac{3}{4}.\end{aligned}$$

△

Veremos agora uma desigualdade muito útil, que diz respeito a funções convexas. Dado um intervalo aberto $I \subseteq \mathbb{R}$, dizemos que $g : I \to \mathbb{R}$ é uma função *convexa* se

$$g(ax + by) \leqslant ag(x) + bg(y)$$

para quaisquer $x, y \in I$ e $a, b \in [0, 1]$ com $a + b = 1$. Essa condição de convexidade pode ser reescrita da seguinte maneira: para todos $x < z < y$ em I, vale

$$\frac{g(z) - g(x)}{z - x} \leqslant \frac{g(y) - g(z)}{y - z}. \tag{6.22}$$

Podemos obter outras caracterizações de convexidade explorando a possível diferenciabilidade de g. Se g' existe e é não-decrescente em todo I, então pelo Teorema do Valor Médio g satisfaz (6.22) e portanto é convexa. Em particular, se g'' existe e é não-negativa em todo I, então g é convexa. São convexas em $\mathbb{R}$ as funções $g(x) = x$, $g(x) = e^x$ e $g(x) = |x|^p$ com $p \geqslant 1$. As funções $g(x) = x^{-1}$, $g(x) = -\sqrt{x}$ e $g(x) = -\log x$ são convexas em $(0, +\infty)$.

Teorema 6.23 (Desigualdade de Jensen)**.** *Seja $I \subseteq \mathbb{R}$ um intervalo aberto, $g : I \to \mathbb{R}$ uma função convexa, e X uma variável aleatória integrável assumindo valores em I. Então $\mathbb{E}[g(X)]$ está definida e*

$$\mathbb{E}[g(X)] \geqslant g(\mathbb{E}X).$$

Observamos que o teorema acima não exclui a possibilidade de $\mathbb{E}[g(x)] = +\infty$.

Demonstração. A prova é ilustrada na Figura 6.2. Preliminarmente,

afirmamos que, para cada $z \in I$ fixo, existe $c \in \mathbb{R}$ tal que

$$g(w) \geqslant g(z) + c(w - z)$$

para todo $w \in I$ (caso g seja diferenciável, podemos tomar $c = g'(z)$ e estamos falando que o gráfico de g está acima de suas retas tangentes). Com efeito, considerando os possíveis valores dos lados esquerdo e direito de (6.22) e usando o Teorema A.1, obtemos $c \in \mathbb{R}$ tal que $\frac{g(z)-g(x)}{z-x} \leqslant c \leqslant \frac{g(y)-g(z)}{y-z}$ para todo $x < z$ e todo $y > z$, provando a afirmação.

Finalmente, tomando $z = \mathbb{E}X$ e usando X no lugar de w, obtemos

$$\mathbb{E}[g(X)] \geqslant \mathbb{E}\big[g(\mathbb{E}X) + c(X - \mathbb{E}X)\big] = g(\mathbb{E}X) + c\,\mathbb{E}X - c\,\mathbb{E}X = g(\mathbb{E}X),$$

o que conclui a demonstração. □

Vejamos alguns exemplos comuns de uso da Desigualdade de Jensen.

Exemplo 6.24. Se X é integrável e $p \geqslant 1$, então

$$\mathbb{E}|X|^p \geqslant (\mathbb{E}|X|)^p \geqslant |\mathbb{E}X|^p \quad \text{e} \quad \mathbb{E}[e^X] \geqslant e^{\mathbb{E}X}.$$

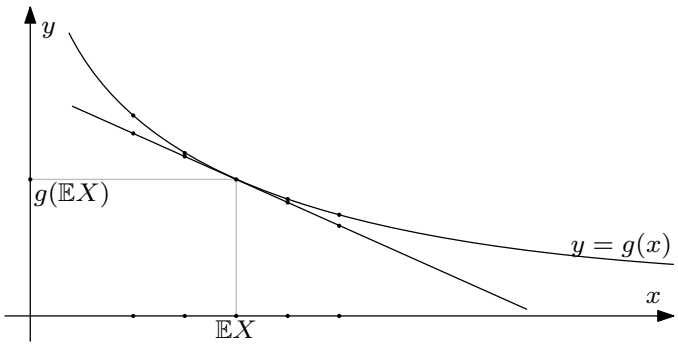


Figura 6.2. Prova da desigualdade de Jensen.

Se X é integrável e positiva, então

$$\mathbb{E}[\tfrac{1}{X}] \geqslant \tfrac{1}{\mathbb{E}X} \quad \text{e} \quad \mathbb{E}[\log X] \leqslant \log(\mathbb{E}X).$$

Com efeito, a primeira desigualdade é obtida usando-se a desigualdade de Jensen com $|X|$ no lugar de X e $g(x) = |x|^p$, a segunda usa $g(x) = |x|$, e as outras são imediatas da Desigualdade de Jensen. $\triangle$

O próximo teorema é uma importante aplicação da Desigualdade de Jensen.

Teorema 6.25 (Desigualdade de Lyapunov)**.** *Sejam X uma variável aleatória e $0 < q \leqslant p$. Então*

$$\big(\mathbb{E}|X|^q\big)^{\frac{1}{q}} \leqslant \big(\mathbb{E}|X|^p\big)^{\frac{1}{p}}.$$

Demonstração. Se $\mathbb{E}|X|^p = +\infty$ a desigualdade vale trivialmente. Suponha que $\mathbb{E}|X|^p < \infty$. Como $|x|^q \leqslant 1 + |x|^p$, segue que $|X|^q$ é integrável.

Observando que a função $g(x) = |x|^{p/q}$ é convexa, temos pela Desigualdade de Jensen que $(\mathbb{E}|X|^p) = \mathbb{E}[(|X|^q)^{p/q}] \geqslant (\mathbb{E}|X|^q)^{p/q}$. Elevando todos os termos a $1/p$, obtemos a desigualdade desejada. $\square$

Terminamos esta seção com a Desigualdade de Cauchy-Schwarz e uma de suas principais aplicações, a Desigualdade de Paley-Zygmund. Apesar de fundamentais em diversas aplicações, essas desigualdades não serão usadas no restante deste livro e podem ser omitidas em um curso introdutório.

Teorema 6.26 (Desigualdade de Cauchy-Schwarz)**.** *Se X e Y têm segundo momento finito, então XY é integrável e*

$$\mathbb{E}[XY] \leqslant \sqrt{\mathbb{E}X^2}\,\sqrt{\mathbb{E}Y^2}.$$

Ainda, se $\mathbb{E}[XY] = \sqrt{\mathbb{E}X^2}\,\sqrt{\mathbb{E}Y^2}$, então existe $c \geqslant 0$ tal que $\mathbb{P}(Y = cX) = 1$, ou então $\mathbb{P}(X = 0) = 1$.

Demonstração. Primeiro veja que XY é integrável porque $|XY| \leqslant X^2 + Y^2$. Sejam $a = \sqrt{\mathbb{E}X^2}$ e $b = \sqrt{\mathbb{E}Y^2}$. Se $a = 0$ ou $b = 0$, o teorema vale trivialmente.

Assumimos então que $0 < a < \infty$ e $0 < b < \infty$. Observamos que

$$0 \leqslant \mathbb{E}\left(\frac{X}{a} - \frac{Y}{b}\right)^2 = \mathbb{E}\left(\frac{X^2}{a^2} - 2\frac{XY}{ab} + \frac{Y^2}{b^2}\right) = 2 - \frac{2\,\mathbb{E}[XY]}{ab},$$

donde

$$\mathbb{E}[XY] \leqslant ab = \sqrt{\mathbb{E}X^2}\,\sqrt{\mathbb{E}Y^2}.$$

Se $\mathbb{E}[XY] = ab$, vale a igualdade na equação acima, donde $\mathbb{E}\left(\frac{X}{a} - \frac{Y}{b}\right)^2 = 0$, logo $\mathbb{P}(\frac{X}{a} - \frac{Y}{b} = 0) = 1$ e portanto $\mathbb{P}(Y = cX) = 1$ com $c = \frac{b}{a}$. □

Munidos dessa desigualdade, podemos finalmente provar a Proposição 6.18.

Demonstração da Proposição 6.18. Sejam $\tilde{X}$ e $\tilde{Y}$ as padronizações de X e Y, respectivamente. Então,

$$\rho(X, Y) = \mathbf{Cov}(\tilde{X}, \tilde{Y}) = \mathbb{E}[\tilde{X}\tilde{Y}] \leqslant \sqrt{\mathbb{E}\tilde{X}^2}\,\sqrt{\mathbb{E}\tilde{Y}^2} = 1,$$

onde a última desigualdade é a Desigualdade de Cauchy-Schwarz. Suponha que $\rho(X, Y) = 1$. Neste caso, vale a igualdade na equação acima. Pela recíproca da Desigualdade de Cauchy-Schwarz, $\tilde{X} = 0$ q.c. ou $\tilde{Y} = c\tilde{X}$ q.c. com $c \geqslant 0$. Por outro lado, como $\mathbb{V}\tilde{X} = \mathbb{V}\tilde{Y} = 1$, segue que $\tilde{Y} = \tilde{X}$ q.c. e, portanto, $Y = aX + b$ q.c. com algum $a > 0$ e $b \in \mathbb{R}$. Reciprocamente, se $Y = aX + b$ q.c. com algum $a > 0$ e $b \in \mathbb{R}$, então $\tilde{Y} = \tilde{X}$ q.c., donde $\rho(X, Y) = 1$. Repetindo-se o mesmo argumento com $-X$ no lugar de X, obtemos que $\rho(X, Y) \geqslant -1$ valendo a igualdade se, e somente se, $Y = aX + b$ q.c. com algum $a < 0$ e $b \in \mathbb{R}$. □

Exemplo 6.27. Sejam $X = \mathbb{1}_A$ e $Y = \mathbb{1}_B$ variáveis aleatórias de Bernoulli com parâmetro p, onde A e B são eventos independentes. Então,

$$\mathbb{E}[XY] = \mathbb{E}[\mathbb{1}_A\mathbb{1}_B] = \mathbb{E}[\mathbb{1}_{A\cap B}] = \mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B) = p^2.$$

Por outro lado,

$$\sqrt{\mathbb{E}X^2}\,\sqrt{\mathbb{E}Y^2} = \sqrt{\mathbb{E}X}\,\sqrt{\mathbb{E}Y} = p.$$

Como $p^2 \leqslant p$, a Desigualdade de Cauchy-Schwarz é satisfeita, valendo a igualdade nos casos extremos $p = 0$ e $p = 1$. $\triangle$

As desigualdades a seguir cotam as probabilidades de uma variável aleatória ser grande ou pequena em função dos seus dois primeiros momentos. Elas são também conhecidas como método do primeiro e segundo momentos.

Teorema 6.28. *Seja N uma variável aleatória assumindo valores inteiros e não-negativos. Então,*

$$\mathbb{P}\big(N > 0\big) \leqslant \mathbb{E}N.$$

Demonstração. Como $\{N > 0\} = \{N \geqslant 1\}$, aplicando a Desigualdade de Markov, obtemos $\mathbb{P}(N > 0) = \mathbb{P}(N \geqslant 1) \leqslant \mathbb{E}N$. $\square$

Teorema 6.29 (Desigualdade de Paley-Zygmund)**.** *Seja X uma variável aleatória não-negativa com segundo momento finito. Para todo $0 \leqslant a < 1$ vale*

$$\mathbb{P}(X > a\,\mathbb{E}X) \geqslant (1-a)^2 \frac{(\mathbb{E}X)^2}{\mathbb{E}X^2}.$$

Demonstração. Basta escrever

$$\mathbb{E}X = \mathbb{E}\Big[X \cdot \mathbb{1}_{\{X \leqslant a\,\mathbb{E}X\}}\Big] + \mathbb{E}\Big[X \cdot \mathbb{1}_{\{X > a\,\mathbb{E}X\}}\Big] \leqslant a\,\mathbb{E}X + \mathbb{E}\Big[X \cdot \mathbb{1}_{\{X > a\,\mathbb{E}X\}}\Big].$$

Aplicando a Desigualdade de Cauchy-Schwarz ao último termo, obtemos

$$\mathbb{E}X \leqslant a\,\mathbb{E}X + \Big(\mathbb{E}X^2 \cdot \mathbb{E}\Big[\mathbb{1}^2_{\{X > a\,\mathbb{E}X\}}\Big]\Big)^{\frac{1}{2}}.$$

Logo,

$$(1-a)\,\mathbb{E}X \leqslant \Big(\mathbb{E}X^2 \cdot \mathbb{P}(X > a\,\mathbb{E}X)\Big)^{\frac{1}{2}}.$$

O que conclui a prova do teorema. $\square$

O caso especial $a = 0$ já é extremamente interessante, pois dá uma cota para que a variável aleatória assuma valores não-nulos a partir da estimativa de seu segundo momento em termos do quadrado do seu primeiro momento.

6.4 Exercícios

§6.1

1. Calcule $\mathbb{V}X$, onde:

(a) $X \sim \text{Geom}(\lambda)$.
(b) $X \sim \text{Poisson}(\lambda)$.
(c) $X \sim \text{Exp}(\lambda)$.
(d) $X \sim \text{Laplace}(a, b)$.

2. Sejam $X \sim \mathcal{U}[0, 3]$ e $Y = \max\{X, 1\}$, calcule $\mathbb{V}Y$.

3. Sejam $X \sim \mathcal{U}[0, 2\pi]$, $Y = \operatorname{sen} X$ e $Z = \cos X$ variáveis aleatórias. Calcule $\mathbf{Cov}(Y, Z)$. As variáveis Y e Z são independentes?

4. Sejam X uma variável aleatória e $a < b$ tais que $\mathbb{P}(a \leqslant X \leqslant b) = 1$.

(a) Mostre que $\mathbb{V}X \leqslant (\mathbb{E}X - a)(b - \mathbb{E}X)$.
(b) Mostre que vale a igualdade na desigualdade acima se, e somente se, $\mathbb{P}(X = a) + \mathbb{P}(X = b) = 1$.

Sugestão: Faça primeiro supondo que $a = 0$ e $b = 1$.

5. Mostre que, se X é integrável, então o mínimo de $\mathbb{E}(X - c)^2$ é atingido quando $c = \mathbb{E}X$.

6. Dizemos que o número real m é uma *mediana* para a variável aleatória X, se $\mathbb{P}(X \geqslant m) \geqslant \frac{1}{2}$ e $\mathbb{P}(X \leqslant m) \geqslant \frac{1}{2}$. Mostre que, se m uma mediana de X, então o mínimo de $\mathbb{E}|X - c|$ é atingido quando $c = m$.

7. Seja X uma variável aleatória discreta com função de probabilidade

$$p_X(n) = \frac{n\lambda^n e^{-\lambda}}{\lambda\, n!}, \ n = 0, 1, 2, 3, \ldots$$

Calcule $\mathbb{V}X$. Dica: Desenvolver $(n-1)(n-2+1)+2(n-1)+1$.

8. Seja X uma variável aleatória absolutamente contínua com função densidade dada por

$$f_X(x) = \begin{cases} \frac{2}{\pi r^2}\sqrt{r^2-(x-a)^2}, & \text{se } x \in [a-r, a+r] \\ 0, & \text{se } x \notin [a-r, a+r]. \end{cases}$$

onde r e a são números reais com $r > 0$.

(a) Calcule $\mathbb{E}X$ e $\mathbb{V}X$.
(b) Deduza uma fórmula para o k-ésimo momento central de X.

9. Seja (X, Y) um vetor aleatório com distribuição uniforme no círculo unitário $C = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 < 1\}$, isto é, com densidade conjunta $f_{X,Y}(x, y) = \frac{1}{\pi}\mathbb{1}_C(x, y)$. Calcule $\mathbf{Cov}(X, Y)$.

10. Sejam X e Y variáveis aleatórias independentes com quarto momento finito. Mostre que

$$\mathbb{E}(X+Y-\mathbb{E}[X+Y])^4 = \mathbb{E}(X-\mathbb{E}X)^4 + \mathbb{E}(Y-\mathbb{E}Y)^4 + 6 \cdot \mathbb{V}X \cdot \mathbb{V}Y.$$

11. Sejam X e Y variáveis aleatórias independentes, ambas com distribuição Bernoulli($\frac{1}{2}$)

(a) Mostre que $X + Y$ e $|X - Y|$ são não-correlacionadas.
(b) Elas são independentes?

§6.2

12. Sejam $X_1, \ldots, X_n$ variáveis aleatórias i.i.d. não degeneradas com segundo momento finito. Calcule $\rho(X_1 + \cdots + X_n, X_1)$.

13. Sejam X e Y variáveis aleatórias i.i.d.

(a) Calcule $\rho(X, 2X)$ e $\rho(X, X + Y)$.
(b) Saberia explicar, sem fazer o cálculo do item anterior, por quê os valores de $\rho(X, 2X)$ e $\rho(X, X + Y)$ são iguais ou diferentes?

§6.3

14. Prove que $\mathbb{E}|X| \leqslant \sqrt{\mathbb{E}X^2}$ sem usar as desigualdades de Jensen e Lyapunov.

15. Suponha que X seja uma variável aleatória tal que $\mathbb{E}X = 10$, $\mathbb{P}(X \leqslant 7) = \frac{2}{10}$ e $\mathbb{P}(X \geqslant 13) = \frac{3}{10}$. Prove que $\mathbb{V}X \geqslant \frac{9}{2}$.

16. Considere uma sequência de variáveis aleatórias $X_1, X_2, X_3, \ldots$ i.i.d. com distribuição Bernoulli(p). Encontre um número n para o qual a média observada, dada por

$$\bar{X}_n(\omega) = \frac{1}{n}\sum_{j=1}^{n} X_j(\omega),$$

não difira da média p por mais de 0,01, com probabilidade mínima de 0,95.

17. Suponha que X seja uma variável aleatória tal que $\mathbb{P}(X \geqslant 0) = 1$ e $\mathbb{P}(X \geqslant 10) = \frac{1}{5}$. Mostre que $\mathbb{E}X \geqslant 2$.

18. Se X é não-degenerada, de quadrado integrável, e $X \geqslant 0$ q.c., prove que

$$\mathbb{P}(X = 0) \leqslant \frac{\mathbb{V}X}{\mathbb{E}X^2} \leqslant \frac{\mathbb{V}X}{(\mathbb{E}X)^2}.$$

19. Sejam X e Y variáveis aleatórias com $\mathbb{E}X = \mathbb{E}Y = 0, \mathbb{V}X = \mathbb{V}Y = 1$ e $\rho = \mathbf{Cov}(X, Y)$.

(a) Mostre que $\mathbb{E}[\max\{X^2, Y^2\}] \leqslant 1 + \sqrt{1-\rho^2}$.

(b) (Tchebyshev bi-dimensional). Conclua que, para todo $\varepsilon > 0$,

$$\mathbb{P}(|X| \geqslant \varepsilon \text{ ou } |Y| \geqslant \varepsilon) \leqslant \frac{1+\sqrt{1-\rho^2}}{\varepsilon^2}.$$

20. Sejam X e Y variáveis aleatórias. Prove que $[F_{X,Y}(x,y)]^2 \leqslant F_X(x)F_Y(y)$ para todos $x, y \in \mathbb{R}$.

Capítulo 7

Convergência de Variáveis Aleatórias

Considere uma sequência de variáveis aleatórias $X_1, X_2, X_3, \ldots$. Em inúmeras situações teóricas e práticas, uma pergunta natural é qual o comportamento de longo prazo da sequência $(X_n)_n$. Dito de outra forma: como se comporta X_n quando n é suficientemente grande?

Tratando-se de variáveis aleatórias, o conceito de convergência é uma generalização do conceito de convergência para números reais. Entretanto, existem várias formas de se fazer essa generalização, e cada forma é a mais natural em determinado contexto. No caso de variáveis aleatórias degeneradas, todas as definições serão equivalentes à convergência de números reais.

7.1 Modos de convergência

Sejam X e $(X_n)_{n\in\mathbb{N}}$ variáveis aleatórias definidas num mesmo espaço de probabilidade $(\Omega, \mathcal{F}, \mathbb{P})$.

Definição 7.1 (Convergência em probabilidade)**.** Dizemos que a sequência

$(X_n)_n$ *converge em probabilidade para* X, denotado por $X_n \xrightarrow{\mathbb{P}} X$, se, para todo $\varepsilon > 0$, vale

$$\mathbb{P}\big(\{\omega \in \Omega : |X_n(\omega) - X(\omega)| \geqslant \varepsilon\}\big) \to 0 \text{ quando } n \to \infty.$$

Exemplo 7.2. Sejam $X_1, X_2, \ldots$ variáveis aleatórias independentes, tais que $X_n \sim \text{Bernoulli}(\frac{1}{n})$. Temos para $\varepsilon < 1$ que

$$\mathbb{P}\big(|X_n - 0| \geqslant \varepsilon\big) = \mathbb{P}(X_n = 1) = \tfrac{1}{n} \to 0,$$

e portanto $X_n \xrightarrow{\mathbb{P}} 0$. △

Exemplo 7.3. Sejam $X_1, X_2, \ldots$ variáveis aleatórias independentes, identicamente distribuídas com distribuição Exp(1) e tome

$$Y_n = \frac{X_n}{\log n}.$$

Então

$$\mathbb{P}\Big(\big|\tfrac{X_n}{\log n} - 0\big| \geqslant \varepsilon\Big) = \mathbb{P}(X_n \geqslant \varepsilon \log n) = n^{-\varepsilon} \to 0,$$

para todo $\varepsilon > 0$, portanto $\frac{X_n}{\log n} \xrightarrow{\mathbb{P}} 0$. △

Exemplo 7.4 (Onda dançante). Sejam $U \sim \mathcal{U}(0,1]$ e $(I_{n,k})_{n,k}$ a sequência de intervalos dada por $I_{n,k} = (\frac{k}{2^n}, \frac{k+1}{2^n}]$, $n \in \mathbb{N}$, $k = 0, \ldots, 2^n - 1$. Definimos as variáveis aleatórias $X_{n,k} = \mathbb{1}_{I_{n,k}}(U)$, que assemelham-se a uma onda dançante (veja Figura 7.1). À medida que n cresce, os intervalos $I_{n,k}$ vão se tornando mais estreitos, de modo que dado qualquer $0 < \varepsilon < 1$,

$$\mathbb{P}(|X_{n,k}| \geqslant \varepsilon) = \mathbb{P}\big(U \in I_{n,k}\big) = 2^{-n} \to 0 \text{ quando } n \to \infty.$$

Podemos ordenar os pares da forma (n,k) sequencialmente da seguinte maneira. Para todo $j \in \mathbb{N}$, existe um único par $(n(j), k(j))$ tal que $j = 2^{n-1} + k$. Tomamos $Y_j = X_{n(j),k(j)}$. Assim, $Y_j \xrightarrow{\mathbb{P}} 0$ quando $j \to \infty$. △

No exemplo acima, para todo $\omega \in \Omega$ fixado, $U(\omega) \in I_{n,k}$ para infinitos valores

de índices (n, k). Com efeito, para cada $n \in \mathbb{N}$ existirá um único valor de $k \in \{0, \ldots, 2^n - 1\}$ tal que $U(\omega) \in I_{n,k}$. Ou seja, $\{\omega \in \Omega : Y_j(\omega) \to 0\} = \emptyset$ apesar de que $Y_j \xrightarrow{\mathbb{P}} 0$. Isto é, a convergência em probabilidade tem suas debilidades, o que nos motiva a fazer a próxima definição.

Definição 7.5 (Convergência quase certa)**.** Dizemos que X_n *converge quase certamente para* X, denotado por $X_n \xrightarrow{\text{q.c.}} X$, se

$$\mathbb{P}\big(\{\omega \in \Omega : X_n(\omega) \to X(\omega) \text{ quando } n \to \infty\}\big) = 1.$$

A convergência quase certa é uma convergência pontual num conjunto de probabilidade 1, ou seja, $X_n(\omega) \to X(\omega)$ para todo ω, exceto em um conjunto de probabilidade nula. Por outro lado, convergência em probabilidade não diz respeito à convergência pontual. Ela apenas afirma que, para valores grandes de n, as variáveis X_n e X são aproximadamente iguais com probabilidade muito alta, conforme ilustrado no Exemplo 7.4.

Exemplo 7.6. Sejam $U \sim \mathcal{U}[0, 1]$ e $(X_n)_n$ a sequência de variáveis aleatórias dada por $X_n = U + U^n$. Afirmamos que $X_n \xrightarrow{\text{q.c.}} U$. Com efeito,

$$\mathbb{P}(X_n \to U) = \mathbb{P}(U^n \to 0) = \mathbb{P}(U \in [0, 1)) = 1$$

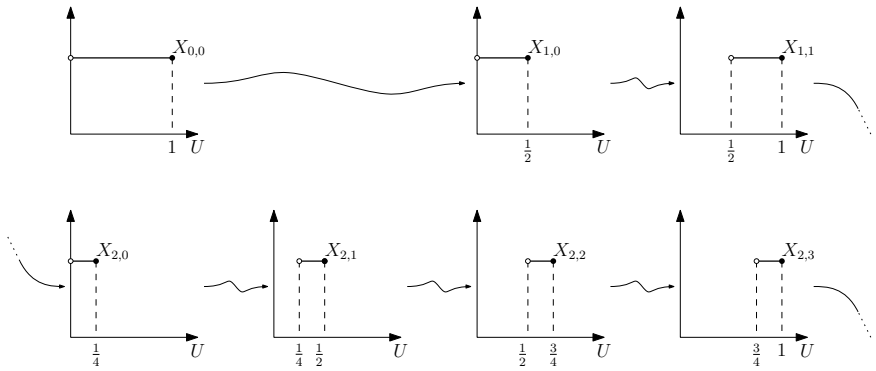


Figura 7.1. Primeiros elementos do contra-exemplo da "onda dançante".

e, portanto, $X_n \overset{\text{q.c.}}{\to} U$. $\triangle$

Definição 7.7 (Convergência em $\mathcal{L}^p$)**.** Dado $p \geqslant 1$, dizemos que X_n *converge para* X *em* $\mathcal{L}^p$, que denotamos por $X_n \overset{\mathcal{L}^p}{\to} X$, se $\mathbb{E}|X_n|^p < \infty$ para todo n e

$$\lim_{n\to\infty} \mathbb{E}|X_n - X|^p = 0.$$

Exemplo 7.8 (Da volta à onda dançante)*.* Seja $(X_{n,k})$ a onda dançante definida no Exemplo 7.4, definimos a "onda dançante crescente" $(\tilde{X}_{n,k})$, onde $\tilde{X}_{n,k} = 2^{n/2} X_{n,k}$, e definimos $(\tilde{Y}_j)_j$ como a correspondente sequência, segundo o mesmo ordenamento feito no Exemplo 7.4. Podemos observar que no caso $p = 1$, vale

$$\mathbb{E}|\tilde{X}_{n,k}| = 2^{-n/2} \to 0 \text{ quando } n \to \infty,$$

logo $\tilde{Y}_j \overset{\mathcal{L}^1}{\to} 0$ quando $j \to \infty$. Por outro lado, no caso $p = 2$, vale

$$\mathbb{E}|\tilde{X}_{n,k}|^2 = 1 \not\to 0,$$

portanto $(\tilde{Y}_j)_j$ não converge para 0 em $\mathcal{L}^2$ quando $j \to \infty$. $\triangle$

Proposição 7.9. *Se* $X_n \overset{\mathcal{L}^p}{\to} X$ *para algum* $p \geqslant 1$*, então* $\mathbb{E}X_n \to \mathbb{E}X$*.*

Demonstração. Pela Desigualdade de Lyapunov,

$$\big|\mathbb{E}X_n - \mathbb{E}X\big| \leqslant \mathbb{E}|X_n - X| \leqslant (\mathbb{E}|X_n - X|^p)^{\frac{1}{p}} \to 0. \qquad \square$$

Definição 7.10 (Convergência em distribuição)**.** Sejam $(X_n)_n$ e X variáveis aleatórias, $(F_{X_n})_{n\geqslant 1}$ e F_X suas respectivas funções de distribuição. Dizemos que $(X_n)_{n\geqslant 1}$ *converge para* X *em distribuição*, denotado por $X_n \overset{\text{d}}{\to} X$, se

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x),$$

para todo $x \in \mathbb{R}$ ponto de continuidade da função F_X.

Exemplo 7.11*.* Seja $(X_n)_n$ uma sequência de variáveis aleatórias com $X_n \sim$

Geom(p_n), neste caso

$$F_{X_n}(x) = \begin{cases} 0, & \text{se } x < 0, \\ 1 - (1-p_n)^{\lfloor x \rfloor}, & \text{se } x \geqslant 0. \end{cases}$$

Além disso, assuma que $np_n \to \lambda > 0$, quando $n \to \infty$. Mostraremos que a sequência $(\frac{X_n}{n})_n$ converge em distribuição para uma distribuição exponencial de parâmetro λ. Para todos $x \in [0, \infty)$ e $n \geqslant 1$, podemos escrever

$$\begin{aligned} F_{\frac{X_n}{n}}(x) &= \mathbb{P}(X_n \leqslant nx) = \mathbb{P}(X_n \leqslant \lfloor nx \rfloor) \\ &= 1 - (1-p_n)^{\lfloor nx \rfloor} = 1 - (1 - \frac{np_n}{n})^{nx}(1 - \frac{np_n}{n})^{\lfloor nx \rfloor - nx}. \end{aligned}$$

Agora, observando que

$$(1 - \frac{np_n}{n})^{nx} \to e^{-\lambda x} \quad \text{e} \quad (1 - \frac{np_n}{n})^{\lfloor nx \rfloor - nx} \to 1,$$

obtemos que $F_{\frac{X_n}{n}}(x) \to 1 - e^{\lambda x}$ quando $n \to \infty$ para todo $x \geqslant 0$. Como já sabemos, essa expressão é a da função de distribuição de uma variável exponencial de parâmetro λ. $\triangle$

Exemplo 7.12. Seja $X_n = \frac{1}{n}$ para $n \geqslant 1$ e $X = 0$. Então $X_n \xrightarrow{\text{d}} X$, embora $\lim_{n\to\infty} F_n(0) = 0 \neq 1 = F(0)$. Mas como 0 não é ponto de continuidade de F, isto não é problema. $\triangle$

No exemplo acima, temos variáveis aleatórias degeneradas que assumem os valores $\frac{1}{n}$ e 0. Qualquer critério de convergência minimamente razoável deveria incluir esse caso. Por isso, não poderíamos pedir que $F_{X_n}(x) \to F_X(x)$ nos pontos de descontinuidade de F_X.

Gostaríamos de ressaltar que, na convergência em distribuição, não é necessário que as variáveis aleatórias $(X_n)_n$ e X estejam definidas no mesmo espaço de probabilidade, pois essa noção de convergência leva em conta apenas as suas respectivas funções de distribuição.

Vejamos agora um critério para a convergência em distribuição de variáveis

aleatórias assumindo valores em $\mathbb{N}_0$.

Proposição 7.13. *Sejam $(X_n)_n$ e X variáveis aleatórias tomando valores em $\mathbb{N}_0$. Então $X_n \xrightarrow{\mathrm{d}} X$ se, e somente se, $p_{X_n}(k) \to p_X(k)$ para todo $k \in \mathbb{N}_0$.*

Demonstração. Observemos que $F_{X_n}(x) = \sum_{k=0}^{\lfloor x \rfloor} p_{X_n}(k)$ e $F_X(x) = \sum_{k=0}^{\lfloor x \rfloor} p_X(k)$ para todo $x \in \mathbb{R}$. Supondo que $p_{X_n}(k) \to p_X(k)$ para todo $k \in \mathbb{N}_0$, isso nos dá $F_n(x) \to F(x)$ para todo $x \in \mathbb{R}$ e, portanto, $X_n \xrightarrow{\mathrm{d}} X$.

Reciprocamente, suponha que $X_n \xrightarrow{\mathrm{d}} X$ e seja $k \in \mathbb{N}_0$. Como $k \pm \frac{1}{2}$ são pontos de continuidade de F_X,

$$p_{X_n}(k) = F_{X_n}(k + \tfrac{1}{2}) - F_{X_n}(k - \tfrac{1}{2}) \to F_X(k + \tfrac{1}{2}) - F_X(k - \tfrac{1}{2}) = p_X(k)$$

quando $n \to \infty$, concluindo a prova. □

Exemplo 7.14. Sejam $(X_n)_n$ e X variáveis aleatórias com $X \sim \text{Poisson}(\lambda)$ e $X_n \sim \text{Binom}(n, \frac{\lambda}{n})$. Conforme vimos na Seção 3.2, $p_{X_n}(k) \to p_X(k)$ para todo $k \in \mathbb{N}_0$. Logo, pela Proposição 7.13, $X_n \xrightarrow{\mathrm{d}} X$. △

Finalizamos esta seção discutindo a unicidade dos limites de sequências de variáveis aleatórias. Se $X_n \xrightarrow{\mathrm{d}} Y$ e $X_n \xrightarrow{\mathrm{d}} Z$, então $Y \sim Z$, veremos a prova na Seção 7.4. Se $X_n \to Y$ e $X_n \to Z$ quase certamente, em probabilidade ou em $\mathcal{L}^p$, então $Y = Z$ quase certamente, deixamos a prova como exercício.

7.2 Lema de Borel-Cantelli

Para estudarmos a convergência de variáveis aleatórias, na "maioria" ou em "quase todos" os pontos $\omega \in \Omega$, precisaremos definir certos eventos que nos permitirão representar a convergência $X_n \to X$ de modo mais tratável.

Definição 7.15 (Limite superior de uma sequência de conjuntos)**.** Seja $(A_n)_n$ uma sequência de subconjuntos de Ω. Definimos o evento $\limsup_{n\to\infty} A_n$,

também denotado por A_n *infinitas vezes*, ou simplesmente A_n *i.v.* como

$$\limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k.$$

Em palavras, o evento definido acima é o conjunto de todos ω tais que, para todo $n \in \mathbb{N}$ existe $k \geqslant n$ para o qual $\omega \in A_k$, isto é, $\omega \in A_n$ para infinitos eventos da sequência $(A_n)_n$, daí a nomenclatura A_n infinitas vezes.

Vejamos como a noção de uma sequência de eventos ocorrer infinitas vezes se relaciona com a convergência de variáveis aleatórias. Relembremos que uma sequência de números reais $(x_n)_n$ converge para o número real x se, para todo $\varepsilon > 0$, temos que $|x_n - x| < \varepsilon$ para todo n suficientemente grande, ou de modo equivalente $|x_n - x| < \varepsilon$ para todo n, exceto para uma quantidade finita de valores de n. Portanto, $x_n \not\to x$ se, e somente se, existe $\varepsilon > 0$ tal que $|x_n - x| \geqslant \varepsilon$ para infinitos valores do índice n.

Proposição 7.16. *Seja $(X_n)_n$ uma sequência de variáveis aleatórias. Então $X_n \overset{\text{q.c.}}{\to} X$ se, e somente se, para todo $\varepsilon > 0$,*

$$\mathbb{P}\big(|X_n - X| \geqslant \varepsilon \textit{ infinitas vezes}\big) = 0.$$

Demonstração. Observemos inicialmente a seguinte igualdade de eventos

$$\{X_n \not\to X\} = \{\exists k \in \mathbb{N}, |X_n - X| \geqslant \tfrac{1}{k} \text{ i.v.}\} = \cup_{k\in\mathbb{N}}\{|X_n - X| \geqslant \tfrac{1}{k} \text{ i.v.}\}.$$

Portanto, dada a definição de convergência quase certa, dizer que $X_n \overset{\text{q.c.}}{\to} X$ é equivalente a afirmar que $\mathbb{P}(\cup_{k\in\mathbb{N}}\{|X_n - X| \geqslant \frac{1}{k} \text{ i.v.}\}) = 0$, o que por sua vez é equivalente a dizer que $\mathbb{P}(|X_n - X| \geqslant \frac{1}{k} \text{ i.v.}) = 0$ para todo $k \in \mathbb{N}$.

Dado qualquer $\varepsilon > 0$, sempre existe $k \in \mathbb{N}$ tal que $\varepsilon \geqslant \frac{1}{k}$. Sendo assim, as afirmações $\mathbb{P}(|X_n - X| \geqslant \frac{1}{k} \text{ i.v.}) = 0$ para todo $k \in \mathbb{N}$ e $\mathbb{P}(|X_n - X| \geqslant \varepsilon \text{ i.v.}) = 0$ para todo $\varepsilon > 0$ são equivalentes, o que completa esta demonstração. $\square$

Uma pergunta que surge naturalmente a partir da proposição acima é: como

estimar a probabilidade de uma sequência de eventos ocorrer infinitas vezes? O Lema de Borel-Cantelli, que enunciaremos a seguir, fornece respostas precisas para esta pergunta.

Teorema 7.17 (Lema de Borel-Cantelli)**.** *Sejam* $(\Omega, \mathcal{F}, \mathbb{P})$ *um espaço de probabilidade e* $(A_n)_n$ *uma sequência de eventos.*

(a) *Se* $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$, *então* $\mathbb{P}(A_n \ i.v.) = 0$.
(b) *Se* $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = +\infty$ *e os eventos* $(A_n)_n$ *são independentes, então* $\mathbb{P}(A_n \ i.v.) = 1$.

Demonstração. Para a primeira parte, observe inicialmente que $(\cup_{k=n}^{\infty} A_k) \downarrow \{A_n \text{ i.v.}\}$. Portanto,

$$\begin{aligned}
\mathbb{P}(\{A_n \text{ i.v.}\}) &= \mathbb{P}\left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k\right) \\
&= \lim_{n\to\infty} \mathbb{P}\left(\bigcup_{k=n}^{\infty} A_k\right) \\
&\leqslant \lim_{n\to\infty} \sum_{k=n}^{\infty} \mathbb{P}(A_k) = 0.
\end{aligned}$$

O último limite é zero devido à convergência da série $\sum_{n=1}^{\infty} \mathbb{P}(A_n)$.

Para a segunda parte, consideremos o evento $\{A_n \ i.v.\}^c$. Pela Lei de De Morgan,

$$\{A_n \text{ i.v.}\}^c = \left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k\right)^c = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k^c.$$

Observe inicialmente que $(\cap_{k=n}^{\infty} A_k^c) \uparrow \{A_n \text{ i.v.}\}^c$, logo

$$\begin{aligned}
\mathbb{P}(\{A_n \text{ i.v.}\}^c) &= \mathbb{P}\left(\bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k^c\right) = \lim_{n\to\infty} \mathbb{P}\left(\bigcap_{k=n}^{\infty} A_k^c\right) \\
&= \lim_{n\to\infty} \lim_{m\to\infty} \mathbb{P}\left(\bigcap_{k=n}^{m} A_k^c\right) = \lim_{n\to\infty} \lim_{m\to\infty} \prod_{k=n}^{m} (1 - \mathbb{P}(A_k))
\end{aligned}$$

$$\leqslant \lim_{n\to\infty} \lim_{m\to\infty} \prod_{k=n}^{m} e^{-\mathbb{P}(A_k)} = \lim_{n\to\infty} \lim_{m\to\infty} e^{-\sum_{k=n}^{m} \mathbb{P}(A_k)} = 0,$$

onde a quarta igualdade é devida à independência de $(A_n)_n$, e por conseguinte de $(A_n^c)_n$, a cota superior segue da desigualdade $1 - x \leqslant e^{-x}$ para todo $x \in \mathbb{R}$ e a igualdade final segue da hipótese $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = +\infty$. □

Corolário 7.18 (Lei 0-1 de Borel)**.** *Se $(A_n)_n$ é uma sequência de eventos independentes, então $\mathbb{P}(\{A_n \text{ i.v.}\}) = 0$ ou 1.*

Observação 7.19. A hipótese de independência é crucial na segunda parte do Lema de Borel-Cantelli. Com efeito, seja A um evento qualquer com $\mathbb{P}(A) \in (0,1)$ e defina $A_n = A$ para cada $n \in \mathbb{N}$. Neste caso, $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = +\infty$ e $\mathbb{P}(A_n \text{ i.v.}) = \mathbb{P}(A) < 1$. △

Exemplo 7.20. Helena começa com um baralho comum (52 cartas), ela retira de modo equiprovável uma carta do baralho, observa sua face, repõe a carta retirada e acrescenta mais outra carta com um coringa na face. Tal procedimento é repetido indefinidamente: embaralham-se as cartas, uma carta é retirada, sua face é observada, esta retorna ao baralho juntamente com um coringa. De modo que o baralho passa a ter uma carta a mais a cada rodada. Sejam $(A_n)_{n\geqslant 1}$ e $(B_n)_{n\geqslant 2}$, sequências de eventos definidos como

$$A_n = \{\text{Helena tira o } 4\clubsuit \text{ na } n\text{-ésima rodada}\} \text{ e } B_n = A_n \cap A_{n-1};$$

observe que $\mathbb{P}(A_n) = \frac{1}{51+n}$ e $\mathbb{P}(B_n) = \frac{1}{(50+n)(51+n)}$. Pelo Lema de Borel-Cantelli,

$$\mathbb{P}(A_n \text{ i.v.}) = 1 \quad \text{e} \quad \mathbb{P}(B_n \text{ i.v.}) = 0$$

Isto é, apesar de Helena retirar o $4\clubsuit$ infinitas vezes com probabilidade um, ela irá retirá-lo duas vezes consecutivas no máximo finitas vezes com probabilidade um. Observe que foi necessário que os eventos $(A_n)_n$ fossem independentes, embora os eventos $(B_n)_n$ não possuam tal propriedade. △

O uso mais frequente que faremos do Lema de Borel-Cantelli será para obter

a convergência quase certa. Se conseguirmos mostrar que

$$\sum_{n=1}^{\infty} \mathbb{P}(|X_n - X| \geqslant \varepsilon) < \infty$$

para todo $\varepsilon > 0$, podemos concluir, pela Proposição 7.16, que $X_n \overset{\text{q.c.}}{\to} X$.

Exemplo 7.21. Sejam $(\Omega, \mathcal{F}, \mathbb{P})$ um espaço de probabilidade e $(X_n)_n$ uma sequência de variáveis aleatórias i.i.d, com distribuição uniforme em $[0,1]$. Defina a sequência de variáveis aleatórias $(Y_n)_n$ como

$$Y_n = \min\{X_1, \ldots, X_n\}, \ n \geqslant 1.$$

Mostraremos que $Y_n \overset{\text{q.c.}}{\to} 0$. Para todos $n \in \mathbb{N}$ e $\varepsilon \in (0,1)$ temos

$$\begin{aligned}
\mathbb{P}(|Y_n| > \varepsilon) &= \mathbb{P}(X_1 > \varepsilon, \ldots, X_n > \varepsilon) \\
&= \prod_{k=1}^{n} \mathbb{P}(X_k > \varepsilon) \\
&= \mathbb{P}(X_1 > \varepsilon)^n \\
&= (1-\varepsilon)^n,
\end{aligned}$$

onde as quatro igualdades acima seguem da definição de Y_n e do fato de que as $(X_n)_n$ são independentes e identicamente distribuídas com distribuição $\mathcal{U}[0,1]$. Somando as probabilidades,

$$\sum_{n=1}^{\infty} \mathbb{P}(|Y_n| \geqslant \varepsilon) = \sum_{n=1}^{\infty} (1-\varepsilon)^n < \infty.$$

Logo, para cada $\varepsilon \in (0,1)$, segue do Lema de Borel-Cantelli que

$$\mathbb{P}(\{|Y_n| > \varepsilon \text{ i.v.}\}) = 0,$$

e consequentemente $Y_n \overset{\text{q.c.}}{\to} 0$ devido à Proposição 7.16. $\triangle$

Exemplo 7.22. Sejam $(a_n)_n$ e $(p_n)_n$ sequências tais que $a_n \uparrow \infty$, $p_n \downarrow 0$ e

$\sum_n p_n = \infty$, e seja $(X_n)_n$ uma sequência de variáveis aleatórias independentes, distribuídas como

$$\mathbb{P}(X_n = a_n) = p_n \quad \text{e} \quad \mathbb{P}(X_n = 0) = 1 - p_n, \quad n \geqslant 1.$$

Pelo Lema de Borel-Cantelli, $\mathbb{P}(X_n \to 0) = 0$, o que contrasta com o fato de que $X_n \xrightarrow{\mathbb{P}} 0$ quando $n \to \infty$. $\triangle$

Como já vimos anteriormente, devemos ser sempre cautelosos ao utilizar a segunda parte do Lema de Borel-Cantelli. A hipótese de independência deve sempre ser verificada, conforme ilustrado no próximo exemplo.

Exemplo 7.23. Sejam U e $(X_n)_n$ variáveis aleatórias definidas como $U \sim \mathcal{U}[0,1]$ e $X_n = \mathbb{1}_{\{U \leqslant \frac{1}{n}\}}$ para $n \geqslant 1$. Observe que $X_n \xrightarrow{\text{q.c.}} 0$, pois $\mathbb{P}(X_n \to 0) = \mathbb{P}(U \in (0,1]) = 1$. Entretanto, como $\mathbb{P}(X_n = 1) = \frac{1}{n}$ e $\sum_{n=1}^{\infty} \frac{1}{n}$ diverge, poderíamos ser tentados a utilizar a segunda parte do Lema de Borel-Cantelli como no exemplo anterior e concluir que $\mathbb{P}(X_n \to 0) = 0$, o que obviamente não é verdadeiro. Defina os eventos $A_n = \{X_n = 1\}$ para todo n. Observe que os eventos $(A_n)_n$ não são independentes, logo não podemos utilizar a segunda parte do Lema de Borel-Cantelli. $\triangle$

Pela segunda parte do Lema de Borel-Cantelli, quando os eventos $(A_n)_n$ são independentes, $\mathbb{P}(A_n \text{ i.v.}) = 0$ implica $\sum_n \mathbb{P}(A_n) < \infty$. Na ausência de independência isso pode ser falso, mas ainda vale que $\mathbb{P}(A_n) \to 0$.

Proposição 7.24. *Se $\mathbb{P}(A_n$ i.v.$) = 0$, então $\mathbb{P}(A_n) \to 0$ quando $n \to \infty$.*

Demonstração. Tome $B_n = \cup_{k \geqslant n} A_k$. Assim, $B_n \downarrow \{A_n \text{ i.v.}\}$ quando $n \to \infty$. Como $B_n \supseteq A_n$, vale que $\mathbb{P}(A_n) \leqslant \mathbb{P}(B_n) \to \mathbb{P}(A_n \text{ i.v.}) = 0$. $\square$

7.3 Relações entre os modos de convergência

Veremos nesta seção que alguns dos modos de convergência, definidos na Seção 7.1, são mais fortes que outros. Mostraremos todas as implicações

possíveis e ilustraremos com exemplos quando um determinado modo de convergência não implica algum outro.

Proposição 7.25 (q.c. $\Rightarrow \mathbb{P}$)**.** *Se* $X_n \overset{\text{q.c.}}{\to} X$*, então* $X_n \overset{\mathbb{P}}{\to} X$*.*

Demonstração. Para qualquer $\varepsilon > 0$, pela Proposição 7.16,

$$\mathbb{P}(|X_n - X| \geqslant \varepsilon \text{ i.v.}) = 0.$$

Pela Proposição 7.24 segue que $\mathbb{P}(|X_n - X| \geqslant \varepsilon) \to 0$, ou seja, $X_n \overset{\mathbb{P}}{\to} X$. □

Contra-exemplo 7.26 ($\mathcal{L}^p \not\Rightarrow$ q.c.)*.* No Exemplo 7.3,

$$\mathbb{P}\big(\tfrac{X_n}{\log n} \geqslant \varepsilon \text{ infinitas vezes}\big) = 1$$

para $0 < \varepsilon < 1$. Portanto não vale que $\frac{X_n}{\log n} \overset{\text{q.c.}}{\to} 0$. Por outro lado, $\mathbb{E}|\frac{X_n}{\log n}|^p = \mathbb{E}|\frac{X_1}{\log n}|^p \to 0$, donde concluímos que $\frac{X_n}{\log n} \to 0$ em $\mathcal{L}^p$. △

Proposição 7.27 ($\mathcal{L}^p \Rightarrow \mathbb{P}$)**.** *Se* $X_n \overset{\mathcal{L}^p}{\to} X$ *para algum* $p \geqslant 1$*, então* $X_n \overset{\mathbb{P}}{\to} X$*.*

Demonstração. Pela desigualdade de Markov,

$$\mathbb{P}(|X_n - X| \geqslant \varepsilon) \leqslant \frac{\mathbb{E}|X_n - X|^p}{\varepsilon^p} \to 0. \qquad \square$$

Proposição 7.28 ($\mathcal{L}^q \Rightarrow \mathcal{L}^p$)**.** *Se* $q \geqslant p \geqslant 1$ *e* $X_n \overset{\mathcal{L}^q}{\to} X$*, então* $X_n \overset{\mathcal{L}^p}{\to} X$*.*

Demonstração. Pela Desigualdade de Lyapunov,

$$\Big(\mathbb{E}\big|X_n - X\big|^p\Big)^{\frac{1}{p}} \leqslant \Big(\mathbb{E}\big|X_n - X\big|^q\Big)^{\frac{1}{q}} \to 0$$

e $\mathbb{E}|X_n|^p \leqslant (\mathbb{E}|X_n|^q)^{p/q} < \infty$. □

Contra-exemplo 7.29 (q.c. $\not\Rightarrow \mathcal{L}^p$)*.* Suponha que $\mathbb{P}(X_n = n^3) = \frac{1}{n^2} = 1 - \mathbb{P}(X_n = 0)$. Então, para $\varepsilon > 0$, temos $\mathbb{P}(X_n \geqslant \varepsilon) \leqslant \frac{1}{n^2}$, portanto $X_n \overset{\text{q.c.}}{\to} 0$.

Entretanto, $\mathbb{E}X_n = n$, logo não podemos ter $X_n \xrightarrow{\mathcal{L}^1} 0$, e pela proposição acima não podemos ter convergência em $\mathcal{L}^p$ para nenhum $p \geqslant 1$. $\triangle$

Contra-exemplo 7.30 ($\mathcal{L}^p \not\Rightarrow$ q.c.). No Exemplo 7.2,

$$\mathbb{E}|X_n - 0|^p = \mathbb{E}X_n^p = \mathbb{P}(X_n = 1) = \frac{1}{n} \to 0,$$

portanto $X_n \xrightarrow{\mathcal{L}^p} 0$ para todo p. No entanto,

$$\mathbb{P}(X_n = 1 \text{ infinitas vezes}) = 1$$

pelo Lema de Borel-Cantelli. Portanto, não vale $X_n \xrightarrow{\text{q.c.}} 0$. $\triangle$

Exemplo 7.31 (q.c. $\not\Rightarrow \mathcal{L}^1$). Sejam $U \sim \mathcal{U}[0,1]$ e $X_n = n\mathbb{1}_{\{U \leqslant \frac{1}{n}\}}$. Então $X_n \to 0$ q.c., embora X_n não convirja a 0 em $\mathcal{L}^1$, pois $\mathbb{E}|X_n| = 1$. $\triangle$

Os Exemplos 7.8 e 7.31 ilustram que convergência q.c. não implica convergência em $\mathcal{L}^p$ e nem convergência em $\mathcal{L}^p$ implica convergência q.c. Observando as Proposições 7.25 e 7.27, convergência em probabilidade não implica convergência q.c. ou $\mathcal{L}^p$.

Proposição 7.32 (Convergência por subsequências)**.** *Se $X_n \xrightarrow{\mathbb{P}} X$ então existe uma subsequência $n_k \to \infty$ tal que $X_{n_k} \xrightarrow{\text{q.c.}} X$.*

Demonstração. Como $\mathbb{P}(|X_n - X| \geqslant \varepsilon) \to 0$ para todo $\varepsilon > 0$, podemos tomar $n_1 > 0$ tal que $\mathbb{P}(|X_{n_1} - X| \geqslant 1) < \frac{1}{2}$. Novamente, podemos tomar $n_2 > n_1$ tal que $\mathbb{P}(|X_{n_2} - X| \geqslant \frac{1}{2}) < \frac{1}{4}$. Sucessivamente, podemos tomar $n_k > n_{k-1}$ tal que $\mathbb{P}(|X_{n_k} - X| \geqslant \frac{1}{k}) < \frac{1}{2^k}$.

Vamos ver que essa sequência n_k satisfaz $X_{n_k} \xrightarrow{\text{q.c.}} X$. Seja $\varepsilon > 0$. Temos que $\mathbb{P}(|X_{n_k} - X| \geqslant \varepsilon) \leqslant \mathbb{P}(|X_{n_k} - X| \geqslant \frac{1}{k})$ para todo $k \geqslant \varepsilon^{-1}$. Por outro lado, $\sum_k \mathbb{P}(|X_{n_k} - X| \geqslant \frac{1}{k}) < \infty$, logo $\sum_k \mathbb{P}(|X_{n_k} - X| \geqslant \varepsilon) < \infty$. Pelo Lema de Borel-Cantelli, $X_{n_k} - X \xrightarrow{\text{q.c.}} 0$, ou seja, $X_{n_k} \xrightarrow{\text{q.c.}} X$. $\square$

Teorema 7.33 (Teorema da Convergência Dominada em $\mathcal{L}^p$)**.** *Seja $p \geqslant 1$. Se $X_n \xrightarrow{\mathbb{P}} X$ e existe uma variável aleatória Y tal que $\mathbb{E}Y^p < \infty$ e $|X_n| \leqslant Y$ q.c. para todo n, então $X_n \xrightarrow{\mathcal{L}^p} X$.*

Demonstração. Suponha inicialmente que $X_n \to X$ quase certamente. Então $|X| \leqslant Y$ q.c. e $|X_n - X|^p \leqslant (|X_n| + |X|)^p \leqslant (2Y)^p$, que é integrável. Por outro lado, como $|X_n - X|^p \xrightarrow{\text{q.c.}} 0$, pelo Teorema da Convergência Dominada, $\mathbb{E}|X_n - X|^p \to 0$.

Finalmente, suponhamos que $X_n \to X$ em probabilidade. Queremos mostrar que a sequência numérica $(\mathbb{E}|X_n - X|^p)_n$ converge para zero. Isto é equivalente a dizer que qualquer subsequência tem uma subsubsequência que converge a zero (Teorema A.8). Seja $(X_{n_k})_k$ uma subsequência qualquer de $(X_n)_n$. Como $\lim_n X_n = X$ em probabilidade, segue que $\lim_k X_{n_k} = X$ em probabilidade e, pela Proposição 7.32, podemos tomar uma subsubsequência $(X_{n_{k_j}})_j$ tal que $\lim_j X_{n_{k_j}} = X$ q.c. Aplicando o caso anterior, obtemos $\lim_j \mathbb{E}|X_{n_{k_j}} - X|^p = 0$, concluindo a prova. □

Proposição 7.34 ($\mathbb{P} \Rightarrow \mathrm{d}$)**.** *Se $X_n \xrightarrow{\mathbb{P}} X$, então $X_n \xrightarrow{\mathrm{d}} X$.*

Demonstração. Seja $x \in \mathbb{R}$ ponto de continuidade de F_X. Temos que mostrar que $F_{X_n}(x) \to F_X(x)$ quando $n \to \infty$. Seja $\varepsilon > 0$. Tome $\delta > 0$ tal que

$$F_X(x) - \varepsilon < F_X(x - \delta) \leqslant F_X(x) \leqslant F_X(x + \delta) < F_X(x) + \varepsilon.$$

Como $\{X_n \leqslant x\} \subseteq \{|X_n - X| \geqslant \delta\} \cup \{X \leqslant x + \delta\}$ e $X_n \xrightarrow{\mathbb{P}} X$, temos, para todo n suficientemente grande,

$$F_{X_n}(x) \leqslant F_X(x + \delta) + \varepsilon < F_X(x) + 2\varepsilon.$$

Da mesma forma, como $\{X \leqslant x - \delta\} \subseteq \{|X_n - X| \geqslant \delta\} \cup \{X_n \leqslant x\}$ e $X_n \xrightarrow{\mathbb{P}} X$, temos, para todo n suficientemente grande,

$$F_{X_n}(x) \geqslant F_X(x - \delta) - \varepsilon > F_X(x) - 2\varepsilon.$$

Portanto, $F_{X_n}(x) \to F_X(x)$, como queríamos demonstrar. $\square$

A convergência em probabilidade implica a convergência em distribuição, mas a recíproca é falsa. De fato a recíproca nem faz muito sentido já que convergência em distribuição não necessita que as variáveis estejam definidas no mesmo espaço de probabilidade.

Exemplo 7.35 ($d \not\Rightarrow \mathbb{P}$). Sejam $X, (X_n)_n$ variáveis aleatórias i.i.d. com distribuição normal padrão, todas definidas em um mesmo espaço de probabilidade. Observe que $X_n \xrightarrow{\mathrm{d}} X$ quando $n \to \infty$ trivialmente. Por outro lado, como $X_n - X \sim \mathcal{N}(0,2)$ para todo $n \in \mathbb{N}$,

$$\mathbb{P}(|X_n - X| \geqslant \varepsilon) = \mathbb{P}(|\mathcal{N}(0,2)| \geqslant \varepsilon) > 0, \quad \text{para todo } \varepsilon > 0.$$

Logo, $(X_n)_n$ não converge em probabilidade para X. $\triangle$

Conforme vimos nos exemplos acima, em geral convergência quase certa não implica convergência em $\mathcal{L}^p$, e vice-versa, mas a implicação pode valer sob condições particulares. O mesmo vale para a relação entre convergência em distribuição e em probabilidade.

Proposição 7.36. *Seja $(X_n)_n$ uma sequência de variáveis aleatórias definidas no mesmo espaço de probabilidade. Se $X_n \xrightarrow{\mathrm{d}} c$ para $c \in \mathbb{R}$ constante, então $X_n \xrightarrow{\mathbb{P}} c$.*

$$\begin{tikzcd}
 & \text{q.c.} \arrow[dr, Rightarrow] & \\
 & & \mathbb{P} \arrow[ul, dotted, "\text{subsequência}"'] \arrow[r, Rightarrow] \arrow[dl, dotted, "\text{caso dominado}"'] & d \arrow[ul, dotted, bend right, "\text{constante}"'] \\
\mathcal{L}^{p+s} \arrow[r, Rightarrow] & \mathcal{L}^p \arrow[ur, Rightarrow] &
\end{tikzcd}$$

Figura 7.2. Diagrama de implicações entre os tipos de convergência.

Demonstração. Convergência em distribuição a uma variável constante quer dizer que $\lim_n F_{X_n}(t) = 0$ se $t < c$ e $\lim_n F_{X_n}(t) = 1$ se $t > c$. Seja $\varepsilon > 0$. Veja que

$$\begin{aligned}\mathbb{P}(|X_n - c| \geqslant \varepsilon) \leqslant \mathbb{P}(X_n \leqslant c - \varepsilon) + \mathbb{P}(X_n > c + \tfrac{\varepsilon}{2}) = \\ = F_{X_n}(c - \varepsilon) + 1 - F_{X_n}(c + \tfrac{\varepsilon}{2}) \to 0 \text{ quando } n \to \infty.\end{aligned}$$

Como isso vale para todo $\varepsilon > 0$, isso conclui a demonstração. □

Completamos assim o diagrama de implicações da Figura 7.2.

7.4 Mais sobre a convergência em distribuição

Nesta seção faremos um estudo mais profundo da convergência em distribuição. Os teoremas aqui contidos são de natureza mais técnica e serão usados nas Seções 9.3, 9.4 e 10.3.

O lema abaixo nos fornece uma informação interessante sobre o conjunto dos pontos de descontinuidade de funções monótonas, em particular vale para funções de distribuição.

Lema 7.37. *O conjunto dos pontos de descontinuidade de qualquer função monótona $f : I \to \mathbb{R}$, onde I é um intervalo não-degenerado da reta, é enumerável. Em particular, qualquer intervalo não-degenerado $J \subseteq I$ contém pontos onde f é contínua.*

Demonstração. Tome uma sequência de intervalos fechados $([a_k, b_k])_k$, tais que $[a_k, b_k] \uparrow I$ quando $k \to \infty$. Suponhamos que f seja não-decrescente; o caso não-crescente é totalmente análogo. Defina $D = \{x \in \mathbb{R} : f(x+) - f(x-) > 0\}$, o conjunto dos pontos de descontinuidade de f, e $D_{n,k} = \{x \in [a_k, b_k] : f(x+) - f(x-) > \frac{1}{n}\}$ para todo n. Temos que $\#D_{n,k} \leqslant n(f(b_k) - f(a_k))$, logo é finito. Como $D = \cup_{n=1}^{\infty} \cup_{k=1}^{\infty} D_{n,k}$, segue que D é enumerável, pois é união enumerável de conjuntos enumeráveis. Por outro

lado, pelo Corolário A.13, intervalos não-degenerados não são enumeráveis, portanto eles devem conter pontos de continuidade de f. $\square$

Proposição 7.38 (Unicidade do limite em distribuição)**.** *O limite em distribuição é único, isto é, se $X_n \xrightarrow{\mathrm{d}} X$ e $X_n \xrightarrow{\mathrm{d}} Y$ então $X \sim Y$.*

Demonstração. Sejam F_X e F_Y as funções de distribuição de X e Y, D_X e D_Y seus respectivos pontos de descontinuidade. Por hipótese, $F_{X_n}(z) \to F_X(z)$ e $F_{X_n}(z) \to F_Y(z)$ para todo $z \in (D_X \cup D_Y)^c$, logo $F_X(z) = F_Y(z)$, $\forall z \in (D_X \cup D_Y)^c$. Pelo Lema 7.37, $D_X \cup D_Y$ é enumerável, logo para todo $t \in D_X \cup D_Y$, existem $z_1, \ldots, z_n, \cdots \in (D_X \cup D_Y)^c$ tais que $z_n \downarrow t$. Como F_X e F_Y são contínuas à direita,

$$F_X(t) = \lim_{n\to\infty} F_X(z_n) = \lim_{n\to\infty} F_Y(z_n) = F_Y(t).$$

Portanto $F_X(z) = F_Y(z)$, $\forall z \in \mathbb{R}$. $\square$

Conforme vimos na seção anterior, a convergência em distribuição é a mais fraca de todas, enquanto a convergência quase certa é uma das mais fortes. Sendo assim, o próximo teorema é um tanto surpreendente, pois nos garante que quando há a convergência em distribuição de uma sequência de variáveis aleatórias para uma outra variável limite, existem cópias com as mesmas distribuições da sequência original e da variável limite para as quais a convergência é quase certa.

Teorema 7.39 (Acoplamento de Skorokhod)**.** *Se $X_n \xrightarrow{\mathrm{d}} X$, então existe um espaço de probabilidade $(\Omega', \mathcal{F}', \mathbb{P}')$ onde estão definidas variáveis aleatórias Y e $(Y_n)_n$ tais que $Y_n \sim X_n$ para todo n, $Y \sim X$, e $Y_n \xrightarrow{\mathrm{q.c.}} Y$.*

Demonstração. Usaremos a função quantil F^{-1} definida na Seção 3.1.3 como

$$F^{-1}(u) = \min\{x \in \mathbb{R} : F(x) \geqslant u\},$$

que permite construir uma variável aleatória Y com a distribuição desejada F a partir de uma uniforme U, tomando-se $Y = F^{-1}(U)$. Para construir as variáveis aleatórias como no enunciado do teorema, usaremos uma mesma variável aleatória U para todas as F_{X_n}. Isto é, tomamos $Y = F_X^{-1}(U)$ e $Y_n = F_{X_n}^{-1}(U)$. Vamos mostrar que $F_{X_n}^{-1}(u) \to F_X^{-1}(u)$ para todo $u \in (0,1)$ ponto de continuidade de F_X^{-1}. Pelo Lema 7.37, os pontos de descontinuidade de F_X^{-1} são enumeráveis, logo $F_{X_n}^{-1}(U) \overset{\text{q.c.}}{\to} F_X^{-1}(U)$, concluindo a prova.

Seja u ponto de continuidade de F_X^{-1}, e seja $\varepsilon > 0$. Defina $x = F_X^{-1}(u)$. Por continuidade, existem v e w tais que $0 < v < u < w < 1$ e $x - \varepsilon < F_X^{-1}(v) \leqslant x \leqslant F_X^{-1}(w) < x + \varepsilon$. Pelo Lema 7.37, existem y e z pontos de continuidade de F_X tais que $x - \varepsilon < y < F_X^{-1}(v) \leqslant x \leqslant F_X^{-1}(w) < z < x + \varepsilon$. Observe que $F_X(y) < v < u < w \leqslant F_X(z)$, onde na primeira e última desigualdades utilizamos (3.15). Como $F_{X_n}(y) \to F_X(y) < u$ e $F_{X_n}(z) \to F_X(z) > u$, segue que $F_{X_n}(y) < u < F_{X_n}(z)$ para todo n suficientemente grande. Por (3.15), esta última desigualdade implica que $x - \varepsilon < y < F_{X_n}^{-1}(u) \leqslant z < x + \varepsilon$. Portanto, $F_{X_n}^{-1}(u) \to F_X^{-1}(u)$, que é o que queríamos mostrar. □

Com o acoplamento de Skorokhod, estamos aptos a provar de modo simples um critério para convergência em distribuição. Esse critério será usado na demonstração do Teorema do Limite Central que veremos no Capítulo 9.

Teorema 7.40 (Teorema de Helly-Bray)**.** *Sejam $(X_n)_n$ e X variáveis aleatórias definidas em um mesmo espaço de probabilidade. São equivalentes:*

(i) $X_n \overset{\text{d}}{\to} X$

(ii) $\mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)]$ *para toda função* $f : \mathbb{R} \to \mathbb{R}$ *contínua e limitada.*

(iii) $\mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)]$ *para toda função* $f : \mathbb{R} \to \mathbb{R}$ *tal que* f, f', f'', f''' *são contínuas e limitadas.*

Demonstração. Começamos por *(i)* ⇒ *(ii)*. Suponha *(i)* e seja $f : \mathbb{R} \to \mathbb{R}$ contínua e limitada. Pelo acoplamento de Skorokhod, existem $Y \sim X$ e $Y_n \sim X_n$ tais que $Y_n \overset{\text{q.c.}}{\to} Y$. Como f é contínua, $f(Y_n) \overset{\text{q.c.}}{\to} f(Y)$. Como f é

limitada, podemos aplicar o Teorema da Convergência Dominada, obtendo

$$\mathbb{E}[f(X_n)] = \mathbb{E}[f(Y_n)] \to \mathbb{E}[f(Y)] = \mathbb{E}[f(X)],$$

o que prova *(ii)*. A implicação *(ii)* $\Rightarrow$ *(iii)* é trivial.

Agora suponha *(iii)*. Seja c um ponto de continuidade de F_X, e seja $\varepsilon > 0$. Tome ε' tal que $F_X(c - \varepsilon') > F_X(c) - \varepsilon$ e $f : \mathbb{R} \to [0, 1]$ uma função tal que f, f', f'', f''' são contínuas e limitadas, $f(x) = 1$ para todo $x \leqslant c - \varepsilon'$ e $f(x) = 0$ para todo $x \geqslant c$ (por exemplo, há um polinômio apropriado de grau 7 em $[c - \varepsilon, c]$). Então

$$\begin{aligned}\liminf_n F_{X_n}(c) &= \liminf_n \mathbb{E}[\mathbb{1}_{(-\infty,c]}(X_n)] \geqslant \liminf_n \mathbb{E}[f(X_n)] = \\ &= \mathbb{E}[f(X)] \geqslant \mathbb{E}[\mathbb{1}_{(-\infty,c-\varepsilon']}(X)] = F_X(c - \varepsilon') \geqslant F_X(c) - \varepsilon.\end{aligned}$$

Como isso vale para todo $\varepsilon > 0$, concluímos que $\liminf_n F_{X_n}(c) \geqslant F_X(c)$. De forma análoga (trocando $c - \varepsilon'$ por $c + \varepsilon'$, e invertendo as desigualdades), podemos obter $\limsup_n F_{X_n}(c) \leqslant F_X(c)$, o que prova *(i)*. $\square$

Observamos que em alguns textos a afirmação do item *(ii)* do teorema acima é tomada como a definição de convergência em distribuição. Isto permite estender a definição de convergência em distribuição para outros contextos, como, por exemplo, para vetores aleatórios.

Outra observação é que o requisito de tomarmos até a derivada terceira no teorema acima não tem nada de especial. Poderíamos tê-lo feito considerando derivadas até uma determinada ordem n. Escolhemos derivada terceira porque será suficiente quando provarmos as versões dos Teoremas do Limite Central de Lyapunov e de Lindeberg nas Seções 9.3 e 9.4.

7.5 Exercícios

§7.1

1. Seja $(A_n)_n$ uma sequência de eventos e $(\mathbb{1}_{A_n})_n$ a sequência de variáveis aleatórias indicadoras das ocorrências dos eventos correspondentes. Encontre uma condição sobre as probabilidades $\mathbb{P}(A_n)$ para que $\mathbb{1}_{A_n} \xrightarrow{\mathbb{P}} 0$.

2. Suponha que $\mathbb{E}X_n \to 0$ e $\mathbb{V}X_n \to 0$. Prove que $X_n \xrightarrow{\mathbb{P}} 0$.

3. Prove a unicidade quase certa do limite quase certo. Isto é, se $X_n \xrightarrow{\text{q.c.}} Y$ e $X_n \xrightarrow{\text{q.c.}} Z$ então $Y = Z$ q.c.

4. Prove a unicidade quase certa do limite em probabilidade. Isto é, se $X_n \xrightarrow{\mathbb{P}} Y$ e $X_n \xrightarrow{\mathbb{P}} Z$ então $Y = Z$ q.c.

5. Seja $U \sim \mathcal{U}[0,1]$ e defina $X_n = n\mathbb{1}_{\{U<\frac{1}{n}\}}$. Verifique respectivamente se existe uma variável aleatória X tal que $X_n \to X$

(a) em distribuição.
(b) em probabilidade.
(c) quase certamente.
(d) em $\mathcal{L}^1$.
(e) em $\mathcal{L}^2$.

6. Sejam $(X_n)_n$ uma sequência de variáveis aleatórias com $X_n \sim \mathcal{U}[0, 1 + \frac{1}{n}]$ e $X \sim \mathcal{U}[0,1]$.

(a) Mostre que $X_n \xrightarrow{\text{d}} X$.
(b) Se $(X_n)_n$ forem independentes, mostre que X_n não converge em probabilidade para X.
(c) Dê um exemplo mostrando que, se a hipótese de independência for retirada, é possível que $X_n \xrightarrow{\text{q.c.}} X$.

7. Suponha que X_n tem distribuição uniforme no conjunto $\{\frac{1}{n}, \frac{2}{n}, \ldots, \frac{n-1}{n}, 1\}$ para todo $n \in \mathbb{N}$. Mostre que $(X_n)_n$ converge em distribuição para uma variável com distribuição uniforme em $[0,1]$.

8. Sejam $(X_n)_n$ uma sequência de variáveis aleatórias i.i.d. com distribuição $\text{Exp}(1)$ e $Z_n = \max\{X_1, \ldots X_n\} - \log n$. Mostre que Z_n converge em distribuição para a distribuição de Gumbel.

9. Sejam $(X_n)_n$ uma sequência de variáveis aleatórias independentes com distribuição uniforme em $[0,1]$, e $Y_n = \max\{X_1, \ldots, X_n\}$. Encontre a função de distribuição de Y_n e o limite em distribuição desta sequência.

10. Suponha que $X_n \xrightarrow{\mathcal{L}^1} X$ e $|Z| \leqslant 1$ q.c. Mostre que $X_n Z \xrightarrow{\mathcal{L}^1} XZ$.

§7.2

11. Seja $(X_n)_{n\in\mathbb{N}}$ uma sequência de variáveis aleatórias independentes tais que $X_n \sim \mathrm{Exp}(\lambda_n)$, onde $\lambda_n = n^3$. Prove que $\mathbb{P}\big(\sum_{n=1}^{\infty} X_n < \infty\big) = 1$.

12. Sejam $(a_n)_n$ uma sequência de números reais e $(X_n)_n$ uma sequência de variáveis aleatórias com distribuição dada por $\mathbb{P}(X_n = a_n) = p_n = 1 - \mathbb{P}(X_n = 0)$ para todo $n \in \mathbb{N}$. Mostre que, se $\sum_{n=1}^{\infty} p_n < \infty$, então $X_n \xrightarrow{\text{q.c.}} 0$.

13. Sejam X_n, $n \in \mathbb{N}$, variáveis aleatórias independentes tais que $X_n \sim \mathrm{Bernoulli}(p_n)$. Estude as condições sobre (p_n) para que:

(a) $X_n \xrightarrow{\mathbb{P}} 0$.
(b) $X_n \xrightarrow{\text{q.c.}} 0$.

14. Definimos o *limite inferior da sequência de conjuntos* $(A_n)_n$ como o conjunto $\liminf_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k$. Mostre que:

(a) $\liminf_{n\to\infty} A_n \subseteq \limsup_{n\to\infty} A_n$.
(b) Dê um exemplo onde vale a inclusão estrita no item anterior.

15. Seja $(X_n)_n$ uma sequência de variáveis aleatórias independentes com $X_n \sim \mathcal{U}[-n, n]$. Calcule $\mathbb{P}(|X_n| \to +\infty)$ e $\mathbb{P}(|X_{n^2}| \to +\infty)$.

16. Seja $(X_n)_n$ uma sequência i.i.d. Mostre que

$$\frac{X_n}{n} \xrightarrow{\text{q.c.}} 0$$

se, e somente se, X_1 for integrável.

17. Seja $(X_n)_n$ uma sequência i.i.d. Mostre que

$$\frac{X_n}{\sqrt{n}} \xrightarrow{\text{q.c.}} 0$$

se, e somente se, $\mathbb{E}X_1^2 < \infty$.

18. Seja $(X_n)_n$ uma sequência i.i.d. com distribuição Exp(1). Mostre que $\mathbb{P}(X_n \geqslant 1 \log n \text{ i.v.}) = 1$ e $\mathbb{P}(X_n \geqslant 2 \log n \text{ i.v.}) = 0$.

19. Seja $(X_n)_n$ uma sequência i.i.d. com distribuição Poisson(λ). Mostre que

$$\frac{X_n}{\log n} \overset{\text{q.c.}}{\to} 0.$$

Sugestão: Mostre antes que $\mathbb{E}[e^{X_1/\varepsilon}] < \infty$.

20. Seja $(X_n)_n$ uma sequência i.i.d. de variáveis aleatórias não-negativas com $\mathbb{E}X_1^2 < \infty$. Mostre que

$$\mathbb{P}\left(\sum_{n=1}^{\infty} \frac{X_n}{n^2} < \infty\right) = 1.$$

21.

(a) Seja X uma variável aleatória com distribuição normal padrão. Mostre que

$$\frac{\mathbb{P}(X > x)}{\frac{1}{x\sqrt{2\pi}} e^{-x^2/2}} \to 1 \quad \text{quando} \quad x \to +\infty.$$

(b) Seja $(X_n)_n$ uma sequência de variáveis i.i.d. com distribuição normal padrão. Encontre uma sequência de números reais (a_n) tais que

$$\mathbb{P}(X_n \geqslant (1-\varepsilon)a_n \text{ i.v.}) = 1 \text{ e } \mathbb{P}(X_n \geqslant (1+\varepsilon)a_n \text{ i.v.}) = 0$$

para todo $\varepsilon > 0$.

22. Seja $(X_n)_n$ uma sequência qualquer de variáveis aleatórias. Mostre que sempre existe uma sequência de números reais $(a_n)_n$ tal que $\frac{X_n}{a_n} \overset{\text{q.c.}}{\to} 0$.

§7.3

23. Dê uma solução alternativa ao Exercício 4, usando a Proposição 7.32.

24. Seja $(X_n)_n$ uma sequência de variáveis aleatórias tais que $\mathbb{E}X_n \to c$ quando $n \to \infty$, onde $c \in \mathbb{R}$.

(a) Mostre que, se $\sum_{n=1}^{\infty} \mathbb{V}X_n$ converge, então $X_n \overset{\text{q.c.}}{\to} c$.
(b) Mostre que, se $\mathbb{V}X_n \to 0$, então $X_n \overset{\mathbb{P}}{\to} c$.
(c) Se $\mathbb{V}X_n \to 0$, então necessariamente $X_n \overset{\text{q.c.}}{\to} c$?

25. Seja $(X_n)_n$ uma sequência de variáveis aleatórias i.i.d. com distribuição $\mathcal{U}[0,1]$ e $Y_n = n^{-X_n}$ para todo $n \in \mathbb{N}$. Mostre que $Y_n \to 0$ em probabilidade mas não quase certamente.

26. Seja X uma variável aleatória integrável, e seja $(A_n)_n$ uma sequência de eventos tais que $\mathbb{P}(A_n) \to 0$. Prove que $\mathbb{E}[X\mathbb{1}_{A_n}] \to 0$.

§7.4

27. Sejam X e $(X_n)_n$ variáveis aleatórias tais que $X_n \overset{\text{d}}{\to} X$, e seja $g : \mathbb{R} \to \mathbb{R}$ uma função contínua. Mostre que $g(X_n) \overset{\text{d}}{\to} g(X)$.

28. Dê uma prova alternativa da Proposição 7.34 usando o Teorema de Helly-Bray. Sugestão: Primeiro prove supondo convergência quase certa, e depois use um argumento semelhante ao da prova do Teorema 7.33.

Capítulo 8

Lei dos Grandes Números

Um dos principais tópicos da Teoria da Probabilidade é o estudo da chamada Lei dos Grandes Números. Ela diz que a soma de muitas variáveis aleatórias independentes (ou não-correlacionadas, etc.) tende a estar próxima da sua esperança. Mais precisamente, a média observada $\frac{X_1+\cdots+X_n}{n}$, que é aleatória, se aproxima da média teórica $\mathbb{E}[\frac{X_1+\cdots+X_n}{n}]$, que é determinística.

Sua manifestação mais simples é na frequência relativa. Imagine que realizamos o mesmo experimento muitas vezes, sob as mesmas condições, e contamos quantos desses experimentos resultaram em sucesso e quantos resultaram em fracasso. Tomando $X_n = 1$ para indicar sucesso na n-ésima tentativa em $X_n = 0$ para indicar fracasso, a frequência relativa observada é dada justamente por $\frac{X_1+\cdots+X_n}{n}$. Enquanto escreviam este preâmbulo, os autores simularam o lançamento de uma moeda honesta um milhão de vezes, e obtiveram coroa 499.947 vezes. Repetindo o mesmo procedimento, obtiveram coroa 499.508 vezes, depois 500.318 vezes, e depois 500.512 vezes. Como previsto pela Lei dos Grandes Números, a frequência relativa ficou sempre muito próxima da probabilidade de se obter coroa em cada lançamento, que é exatamente $\frac{1}{2}$. Intuitivamente, tendemos a associar probabilidade de sucesso a frequência relativa de sucessos, e essa associação já está quase naturalizada no nosso pensamento. Entretanto, no estudo da Teoria da

Probabilidade como uma teoria axiomática, isso será um *teorema* e não uma *definição*. Se, por um lado, não definimos a probabilidade de se obter coroa no lançamento da moeda como sendo dada pela frequência relativa, por outro lado, é reconfortante que a teoria seja capaz de autojustificar-se ao estabelecer que tal frequência relativa deve se aproximar do valor teórico quando o número de realizações do experimento aumenta.

De forma mais geral, o resultado relevante de cada experimento não precisa ser 0 ou 1 para representar fracasso ou sucesso, de fato pode ser qualquer variável aleatória. Antes de escrever este parágrafo, os autores lançaram um dado dez vezes, e a soma dos valores obtidos foi de 38. Repetindo o mesmo procedimento, obtiveram 33, 28, 37 e por último 43. Não pareceu que esta soma estivesse muito bem concentrada próxima de algum valor determinístico. Entretanto, os autores depois simularam dez mil lançamentos do dado, e a soma dos resultados foi 35.082. Repetindo o mesmo procedimento, obtiveram como soma 34.769, depois 35.419, e depois 34.691. Como previsto pela Lei dos Grandes Números, quando o número de lançamentos foi grande, a média observada se aproximou da média teórica, dada por $\mathbb{E}X_1 = \frac{7}{2}$.

Nas seções seguintes, vamos provar que vale a Lei dos Grandes Números sob diferentes hipóteses, com demonstrações que vão aumentando em nível de complexidade. Ademais, como visto no capítulo anterior, a noção de "aproximar-se" pode ter mais de um significado quando falamos de quantidades aleatórias, o que se traduz em distintas formulações da Lei dos Grandes Números. Estudaremos hipóteses sob as quais essa aproximação se dará em diferentes sentidos. Finalizamos com algumas aplicações na última seção.

8.1 Lei Fraca dos Grandes Números

Sejam $X_1, X_2, \ldots$ variáveis aleatórias integráveis em $(\Omega, \mathcal{F}, \mathbb{P})$ e $S_1, S_2, \ldots$ suas somas parciais dadas por

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

A *Lei Fraca dos Grandes Números* diz que, sob certas hipóteses,

$$\frac{S_n - \mathbb{E}S_n}{n} \xrightarrow{\mathbb{P}} 0.$$

Teorema 8.1 (Lei dos Grandes Números de Bernoulli)**.** *Considere uma sequência de ensaios independentes tendo probabilidade p de sucesso em cada ensaio. Se S_n é o número de sucessos nos primeiros n ensaios, então*

$$\frac{S_n}{n} \xrightarrow{\mathbb{P}} p.$$

A Lei dos Grandes Números de Bernoulli tem uma importância histórica inestimável[12]. De certa forma, esse teorema justifica o conceito de probabilidade como sendo a frequência relativa de ocorrência de um evento, isto é,

$$p \approx \frac{\text{número de experimentos em que o evento é observado}}{\text{número total de experimentos realizados}},$$

onde a ideia de aproximação passa a ter um significado mais preciso, o da convergência em probabilidade. Não veremos a demonstração original de Bernoulli, o teorema abaixo é mais geral.

Teorema 8.2 (Lei dos Grandes Números de Tchebyshev)**.** *Sejam $X_1, X_2, \ldots$ variáveis aleatórias não-correlacionadas. Suponha que existe M finito tal que $\mathbb{V}X_n < M$ para todo n. Então*

$$\frac{S_n - \mathbb{E}S_n}{n} \xrightarrow{\mathbb{P}} 0.$$

Demonstração. Pela Desigualdade de Tchebyshev,

$$\mathbb{P}\left(\left|\frac{S_n - \mathbb{E}S_n}{n}\right| \geqslant \varepsilon\right) \leqslant \frac{\mathbb{V}(\frac{S_n}{n})}{\varepsilon^2} = \frac{\mathbb{V}S_n}{\varepsilon^2 n^2} = \frac{\sum_{j=1}^n \mathbb{V}X_j}{\varepsilon^2 n^2} \leqslant \frac{n \cdot M}{\varepsilon^2 n^2} \to 0. \qquad \square$$

[12] A Lei dos Grande Números de Bernoulli aparece pela primeira vez sob o nome de *Teorema Áureo* em seu livro *Ars Conjectandi* publicado postumamente em 1713.

Posteriormente, foi provada por A.Y. Khintchine a versão abaixo que retira a hipótese de variância finita, mas precisa que as variáveis sejam i.i.d.

Teorema 8.3 (Lei dos Grandes Números de Khintchine)**.** *Sejam $X_1, X_2, \ldots$ variáveis aleatórias independentes, identicamente distribuídas e integráveis, com média μ. Então*

$$\frac{S_n}{n} \xrightarrow{\mathbb{P}} \mu.$$

A demonstração original de Khintchine foi feita usando o método de truncamento, aparentemente introduzido por Markov, e utilizado em seguida por Kolmogorov na prova da Lei Forte dos Grandes Números. Este teorema é corolário da Lei Forte dos Grandes Números de Kolmogorov, que será enunciada na próxima seção e provada na seção seguinte. Uma prova alternativa usando funções características será dada na Seção 10.2.

8.2 Lei Forte dos Grandes Números

Considere uma sequência $(X_n)_n$ de variáveis aleatórias, todas com média μ. A Lei Fraca dos Grandes Números diz que, se tomamos um valor de n grande o suficiente, então, como alta probabilidade, $\frac{S_n}{n}$ estará bem próximo de μ. Observe que essa a afirmação diz respeito ao comportamento estatístico de S_n para qualquer valor de n previamente selecionado, desde que seja grande. Mas nada impede que $\frac{S_n}{n}$ esteja longe de μ para infinitos de valores de n posteriores àquele previamente selecionado. Em contraste, a Lei Forte dos Grandes Números, que definimos abaixo, se refere ao comportamento da sequência aleatória $(\frac{S_n}{n})_{n\in\mathbb{N}}$ como um todo. Ela diz que, quase certamente, a média observada $\frac{S_n}{n}$ se aproxima de μ quando n cresce, não deixando aberta a possibilidade de que ela siga oscilando indefinidamente.

A *Lei Forte dos Grandes Números* diz que, sob certas hipóteses,

$$\frac{S_n - \mathbb{E}S_n}{n} \xrightarrow{\text{q.c.}} 0.$$

Na Seção 7.3, vimos que convergência quase certa implica convergência em probabilidade. Portanto, a Lei Forte implica a Lei Fraca.

Teorema 8.4 (Lei dos Grandes Números de Borel)**.** *Considere uma sequência de ensaios independentes tendo a mesma probabilidade p de sucesso em cada ensaio. Se S_n é o número de sucessos nos primeiros n ensaios, então*

$$\frac{S_n}{n} \overset{\text{q.c.}}{\to} p.$$

Não veremos a demonstração original de Borel. A seguir, enunciaremos e provaremos a Lei dos Grandes Números de Cantelli, que é mais geral que a de Borel.

Teorema 8.5 (Lei dos Grandes Números de Cantelli)**.** *Sejam $X_1, X_2, \ldots$ variáveis aleatórias independentes com quarto momento limitado. Então*

$$\frac{S_n - \mathbb{E}S_n}{n} \overset{\text{q.c.}}{\to} 0.$$

Demonstração. Podemos supor que $\mathbb{E}X_n = 0$ (senão consideramos $X_n - \mathbb{E}X_n$ e usamos as desigualdades $|x| \leqslant 1 + x^4$ e $(x-c)^4 \leqslant 16x^4 + 16c^4$ para mostrar que essa sequência também tem quarto momento limitado). Observe que

$$\begin{aligned} S_n^4 = (X_1 + \cdots + X_n)^4 = \sum_{r,j,k,l} X_r X_j X_k X_l = \sum_r X_r^4 + \frac{4!}{2!2!} \sum_{r<j} X_r^2 X_j^2 + \\ + \frac{4!}{3!} \sum_{r \neq k} X_r^3 X_k + \frac{4!}{2!} \sum_{\substack{j<k \\ r \neq j,k}} X_r^2 X_j X_k + 4! \sum_{r<j<k<l} X_r X_j X_k X_l. \end{aligned}$$

Por independência,

$$\begin{aligned} \mathbb{E}S_n^4 = \sum_r \mathbb{E}X_r^4 + 6 \sum_{r<j} \mathbb{E}[X_r^2 X_j^2] + \\ + \sum_k \mathbb{E}\Big[4 \sum X_r^3 + 12 \sum X_r^2 X_j + 24 \sum X_r X_j X_l\Big] \mathbb{E}X_k. \end{aligned}$$

Como assumimos que $\mathbb{E}X_k = 0$, a segunda linha é igual a zero. Tome M tal que $\mathbb{E}X_k^4 \leqslant M$ para todo k. Para o segundo termo, observamos que, como $x^2z^2 \leqslant x^4 + z^4$, temos $\mathbb{E}[X_r^2 X_j^2] \leqslant 2M$. Assim,

$$\mathbb{E}S_n^4 \leqslant nM + 12\tbinom{n}{2}M = (6n^2 - 5n)M \leqslant 6n^2 M.$$

Pela Desigualdade de Markov,

$$\mathbb{P}\left(\left|\frac{S_n}{n}\right| \geqslant \varepsilon\right) \leqslant \frac{\mathbb{E}S_n^4}{\varepsilon^4 n^4} \leqslant \frac{6M}{\varepsilon^4 n^2},$$

e pelo Lema de Borel-Cantelli segue que $\frac{S_n}{n} \overset{\text{q.c.}}{\to} 0$. □

Teorema 8.6 (Lei dos Grandes Números de Kolmogorov)**.** *Sejam $X_1, X_2, \ldots$ variáveis aleatórias independentes duas a duas, identicamente distribuídas e integráveis. Então*

$$\frac{S_n}{n} \overset{\text{q.c.}}{\to} \mathbb{E}X_1.$$

O teorema como enunciado acima é um caso particular do Teorema 8.11, visto na Seção 8.3, ou do Teorema Ergódico de Birkhoff, visto na Seção 14.3.

Enfatizamos que a hipótese de X_1 ser integrável é a mais fraca possível para que valha a Lei Forte dos Grandes Números quando a sequência $(X_n)_n$ for i.i.d. Isto é, X_1 ser integrável é equivalente a $\frac{X_1+\cdots+X_n}{n} \overset{\text{q.c.}}{\to} \mu$ para algum $\mu \in \mathbb{R}$, conforme mostramos na seguinte proposição.

Proposição 8.7. *Seja $(X_n)_n$ uma sequência de variáveis aleatórias i.i.d. tal que $\frac{X_1+\cdots+X_n}{n} \overset{\text{q.c.}}{\to} \mu$, onde μ é um número real. Então $\mathbb{E}|X_1| < \infty$ e $\mu = \mathbb{E}X_1$.*

Demonstração. Observe que, por hipótese,

$$\frac{X_n}{n} = \frac{X_1 + \cdots + X_n}{n} - \frac{n-1}{n} \cdot \frac{X_1 + \cdots + X_{n-1}}{n-1} \overset{\text{q.c.}}{\to} 0.$$

Pela Proposição 7.16, $\mathbb{P}(\{|X_n| \geqslant n \text{ i.v.}\}) = 0$. Como a sequência $(X_n)_n$ é independente, segue do Lema de Borel-Cantelli que $\sum_n \mathbb{P}(|X_n| \geqslant n) < \infty$

Como $(X_n)_n$ é identicamente distribuída, $\sum_n \mathbb{P}(|X_1| \geqslant n) < \infty$. Isso que garante a integrabilidade de X_1 pela Proposição 5.35. $\square$

8.3 Leis Fortes de Kolmogorov

Nesta seção provaremos outras duas versões de Leis Forte dos Grandes Números com hipóteses mais generosas que aquelas dos Teoremas 8.5 e 8.6.

A Primeira Lei Forte de Kolmogorov, provada em 1930, permite que as variáveis não tenham a mesma distribuição, mas tem uma condição sobre as variâncias. A chamada Lei Forte de Kolmogorov, provada em 1933, estabelece que a hipótese de integrabilidade é suficiente no caso de variáveis i.i.d., e é baseada na Primeira Lei Forte combinada com o método de truncamento.

A prova que daremos para a Primeira Lei Forte de Kolmogorov usa a elegante ideia, introduzida por Etemadi em 1981, de considerar variáveis não-negativas e estudar uma subsequência que cresce exponencialmente rápido porém com expoente pequeno. Com essa ideia precisamos apenas que as variáveis sejam independentes duas a duas, mas temos que supor uma condição sobre o primeiro momento que não era usada na demonstração original. Isso também permite provar a Lei Forte de Kolmogorov supondo apenas que as variáveis sejam independentes duas a duas.

Teorema 8.8 (Primeira Lei Forte dos Grandes Números)**.** *Sejam* $(Y_n)_n$ *variáveis independentes duas a duas. Suponha que* $\sup_n \mathbb{E}|Y_n| < \infty$ *e* $\sum_{n=1}^{\infty} \frac{\mathbb{E}Y_n^2}{n^2} < \infty$*. Então*

$$\frac{Y_1 + \cdots + Y_n - \mathbb{E}[Y_1 + \cdots + Y_n]}{n} \overset{\text{q.c.}}{\to} 0.$$

Demonstração. Utilizando a decomposição $Y_n = Y_n^+ - Y_n^-$, podemos supor, sem perda de generalidade, que Y_n é não-negativa para todo $n \in \mathbb{N}$.

Denotamos as somas por $T_n = Y_1 + \cdots + Y_n$. Mostraremos primeiro que

$$\frac{T_{k_n} - \mathbb{E}T_{k_n}}{k_n} \overset{\text{q.c.}}{\to} 0$$

para uma família especial de subsequências $(k_n)_n$. Seja $\alpha > 1$. Defina $k_n = \lfloor \alpha^n \rfloor$. Note que $k_n^{-2} \leqslant 4\alpha^{-2n}$ (pois $x \leqslant 2\lfloor x \rfloor$ para todo $x \geqslant 1$) e $\frac{k_{n+1}}{k_n} \to \alpha$. Para cada $r \in \mathbb{N}$, escrevemos $n_r = \min\{n \in \mathbb{N} : k_n \geqslant r\} \geqslant \log_\alpha r$. Estimando a partir da Desigualdade de Tchebyshev, obtemos

$$\begin{aligned}
\sum_{n=1}^{\infty} \mathbb{P}(|T_{k_n} - \mathbb{E}T_{k_n}| \geqslant \varepsilon k_n) &\leqslant \sum_{n=1}^{\infty} \frac{\mathbb{V}T_{k_n}}{\varepsilon^2 k_n^2} = \varepsilon^{-2} \sum_{n=1}^{\infty} k_n^{-2} \sum_{r=1}^{k_n} \mathbb{V}Y_r \\
&= \varepsilon^{-2} \sum_{r=1}^{\infty} \mathbb{V}Y_r \sum_{n:k_n \geqslant r} k_n^{-2} = \varepsilon^{-2} \sum_{r=1}^{\infty} \mathbb{V}Y_r \sum_{n=n_r}^{\infty} k_n^{-2} \\
&\leqslant \varepsilon^{-2} \sum_{r=1}^{\infty} \mathbb{V}Y_r \frac{4\alpha^{-2n_r}}{1-\alpha^{-2}} \leqslant \frac{4\varepsilon^{-2}}{1-\alpha^{-2}} \sum_{r=1}^{\infty} \frac{\mathbb{V}Y_r}{r^2} < \infty,
\end{aligned}$$

onde na primeira igualdade usamos a hipótese de que as variáveis $(Y_n)_n$ são independentes duas a duas. Pelos Lema de Borel-Cantelli e Proposição 7.16,

$$\frac{T_{k_n} - \mathbb{E}T_{k_n}}{k_n} \overset{\text{q.c.}}{\to} 0. \tag{8.9}$$

Para concluir a prova, falta preencher as lacunas, controlando a diferença

$$\frac{T_r - \mathbb{E}T_r}{r}$$

para todo $r \in \mathbb{N}$ grande e não apenas na subsequência k_n. É aqui que usamos a hipótese de que as variáveis Y_n são não-negativas. Para cada $r \geqslant \alpha$, tomamos $n = n_r - 1$, de forma que $1 \leqslant k_n < r \leqslant k_{n+1}$. Usando o fato de que T_r e $\mathbb{E}T_r$ são não-decrescentes e $M = \sup_r \mathbb{E}Y_r < \infty$, vamos trabalhar a expressão acima para reduzir o problema a (8.9). Para isso, primeiro majoramos, depois alinhamos os índices do numerador, e finalmente os do

denominador:

$$\begin{aligned}\frac{T_r - \mathbb{E}T_r}{r} &\leqslant \frac{T_{k_{n+1}} - \mathbb{E}T_{k_n}}{r} \leqslant \frac{T_{k_{n+1}}}{k_n} - \frac{\mathbb{E}T_{k_n}}{k_{n+1}} \\ &= \frac{T_{k_{n+1}}}{k_n} - \frac{\mathbb{E}T_{k_{n+1}}}{k_n} + (\tfrac{1}{k_n} - \tfrac{1}{k_{n+1}})\mathbb{E}T_{k_{n+1}} + \frac{\mathbb{E}[T_{k_{n+1}} - T_{k_n}]}{k_{n+1}} \\ &\leqslant \frac{k_{n+1}}{k_n} \cdot \frac{T_{k_{n+1}} - \mathbb{E}T_{k_{n+1}}}{k_{n+1}} + (\tfrac{1}{k_n} - \tfrac{1}{k_{n+1}})k_{n+1}M + \frac{k_{n+1} - k_n}{k_{n+1}}M.\end{aligned}$$

Usando (8.9), vemos que $\limsup_r \frac{T_r - \mathbb{E}T_r}{r} \leqslant \alpha \cdot 0 + (\alpha - 1)M + (1 - \alpha^{-1})M$ q.c. Como isso vale para todo $\alpha > 1$, $\limsup_r \frac{T_r - \mathbb{E}T_r}{r} \leqslant 0$ q.c. De forma análoga, mostramos que $\liminf_r \frac{T_r - \mathbb{E}T_r}{r} \geqslant 0$ q.c., o que conclui a prova. □

Lema 8.10. *Seja $(X_n)_n$ uma sequência de variáveis aleatórias identicamente distribuídas e integráveis. Para cada $n \in \mathbb{N}$, defina $Y_n = X_n \mathbb{1}_{\{|X_n| \leqslant n\}}$. Então $\sum_{n=1}^{\infty} \frac{\mathbb{E}Y_n^2}{n^2} < \infty$.*

Demonstração. Escrevendo $X = |X_1|$ por simplicidade,

$$\begin{aligned}\sum_{n=1}^{\infty} n^{-2}\mathbb{E}Y_n^2 &= \sum_{n=1}^{\infty} n^{-2}\mathbb{E}[X_n^2 \mathbb{1}_{\{|X_n| \leqslant n\}}] = \sum_{n=1}^{\infty} n^{-2}\mathbb{E}[X^2 \mathbb{1}_{\{X \leqslant n\}}] \\ &= \mathbb{E}\Big[X^2 \sum_{n=1}^{\infty} n^{-2}\mathbb{1}_{\{X \leqslant n\}}\Big] \leqslant \mathbb{E}\Big[X^2(X^{-1} + X^{-2})\Big] = 1 + \mathbb{E}X < \infty,\end{aligned}$$

onde usamos o Teorema da Convergência Monótona para comutar soma e esperança, além da estimativa $\sum_{n \geqslant \lceil x \rceil} n^{-2} \leqslant x^{-1} + x^{-2}$ para $x > 0$. □

Teorema 8.11 (Segunda Lei Forte dos Grandes Números)**.** *Seja $(X_n)_n$ sequência de variáveis aleatórias, independentes duas a duas, identicamente distribuídas e integráveis. Então*

$$\frac{X_1 + \cdots + X_n}{n} \xrightarrow{\text{q.c.}} \mathbb{E}X_1.$$

Demonstração. Seja $Y_n = X_n \mathbb{1}_{\{|X_n| \leqslant n\}}$ para cada $n \in \mathbb{N}$. Combinando o Lema 8.10 com a Primeira Lei Forte dos Grandes Números, obtemos

$$\frac{Y_1 + \cdots + Y_n - \mathbb{E}[Y_1 + \cdots + Y_n]}{n} \overset{\text{q.c.}}{\to} 0.$$

Por outro lado, do Teorema da Convergência Dominada,

$$\mathbb{E}Y_n = \mathbb{E}[X_n \mathbb{1}_{\{|X_n| \leqslant n\}}] = \mathbb{E}[X_1 \mathbb{1}_{\{|X_1| \leqslant n\}}] \to \mathbb{E}X_1$$

e, por conseguinte,

$$\frac{\mathbb{E}[Y_1 + \cdots + Y_n]}{n} \to \mathbb{E}X_1.$$

Combinando estes limites, obtemos

$$\frac{Y_1 + \cdots + Y_n}{n} \overset{\text{q.c.}}{\to} \mathbb{E}X_1.$$

Finalmente, escrevemos $X_n = Y_n + Z_n$, onde $Z_n = X_n \mathbb{1}_{\{|X_n| > n\}}$. Veja que

$$\sum_{n=1}^{\infty} \mathbb{P}(Z_n \neq 0) = \sum_{n=1}^{\infty} \mathbb{P}(|X_n| > n) = \sum_{n=1}^{\infty} \mathbb{P}(|X_1| > n) < \infty$$

pela Proposição 5.35. Pelo Lema de Borel-Cantelli, $\mathbb{P}(Z_n \neq 0 \text{ i.v.}) = 0$, logo $Z_n \overset{\text{q.c.}}{\to} 0$ e portanto

$$\frac{Z_1 + \cdots + Z_n}{n} \overset{\text{q.c.}}{\to} 0,$$

concluindo a prova. □

8.4 Algumas aplicações

Nesta seção faremos quatro aplicações da Lei dos Grandes Números. Elas podem ser lidas em qualquer ordem. A terceira é aplicação da Lei Fraca, a primeira e a última, da Lei Forte. A segunda é na verdade uma aplicação da Desigualdade de Tchebyshev, como feito na demostração da Lei Fraca.

Teorema de Glivenko-Cantelli

O teorema abaixo tem um papel central em Estatística Matemática. Ele relaciona a chamada *função de distribuição empírica*, construída a partir de observações de variáveis aleatórias i.i.d., com a função de distribuição comum a essas variáveis, em princípio desconhecida. Suas aplicações ocorrem em várias áreas, como por exemplo Econometria e Aprendizado de Máquina.

Teorema 8.12 (Teorema de Glivenko-Cantelli)**.** *Sejam $(X_n)_n$ variáveis aleatórias i.i.d. com função de distribuição F_X, e defina*

$$\hat{F}_n(x) = \frac{1}{n}\sum_{k=1}^{n} \mathbb{1}_{\{X_n \leqslant x\}}.$$

Então

$$\mathbb{P}\Big(\lim_{n\to\infty} \sup_{x\in\mathbb{R}} |\hat{F}_n(x) - F_X(x)| = 0\Big) = 1.$$

Demonstração. Defina $G_X(x) = F_X(x-) = \mathbb{P}(X < x)$ e $\hat{G}_n(x) = \hat{F}_n(x-) = n^{-1}\sum_{k=1}^{n} \mathbb{1}_{\{X_k<x\}}$, as funções análogas a $F_X(x)$ e $\hat{F}_n(x)$ porém contínuas à esquerda.

Seja $\varepsilon > 0$. Tome $k \in \mathbb{N}$ e $x_1 < \cdots < x_k$ tais que $G_X(x_1) < \varepsilon$, $F_X(x_k) > 1-\varepsilon$ e $G_X(x_j) < F_X(x_{j-1}) + \varepsilon$ para $j = 2, \ldots, k$. Isso pode ser feito a partir da função quantil F_X^{-1}, tomando uma coleção de pontos em $(0, 1)$ com separação menor que ε (a condição $G_X(x_j) < F_X(x_{j-1}) + \varepsilon$ pode parecer incompatível com o fato de F_X ter saltos, mas isso não é um problema se os pontos onde F_X salta mais que ε estiverem entre os $x_1, \ldots, x_k$).

Seja $\delta > 0$. Pela Lei dos Grandes Números de Borel, quase certamente existem $N_1, \ldots, N_k$ aleatórios tais que $|\hat{F}_n(x_j) - F_X(x_j)| < \delta$ e $|\hat{G}_n(x_j) - G_X(x_j)| < \delta$ para todo $j = 1, \ldots, k$ e todo $n \geqslant N_j$. Seja $N_0 = \max\{N_1, \ldots, N_k\}$.

Agora seja x tal que $x_{j-1} < x < x_j$ para algum j. Temos

$$\hat{F}_n(x_{j-1}) \leqslant \hat{F}_n(x) \leqslant \hat{G}_n(x_j) \text{ e } F_X(x_{j-1}) \leqslant F_X(x) \leqslant G_X(x_j).$$

Subtraindo, obtemos

$$\hat{F}_n(x) - F_X(x) \leqslant \hat{G}_n(x_j) - F_X(x_{j-1}) \leqslant \hat{G}_n(x_j) - G_X(x_j) + \varepsilon$$

e

$$\hat{F}_n(x) - F_X(x) \geqslant \hat{F}_n(x_{j-1}) - G_X(x_j) \geqslant \hat{F}_n(x_{j-1}) - F_X(x_{j-1}) - \varepsilon.$$

Logo, para $n \geqslant N_0$, vale $|\hat{F}_n(x) - F_X(x)| < \delta + \varepsilon$. Argumento similar vale para $x < x_1$ e também para $x > x_k$. Para $x = x_1, \ldots, x_k$, vale $|\hat{F}_n(x) - F_X(x)| < \delta$. Ou seja, $\sup_{x\in\mathbb{R}} |\hat{F}_n(x) - F_X(x)| \leqslant \delta + \varepsilon$ para todo $n \geqslant N_0$. Portanto,

$$\mathbb{P}\Big(\limsup_{n\to\infty} \sup_{x\in\mathbb{R}} |\hat{F}_n(x) - F_X(x)| \leqslant \delta + \varepsilon \Big) = 1.$$

Fazendo $\varepsilon = \delta = \frac{1}{m}$ e tomando a interseção sobre $m \in \mathbb{N}$, obtemos a convergência do enunciado. $\square$

Teorema de Weierstrass

A esta altura está claro que Análise Real e Teoria da Medida são ferramentas mais que importantes para o estudo de Probabilidade. Porém, um fato notável é que temos uma via de mão dupla. Há teoremas da Análise que podem ser obtidos com demonstrações probabilísticas. O Teorema de Weierstrass diz que toda função $f : [a, b] \to \mathbb{R}$ contínua, por mais complicada que seja, pode ser aproximada uniformemente por uma sequência de polinômios.[13] Observe que trata-se de um teorema puramente de Análise Real.

Teorema 8.13 (Teorema de Weierstrass)**.** *Sejam $f : [a, b] \to \mathbb{R}$ uma função contínua e $\varepsilon > 0$. Existe um polinômio g tal que $|g(x) - f(x)| < \varepsilon$ para todo $x \in [a, b]$.*

Sejam f e ε como no enunciado. Podemos supor que $a = 0$ e $b = 1$, pois há

[13] O Teorema 8.13 foi provado originalmente por Weierstrass em 1885. A demonstração probabilística que fornecemos é devida a Bernstein, publicada em 1912.

uma bijeção entre $[0,1]$ e $[a,b]$ em forma de polinômio. Como f é contínua no intervalo fechado $[0,1]$, existe M tal que $|f(p)| \leqslant M$ para todo $p \in [0,1]$. Ademais, existe $\delta > 0$ tal que $|f(p) - f(q)| < \frac{\varepsilon}{2}$ para todos $p, q \in [0,1]$ tais que $|p - q| < \delta$, porque f é uniformemente contínua (Teorema A.17).

O polinômio que vai aproximar f será a esperança de uma variável aleatória cuja distribuição é parametrizada por p.

Fixe algum $n > \frac{M}{\varepsilon\delta^2}$. Para $p \in [0,1]$, defina uma variável aleatória $Z_p \sim \mathrm{Binom}(n,p)$ e $X_p = \frac{Z_p}{n}$. Note que X_p toma valores em $[0,1]$ e tem média p.

O polinômio desejado será dado por

$$g(p) = \mathbb{E}[f(X_p)] = \sum_{k=0}^{n} f(\tfrac{k}{n})\tbinom{n}{k} p^k (1-p)^{n-k},$$

que é chamado *polinômio de Bernstein*.

Estimamos $g(p) - f(p)$ usando a Desigualdade de Tchebyshev:

$$\begin{aligned}
|g(p) - f(p)| &= \big|\mathbb{E}[f(X_p) - f(p)]\big| \leqslant \mathbb{E}|f(X_p) - f(p)| \\
&= \mathbb{E}\big[|f(X_p) - f(p)|\mathbb{1}_{\{|X_p - p| < \delta\}}\big] + \mathbb{E}\big[|f(X_p) - f(p)|\mathbb{1}_{\{|X_p - p| \geqslant \delta\}}\big] \\
&\leqslant \tfrac{\varepsilon}{2} + 2M\mathbb{P}(|X_p - p| \geqslant \delta) \leqslant \tfrac{\varepsilon}{2} + 2M\tfrac{\mathbb{V}X_p}{\delta^2} \\
&= \tfrac{\varepsilon}{2} + 2M\tfrac{p(1-p)}{n\delta^2} \leqslant \tfrac{\varepsilon}{2} + \tfrac{M}{2n\delta^2} < \varepsilon,
\end{aligned}$$

o que conclui a prova do Teorema de Weierstrass.

Método de Monte Carlo

No problema da agulha de Buffon (Seção 1.1.3), vimos que ao lançarmos aleatoriamente uma agulha de comprimento ℓ sobre um piso cortado por um feixe de retas paralelas e equidistantes também de ℓ, a probabilidade de a agulha cruzar uma das retas do piso é de $\frac{2}{\pi}$. Sendo assim, se lançarmos uma agulha sucessivas vezes e de modo independente, a Lei Forte dos Grandes

Números nos diz que

$$\frac{X_1 + \cdots + X_n}{n} \overset{\text{q.c.}}{\to} \frac{2}{\pi},$$

onde X_k é a função indicadora do evento que a agulha cruza alguma reta do piso na k-ésima tentativa. Portanto, a Lei dos Grandes Números pode ser usada em simulações de lançamentos de agulha para estimarmos o valor de π. Isto é, $\frac{2n}{X_1+\cdots+X_n}$ se aproxima de π quando o número de tentativas, n, é grande. Mais que isto, a Desigualdade de Tchebyshev nos diz que

$$\mathbb{P}\Big(\Big|\frac{X_1 + \cdots + X_n}{n} - \frac{2}{\pi}\Big| > \varepsilon\Big) \leqslant \frac{\mathbb{V}X_1}{n\varepsilon^2} < \frac{1}{4n\varepsilon^2}.$$

Ou seja, temos inclusive um certo controle do quão grande deve ser o número de lançamentos, n, para que uma aproximação com margem de erro ε tenha determinado nível de confiabilidade (no caso, $1 - \frac{1}{4n\varepsilon^2}$).[14]

Esta mesma ideia pode ser utilizada para se estimar, via Lei dos Grandes Números, diversas outras quantidades como no exemplo a seguir.

Exemplo 8.14. Seja $f : [0,1] \to [0,1]$ uma função integrável, gostaríamos de determinar quanto vale a integral $\int_0^1 f(x)\mathrm{d}x$. Sejam $(X_n)_n$ e $(Y_n)_n$ duas sequências variáveis aleatórias independentes, todas elas com distribuição uniforme no intervalo $[0,1]$ e $(Z_n)_n$ a sequência dada por $Z_n = \mathbb{1}_{\{f(X_n)>Y_n\}}$. Isto é, para todo $n \in \mathbb{N}$, (X_n, Y_n) são as coordenadas de um ponto sorteado uniformemente no quadrado $[0,1]^2$ e Z_n é a variável que indica se tal ponto se encontra abaixo do gráfico da função f. Como a sequência $(Z_n)_n$ é i.i.d., pela Lei Forte dos Grandes Números,

$$\frac{Z_1 + \cdots + Z_n}{n} \overset{\text{q.c.}}{\to} \mathbb{E}Z_1 = \int_0^1 f(x)\mathrm{d}x.$$

Ou seja, simulando distribuições uniformes no intervalo $[0,1]$, podemos aproximar a integral definida $\int_0^1 f(x)\mathrm{d}x$ por mais complicada que seja a

[14] Este experimento já foi realizado por diversas personalidades ao longo do tempo, sendo que a provável primeira aproximação foi realizada pelo astrônomo suíço R. Wolf em 1850 que ao lançar uma agulha 5.000 vezes encontrou a aproximação $\pi \approx 3{,}1596$.

função f. $\triangle$

Damos o nome de *Método de Monte Carlo* a qualquer cálculo que façamos via simulação (seja com computadores, lançando agulhas, dados, etc.), onde utilizamos a Lei dos Grandes Números para justificar o referido cálculo, como no exemplo acima. Tal ideia é utilizada em diversos campos do conhecimento, podendo ser aplicada desde o cálculo de velocidade de moléculas a probabilidades relacionadas a campeonatos de futebol.

Números normais

Suponha que um número x tenha sido sorteado de modo uniforme em $[0, 1)$. O que podemos dizer sobre a frequência relativa com que cada dígito aparece na expansão decimal de x?

De modo mais formal, seja U uma variável aleatória com distribuição uniforme em $[0, 1)$ e defina $(X_n)_n$, tomando valores em $\{0, 1, \ldots, 9\}$ como sendo a sequência de dígitos na expansão decimal de U.

Analogamente ao Lema 4.28, a sequência $(X_n)_n$ é i.i.d. com distribuição uniforme no conjunto $\{0, 1, \ldots, 9\}$. Portanto, para todo dígito $k \in \{0, 1, \ldots, 9\}$, a sequência $(Z_n^k)_n$, dada por $Z_n^k = \mathbb{1}_{\{X_n = k\}}$, também é i.i.d. com distribuição Bernoulli($\frac{1}{10}$). Pela Lei Forte dos Grandes Números,

$$\frac{Z_1^k + \cdots + Z_n^k}{n} \overset{\text{q.c.}}{\to} \frac{1}{10}, \text{ para todo } k \in \{0, 1, \ldots, 9\}.$$

Dizemos que um número é *simplesmente normal na base* 10 se todos os dígitos de sua expansão decimal aparecem com frequência relativa igual a $\frac{1}{10}$. Observe que um número racional é simplesmente normal na base 10 se, e somente se, sua dízima periódica contém todos os 10 dígitos e todos eles têm a mesma frequência de $\frac{1}{10}$. Pela afirmação acima, um número sorteado uniformemente em $[0, 1)$ é simplesmente normal na base 10, quase certamente. Acredita-se que e, π e $\sqrt{2}$ sejam normais apesar de não haver ainda uma prova.

Dizemos que um número é *normal na base* 10 se, para todo $l \in \mathbb{N}$, todas as 10^l sequências de l dígitos aparecem em sua expansão decimal com frequência relativa igual a $\frac{1}{10^l}$. Utilizando a Lei Forte dos Grandes Números e com algum esforço adicional, pode-se mostrar que um número sorteado uniformemente em $[0, 1)$ é quase certamente normal na base 10. A prova será dada como exercício guiado ao final deste capítulo.

Dado um número natural $b \geqslant 2$, dizemos que um número real é *simplesmente normal na base* b se todos os dígitos de sua expansão na base b aparecem com frequência relativa igual a $\frac{1}{b}$, e dizemos que um número real é *normal na base* b se, para todo $l \in \mathbb{N}$, todas as b^l sequências de l dígitos aparecem em sua expansão decimal com frequência relativa igual a $\frac{1}{b^l}$. No mesmo exercício guiado vamos mostrar que um número sorteado uniformemente em $[0, 1)$ é quase certamente normal na base b.

Finalmente, dizemos que um número é *normal* se ele é normal em toda base. Observe que um número sorteado uniformemente em $[0, 1)$ é quase certamente normal, pois o evento de ser normal é interseção enumerável de eventos com probabilidade 1. Apesar disso, a construção explícita de um número normal é algo bastante desafiador.

8.5 Exercícios

§8.1

1. Se, no jogo de roleta, apostamos 1 ficha no número 27, o nosso ganho esperado é de $-\frac{2}{38}$ fichas. O que diz a Lei dos Grandes Números sobre o ganho acumulado de um jogador que repete esta aposta indefinidamente? E se ele variar o número apostado?

2. Temos dois dados honestos, um cúbico numerado de 1 a 6 e um octaédrico numerado de 1 a 8. Um destes dados é escolhido aleatoriamente e lançado n vezes.

(a) A Lei dos Grandes Números pode ser usada para prever a frequência relativa com que aparece o número 6?
(b) Descreva um critério que permita inferir qual dado foi escolhido sabendo-se apenas a frequência relativa do número 6.
(c) Encontre um valor de n para o qual seja possível afirmar que a probabilidade de que o critério descrito no item anterior acerte seja de pelo menos $\frac{9}{10}$.

3. Exiba um exemplo de sequência de variáveis aleatórias ilustrando que a hipótese de variáveis não-correlacionadas não pode ser retirada na Lei Fraca de Tchebyshev.

4. Exiba um exemplo de sequência de variáveis aleatórias ilustrando que a hipótese de variâncias limitadas não pode ser retirada na Lei Fraca de Tchebyshev. Dica: Tente com variáveis normais.

§8.2

5. Sejam $(X_n)_n$ uma sequência de variáveis aleatórias i.i.d. com distribuição normal padrão. Calcule quanto vale quase certamente o limite

$$\lim_{n\to\infty} \frac{(aX_1+b)^2+\cdots+(aX_n+b)^2}{X_1^2+\cdots+X_n^2}, \quad a,b\in\mathbb{R}.$$

6. Sejam $(X_n)_n$ uma sequência de variáveis aleatórias i.i.d. com distribuição $\mathcal{U}[0,1]$, calcule o limite quase certo da média geométrica

$$\sqrt[n]{X_1\cdots X_n}.$$

7. Seja $(X_n)_n$ uma sequência de variáveis aleatórias independentes com $X_n \sim \mathcal{U}[0,n]$. Defina $Z_n = \mathbb{1}_{\{X_{2n}>X_{2n+1}\}}$ e $S_n = Z_1+\cdots+Z_n$.

(a) Calcule $\lim_{n\to\infty}\frac{S_n}{n}$.
(b) A convergência acima vale quase certamente ou apenas em probabilidade? Justifique.

8. Seja $(X_n)_n$ uma sequência de variáveis aleatórias independentes com $\mathbb{P}(X_n = n^2) = \frac{1}{n^3} = 1 - \mathbb{P}(X_n = 0)$ para todo $n \in \mathbb{N}$. Vale a Lei Fraca dos Grandes Números ou a Lei Forte?

9. Seja $(X_n)_n$ uma sequência de variáveis aleatórias independentes com $\mathbb{P}(X_n = n^2) = \frac{1}{n^2} = 1 - \mathbb{P}(X_n = 0)$ para todo $n \in \mathbb{N}$. Vale a Lei Fraca dos Grandes Números ou a Lei Forte?

10. Seja $(X_n)_n$ uma sequência de variáveis aleatórias independentes com $\mathbb{P}(X_n = n) = \mathbb{P}(X_n = -n) = \frac{1}{2n \log n}$ e $\mathbb{P}(X_n = 0) = 1 - \frac{1}{n \log n}$ para todo $n \in \mathbb{N}$. Mostre que vale a Lei Fraca dos Grandes Números mas não vale a Lei Forte.

§8.3

11. Seja $(X_n)_n$ uma sequência de variáveis aleatórias independentes duas a duas. Diga se vale a Lei Fraca dos Grandes Números ou a Lei Forte nos seguintes casos:

(a) $\mathbb{P}(X_n = n) = \mathbb{P}(X_n = -n) = \frac{1}{2}$.
(b) $\mathbb{P}(X_n = 2^n) = \mathbb{P}(X_n = -2^n) = 2^{-2n-1}$ e $\mathbb{P}(X_n = 0) = 1 - 2^{-2n}$.
(c) $\mathbb{P}(X_n = 2^n) = 2^{-n}$ e $\mathbb{P}(X_n = 0) = 1 - 2^{-n}$.
(d) $\mathbb{P}(X_n = n^2) = \mathbb{P}(X_n = -n^2) = (n+1)^{-3}$ e $\mathbb{P}(X_n = 0) = 1 - 2(n+1)^{-3}$.

12. Seja $(X_n)_n$ uma sequência de variáveis aleatórias i.i.d. tais que $\mathbb{E}X_1 = +\infty$. Mostre que $\frac{S_n}{n} \overset{\text{q.c.}}{\to} +\infty$.

13 (Lei Forte para segundo momento limitado)**.** Seja $(X)_n$ uma sequência de variáveis aleatórias não-correlacionadas, tais que $\mathbb{E}X_n^2 < C$ para algum $C > 0$, para todo $n \in \mathbb{N}$. Vamos mostrar que

$$\frac{S_n - \mathbb{E}S_n}{n} \overset{\text{q.c.}}{\to} 0,$$

onde $S_n = X_1 + \cdots + X_n$ para todo $n \in \mathbb{N}$, seguindo os passos abaixo.

(a) Mostre que, sem perda de generalidade, podemos supor que $\mathbb{E}X_n = 0$ para todo $n \in \mathbb{N}$.

(b) Mostre que $\frac{S_{m^2}}{m^2} \overset{\text{q.c.}}{\to} 0$ quando $m \to \infty$.

(c) Dado $n \in \mathbb{N}$, seja m o único número $m \in \mathbb{N}$ tal que $m^2 \leqslant n < (m+1)^2$ e defina $r_n = m^2$. Mostre que $\frac{S_{r_n}}{n} \overset{\text{q.c.}}{\to} 0$.

(d) Mostre que

$$\mathbb{E}\Big(\frac{S_n - S_{r_n}}{n}\Big)^2 \leqslant \frac{2C}{n^{3/2}} + \frac{C}{n^2}.$$

(e) Mostre que $\frac{S_n - S_{r_n}}{n} \overset{\text{q.c.}}{\to} 0$.

(f) Conclua o exercício.

§8.4

14. Seguindo os passos abaixo, mostre que, quase certamente, um número X sorteado uniformemente no intervalo $[0,1]$ é normal na base b.

(a) Seja $(X_n)_n$ a sequência dos dígitos de X na base b, de forma que $(X_n)_n$ é i.i.d. com distribuição uniforme discreta em $\{0,\dots,b-1\}$. Dada um par ordenado $(z_1,z_2) \in \{0,1,\dots,b-1\}^2$, defina a sequência de variáveis aleatórias $(Y_n)_n$, onde $Y_n = \mathbb{1}_{\{(X_n,X_{n+1})=(z_1,z_2)\}}$ para todo $n \in \mathbb{N}$. Verifique que a sequência $(Y_n)_n$ é identicamente distribuída, não é independente, mas a sequência $(Y_{2n-1})_n$ é independente, assim como a sequência $(Y_{2n})_n$.

(b) Mostre que

$$\lim_n \frac{Y_1 + Y_3 + \cdots + Y_{2n-1}}{n} = \lim_n \frac{Y_2 + Y_4 + \cdots + Y_{2n}}{n} = \frac{1}{b^2},$$

portanto

$$\lim_n \frac{Y_1 + Y_2 + \cdots + Y_{2n}}{2n} = \frac{1}{b^2}.$$

(c) Mostre que

$$\left|\frac{Y_1 + Y_2 + \cdots + Y_{2n}}{2n} - \frac{Y_1 + Y_2 + \cdots + Y_{2n-1}}{2n-1}\right| \leqslant \frac{1}{2n-1}$$

e conclua que todo par ordenado $(z_1,z_2) \in \{0,1,\dots,b-1\}^2$ aparece na expansão de X na base b com frequência relativa igual a $\frac{1}{b^2}$ quase

certamente.

(d) Generalize os itens anteriores, considerando $l \geqslant 2$ inteiro e l-uplas ordenadas da forma $(z_1, \ldots, z_l) \in \{0, 1, \ldots, b-1\}^l$, e conclua que X é normal na base b quase certamente.

Capítulo 9

Teorema do Limite Central

Seja $(X_n)_n$ uma sequência i.i.d. de variáveis aleatórias. Pela Lei dos Grandes Números sabemos que a média observada $\frac{S_n}{n} = \frac{X_1+\cdots+X_n}{n}$ se aproxima de sua média μ para valores grandes de n, isto é,

$$\frac{S_n}{n} \to \mu.$$

Porém, não é razoável esperar que $\frac{S_n}{n}$ seja exatamente igual a μ. Então a primeira pergunta que surge é sobre a flutuação da média observada, $\frac{S_n}{n}$, em torno de sua média μ. Tipicamente, essa diferença $\frac{S_n}{n} - \mu$ ocorre em qual escala? Nessa escala, podemos dizer como se distribui essa flutuação?

9.1 Teorema do Limite Central

Continuando a discussão do preâmbulo deste capítulo, não é difícil adivinhar a escala em que ocorre essa flutuação. De fato, sabemos que $\mathbb{E}S_n = n\,\mathbb{E}X_1 = n\mu$ e, denotando $\sigma^2 = \mathbb{V}X_1$, temos $\mathbb{V}S_n = n\,\mathbb{V}X_1 = n\sigma^2$, logo

$$\sigma(S_n) = \sigma\sqrt{n}.$$

Ou seja, a esperança da média observada é μ e seu desvio-padrão é $\frac{\sigma}{\sqrt{n}}$. Isso é uma indicação de que tipicamente as flutuações assumem valores da ordem $\frac{\sigma}{\sqrt{n}}$. Vamos supor que esse argumento está correto para tentar entender qual poderia ser o comportamento estatístico das flutuações nessa escala. Primeiro escrevemos

$$\frac{S_n}{n} = \mu + \frac{\sigma}{\sqrt{n}}\, Y_n,$$

de forma que $\mathbb{E}Y_n = 0$ e $\mathbb{V}Y_n = 1$. Será que a distribuição de Y_n se aproxima de alguma distribuição que não depende de n? Observe que, se $X_n \sim \mathcal{N}(\mu, \sigma^2)$, vimos na Seção 4.3 que $S_n \sim \mathcal{N}(n\mu, n\sigma^2)$, logo $Y_n \sim \mathcal{N}(0,1)$. Ou seja, pelo menos neste caso, $Y_n \xrightarrow{\mathrm{d}} \mathcal{N}(0,1)$. Uma das razões pelas quais a distribuição normal é tão importante é que essa aproximação vale para qualquer distribuição com segundo momento finito.

Teorema 9.1 (Teorema do Limite Central, caso i.i.d.)**.** *Seja $(X_n)_{n\in\mathbb{N}}$ uma sequência de variáveis aleatórias i.i.d. não-degeneradas com segundo momento finito. Tome $S_n = X_1 + X_2 + \cdots + X_n$. Então,*

$$\frac{S_n - \mathbb{E}S_n}{\sqrt{\mathbb{V}S_n}} \xrightarrow{\mathrm{d}} \mathcal{N}(0,1). \tag{9.2}$$

Podemos reescrever o limite acima e obter, informalmente,

$$\frac{S_n}{n} \approx \mu + \frac{\sigma}{\sqrt{n}}\, Z,$$

com $Z \sim \mathcal{N}(0,1)$. Ou seja, a escala em que a média observada $\frac{S_n}{n}$ flutua em torno de μ é de fato dada por $\frac{\sigma}{\sqrt{n}}$. Ademais, seu comportamento nessa escala possui forte regularidade estatística, e sua distribuição se aproxima de uma normal padrão. Dito de outra forma, a distribuição de S_n pode ser aproximada por uma normal com mesma média e variância: $S_n \approx \mathcal{N}(\mathbb{E}S_n, \mathbb{V}S_n)$.

Enfatizamos que o teorema acima é muito abrangente, pois ele vale para *qualquer distribuição com segundo momento finito.* Ainda assim, ele não dá uma ideia do verdadeiro nível de generalidade com que o fenômeno

de aproximação à normal é observado na prática. Apesar de a hipótese $\mathbb{E}X_k^2 < \infty$ ser necessária para que (9.2) faça sentido, e abandonar a hipótese de independência estaria acima dos nossos objetivos, podemos questionar a necessidade de as variáveis da sequência $(X_n)_n$ terem a mesma distribuição.

Como regra geral, se n é grande e cada uma das X_k é pequena se comparada a $\sqrt{\mathbb{V}S_n}$, então a distribuição de $\frac{S_n - \mathbb{E}S_n}{\sqrt{\mathbb{V}S_n}}$ é aproximadamente uma normal padrão. Esse fenômeno ocorre em diversas situações, quando uma grandeza é determinada pela contribuição de muitíssimos fatores, cada um deles pequeno se comparado com o todo, e ocorre mesmo que esses fatores não tenham a mesma distribuição. Por isso a distribuição normal surge em tantos contextos: erros de medidas em experimentos feitos com cuidado profissional, altura das pessoas em uma população homogênea em sexo e etnia, erros de horário em relógios de alto padrão, etc...

O enunciado abaixo, ainda simples, vai além do caso de distribuições idênticas, aceitamos em troca a hipótese de terceiros momentos limitados.

Teorema 9.3. *Seja $(X_n)_{n\in\mathbb{N}}$ uma sequência de variáveis aleatórias independentes tais que, para algum $\varepsilon > 0$ e algum $M < \infty$, vale $\mathbb{E}|X_n^3| \leqslant M$ e $\mathbb{V}X_n \geqslant \varepsilon$ para todo n. Então,*

$$\frac{S_n - \mathbb{E}S_n}{\sqrt{\mathbb{V}S_n}} \xrightarrow{\mathrm{d}} \mathcal{N}(0,1).$$

O enunciado acima é um corolário do Teorema do Limite Central de Lyapunov, que será visto na Seção 9.3. Esse teorema é suficiente para a grande maioria das aplicações práticas. Sua prova é simples e bastante transparente.

Ainda sem questionar a hipótese de independência, poderíamos perguntar-nos se as hipóteses do Teorema de Lyapunov com respeito a variância e terceiro momento são realmente necessárias. A resposta é dada pelo Teorema do Limite Central de Lindeberg, visto na Seção 9.4. Ali vamos demonstrar o limite (9.2) sob hipóteses mínimas. Concluímos este capítulo com a Seção 9.5, onde provamos os Teoremas de Slutsky e o Método Delta, que têm importantes aplicações em Estatística.

Se ao Teorema 9.1 agregarmos a hipótese de terceiro momento finito, ele passa a ser um corolário do Teorema 9.3. Sem essa hipótese, será obtido como aplicação do Teorema do Limite Central de Lindeberg na Seção 9.4, ou como aplicação do Teorema da Continuidade de Lévy para funções características, que será vista mais adiante, no Capítulo 10.

O caso mais simples do Teorema do Limite Central é para distribuição de Bernoulli, conhecido como Teorema de De Moivre-Laplace. Sua demonstração requer apenas conhecimentos de Cálculo e alguma perseverança com contas, e este será o primeiro que veremos neste capítulo.

As três seções seguintes são independentes entre si, é possível ler qualquer uma delas mesmo tendo saltado as anteriores.

9.2 Teorema de De Moivre-Laplace

O exemplo mais simples da aproximação $S_n \approx \mathcal{N}(n\mu, n\sigma^2)$ é quando lançamos uma moeda honesta n vezes e contamos o número S_n de caras. Neste caso S_n

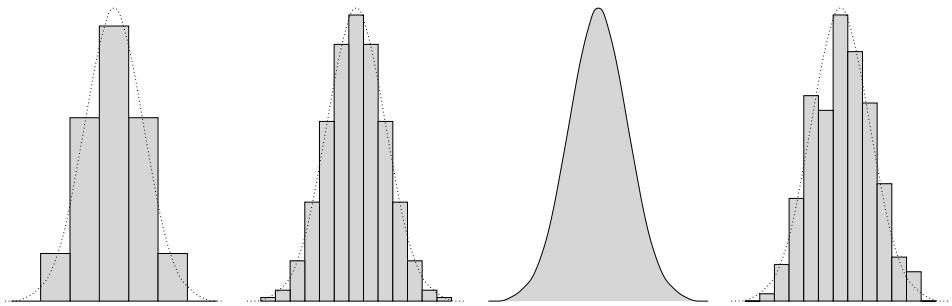


Figura 9.1. Função de probabilidade de $\frac{S_n - n\mu}{\sigma\sqrt{n}}$ para S_n com distribuições Binom$(4, \frac{1}{2})$ e Binom$(16, \frac{1}{2})$ para valores entre -3 e 3. A área de cada retângulo é dada pela função de probabilidade. O terceiro gráfico é a função de densidade de uma normal padrão, assim como as linhas pontilhadas. O quarto gráfico representa as frequências relativas de $\frac{S_n - n\mu}{\sigma\sqrt{n}}$ para S_n com distribuição Binom$(16, \frac{1}{2})$, em um experimento real com 200 amostras.

tem distribuição Binom$(n, \frac{1}{2})$. Na Figura 9.1 vemos como essa distribuição, devidamente normalizada, se aproxima da distribuição normal padrão.

Teorema 9.4 (Teorema de De Moivre-Laplace)**.** *Seja* $(X_n)_{n\in\mathbb{N}}$ *uma sequência de variáveis aleatórias independentes, com distribuição* Bernoulli(p)*, onde* $p = 1 - q \in (0, 1)$*, e tome* $S_n = X_1 + \cdots + X_n$*. Então para todos* $a < b$

$$\mathbb{P}\left(a < \frac{S_n - np}{\sqrt{npq}} \leqslant b\right) \to \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}}\, \mathrm{d}x.$$

O teorema foi provado inicialmente por De Moivre supondo que $p = \frac{1}{2}$ e generalizado por Laplace para $0 < p < 1$. De fato, ele segue de uma aproximação muito mais fina:[15]

$$\frac{n!}{k!\,(n-k)!}\, p^k\, q^{n-k} \sim \frac{1}{\sqrt{2\pi npq}} e^{-\frac{x_k^2}{2}}, \tag{9.5}$$

onde

$$x_k = x_{n,k} = \frac{k - np}{\sqrt{npq}}.$$

Nesta seção, denotaremos por $a_{n,k} \sim b_{n,k}$ a propriedade de que

$$\max_{k:|x_k|\leqslant M} \frac{a_{n,k}}{b_{n,k}} \to 1 \quad \text{e} \quad \min_{k:|x_k|\leqslant M} \frac{a_{n,k}}{b_{n,k}} \to 1 \tag{9.6}$$

quando $n \to \infty$, para todo $M < \infty$ fixo.

Essa aproximação é muito mais fina porque diz não apenas que a probabilidade de a flutuação estar dentro de um certo intervalo é bem aproximada pela normal, mas também que a função de probabilidade de cada um dos possíveis valores dentro de um intervalo fixo é aproximada pela densidade da normal.

Para entender de onde vem essa aproximação, primeiro precisamos de uma expressão mais palpável para $n!$. Qual a probabilidade de obtermos

[15] Essa aproximação é conhecida como Teorema do Limite Local e é aparentemente a primeira vez na Matemática em que aparece a famosa função $e^{-x^2/2}$.

exatamente 60 caras se lançamos uma moeda 120 vezes? A resposta é fácil:

$$\mathbb{P}(S_{120} = 60) = \frac{120!}{60!\,60!} \times \frac{1}{2^{120}}.$$

Essa expressão é simples e matematicamente perfeita. Porém, quanto vale essa probabilidade? Mais de 0,15? Menos de 0,05? Entre 0,05 e 0,10? Boas calculadoras científicas não têm capacidade de calcular 120!. Num computador esse cálculo resulta em 0,072684979. Mas e se fossem 40.000 lançamentos da moeda? E se estivéssemos interessados em calcular $\mathbb{P}(S_{40.000} \leqslant 19.750)$, faríamos um cálculo semelhante 250 vezes para depois somar? As expressões com fatorial são perfeitas para a combinatória, mas impraticáveis para se fazer estimativas. Nosso socorro será a *fórmula de Stirling*

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

provada no Apêndice B. Essa aproximação de $n!$ pela fórmula de Stirling é muito boa mesmo sem tomar n grande. Ela aproxima 1! por 0,92, 2! por 1,92, 4! por 23,5, e a partir de $9! = 362.880$, que é aproximado por 359.537, o erro é menor que 1%. À primeira vista $n^n\, e^{-n}\, \sqrt{2\pi n}$ não parece mais agradável do que $n!$. Mas vejamos como isso ajuda com o cálculo anterior. Temos:

$$\frac{(2k)!}{k!\,k!} \times \frac{1}{2^{2k}} \sim \frac{(2k)^{2k} e^{-2k} \sqrt{4\pi k}}{(k^k e^{-k} \sqrt{2\pi k})^2} \times \frac{1}{2^{2k}} = \frac{1}{\sqrt{\pi k}}$$

que para $k = 60$ pode até ser calculado à mão, obtendo-se $0{,}0728\ldots$. Mais do que isso, acabamos de obter a aproximação (9.5) no caso particular em que $p = q = \frac{1}{2}$ e $x_k = 0$.

Vamos mostrar (9.5) para M fixo. Aplicando a fórmula de Stirling obtemos

$$\begin{aligned}\frac{n!}{k!\,(n-k)!}\; p^k\; q^{n-k} &\sim \frac{n^n e^{-n} \sqrt{2\pi n}\, p^k\, q^{n-k}}{k^k e^{-k} \sqrt{2\pi k} (n-k)^{n-k} e^{k-n} \sqrt{2\pi (n-k)}} \\ &= \frac{(\frac{np}{k})^k (\frac{nq}{n-k})^{n-k}}{\sqrt{2\pi k (n-k)/n}}.\end{aligned}$$

Observe que, com a restrição $|x_k| < M$, vale

$$k = np + \sqrt{npq}\, x_k \sim np \qquad \text{e} \qquad n - k = nq - \sqrt{npq}\, x_k \sim nq.$$

Daí obtemos

$$\frac{n!}{k!\,(n-k)!}\; p^k\; q^{n-k} \sim \frac{\left(\frac{np}{k}\right)^k \left(\frac{nq}{n-k}\right)^{n-k}}{\sqrt{2\pi npq}} = \frac{f(n,k)}{\sqrt{2\pi npq}}.$$

Vamos estudar $\log f(n,k)$. Reescrevendo cada termo,

$$\frac{np}{k} = 1 - \frac{\sqrt{npq}\, x_k}{k} \qquad \text{e} \qquad \frac{nq}{n-k} = 1 + \frac{\sqrt{npq}\, x_k}{n-k}.$$

A expansão de Taylor de $\log(1+x)$ até o termo de segunda ordem nos dá

$$\log(1+x) = x - \frac{x^2}{2} + r(x), \qquad \text{onde } \frac{r(x)}{x^2} \to 0 \text{ quando } x \to 0.$$

Assim,

$$\begin{aligned}\log f(n,k) = k\left(-\frac{\sqrt{npq}\, x_k}{k} - \frac{npqx_k^2}{2k^2} + r\left(\tfrac{\sqrt{npq}\, x_k}{k}\right)\right) + \\ + (n-k)\left(\frac{\sqrt{npq}\, x_k}{n-k} - \frac{npqx_k^2}{2(n-k)^2} + r\left(\tfrac{\sqrt{npq}\, x_k}{n-k}\right)\right).\end{aligned}$$

Note que os primeiros termos se cancelam e, quando $n \to \infty$,

$$\log f(n,k) \sim -\frac{npqx_k^2}{2k} - \frac{npqx_k^2}{2(n-k)} \sim -\frac{qx_k^2}{2} - \frac{px_k^2}{2} = -\frac{x_k^2}{2}.$$

Daí segue que

$$f(n,k) \sim e^{-\frac{x_k^2}{2}}$$

pois, sob a restrição $|x_k| < M$, podemos tomar a exponencial de ambos os lados e ainda manter a equivalência assintótica. Isso termina a prova de (9.5).

Somando sobre os possíveis valores de S_n, obtemos

$$\mathbb{P}\Big(a < \tfrac{S_n - np}{\sqrt{npq}} \leqslant b\Big) = \sum_{a < x_k \leqslant b} \mathbb{P}(S_n = k) = \sum_{a < x_k \leqslant b} \frac{n!}{k!\,(n-k)!}\, p^k\, q^{n-k},$$

onde os somatórios são sobre k com a condição sobre x_k, que é dado por $x_k = \frac{k-np}{\sqrt{npq}}$. Observando que

$$x_{k+1} - x_k = \frac{1}{\sqrt{npq}},$$

e substituindo (9.5), obtemos

$$\mathbb{P}\Big(a < \tfrac{S_n - np}{\sqrt{npq}} \leqslant b\Big) \sim \sum_{a < x_k \leqslant b} \frac{e^{-\frac{x_k^2}{2}}}{\sqrt{2\pi npq}} = \frac{1}{\sqrt{2\pi}} \sum_{a < x_k \leqslant b} e^{-\frac{x_k^2}{2}} \cdot [x_{k+1} - x_k].$$

Aqui há um detalhe sutil a observar: a equivalência assintótica está relacionada com um certo limite quando $n \to \infty$ e, como a quantidade de valores de k que entram no somatório acima varia com n, é importante que estamos usando (9.6) como definição de equivalência assintótica.

Finalmente, observamos que o somatório acima é uma soma de Riemann. Pelo Teorema A.3, ela se aproxima de $\frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}}\, \mathrm{d}x$ quando $n \to \infty$. Isso termina a prova do Teorema 9.4.

9.3 Teorema do Limite Central de Lyapunov

Nesta seção apresentaremos uma versão de Teorema do Limite Central que, ainda sob a hipótese de independência, é a primeira versão que permite variáveis aleatórias com distribuições diferentes.

Teorema 9.7 (Teorema do Limite Central de Lyapunov)**.** *Sejam* $(X_n)_n$

variáveis aleatórias independentes com terceiros momentos finitos satisfazendo

$$\frac{\sum_{j=1}^{n} \mathbb{E}|X_j - \mathbb{E}X_j|^3}{\left(\sum_{j=1}^{n} \mathbb{V}X_j\right)^{3/2}} \to 0, \text{ quando } n \to \infty.$$

Então,

$$\frac{S_n - \mathbb{E}S_n}{\sqrt{\mathbb{V}S_n}} \xrightarrow{\mathrm{d}} \mathcal{N}(0,1).$$

Demonstração. Pelo Teorema de Helly-Bray, basta mostrar que

$$\mathbb{E}\left[f\left(\frac{S_n - \mathbb{E}S_n}{\sqrt{\mathbb{V}S_n}}\right)\right] \to \int_{-\infty}^{\infty} f(x)\frac{e^{-x^2/2}}{\sqrt{2\pi}}\mathrm{d}x \quad \text{quando } n \to \infty,$$

para qualquer f tal que f, f', f'' e f''' sejam limitadas.

Podemos assumir, sem perda de generalidade, que $\mathbb{E}X_k = 0$. Escreva $\sigma_k = \sqrt{\mathbb{V}X_k}$. Considere $(Y_n)_n$ independentes, e também independentes de $(X_n)_n$, com a distribuição $\mathcal{N}(0, \sigma_k^2)$. (Aceitaremos sem prova existência de tal sequência no mesmo espaço de probabilidade.)

Escreva

$$W_n = \frac{Y_1 + \cdots + Y_n}{\sqrt{\mathbb{V}S_n}} \sim \mathcal{N}(0,1) \quad \text{e} \quad Z_n = \frac{X_1 + \cdots + X_n}{\sqrt{\mathbb{V}S_n}}.$$

Fixe n, defina $Z^0 = W_n$ e, para $k = 1, \ldots, n$, defina

$$U^k = Z^{k-1} - \frac{Y_k}{\sqrt{\mathbb{V}S_n}} \quad \text{e} \quad Z^k = U^k + \frac{X_k}{\sqrt{\mathbb{V}S_n}},$$

de modo que $Z^n = Z_n$. Agora,

$$\mathbb{E}[f(Z_n)] - \mathbb{E}[f(W_n)] = \mathbb{E}[f(Z^n)] - \mathbb{E}[f(Z^0)] = \sum_{k=1}^{n} \mathbb{E}[f(Z^k)] - \mathbb{E}[f(Z^{k-1})].$$

Usando a expansão de Taylor de f com resto de Lagrange,

$$\begin{aligned} f(Z^k) &= f(U_k + \tfrac{X_k}{\sqrt{\mathbb{V}S_n}}) \\ &= f(U^k) + f'(U^k)\frac{X_k}{\sqrt{\mathbb{V}S_n}} + + \frac{f''(U_k)}{2}\frac{X_k^2}{\mathbb{V}S_n} + \frac{f'''(\theta^k)}{6}\frac{X_k^3}{(\mathbb{V}S_n)^{3/2}}, \end{aligned}$$

e

$$\begin{aligned} f(Z^{k-1}) &= f(U_k + \tfrac{Y_k}{\sqrt{\mathbb{V}S_n}}) \\ &= f(U^k) + f'(U^k)\frac{Y_k}{\sqrt{\mathbb{V}S_n}} + \frac{f''(U_k)}{2}\frac{Y_k^2}{\mathbb{V}S_n} + \frac{f'''(\tilde{\theta}^k)}{6}\frac{Y_k^3}{(\mathbb{V}S_n)^{3/2}}, \end{aligned}$$

onde θ_k e $\tilde{\theta}_k$ vêm da expansão de Taylor e dependem de U_k, X_k e Y_k.

Observe que $\mathbb{E}X_k = \mathbb{E}Y_k$, $\mathbb{E}X_k^2 = \mathbb{E}Y_k^2$, que U_k é independente de X_k e Y_k, e recordemos que f e suas três primeiras derivadas são limitadas. Queremos tomar a esperança acima e subtrair. Como f é limitada, $f(U^k)$ é integrável e este primeiro termo se cancela. Além disso, $f'(U^k)$ é integrável e, usando a independência,

$$\mathbb{E}[f'(U^k)X_k] = \mathbb{E}[f'(U^k)] \cdot \mathbb{E}X_k = \mathbb{E}[f'(U^k)] \cdot \mathbb{E}Y_k = \mathbb{E}[f'(U^k)Y_k],$$

de forma que este termo também se cancela, bem como o terceiro termo pelo mesmo motivo. Portanto,

$$\mathbb{E}[f(Z^k)] - \mathbb{E}[f(Z^{k-1})] = \frac{\mathbb{E}[f'''(\theta_k)X_k^3] - \mathbb{E}[f'''(\tilde{\theta}_k)Y_k^3]}{6(\mathbb{V}S_n)^{3/2}}.$$

Tomando $C = \sup_x |f'''(x)|$, podemos estimar

$$\begin{aligned} \Big|\mathbb{E}[f(Z^k)] - \mathbb{E}[f(Z^{k-1})]\Big| &\leqslant \frac{C\,\mathbb{E}|X_k|^3 + C\mathbb{E}|Y_k|^3}{6(\mathbb{V}S_n)^{3/2}} \\ &\leqslant C\frac{\mathbb{E}|X_k|^3}{(\mathbb{V}S_n)^{3/2}}, \end{aligned}$$

onde na última desigualdade usamos que $\mathbb{E}|Y_k|^3 = \frac{4\sigma_k^3}{\sqrt{2\pi}} \leqslant 2\sigma_k^3$ e que

$$\sigma_k = (\mathbb{E}|X_k|^2)^{\frac{1}{2}} \leqslant (\mathbb{E}|X_k|^3)^{\frac{1}{3}},$$

pela desigualdade de Lyapunov. Finalmente, somando sobre k,

$$\begin{aligned}\Big|\mathbb{E}[f(Z^n)] - \mathbb{E}[f(Z^0)]\Big| &\leqslant \sum_{k=1}^{n}\Big|\mathbb{E}[f(Z^k)] - \mathbb{E}[f(Z^{k-1})]\Big| \\ &\leqslant C\frac{\sum_{k=1}^{n}\mathbb{E}|X_k|^3}{(\sum_{k=1}^{n}\mathbb{V}X_k)^{3/2}} \to 0\end{aligned}$$

quando $n \to \infty$ por hipótese, o que conclui a prova do teorema. □

9.4 Teorema do Limite Central de Lindeberg

Nesta seção provaremos a versão mais geral do Teorema do Limite Central no contexto de sequências de variáveis independentes, o chamado Teorema de Lindeberg.

Ao longo desta seção, sempre suporemos que $(X_n)_n$ é sequência de variáveis aleatórias definidas em um mesmo espaço de probabilidade, para cada $n \in \mathbb{N}$ denotamos $S_n = \sum_{k=1}^{n} X_k$, $\mu_k = \mathbb{E}X_k$, $\sigma_n^2 = \mathbb{V}X_n$, $s_n^2 = \mathbb{V}S_n$ e assumiremos que $s_n \to +\infty$.

Definição 9.8 (Condição de Lindeberg)**.** Dizemos que a sequência de variáveis aleatórias $(X_n)_n$ satisfaz à *condição de Lindeberg* se

$$\lim_{n\to\infty}\frac{1}{s_n^2}\sum_{k=1}^{n}\int_{\{|X_k-\mu_k|>\varepsilon s_n\}}(X_k-\mu_k)^2\,\mathrm{d}\mathbb{P} = 0$$

para todo $\varepsilon > 0$.

Agora estamos prontos para enunciar o Teorema de Lindeberg.

Teorema 9.9 (Teorema de Lindeberg)**.** *Seja $(X_n)_n$ uma sequência de variáveis aleatórias independentes. Se a condição de Lindeberg é satisfeita, então*

$$\frac{S_n - \mathbb{E}S_n}{\sqrt{\mathbb{V}S_n}} \xrightarrow{\mathrm{d}} \mathcal{N}(0,1).$$

Ao final desta seção, enunciaremos uma versão mais geral, em que as variáveis $X_1, \ldots, X_n$ que compõem S_n serão denotadas $X_{n,1}, \ldots, X_{n,n}$ porque podem ser diferentes para cada n.

Demonstração do Teorema 9.1. Basta verificar que, no caso i.i.d. com segundo momento finito, a condição de Lindeberg é satisfeita. Escrevemos $\mu = \mathbb{E}X_1$ e $\sigma^2 = \mathbb{V}X_1$, assim $s_n^2 = n\sigma^2$. Usando o Teorema da Convergência Dominada,

$$\frac{1}{s_n^2}\sum_{k=1}^{n}\int_{\{|X_k-\mu|>\varepsilon s_n\}}(X_k-\mu)^2\,\mathrm{d}\mathbb{P} = \frac{1}{\sigma^2}\int_{\{|X_1-\mu|>\varepsilon\sigma\sqrt{n}\}}(X_1-\mu)^2\,\mathrm{d}\mathbb{P} \to 0,$$

pois $\{|X_1-\mu| > \varepsilon\sigma\sqrt{n}\} \downarrow \emptyset$ quando $n \to \infty$. □

Para o Teorema do Limite Central de Lyapunov, obtemos a versão abaixo que é mais geral do que aquela provada na Seção 9.3.

Corolário 9.10 (Teorema do Limite Central de Lyapunov)**.** *Seja $(X_n)_n$ uma sequência de variáveis aleatórias independentes. Se existe $\delta > 0$ tal que*

$$\lim_{n\to\infty}\frac{1}{s_n^{2+\delta}}\sum_{k=1}^{n}\mathbb{E}|X_k-\mu_k|^{2+\delta} = 0,$$

então

$$\frac{S_n - \mathbb{E}S_n}{\sqrt{\mathbb{V}S_n}} \xrightarrow{\mathrm{d}} \mathcal{N}(0,1).$$

Demonstração. Observe que

$$\frac{1}{s_n^{2+\delta}}\sum_{k=1}^{n}\mathbb{E}|X_k-\mu_k|^{2+\delta} \geqslant \frac{1}{s_n^{2+\delta}}\sum_{k=1}^{n}\int_{\{|X_k-\mu_k|>\varepsilon s_n\}}|X_k-\mu_k|^{2+\delta}\,\mathrm{d}\mathbb{P}$$

$$
\begin{aligned}
&= \frac{1}{s_n^{2+\delta}} \sum_{k=1}^{n} \int_{\{|X_k-\mu_k|>\varepsilon s_n\}} (X_k-\mu_k)^2 |X_k-\mu_k|^\delta \, \mathrm{d}\mathbb{P} \\
&\geqslant \frac{\varepsilon^\delta}{s_n^2} \sum_{k=1}^{n} \int_{\{|X_k-\mu_k|>\varepsilon s_n\}} (X_k-\mu_k)^2 \, \mathrm{d}\mathbb{P}
\end{aligned}
$$

para todo $\varepsilon > 0$ e, portanto, vale a condição de Lindeberg. □

Apesar de ser mais simples de ser verificada na prática, a condição de Lyapunov é mais restritiva que a de Lindeberg, como vemos no seguinte exemplo.

Exemplo 9.11. Seja $(X_n)_n$ uma sequência de variáveis aleatórias i.i.d. com distribuição comum $\mathbb{P}(X = k) = \frac{C}{k^3 \log^2 k}$ para todo $k \in \mathbb{N}$, onde C é a constante tal que $C^{-1} = \sum_{k=1}^{\infty} \frac{1}{k^3 \log^2 k}$. Podemos verificar que $\mathbb{E}X^2 < \infty$, logo a sequência $(X_n)_n$ satisfaz à condição de Lindeberg conforme visto na demonstração do Teorema 9.1. Por outro lado, $\mathbb{E}|X|^{2+\delta} = +\infty$ para todo $\delta > 0$, portanto a condição de Lyapunov nunca é satisfeita. △

A condição de Lindeberg é uma forma de quantificar a ideia de que a contribuição de cada parcela $\frac{X_k-\mu_k}{s_n}$ na soma $\frac{S_n-\mathbb{E}S_n}{s_n}$ é pequena quando n se torna grande. Mais precisamente, ela diz que as contribuições para a esperança do desvio quadrático, mesmo quando somadas, provêm de desvios relativamente pequenos se comparados com o desvio-padrão de S_n. A proposição abaixo diz que, neste caso, a contribuição de cada parcela na variância de S_n é desprezível.

Proposição 9.12. *Seja $(X_n)_n$ uma sequência de variáveis aleatórias satisfazendo à condição de Lindeberg. Então, $\frac{\sigma_n^2}{s_n^2} \to 0$.*

Omitimos a prova pois será um caso particular da Proposição 9.15.

Exemplo 9.13. Seja $(X_n)_n$ uma sequência de variáveis aleatórias independentes, com $X_1 \sim \mathcal{N}(0,1)$ e $X_n \sim \mathcal{N}(0, 2^{n-2})$ para todo $n \geqslant 2$. Como $s_n^2 = 2^{n-1}$,

$$
\lim_{n\to\infty} \frac{\sigma_n^2}{s_n^2} = \lim_{n\to\infty} \frac{2^{n-2}}{2^{n-1}} = \frac{1}{2}.
$$

Pela Proposição 9.12, a condição de Lindeberg não é satisfeita. Entretanto, vale (9.2), pois S_n tem distribuição normal para todo n. $\triangle$

O exemplo acima ilustra que a condição de Lindeberg não é necessária para que valha a conclusão do Teorema do Limite Central. O Teorema de Feller, que enunciaremos abaixo sem prova, diz que esse tipo de exemplo somente pode ocorrer quando σ_n^2 é responsável por uma fração não-desprezível de s_n^2. A prova desse teorema encontra-se na Seção III.4 de Shiryaev (1996) ou na Seção XV.6 de Feller (1971).

Teorema 9.14 (Teorema de Feller)**.** *Seja $(X_n)_n$ uma sequência de variáveis aleatórias independentes. Se $\frac{S_n - \mathbb{E}S_n}{\sqrt{\mathbb{V}S_n}} \xrightarrow{\mathrm{d}} \mathcal{N}(0,1)$ e $\frac{\sigma_n^2}{s_n^2} \to 0$ quando $n \to \infty$, então a condição de Lindeberg é satisfeita.*

Concluímos esta seção com a versão do Teorema de Lindeberg para arranjos triangulares de variáveis aleatórias. Para isso, a partir de agora consideremos uma família de variáveis aleatórias da forma $(X_{k,n})_{n\in\mathbb{N},k=1,\ldots,n}$. Definimos $\mu_{k,n} = \mathbb{E}X_{k,n}$, $\sigma_{k,n}^2 = \mathbb{V}X_{k,n}$, $S_n = \sum_{k=1}^n X_{k,n}$ e $s_n^2 = \mathbb{V}S_n$.

Dizemos que o arranjo $(X_{k,n})_{n,k}$ satisfaz à *condição de Lindeberg* se

$$\lim_{n\to\infty} \frac{1}{s_n^2} \sum_{k=1}^n \int_{\{|X_{k,n}-\mu_{k,n}|>\varepsilon s_n\}} (X_{k,n} - \mu_{k,n})^2 \, \mathrm{d}\mathbb{P} = 0$$

para todo $\varepsilon > 0$.

Observe que, dada uma sequência $(X_n)_{n\in\mathbb{N}}$ de variáveis aleatórias satisfazendo à condição de Lindeberg para sequências, definida no início desta seção, podemos definir a família $(Z_{k,n})_{n,k}$ por $Z_{k,n} = X_k$, e esse arranjo satisfará à condição de Lindeberg para arranjos triangulares, definida logo acima. Assim, o Teorema 9.9 é corolário desse que enunciaremos e provaremos mais abaixo. Para isso, vamos precisar da seguinte generalização da Proposição 9.12.

Proposição 9.15. *Seja $(X_{k,n})_{n,k}$ um arranjo triangular de variáveis*

aleatórias satisfazendo à condição de Lindeberg. Então,

$$\lim_{n\to\infty} \max_{1\leqslant k\leqslant n} \frac{\sigma_{k,n}^2}{s_n^2} = 0.$$

Demonstração. Defina $Y_{k,n} = \frac{X_{k,n}-\mu_{k,n}}{s_n}$ para $k = 1, \ldots, n$. Dado qualquer $\varepsilon > 0$,

$$\begin{aligned}
\max_{1\leqslant k\leqslant n} \frac{\sigma_{k,n}^2}{s_n^2} &= \max_{1\leqslant k\leqslant n} \mathbb{E}Y_{k,n}^2 \\
&\leqslant \varepsilon^2 + \max_{1\leqslant k\leqslant n} \int_{\{|Y_{k,n}|>\varepsilon\}} Y_{k,n}^2 \,\mathrm{d}\mathbb{P} \\
&\leqslant \varepsilon^2 + \sum_{k=1}^{n} \int_{\{|Y_{k,n}|>\varepsilon\}} Y_{k,n}^2 \,\mathrm{d}\mathbb{P} \\
&= \varepsilon^2 + \frac{1}{s_n^2} \sum_{k=1}^{n} \int_{\{|X_{k,n}-\mu_{k,n}|>\varepsilon s_n\}} (X_{k,n} - \mu_{k,n})^2 \,\mathrm{d}\mathbb{P}.
\end{aligned}$$

Como a condição de Lindeberg é satisfeita,

$$\limsup_{n\to\infty} \max_{1\leqslant k\leqslant n} \frac{\sigma_{k,n}^2}{s_n^2} \leqslant \varepsilon^2, \text{ para todo } \varepsilon > 0,$$

o que conclui a prova dessa proposição. □

Teorema 9.16 (Teorema de Lindeberg para arranjos triangulares)**.** *Seja $(X_{k,n})_{n,k}$ um arranjo triangular de variáveis aleatórias independentes. Se a condição de Lindeberg é satisfeita, então*

$$\frac{S_n - \mathbb{E}S_n}{\sqrt{\mathbb{V}S_n}} \xrightarrow{\mathrm{d}} \mathcal{N}(0,1).$$

Demonstração. Sem perda de generalidade, podemos supor que $\mu_{k,n} = 0$ e $s_n = 1$ para todos $n \in \mathbb{N}$ e $k = 1, \ldots, n$. Queremos mostrar que $S_n \xrightarrow{\mathrm{d}} \mathcal{N}(0,1)$. Pelo Teorema de Helly-Bray, é suficiente mostrar que $\mathbb{E}[f(S_n)] \to \mathbb{E}[f(N)]$,

para toda função $f : \mathbb{R} \to \mathbb{R}$ tal que f, f', f'', f''' são contínuas e limitadas, onde $N \sim \mathcal{N}(0,1)$. Fixe f com tais propriedades.

Seja $(N_{k,n})_{n,k}$ um arranjo triangular de variáveis aleatórias independentes entre si, independentes de $(X_{k,n})_{n,k}$ e tais que $N_{k,n} \sim \mathcal{N}(0, \sigma_{k,n}^2)$. Tomando um espaço produto, podemos supor que todas essas variáveis estão definidas no mesmo espaço de probabilidade que $(X_{n,k})_{n,k}$. Observe que, para todo $n \in \mathbb{N}$,

$$\sum_{k=1}^{n} N_{k,n} \sim \mathcal{N}(0,1),$$

donde

$$\mathbb{E}\left[f\left(\sum_{k=1}^{n} N_{k,n}\right)\right] = \mathbb{E}[f(N)].$$

Fixado $n \in \mathbb{N}$, para cada $k = 1, \ldots, n$, defina a variável

$$Z_{k,n} = \sum_{j=1}^{k-1} X_{j,n} + \sum_{j=k+1}^{n} N_{j,n}$$

e observe que $Z_{n,n} + X_{n,n} = S_n$ e $Z_{1,n} + N_{1,n} \sim \mathcal{N}(0,1)$. Observando também que $Z_{k,n} + X_{k,n} = Z_{k+1,n} + N_{k+1,n}$, obtemos

$$\begin{aligned}
|\mathbb{E}[f(S_n)] - \mathbb{E}[f(N)]| &= \big|\mathbb{E}\big[f(Z_{n,n} + X_{n,n})\big] - \mathbb{E}\big[f(Z_{1,n} + N_{1,n})\big]\big| \\
&= \left|\sum_{k=1}^{n} \mathbb{E}\big[f(Z_{k,n} + X_{k,n})\big] - \mathbb{E}\big[f(Z_{k,n} + N_{k,n})\big]\right| \\
&\leqslant \sum_{k=1}^{n} \Big|\mathbb{E}[f(Z_{k,n} + X_{k,n})] - \mathbb{E}[f(Z_{k,n} + N_{k,n})]\Big|.
\end{aligned}$$

A ideia é mostrar que, ao trocarmos as variáveis da sequência $(X_{k,n})$ pelas de $(N_{k,n})$, uma de cada vez para $k = 1, \ldots, n$, a distância entre $Z_{k,n} + X_{k,n}$ e $Z_{k,n} + N_{k,n}$ será pequena, mesmo depois de somar sobre k.

Defina $g : \mathbb{R} \to \mathbb{R}$ como

$$g(h) = \sup_{x \in \mathbb{R}} \left| f(x) - f(x+h) - f'(x)h - \frac{f''(x)h^2}{2} \right|.$$

Pela Fórmula de Taylor com resto de Lagrange de terceira ordem,

$$f(x) - f(x+h) - f'(x)h - \frac{f''(x)}{2}h^2 = \frac{f'''(y)}{3!}h^3,$$

para algum $y \in (x-h, x+h)$. Como f''' é limitada, existe constante $C > 0$, tal que $g(h) \leqslant C|h|^3$ para todo $h \in \mathbb{R}$.

Agora, pela fórmula de Taylor com resto de Lagrange de segunda ordem,

$$f(x) - f(x+h) - f'(x)h - \frac{f''(x)}{2}h^2 = \frac{f''(y) - f''(x)}{2}h^2,$$

para algum $y \in (x-h, x+h)$. Novamente, como f'' é limitada, existe uma outra constante $M > 0$ tal que $g(h) \leqslant M|h|^2$ para todo $h \in \mathbb{R}$. Essa cota é melhor que a anterior quando h não é muito pequeno.

Por outro lado, a partir da definição de g, podemos verificar que

$$\left| f(x+h_1) - f(x+h_2) - f'(x)(h_2 - h_1) - \frac{f''(x)(h_2^2 - h_1^2)}{2} \right| \leqslant g(h_1) + g(h_2).$$

Aplicando à diferença das esperanças que queremos estimar, usaremos as cotas acima com $x = Z_{k,n}$, $h_1 = X_{k,n}$, e $h_2 = N_{k,n}$. Da desigualdade acima,

$$\begin{aligned} |\mathbb{E}[f(Z_{k,n} + X_{k,n}) - f(Z_{k,n} + N_{k,n})]| &\leqslant \mathbb{E}[g(X_{k,n}) + g(N_{k,n})] + \\ &+ |\mathbb{E}[f'(Z_{k,n})(X_{k,n} - N_{k,n})]| + \left| \mathbb{E}\left[\tfrac{1}{2} f''(Z_{k,n}) \Big(X_{k,n}^2 - N_{k,n}^2 \Big) \right] \right|. \end{aligned}$$

Observe que $Z_{k,n}$ é independente de $X_{k,n}$ e $N_{k,n}$. Como $\mathbb{E}[X_{k,n} - N_{k,n}] = \mathbb{E}[X_{k,n}^2 - N_{k,n}^2] = 0$, as parcelas envolvendo f' e f'' na estimativa acima são

nulas. Portanto,

$$|\mathbb{E}[f(S_n)] - \mathbb{E}[f(N)]| \leqslant \sum_{k=1}^{n} \big(\mathbb{E}[g(X_{k,n})] + \mathbb{E}[g(N_{k,n})]\big).$$

Dessa forma, para concluir a demonstração basta mostrar que

$$\lim_{n\to\infty} \sum_{k=1}^{n} \mathbb{E}[g(X_{k,n})] = 0 \quad \text{e} \quad \lim_{n\to\infty} \sum_{k=1}^{n} \mathbb{E}[g(N_{k,n})] = 0.$$

Comecemos pelo primeiro limite. Seja $\varepsilon > 0$. Vamos estimar separando valores de X pequenos dos demais:

$$\begin{aligned}
\mathbb{E}[g(X_{k,n})] &= \int_{\{|X_{k,n}|\leqslant\varepsilon\}} g(X_{k,n})\,\mathrm{d}\mathbb{P} + \int_{\{|X_{k,n}|>\varepsilon\}} g(X_{k,n})\,\mathrm{d}\mathbb{P} \\
&\leqslant C \int_{\{|X_{k,n}|\leqslant\varepsilon\}} |X_{k,n}|^3\,\mathrm{d}\mathbb{P} + M \int_{\{|X_{k,n}|>\varepsilon\}} X_{k,n}^2\,\mathrm{d}\mathbb{P} \\
&\leqslant \varepsilon C \int_{\Omega} X_{k,n}^2\,\mathrm{d}\mathbb{P} + M \int_{\{|X_{k,n}|>\varepsilon\}} X_{k,n}^2\,\mathrm{d}\mathbb{P} \\
&\leqslant \varepsilon C \sigma_{k,n}^2 + M \int_{\{|X_{k,n}|>\varepsilon\}} X_{k,n}^2\,\mathrm{d}\mathbb{P}.
\end{aligned}$$

Na primeira desigualdade, utilizamos a cota $g(h) \leqslant C|h|^3$ para a primeira integral e $g(h) \leqslant Mh^2$ para a segunda. Somando em k, tomando o limite $n \to \infty$ e usando a condição de Lindeberg, obtemos

$$\limsup_{n\to\infty} \sum_{k=1}^{n} \mathbb{E}[g(X_{k,n})] \leqslant \varepsilon C + M \lim_{n\to\infty} \sum_{k=1}^{n} \int_{\{|X_{k,n}|>\varepsilon\}} X_{k,n}^2\,\mathrm{d}\mathbb{P} = \varepsilon C.$$

Como a desigualdade acima é válida para todo $\varepsilon > 0$,

$$\lim_{n\to\infty} \sum_{k=1}^{n} \mathbb{E}[g(X_{k,n})] = 0.$$

Para a estimativa do segundo limite, com $N_{k,n}$ no lugar de $X_{k,n}$, procedemos

de modo idêntico e obtemos

$$\lim_{n\to\infty}\sum_{k=1}^{n}\mathbb{E}[g(N_{k,n})] \leqslant \varepsilon C + M\lim_{n\to\infty}\sum_{k=1}^{n}\int_{\{|N_{k,n}|>\varepsilon\}} N_{k,n}^2\,\mathrm{d}\mathbb{P}.$$

Apesar de não sabermos de antemão se $(N_{k,n})_{n,k}$ satisfaz à condição de Lindeberg, podemos fazer a seguinte estimativa

$$\begin{aligned}\sum_{k=1}^{n}\int_{\{|N_{k,n}|>\varepsilon\}} N_{k,n}^2\,\mathrm{d}\mathbb{P} &\leqslant \frac{1}{\varepsilon}\sum_{k=1}^{n}\mathbb{E}|N_{k,n}|^3 = \frac{\mathbb{E}|N|^3}{\varepsilon}\sum_{k=1}^{n}\sigma_{k,n}^3\\ &\leqslant \frac{\mathbb{E}|N|^3}{\varepsilon}\Big(\max_{1\leqslant k\leqslant n}\sigma_{k,n}\Big)\sum_{k=1}^{n}\sigma_{k,n}^2 = \frac{\mathbb{E}|N|^3}{\varepsilon}\Big(\max_{1\leqslant k\leqslant n}\sigma_{k,n}\Big)\to 0,\end{aligned}$$

pois como vale a condição de Lindeberg para $(X_{k,n})_{n,k}$, podemos evocar a Proposição 9.15. Isto mostra que

$$\lim_{n\to\infty}\sum_{k=1}^{n}\mathbb{E}[g(N_{k,n})] = 0,$$

o que conclui a prova do Teorema de Lindeberg. □

9.5 Teorema de Slutsky e Método Delta

Nesta seção veremos o Teorema de Slutsky e o Método Delta, ambos de suma importância em Inferência Estatística. Começamos pelo Teorema do Mapeamento Contínuo, que é uma propriedade fundamental da convergência em distribuição.

Teorema 9.17 (Teorema do Mapeamento Contínuo)**.** *Sejam $(X_n)_n$ e X variáveis aleatórias e $g:\mathbb{R}\to\mathbb{R}$ uma função mensurável. Suponha que g seja contínua em todos os pontos de algum conjunto $C\in\mathcal{B}$ tal que $\mathbb{P}(X\in C)=1$. Se $X_n \xrightarrow{\mathrm{d}} X$, então $g(X_n)\xrightarrow{\mathrm{d}} g(X)$.*

Demonstração. Pelo acoplamento de Skorokhod, existem $Y\sim X$ e $Y_n\sim X_n$

tais que $\mathbb{P}(\{\omega : Y_n(\omega) \to Y(\omega)\}) = 1$. Por hipótese, $\mathbb{P}(\{\omega : Y(\omega) \in C\}) = 1$. Além disso, para todo ω tal que $Y_n(\omega) \to Y(\omega)$ e tal que $Y(\omega) \in C$, vale $g(Y_n(\omega)) \to g(Y(\omega))$. Portanto, $g(Y_n) \overset{\text{q.c.}}{\to} g(Y)$. Em particular, $g(Y_n) \overset{\text{d}}{\to} g(Y)$. Por outro lado, $g(Y_n) \sim g(X_n)$ e $g(Y) \sim g(X)$ e, pela unicidade do limite em distribuição, $g(X_n) \overset{\text{d}}{\to} g(X)$, concluindo a prova. $\square$

No que segue, usaremos o teorema acima em toda a sua generalidade, e também corolários óbvios como $cX_n \overset{\text{d}}{\to} cX$ e $X_n + c \overset{\text{d}}{\to} X + c$, se $X_n \overset{\text{d}}{\to} X$.

Teorema 9.18 (Teorema de Slutskty)**.** *Sejam X, $(X_n)_n$, $(Y_n)_n$ variáveis aleatórias e $c \in \mathbb{R}$. Se $X_n \overset{\text{d}}{\to} X$ e $Y_n \overset{\text{d}}{\to} c$, então:*

(1) $X_n + Y_n \overset{\text{d}}{\to} X + c$;
(2) $Y_n \cdot X_n \overset{\text{d}}{\to} cX$;
(3) $\frac{X_n}{Y_n} \overset{\text{d}}{\to} \frac{X_n}{c}$, *caso $c \neq 0$ e $Y_n \neq 0$ q.c. para todo n.*

Demonstração. Começamos pelo item (1). Supomos inicialmente que $c = 0$. Vamos usar o Teorema de Helly-Bray três vezes. Seja $g : \mathbb{R} \to \mathbb{R}$ contínua, limitada e com derivada limitada. Tome $M = \sup_{x\in\mathbb{R}} |g(x)|$ e $C = \sup_{x\in\mathbb{R}} |g'(x)|$, de modo que $|g(x) - g(y)| \leqslant C\,|x - y|$ para todos $x, y \in \mathbb{R}$. Desenvolvemos

$$\begin{aligned} |\mathbb{E}[g(X_n + Y_n)] - \mathbb{E}[g(X_n)]| &\leqslant \mathbb{E}|g(X_n + Y_n) - g(X_n)| \\ &\leqslant \mathbb{E}\big[\min\{2M, C\,|Y_n|\}\big] \end{aligned}$$

A última esperança converge a zero pelo Teorema de Helly-Bray, pois $\min\{2M, C\,|y|\}$ é uma função contínua e limitada de y, e $Y_n \overset{\text{d}}{\to} 0$. Da mesma forma, $\mathbb{E}[g(X_n)] \to \mathbb{E}[g(X)]$, pois g é contínua, limitada e $X_n \overset{\text{d}}{\to} X$. Portanto, $\mathbb{E}[g(X_n + Y_n)] \to \mathbb{E}[g(X)]$. Como isso vale para toda função contínua com derivada limitada, novamente pelo Teorema de Helly-Bray concluímos que $X_n + Y_n \overset{\text{d}}{\to} X$. No caso $c \neq 0$, escrevemos $X_n + Y_n = (X_n + c) + (Y_n - c)$. Observe que $X_n + c \overset{\text{d}}{\to} X + c$ e $Y_n - c \overset{\text{d}}{\to} 0$. Aplicando o caso anterior com $X_n + c$ no lugar de X_n e $Y_n - c$ no lugar de Y_n, concluímos que $X_n + Y_n \overset{\text{d}}{\to} X + c$.

Passamos ao item (2). Como no item anterior, vamos supor inicialmente que $c = 0$. Sejam $\varepsilon > 0$ e $\delta > 0$. Tome $K > 0$ tal que $\mathbb{P}(|X| > K) < \delta$ e K e $-K$ sejam pontos de continuidade de F_X (é possível tomar tal K pois $\{|X| > n\} \downarrow \emptyset$ quando $n \to \infty$ e o conjunto de pontos onde F_X é descontínua é enumerável). Vamos mostrar que $X_n Y_n \xrightarrow{\mathbb{P}} 0$. Expandindo:

$$\begin{aligned}\limsup_n \mathbb{P}(|X_n Y_n| \geqslant \varepsilon) &\leqslant \lim_n \mathbb{P}(|X_n| > K) + \lim_n \mathbb{P}(|Y_n| \geqslant \tfrac{\varepsilon}{K}) \\ &= \mathbb{P}(|X| > K) + 0 < \delta,\end{aligned}$$

onde na igualdade acima utilizamos que $Y_n \xrightarrow{\mathrm{d}} 0$. Como isso vale para todo $\delta > 0$, temos $\lim_n \mathbb{P}(|X_n Y_n| \geqslant \varepsilon) = 0$, donde $X_n Y_n \xrightarrow{\mathbb{P}} 0$ e, em particular, $X_n Y_n \xrightarrow{\mathrm{d}} 0$. No caso $c \neq 0$, escrevemos $X_n Y_n = X_n(Y_n - c) + cX_n$. Observe que $cX_n \xrightarrow{\mathrm{d}} cX$ e $Y_n - c \xrightarrow{\mathrm{d}} 0$. Pelo caso anterior com $Y_n - c$ no lugar de Y_n, obtemos $X_n(Y_n - c) \xrightarrow{\mathrm{d}} 0$. Pelo item (1), $Y_n \cdot X_n \xrightarrow{\mathrm{d}} cX$.

Provemos finalmente (3). Considere $g : \mathbb{R} \to \mathbb{R}$ dada por $g(0) = 0$ e $g(y) = y^{-1}$ se $y \neq 0$. Pelo Teorema do Mapeamento Contínuo, $\frac{1}{Y_n} \xrightarrow{\mathrm{d}} \frac{1}{c}$, pois g é descontinua apenas em 0 e $c \neq 0$. Aplicando o item (2) com $\frac{1}{Y_n}$ no lugar de Y_n, concluímos que $\frac{X_n}{Y_n} \xrightarrow{\mathrm{d}} \frac{X}{c}$. □

Corolário 9.19. *Sejam X e $(X_n)_n$ variáveis aleatórias e sejam c e $(c_n)_n$ constantes reais. Se $X_n \xrightarrow{\mathrm{d}} X$ e $c_n \to c$, então $c_n X_n \xrightarrow{\mathrm{d}} cX$ e $X_n + c_n \xrightarrow{\mathrm{d}} X + c$.*

Observação 9.20. A conclusão do teorema acima pode ser falsa se $Y_n \xrightarrow{\mathrm{d}} Y$ com Y aleatória, pois as hipóteses do teorema nada dizem sobre a distribuição conjunta de X_n e Y_n. Com efeito, tome $Z \sim \mathcal{N}(0,1)$, $X_n = Y_n = Z$, $X = Z$ e $Y = -Z$. Observe que $X_n + Y_n \sim \mathcal{N}(0,4)$ e $X + Y = 0$ q.c., de forma que $X_n \xrightarrow{\mathrm{d}} X$, $Y_n \xrightarrow{\mathrm{d}} Y$, mas $X_n + Y_n$ não converge para $X + Y$. △

Exemplo 9.21. Considere uma sequência $(X_n)_n$ de variáveis aleatórias i.i.d. não-degeneradas com segundo momento finito, com média μ e variância σ. Se μ é desconhecido, um estimador para μ após observar os valores $X_1, \ldots, X_n$ é

$$\hat{\mu}_n = \frac{X_1 + \cdots + X_n}{n}.$$

Pela Lei dos Grandes Números de Kolmogorov, $\hat{\mu}_n \overset{\text{q.c.}}{\to} \mu$. Queremos saber qual é a distribuição do erro $\hat{\mu}_n - \mu$. Pelo Teorema do Limite Central,

$$\sqrt{n}\,\frac{\hat{\mu}_n - \mu}{\sigma} \overset{\text{d}}{\to} \mathcal{N}(0,1),$$

ou seja, $\hat{\mu}_n - \mu$ tem distribuição aproximadamente normal com média zero e variância σ^2/n. Porém, gostaríamos de saber a distribuição aproximada de $\hat{\mu}_n - \mu$ sem ter que fazer referência ao parâmetro σ^2, pois este pode ser também desconhecido. Um estimador para σ^2 é dado por[16]

$$\hat{s}_n^2 = \frac{\sum_{k=1}^n (X_k - \frac{1}{n}\sum_{j=1}^n X_j)^2}{n-1}.$$

Reescrevendo e expandindo a expressão acima, chegamos a

$$\hat{s}_n^2 = \frac{n}{n-1}\left[\Big(\frac{1}{n}\sum_{k=1}^n X_k^2\Big) - \Big(\frac{1}{n}\sum_{k=1}^n X_k\Big)^2\right] \overset{\text{q.c.}}{\to} \mathbb{E}X_1^2 - (\mathbb{E}X_1)^2 = \sigma^2$$

pela Lei dos Grandes Números de Kolmogorov. Em particular,

$$\frac{\hat{s}_n}{\sigma} \overset{\text{d}}{\to} 1$$

e, pelo item (3) do Teorema de Slutsky,

$$\sqrt{n}\,\frac{\hat{\mu}_n - \mu}{\hat{s}_n} = \sqrt{n}\,\frac{\hat{\mu}_n - \mu}{\sigma \cdot (\hat{s}_n/\sigma)} \overset{\text{d}}{\to} \mathcal{N}(0,1),$$

ou seja, $\hat{\mu}_n - \mu$ tem distribuição aproximadamente normal com média zero e variância $\hat{s}_n^2/n$ (para que o quociente acima esteja definido, substituímos $\hat{s}_n$ por 1 caso seu valor seja 0; a probabilidade desse evento tende a zero). $\triangle$

No exemplo acima, qual a distribuição aproximada de $\hat{\mu}_n^{-1} - \mu^{-1}$, ou de $\hat{\mu}_n^2 - \mu^2$? A resposta é dada por outra aplicação do Teorema de Slutsky.

Teorema 9.22 (Método Delta)**.** *Seja $(Y_n)_n$ uma sequência de variáveis*

[16] A razão para tomar-se $n-1$ no denominador é para que $\mathbb{E}[\hat{s}_n^2] = \sigma^2$.

aleatórias tais que

$$r_n \cdot (Y_n - \mu) \xrightarrow{\mathrm{d}} Z$$

para alguma sequência de números reais $r_n \to +\infty$*, algum* $\mu \in \mathbb{R}$ *e alguma variável aleatória* Z*. Se* $g : \mathbb{R} \to \mathbb{R}$ *é uma função mensurável e* g *é diferenciável em* μ*, então*

$$r_n \cdot (g(Y_n) - g(\mu)) \xrightarrow{\mathrm{d}} g'(\mu)\, Z.$$

Demonstração. Defina $h : \mathbb{R} \to \mathbb{R}$ como $h(\mu) = g'(\mu)$ e $h(x) = \frac{g(x)-g(\mu)}{x-\mu}$ para $x \neq \mu$, e observe que h é contínua em μ. Reescrevemos

$$r_n \cdot [g(Y_n) - g(\mu)] = r_n \cdot h(Y_n) \cdot (Y_n - \mu).$$

Observe que, pelo Corolário 9.19, $Y_n = \mu + r_n^{-1} \cdot r_n \cdot (Y_n - \mu) \xrightarrow{\mathrm{d}} \mu$. Pelo Teorema do Mapeamento Contínuo, $h(Y_n) \xrightarrow{\mathrm{d}} h(\mu) = g'(\mu)$, pois h é contínua em μ. Por outro lado, $r_n \cdot (Y_n - \mu) \xrightarrow{\mathrm{d}} Z$ e, pelo item (2) do Teorema de Slutsky, $h(Y_n) \cdot r_n \cdot (Y_n - \mu) \xrightarrow{\mathrm{d}} g'(\mu)\, Z$, concluindo a prova. □

Observação 9.23. No contexto do Teorema do Limite Central, se

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

com $S_n = X_1 + \cdots + X_n$, podemos tomar $Y_n = \bar{X}_n = \frac{S_n}{n}$ e $r_n = \sqrt{n}$, de forma a reescrever a convergência como

$$\sqrt{n}\,(\bar{X}_n - \mu) \xrightarrow{\mathrm{d}} \mathcal{N}(0, \sigma^2).$$

Neste caso, o Método Delta nos dá

$$\sqrt{n}\,(g(\bar{X}_n) - g(\mu)) \xrightarrow{\mathrm{d}} \mathcal{N}(\,0, \sigma^2\, g'(\mu)^2\,)$$

supondo que $g'(\mu) \neq 0$ (ou, caso $g'(\mu) = 0$, convencionando que $\mathcal{N}(0,0)$ denota uma variável aleatória degenerada q.c. igual a zero). △

Exemplo 9.24. Sejam $(X_n)_n$ variáveis aleatórias i.i.d. com distribuição $\text{Exp}(\lambda)$ e considere o estimador

$$\hat{\lambda}_n = \frac{n}{X_1 + \cdots + X_n}$$

para o parâmetro λ. Pela Lei dos Grandes Números de Cantelli, $\hat{\lambda}_n \overset{\text{q.c.}}{\to} \lambda$ e gostaríamos de saber a distribuição aproximada do erro $\hat{\lambda}_n - \lambda$. Observe que, pelo Teorema do Limite Central,

$$\sqrt{n}\,(\hat{\lambda}_n^{-1} - \lambda^{-1}) \overset{\text{d}}{\to} \mathcal{N}(0, \lambda^{-2}).$$

Aplicando o Método Delta com $g(x) = x^{-1}$, obtemos

$$\sqrt{n}\,(\hat{\lambda}_n - \lambda) \to \mathcal{N}(0, \lambda^2). \qquad \triangle$$

Exemplo 9.25. Sejam $(X_n)_n$ variáveis aleatórias i.i.d. com média $\mu > 0$ e variância σ^2, e defina

$$\hat{\mu}_n = \frac{X_1 + \cdots + X_n}{n}.$$

Suponha que estamos interessados na grandeza $\mu^2 = (\mathbb{E}X_1)^2$, e usamos $\hat{\mu}_n^2$ para estimá-la. Para o erro desse estimador, o Método Delta nos dá

$$\sqrt{n}\,(\hat{\mu}_n^2 - \mu^2) \overset{\text{d}}{\to} \mathcal{N}(0, 4\sigma^2\mu^2). \qquad \triangle$$

9.6 Exercícios

§ 9.1

1. Um par de dados honestos é lançado 180 vezes por hora.

(a) Qual a probabilidade aproximada de que 25 ou mais lançamentos tenham tido soma 7 na primeira hora?

(b) Qual a probabilidade aproximada de que entre 700 e 750 lançamentos tenham tido soma 7 durante 24 horas?

2. Imagine um modelo idealizado com M eleitores, dos quais M_A pretendem votar no candidato A. Suponha que seja possível sortear um desses eleitores ao acaso, e de forma equiprovável. Definimos

$$X = \begin{cases} 1, & \text{caso o eleitor sorteado vá votar no candidato } A, \\ 0, & \text{caso contrário.} \end{cases}$$

Deseja-se estimar a proporção $p = \frac{M_A}{M}$ de eleitores do candidato A, que é desconhecida. Para isso, repete-se este processo N vezes, obtendo-se $X_1, \ldots, X_N$. Para estimar o valor de p considera-se

$$\widehat{p}_N = \frac{X_1 + \cdots + X_N}{N}.$$

Supomos *a priori* que p é bem próximo de $\frac{1}{2}$, de forma que $\mathbb{V}X \approx \frac{1}{4}$. Se entrevistamos $N = 2500$ eleitores, calcule aproximadamente a probabilidade de essa pesquisa cometer um erro $|\widehat{p}_N - p|$ maior que 0,01.

3. A quantidade de uvas-passas encontradas em cada panetone de uma determinada marca é independente dos demais panetones e segue a distribuição de Poisson com parâmetro $\lambda = 25$ (ou seja, tem esperança igual à variância, igual a λ). Um grupo de estudantes de férias resolve estimar o valor de λ, uma vez que o mesmo é desconhecido para eles. Para isso, vão contar as uvas-passas de uma amostra de $N = 625$ panetones e registrar o resultado de cada contagem $X_1, \ldots, X_N$. A estimativa $\widehat{\lambda}_N$ para o valor de λ que os estudantes vão adotar será dada por

$$\widehat{\lambda}_N = \frac{X_1 + \cdots + X_N}{N}.$$

(a) Qual é o valor aproximado da probabilidade de que o valor $\widehat{\lambda}_N$ esteja entre 24,8 e 25,4?

(b) Para que o erro $\left|\widehat{\lambda}_N - \lambda\right|$ fosse menor que 0,075 com probabilidade pelo menos igual a 0,8664, qual deveria ser o número N de panetones

examinados?

4. Use o Teorema do Limite Central para verificar que

$$\lim_{n\to\infty} e^{-n} \sum_{k=0}^{n} \frac{2n^k}{k!} = 1.$$

5. Se lançamos 10.000 vezes uma moeda honesta, calcule aproximadamente a probabilidade de que o número de vezes que se obtém coroa seja no mínimo 4.893 e no máximo 4.967.

§ 9.2

6. Sejam $(X_n)_n$ i.i.d. com $\mathbb{P}(X_1 = +1) = \mathbb{P}(X_1 = -1) = \frac{1}{2}$ e considere o passeio aleatório dado por $S_0 = 0$ e $S_n = X_1 + \cdots + X_n$. Seja Z_n o número de vezes que o passeio aleatório retorna ao ponto inicial $x = 0$ até o tempo n. Mostre que $\lim_n \mathbb{E}Z_n = +\infty$.

7. Seja τ o primeiro tempo de retorno à origem do passeio aleatório simples simétrico definido no exercício acima. Mostre que, quando $n \to \infty$,

$$\mathbb{P}(\tau > 2n) \sim \frac{1}{\sqrt{\pi n}} \quad \text{e} \quad \mathbb{P}(\tau = 2n) \sim \frac{1}{2\sqrt{\pi} n^{\frac{3}{2}}}.$$

Dica: No Capítulo 1, provamos que $\mathbb{P}(\tau > 2n) = \mathbb{P}(S_{2n} = 0)$.

§ 9.3

8. Seja $(X_n)_n$ uma sequência de variáveis aleatórias independentes, mostre que vale (9.2) nos seguintes casos:

(a) $X_n \sim \mathcal{U}[0, n]$;
(b) $\mathbb{P}(X_n = n) = \mathbb{P}(X_n = -n) = \frac{1}{2}$.

9. Seja $(X_n)_n$ uma sequência de variáveis aleatórias independentes tais que $X_n \sim \mathcal{U}[0, 1]$ para n par e $X_n \sim \mathcal{U}[0, 2]$ para n ímpar. Mostre que vale (9.2).

10. Seja $(Y_n)_n$ uma sequência de variáveis aleatórias independentes, com $Y_n \sim \mathcal{U}[-n, n]$ e defina $X_n = \operatorname{sen} Y_n$. Vale (9.2)?

§ 9.4

11. Sejam $X_1, \ldots, X_n, \ldots$ variáveis aleatórias independentes tais que

$$\mathbb{P}(X_n = n^\alpha) = \mathbb{P}(X_n = -n^\alpha) = \frac{1}{6n^{2(\alpha-1)}} \quad \text{e } \mathbb{P}(X_n = 0) = 1 - \frac{1}{3n^{2(\alpha-1)}}.$$

Mostre que a sequência $(X_n)_n$ satisfaz à condição de Lindeberg se, e somente se, $\alpha < \frac{3}{2}$.

12. Dê um exemplo de uma sequência $(X_n)_n$ de variáveis aleatórias independentes, para a qual vale (9.2) mas não vale a condição de Lindeberg.

Dica: Tente com variáveis normais.

13. Dê um exemplo de uma sequência $(X_n)_n$ de variáveis aleatórias independentes com média zero e variância um, para a qual não vale (9.2).

Dica: Tente com $\mathbb{P}(X_n^2 = n^2) = n^{-2}$.

Capítulo 10

Transformadas

A função geradora de momentos e a função característica estão entre os exemplos mais importantes de *transformadas*. A ideia geral de transformada é mapear certos objetos em objetos de outro tipo e outras propriedades, onde determinadas análises são possivelmente mais fáceis. Isso ficará claro nos exemplos e aplicações. A função geradora de momentos é a instância da *transformada de Laplace* de uma distribuição em $\mathbb{R}$, e a função característica é a *transformada de Fourier*.

10.1 Função geradora de momentos

Definição 10.1. Seja X uma variável aleatória. Define-se a *função geradora de momentos de* X como a função $M_X : \mathbb{R} \to [0, +\infty]$ dada por

$$M_X(t) = \mathbb{E}[e^{tX}].$$

Pelo Teorema 5.36, podemos calcular a função geradora de momentos por

$$M_X(t) = \sum_x e^{tx} \cdot \mathbb{P}(X = x)$$

se X é discreta. Pelo Teorema 5.39,

$$M_X(t) = \int_{-\infty}^{\infty} e^{tx} f_X(x)\,\mathrm{d}x$$

se X é absolutamente contínua com densidade f_X.

Exemplo 10.2 (Bernoulli). Se $X \sim \text{Bernoulli}(p)$, então

$$M_X(t) = pe^t + 1 - p = 1 + p(e^t - 1). \qquad \triangle$$

Exemplo 10.3 (Geométrica). Se $X \sim \text{Geom}(p)$, então

$$M_X(t) = \sum_{n=1}^{\infty} e^{tn} p(1-p)^{n-1} = \begin{cases} \frac{p}{e^{-t}+p-1}, & t < \log \frac{1}{1-p}, \\ +\infty, & t \geqslant \log \frac{1}{1-p}. \end{cases} \qquad \triangle$$

Exemplo 10.4 (Poisson). Se $X \sim \text{Poisson}(\lambda)$, então

$$M_X(t) = \sum_{n=0}^{\infty} e^{tn} \frac{e^{-\lambda}\lambda^n}{n!} = e^{-\lambda} \sum_{n=0}^{\infty} \frac{(\lambda e^t)^n}{n!} = e^{-\lambda} e^{\lambda e^t} = e^{\lambda(e^t - 1)}. \qquad \triangle$$

Proposição 10.5. *Se X e Y são independentes, então*

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

para todo $t \in \mathbb{R}$.

Demonstração. Como X e Y são independentes, e^{tX} e e^{tY} também o são, pela Observação 4.13. Logo, podemos escrever

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

Exemplo 10.6 (Binomial). Se $X \sim \text{Binom}(n, p)$, então X é distribuída como a soma de n variáveis $X_1, \ldots, X_n$ independentes com distribuição Bernoulli(p). Portanto,

$$M_X(t) = M_{X_1}(t) \cdots M_{X_n}(t) = [1 + p(e^t - 1)]^n. \qquad \triangle$$

A proposição abaixo explica o porquê do nome função geradora de momentos.

Proposição 10.7. *Se M_X é finita em algum intervalo $(-\varepsilon, +\varepsilon)$, então os momentos de X podem ser obtidos através das derivadas de M_X por*

$$\mathbb{E}X^k = M_X^{(k)}(0), \quad k = 1, 2, 3, \ldots.$$

Demonstração. Para lembrar da fórmula, fazemos informalmente

$$\tfrac{\mathrm{d}^k}{\mathrm{d}t^k} M_X(t) = \tfrac{\mathrm{d}^k}{\mathrm{d}t^k}\, \mathbb{E}[e^{tX}] = \mathbb{E}\Big[\tfrac{\mathrm{d}^k}{\mathrm{d}t^k}\, e^{tX}\Big] = \mathbb{E}[X^k e^{tX}].$$

No caso de X ser uma variável aleatória simples, essa é a demonstração, pois a esperança é uma soma finita e podemos derivar dentro da soma. No caso geral, há que se justificar passagem da derivada para dentro da esperança.

Para continuar a demonstração, o leitor tem que ter passado pela Seção 5.3, pois vamos usar o Teorema 5.49.

Vamos mostrar, por indução em k, que a fórmula acima vale para todo $t \in (-\varepsilon, +\varepsilon)$. Para $k = 0$, a identidade vale trivialmente. Suponhamos a identidade acima válida para algum $k \in \mathbb{N}_0$. Queremos mostrar que

$$\tfrac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[X^k e^{tX}] = \mathbb{E}[X^{k+1} e^{tX}] \tag{10.8}$$

para todo $t \in (-\varepsilon, \varepsilon)$. Sejam b e δ tais que $0 < b < \delta < \varepsilon$. Definimos $f(t, x) = x^k e^{tx}$, de forma que $\frac{\partial}{\partial t} f(t, x) = x^{k+1} e^{tx}$. Observe que, para $t \in [-b, b]$,

$$\Big|\tfrac{\partial}{\partial t} f(t, X)\Big| = |X^{k+1}|\, e^{t\,X} \leqslant |X^{k+1}|\, e^{b\,|X|} \leqslant C e^{\delta\,|X|} \leqslant C(e^{\delta X} + e^{-\delta X}),$$

onde C depende de k, b e δ. Como o último termo das desigualdades acima é integrável, pelo Teorema 5.49 temos (10.8) para todo $t \in [-b, b]$. Como $b < \varepsilon$ é arbitrário, o mesmo vale para todo $t \in (-\varepsilon, \varepsilon)$, o que conclui a prova. □

Nos exemplos abaixo, ilustramos como calcular $\mathbb{E}X$ e $\mathbb{V}X$ para algumas distribuições cujas funções geradoras de momento já nos são conhecidas.

Exemplo 10.9 (Geométrica). Se $X \sim \text{Geom}(p)$, então

$$\mathbb{E}X = M_X'(0) = \tfrac{1}{p},\ \mathbb{E}X^2 = M_X''(0) = \tfrac{2}{p^2} - \tfrac{1}{p},\ \mathbb{V}X = \mathbb{E}X^2 - (\mathbb{E}X)^2 = \tfrac{1-p}{p^2}. \quad \triangle$$

Exemplo 10.10 (Poisson). Se $X \sim \text{Poisson}(\lambda)$, então

$$\mathbb{E}X = M_X'(0) = \lambda,\ \mathbb{E}X^2 = M_X''(0) = \lambda^2 + \lambda,\ \mathbb{V}X = \mathbb{E}X^2 - (\mathbb{E}X)^2 = \lambda. \quad \triangle$$

O teorema abaixo nos diz que a função geradora de momentos determina a distribuição de modo único. A demonstração envolve aspectos técnicos acima dos nossos objetivos, e de fato é difícil encontrá-la em livros-texto (veja a Seção 30 de Billingsley (1995)).

Teorema 10.11 (Unicidade)**.** *Dadas X e Y variáveis aleatórias, se existe $\varepsilon > 0$ tal que $M_X(t) = M_Y(t) < \infty$ para todo $t \in (-\varepsilon, +\varepsilon)$, então $X \sim Y$.*

Exemplo 10.12 (Soma de variáveis Poisson independentes). Se $X \sim \text{Poisson}(\lambda)$ e $Y \sim \text{Poisson}(\mu)$ são independentes, então

$$M_{X+Y}(t) = M_X(t) \cdot M_Y(t) = e^{\lambda(e^t-1)} e^{\mu(e^t-1)} = e^{(\lambda+\mu)(e^t-1)} = M_Z(t),$$

onde $Z \sim \text{Poisson}(\lambda + \mu)$. Portanto, $X + Y \sim \text{Poisson}(\lambda + \mu)$. $\triangle$

Concluímos esta seção com uma generalização da Proposição 10.7 para o caso $k = 1$, que considera apenas a derivada lateral.

Proposição 10.13. *Se M_X é finita em algum intervalo $(-\varepsilon, 0]$, então a esperança de X pode ser obtida como derivada lateral*

$$\mathbb{E}X = M_X'(0^-).$$

Demonstração. A prova exige familiaridade com as Seções 5.3 e 5.4. Sem perda de generalidade, podemos supor que $\varepsilon > 1$. Seja $(t_n)_n$ uma sequência de números reais estritamente crescente, tal que $t_1 = -1$ e $t_n \to 0$. Temos que mostrar que

$$\frac{\mathbb{E}[e^{t_n X}] - 1}{t_n} \to \mathbb{E}X.$$

Por clareza, vamos considerar primeiro dois casos particulares. Suponhamos inicialmente que $X \geqslant 0$ q.c. Observe que, se $X \geqslant 0$, então a sequência $\left(\frac{e^{t_n X}-1}{t_n}\right)_n$ é não-decrescente, como consequência do fato de que função que leva t em e^{tX} é convexa. Pelo Teorema da Convergência Monótona, obtemos o limite desejado, provando a proposição no caso não-negativo, incluindo o caso em que $\mathbb{E}X = +\infty$.

Suponhamos agora que $X \leqslant 0$ q.c. Observamos que, se $X \leqslant 0$, então a sequência acima é não-crescente, também por convexidade. Assim, os termos da sequência são dominados pelo primeiro termo, $1 - e^{-X}$, que é integrável por hipótese. Pelo Teorema da Convergência Dominada, obtemos o limite desejado também neste caso.

Finalmente, consideramos o caso geral. Aplicando os casos anteriores às esperanças condicionais, obtemos

$$\frac{\mathbb{E}[e^{t_n X}|X < 0] - 1}{t_n} \to \mathbb{E}[X|X < 0]$$

e

$$\frac{\mathbb{E}[e^{t_n X}|X \geqslant 0] - 1}{t_n} \to \mathbb{E}[X|X \geqslant 0].$$

Usando novamente a hipótese, temos que

$$\mathbb{P}(X < 0)\,\mathbb{E}[X|X < 0] = \mathbb{E}X^- \leqslant \mathbb{E}[e^{-X}] < \infty,$$

logo podemos combinar os limites acima e obter

$$\begin{aligned}\frac{\mathbb{E}[e^{t_n X}] - 1}{t_n} &= \frac{\mathbb{P}(X < 0)\mathbb{E}[e^{t_n X}|X < 0] + \mathbb{P}(X \geqslant 0)\mathbb{E}[e^{t_n X}|X \geqslant 0] - 1}{t_n} \\ &\to \mathbb{P}(X < 0)\mathbb{E}[X|X < 0] + \mathbb{P}(X \geqslant 0)\mathbb{E}[X|X \geqslant 0] = \mathbb{E}X,\end{aligned}$$

onde também usamos a Proposição 5.54, concluindo esta prova. □

10.2 Função característica

Do ponto de vista teórico, a função característica é bem mais robusta e funcional que a função geradora de momentos: está definida para qualquer distribuição; sempre determina a distribuição; determina também a convergência em distribuição; não bastasse, ainda gera momentos. Entretanto, a função característica envolve a manipulação de números complexos. Ressaltamos, porém, que o estudo de funções características não requer conhecimentos de cálculo em uma *variável* complexa. Isso porque as integrais são calculadas em $\mathrm{d}x$ para $x \in \mathbb{R}$ e não em $\mathrm{d}z$ para caminhos $\gamma \subseteq \mathbb{C}$.

Definição 10.14 (Variável aleatória complexa)**.** Uma *variável aleatória complexa* é uma função $Z : \Omega \to \mathbb{C}$ tal que $Z = X + i\,Y$, onde X e Y são variáveis aleatórias reais. Se X e Y são integráveis, dizemos que Z é integrável e definimos

$$\mathbb{E}Z = \mathbb{E}X + i\,\mathbb{E}Y.$$

A integração de funções complexas em domínios reais pode ser feita, para todos os fins práticos, como no caso real. Ou seja, se $F : \mathbb{R} \to \mathbb{C}$ satisfaz $\frac{\mathrm{d}}{\mathrm{d}x}F(x) = f(x)$ para $x \in [a, b]$, então

$$\int_a^b f(x)\,\mathrm{d}x = F(b) - F(a). \tag{10.15}$$

Vamos utilizar a *fórmula de Euler*, que define

$$e^{iy} = \cos y + i \operatorname{sen} y$$

para $y \in \mathbb{R}$. Usaremos sem demonstração os seguintes fatos:

$$e^z = \sum_n \frac{z^n}{n!}, \qquad e^{z+w} = e^z e^w,$$

$$(e^g)' = e^g g', \qquad \left(1 + \frac{z_n}{n}\right)^n \to e^z \text{ se } z_n \to z,$$

onde $z, w, z_n \in \mathbb{C}$.

Proposição 10.16. *Se Z e W são variáveis aleatórias complexas integráveis, então $Z + W$ é integrável com $\mathbb{E}[Z + W] = \mathbb{E}Z + \mathbb{E}W$, e para $z \in \mathbb{C}$ tem-se zZ integrável com $\mathbb{E}[zZ] = z\,\mathbb{E}Z$. Se, além disso, Z e W são independentes, então ZW é integrável com $\mathbb{E}[ZW] = \mathbb{E}Z \cdot \mathbb{E}W$.*

Demonstração. Sejam $z = x + iy \in \mathbb{C}$, com $x, y \in \mathbb{R}$, e $Z = X + i\,Y$, $W = A + iB$ com X, Y, A e B variáveis aleatórias reais integráveis. Observe que $Z + W = (X + A) + i(Y + B)$ e $zZ = (xX - yY) + i(xY + yX)$. Logo, a integrabilidade de $Z + W$ e zZ segue da integrabilidade de suas respectivas partes real e imaginária. Portanto,

$$\mathbb{E}[Z+W] = \mathbb{E}[X+A]+i\,\mathbb{E}[Y+B] = (\mathbb{E}X+i\,\mathbb{E}Y)+(\mathbb{E}A+i\,\mathbb{E}B) = \mathbb{E}Z+\mathbb{E}W$$

e

$$\mathbb{E}[zZ] = \mathbb{E}[xX - yY] + i\,\mathbb{E}[xY + yX] = (x + iy)\mathbb{E}X + i(x + iy)\mathbb{E}Y = z\,\mathbb{E}Z.$$

Se Z e W são independentes, então (X, Y) e (A, B) são independentes. Logo

$$\begin{aligned}\mathbb{E}[ZW] &= \mathbb{E}[XA - YB] + i\,\mathbb{E}[XB + YA]\\ &= (\mathbb{E}X \cdot \mathbb{E}A - \mathbb{E}Y \cdot \mathbb{E}B) + i(\mathbb{E}X \cdot \mathbb{E}B + \mathbb{E}Y \cdot \mathbb{E}A)\\ &= \ (\mathbb{E}X + i\,\mathbb{E}Y)(\mathbb{E}A + i\,\mathbb{E}B) = \mathbb{E}Z \cdot \mathbb{E}W,\end{aligned}$$

onde a hipótese de independência foi usada na segunda igualdade. □

Proposição 10.17. *Se $\mathbb{E}|Z| < \infty$, então Z é integrável e $|\mathbb{E}Z| \leqslant \mathbb{E}|Z|$.*

Demonstração. Escreva $Z = X + i\,Y$. Como $|X| \leqslant |Z|$ e $|Y| \leqslant |Z|$, segue que X e Y são integráveis, logo Z é integrável. Se $\mathbb{E}Z \in \mathbb{R}$, então $|\mathbb{E}Z| = |\mathbb{E}X| \leqslant \mathbb{E}|X| \leqslant \mathbb{E}|Z|$. No caso geral, escrevemos $|\mathbb{E}Z| = w\,\mathbb{E}Z$, com $|w| = 1$, de forma que $\mathbb{E}[wZ] = w\,\mathbb{E}Z = |\mathbb{E}Z| \in \mathbb{R}$ e, pelo caso anterior, $|\mathbb{E}Z| = |w\,\mathbb{E}Z| = |\mathbb{E}[wZ]| \leqslant \mathbb{E}|wZ| = \mathbb{E}|Z|$. □

A seguir definiremos a função característica de uma variável aleatória.

Definição 10.18 (Função característica)**.** A *função característica* de uma variável aleatória X, denotada por φ_X, é a função $\varphi_X : \mathbb{R} \to \mathbb{C}$ definida como

$$\varphi_X(t) = \mathbb{E}[e^{itX}] = \mathbb{E}[\cos(tX)] + i\mathbb{E}[\operatorname{sen}(tX)], \quad t \in \mathbb{R}.$$

Observe que a função característica de uma variável aleatória depende de fato apenas de sua distribuição. Quando a variável for absolutamente contínua com função densidade f, sua função característica é simplesmente a transformada de Fourier da função f.

Observação 10.19. Como $|e^{itX}| = 1$, $\varphi_X(t)$ está definida para todo $t \in \mathbb{R}$. $\triangle$

Exemplo 10.20 (Uniforme). Se $X \sim \mathcal{U}[a, b]$, então

$$\begin{aligned}
\varphi_X(t) &= \mathbb{E}[e^{itX}] = \mathbb{E}[\cos(tX)] + i\,\mathbb{E}[\operatorname{sen}(tX)] \\
&= \int_a^b \cos(tx)\frac{1}{b-a}\mathrm{d}x + i\int_a^b \operatorname{sen}(tx)\frac{1}{b-a}\mathrm{d}x \\
&= \frac{\operatorname{sen}(tb) - \operatorname{sen}(ta)}{t(b-a)} - i\frac{\cos(tb) - \cos(ta)}{t(b-a)} \\
&= \frac{-ie^{itb} + ie^{ita}}{t(b-a)} = \frac{e^{itb} - e^{ita}}{it(b-a)}.
\end{aligned}$$

Ou, usando (10.15):

$$\varphi_X(t) = \int_a^b \frac{e^{itx}}{b-a}\mathrm{d}x = \frac{e^{itb} - e^{ita}}{it(b-a)}. \qquad \triangle$$

Exemplo 10.21 (Poisson). Se $X \sim \text{Poisson}(\lambda)$, então:

$$\varphi_X(t) = \mathbb{E}[e^{itX}] = \sum_{n=0}^{\infty} e^{itn}\frac{e^{-\lambda}\lambda^n}{n!} = e^{-\lambda}\sum_{n=0}^{\infty}\frac{(e^{it}\lambda)^n}{n!} = e^{-\lambda}e^{e^{it}\lambda} = e^{\lambda(e^{it}-1)}. \triangle$$

Exemplo 10.22 (Geométrica). Se $X \sim \mathrm{Geom}(p)$, então

$$\varphi_X(t) = \sum_{n=1}^{\infty} e^{itn} \cdot p(1-p)^{n-1} = e^{it} p \sum_{m=0}^{\infty} [(e^{it})(1-p)]^m = \frac{p}{e^{-it} + p - 1}. \triangle$$

Proposição 10.23. *Para todo $t \in \mathbb{R}$, vale $|\varphi_X(t)| \leqslant 1$. Além disso, $\varphi(0) = 1$. Ademais, $\varphi_{aX+b}(t) = e^{itb}\varphi_X(at)$ para $a, b \in \mathbb{R}$.*

Demonstração. Basta observar que

$$|\varphi_X(t)| = |\mathbb{E}[e^{itX}]| \leqslant \mathbb{E}|e^{itX}| = \mathbb{E}[\mathrm{sen}^2(tX) + \cos^2(tX)] = 1,$$

$$\varphi_X(0) = \mathbb{E}[e^0] = 1,$$

e

$$\varphi_{aX+b}(t) = \mathbb{E}[e^{it(aX+b)}] = e^{itb}\,\mathbb{E}[e^{itaX}] = e^{itb}\varphi_X(at). \qquad \square$$

Proposição 10.24 (Independência)**.** *Se X e Y são independentes, então*

$$\varphi_{X+Y}(t) = \varphi_X(t) \cdot \varphi_Y(t)$$

para todo $t \in \mathbb{R}$.

Demonstração. Observe que, se X e Y são independentes, então e^{itX} e e^{itY} também o são. Logo podemos escrever

$$\varphi_{X+Y}(t) = \mathbb{E}[e^{it(X+Y)}] = \mathbb{E}\Big[e^{itX}e^{itY}\Big] = \varphi_X(t) \cdot \varphi_Y(t),$$

onde a independência foi utilizada na última igualdade. $\square$

Proposição 10.25 (Cálculo de Momentos)**.** *Se $\mathbb{E}|X^k| < \infty$, então φ_X tem k-ésima derivada, e ademais*

$$\varphi_X^{(k)}(0) = i^k\,\mathbb{E}X^k.$$

Demonstração. Assim como fizemos para a função geradora de momentos, gostaríamos de derivar dentro da esperança de modo que possamos escrever

$$\varphi_X^{(k)}(t) = \mathbb{E}[i^k X^k e^{itX}], \tag{10.26}$$

obtendo assim $\varphi_X^{(k)}(0) = i^k\,\mathbb{E}X^k$. Vamos mostrar, por indução em k, que, se $\mathbb{E}|X^k| < \infty$, então vale (10.26) para todo $t \in \mathbb{R}$.

O caso $k = 0$ é trivial. Suponha que a afirmação vale para algum k, e seja X tal que $\mathbb{E}|X^{k+1}| < \infty$. Observe que $\mathbb{E}|X^k| < \infty$, pois $|x^k| \leqslant 1 + |x^{k+1}|$, logo vale (10.26). Defina $f(t,x) = i^k X^k e^{itx}$. Observe que

$$\left|\tfrac{\partial}{\partial t} f(t,X)\right| = |X^{k+1}|\, e^{it\,|X|} \leqslant |X^{k+1}|,$$

e este último é integrável por hipótese. Pelo Teorema 5.49,

$$\varphi_X^{(k+1)}(t) = \tfrac{\mathrm{d}}{\mathrm{d}t}\varphi_X^{(k)}(t) = \tfrac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[i^k X^k\, e^{itX}] = \mathbb{E}[i^{k+1} X^{k+1}\, e^{itX}]$$

para todo $t \in \mathbb{R}$, o que conclui a prova por indução. □

Exemplo 10.27 (Normal). Se $X \sim \mathcal{N}(0,1)$, então $\varphi_X(t) = e^{-\frac{t^2}{2}}$. Usando (10.26), podemos derivar dentro da integral, e integrando por partes obtemos

$$\varphi_X'(t) = \int_{-\infty}^{+\infty} \frac{ixe^{itx-\frac{x^2}{2}}}{\sqrt{2\pi}}\mathrm{d}x = \int_{-\infty}^{+\infty} \frac{i^2 te^{itx-\frac{x^2}{2}}}{\sqrt{2\pi}}\mathrm{d}x = -t\,\varphi_X(t).$$

As soluções da equação $\varphi_X'(t) = -t\,\varphi_X(t)$ $\forall t \in \mathbb{R}$ são dadas por $\varphi_X(t) = c\,e^{-\frac{t^2}{2}}$, com $c \in \mathbb{C}$. Avaliando no ponto $t = 0$, obtemos $\varphi_X(t) = e^{-\frac{t^2}{2}}$, como anunciado acima.

Ademais, se $X \sim \mathcal{N}(\mu, \sigma^2)$, então $\varphi_X(t) = e^{it\mu - \frac{\sigma^2 t^2}{2}}$ pela Proposição 10.23.

△

Corolário 10.28 (Expansão de Taylor)**.** *Se* $\mathbb{E}|X^k| < \infty$*, então*

$$\begin{aligned}\varphi_X(t) &= \varphi_X(0) + \varphi_X'(0)\cdot t + \varphi_X''(0)\,\frac{t^2}{2} + \varphi_X'''(0)\,\frac{t^3}{6} + \cdots + \varphi_X^{(k)}\,\frac{t^k}{k!} + r_k(t)\\ &= 1 + i\mathbb{E}X\cdot t - \frac{\mathbb{E}X^2}{2}t^2 - i\frac{\mathbb{E}X^3}{6}t^3 + \cdots + i^k\frac{\mathbb{E}X^k}{k!}t^k + r_k(t),\end{aligned}$$

onde o resto $r_k(t)$ *satisfaz* $\frac{r_k(t)}{t^k} \to 0$ *quando* $t \to 0$*.*

Demonstração. Imediato da Proposição 10.25 e Teorema A.6. □

Exemplo 10.29 (Poisson)*.* Calculando os momentos da Poisson: $\mathbb{E}X = -i\,\varphi_X'(0) = \lambda$, $\mathbb{E}X^2 = -\varphi_X''(0) = \lambda^2 + \lambda$, $\mathbb{V}X = \mathbb{E}X^2 - (\mathbb{E}X)^2 = \lambda$. △

Outra grande utilidade da função característica é que permite determinar a distribuição de determinada variável aleatória. Enunciaremos agora este fato, mas sua prova será dada na seção seguinte.

Teorema 10.30 (Teorema da Unicidade)**.** *Se duas variáveis aleatórias têm a mesma função característica, então têm a mesma distribuição.*

Exemplo 10.31 (Soma de variáveis Poisson independentes)*.* Se $X \sim \text{Poisson}(\lambda)$ e $Y \sim \text{Poisson}(\mu)$ são independentes, então

$$\varphi_{X+Y}(t) = \varphi_X(t)\cdot\varphi_Y(t) = e^{\lambda(e^{it}-1)}e^{\mu(e^{it}-1)} = e^{(\lambda+\mu)(e^{it}-1)} = \varphi_Z(t),$$

onde $Z \sim \text{Poisson}(\lambda+\mu)$. Portanto, $X + Y \sim \text{Poisson}(\lambda+\mu)$. △

A seguir enunciaremos uma versão simplificada do mais importante teorema deste capítulo, o Teorema da Continuidade de Lévy, que relaciona convergência de funções características com a convergência em distribuição estudada no Capítulo 7. Daremos alguns exemplos de sua grande aplicabilidade. Na Seção 10.3, enunciaremos e provaremos uma versão um pouco mais geral que a enunciada abaixo.

Teorema 10.32 (Teorema da Continuidade de Lévy)**.** *Sejam X e $(X_n)_{n\in\mathbb{N}}$ variáveis aleatórias. Então X_n converge em distribuição para X se, e somente se, $\varphi_{X_n}(t) \to \varphi_X(t)$ para todo $t \in \mathbb{R}$.*

Como exemplos da força do Teorema da Continuidade, forneceremos novas provas da convergência de binomial a Poisson, da Lei Fraca dos Grandes Números e do Teorema do Limite Central para o caso i.i.d.

Exemplo 10.33 (Binomial Converge para Poisson). Seja $\lambda > 0$ e para $n > \lambda^{-1}$ considere $X_n \sim \text{Binom}(n, \frac{\lambda}{n})$. Então $X_n \xrightarrow{\text{d}} \text{Poisson}(\lambda)$. Com efeito, analisando a função característica das X_n obtemos $\varphi_{X_n}(t) = [1+\frac{\lambda}{n}(e^{it}-1)]^n \to e^{\lambda(e^{it}-1)} = \varphi_X(t)$, onde $X \sim \text{Poisson}(\lambda)$. △

Exemplo 10.34 (Lei Fraca dos Grandes Números para o caso i.i.d.). Faremos uma demonstração curta do Teorema 8.3. Como as X_n são i.i.d.,

$$\varphi_{\frac{S_n}{n}}(t) = \varphi_{S_n}(\tfrac{t}{n}) = \varphi_{X_1}(\tfrac{t}{n}) \cdots \varphi_{X_n}(\tfrac{t}{n}) = \left[\varphi_{X_1}(\tfrac{t}{n})\right]^n = \left[1 + i\frac{\mu t}{n} + r_1(\tfrac{t}{n})\right]^n,$$

onde $r_1(\cdot)$ é tal que $\frac{r_1(w)}{w} \to 0$ quando $w \to 0$. Segue que $\varphi_{\frac{S_n}{n}}(t) \to e^{it\mu}$ quando $n \to \infty$, para todo $t \in \mathbb{R}$. Pelo Teorema da Continuidade de Lévy, $\frac{S_n}{n} \xrightarrow{\text{d}} \mu$. Como μ é constante, segue da Proposição 7.36 que $\frac{S_n}{n} \xrightarrow{\mathbb{P}} \mu$. △

Exemplo 10.35 (Teorema do Limite Central). Faremos uma demonstração curta do Teorema 9.1. Supomos sem perda de generalidade que $\mu = 0$. Como as X_n são i.i.d.,

$$\varphi_{\frac{S_n}{\sigma\sqrt{n}}}(t) = \varphi_{S_n}(\tfrac{t}{\sigma\sqrt{n}}) = \left[\varphi_{X_1}\left(\tfrac{t}{\sigma\sqrt{n}}\right)\right]^n = \left[1 - \frac{t^2}{2n} + r_2\left(\tfrac{t}{\sigma\sqrt{n}}\right)\right]^n,$$

onde $r_2(\cdot)$ é tal que $\frac{r_2(w)}{w^2} \to 0$ quando $w \to 0$. Segue que $\varphi_{\frac{S_n}{\sigma\sqrt{n}}}(t) \to e^{-\frac{t^2}{2}}$ quando $n \to \infty$, para todo $t \in \mathbb{R}$. Como $e^{-t^2/2}$ é a função característica de uma normal padrão (Exemplo 10.27), pelo Teorema da Continuidade de Lévy, segue que $\frac{S_n}{\sigma\sqrt{n}} \xrightarrow{\text{d}} \mathcal{N}(0,1)$. △

10.3 Unicidade e convergência

Nesta seção provaremos o Teorema da Unicidade e o Teorema da Continuidade de Lévy. A maior parte desta seção será uma continuação da Seção 7.4, em que vamos estabelecer propriedades de convergência em distribuição de variáveis aleatórias análogas às de convergência de números reais.

Teorema 10.36 (Teorema da Seleção de Helly)**.** *Dada uma sequência $(F_n)_n$ de funções de distribuição, existem uma subsequência $(F_{n_k})_k$ e uma função $F : \mathbb{R} \to [0,1]$ não-decrescente e contínua à direita tal que $F_{n_k}(x) \to F(x)$ quando $k \to \infty$, para todo x ponto de continuidade de F.*

Demonstração. Seja $B = \{r_1, r_2, r_3, \dots\}$ um conjunto denso em $\mathbb{R}$. Como $(F_n(r_1))_n$ é uma sequência limitada ao intervalo $[0,1]$, existe uma subsequência $(n_j^1)_j$ de $\mathbb{N}$ tal que $F_{n_j^1}(r_1) \to G(r_1)$, quando $j \to \infty$, para algum número $G(r_1) \in [0,1]$. Da mesma maneira, existe uma subsequência $(n_j^2)_j$ de $(n_j^1)_j$ tal que $F_{n_j^2}(r_2) \to G(r_2)$ para algum número $G(r_2) \in [0,1]$. Por conveniência, podemos supor que $n_1^2 = n_1^1$ e $n_2^2 = n_2^1$. Prosseguindo recursivamente, para todo $k \in \mathbb{N}$ existe $(n_j^k)_j$ subsequência de $(n_j^{k-1})_j$ tal que $n_l^k = n_l^{k-1}$ para $l = 1, \dots, k$ e tal que $F_{n_j^k}(r_k) \to G(r_k)$ para algum número $G(r_k) \in [0,1]$. A subsequência $(n_j)_j$ desejada é definida tomando-se $n_j = n_j^j$ para todo $j \in \mathbb{N}$. Observe que, para cada $k \in \mathbb{N}$, $(n_j)_j$ é subsequência de $(n_j^k)_j$, donde $F_{n_j}(r_k) \to G(r_k)$ quando $j \to \infty$. Ademais, $G : B \to [0,1]$ é não-decrescente.

Defina $F(x) = \inf\{G(r) : r \in B, r > x\}$. Note que F é não-decrescente por inclusão, e é contínua à direita, pois $\inf A_n \downarrow \inf A$ sempre que $A_n \uparrow A \subseteq \mathbb{R}$. Resta mostrar que $F_{n_j}(x) \to F(x)$ para todo ponto x onde F é contínua. Sejam $\varepsilon > 0$ e $x \in \mathbb{R}$ ponto de continuidade. Tome $\delta > 0$ tal que $F(x) - \varepsilon < F(x-\delta) \leqslant F(x+\delta) < F(x) + \varepsilon$, e tome $r', r'' \in B$ tais que $x - \delta < r' < x < r'' < x + \delta$. Pela definição de F e como G é não-decrescente, $F(x-\delta) \leqslant G(r') \leqslant F(x) \leqslant G(r'') \leqslant F(x+\delta)$. Daí segue que, para todo j suficientemente grande, $F(x) - \varepsilon < F_{n_j}(r') \leqslant F_{n_j}(r'') < F(x) + \varepsilon$ e, como a função F_{n_j} é não-decrescente, $F(x) - \varepsilon < F_{n_j}(x) < F(x) + \varepsilon$; o que conclui esta demonstração. $\square$

Observe que o Teorema de Seleção de Helly não garante que a função limite F seja uma função de distribuição. Intuitivamente, isso pode falhar porque é possível deixar massa escapar ao infinito. Por exemplo, se $X_n \sim \mathcal{U}[-n, 2n]$, então $F_{X_n}(x) \to \frac{1}{3}$ para todo $x \in \mathbb{R}$. A função limite F é não-decrescente e contínua à direita, mas não é função de distribuição pois $\lim_{x\to\pm\infty} F(x) = \frac{1}{3}$. Nesse exemplo, podemos pensar que $\frac{1}{3}$ da massa escapou para $-\infty$ e $\frac{2}{3}$ da massa escaparam para $+\infty$. Por isso introduzimos a seguinte definição.

Definição 10.37. Dada uma sequência $(X_n)_n$ de variáveis aleatórias, dizemos que a sequência está *confinada* se, para todo $\varepsilon > 0$, existe um conjunto compacto K tal que $\mathbb{P}(X_n \notin K) < \varepsilon$ para todo $n \in \mathbb{N}$.

Podemos agora enunciar o seguinte corolário do Teorema de Seleção de Helly.

Corolário 10.38. *Seja $(X_n)_n$ uma sequência confinada de variáveis aleatórias. Então existem uma subsequência $(X_{n_k})_k$ e uma variável aleatória X tais que $X_{n_k} \xrightarrow{\mathrm{d}} X$.*

Demonstração. Sejam $(F_n)_n$ as respectivas funções de distribuição de $(X_n)_n$. Pelo Teorema da Seleção de Helly, existem uma subsequência $(F_{n_k})_k$ e uma função $F : \mathbb{R} \to [0, 1]$ não-decrescente e contínua à direita tais que $F_{n_k}(x) \to F(x)$ para todo x ponto de continuidade de F. Para que F seja de fato uma função de distribuição, basta que ela satisfaça $\lim_{x\to+\infty}(F(x) - F(-x)) = 1$. Seja $\varepsilon > 0$. Tome M tal que $\mathbb{P}(|X_n| > M) \leqslant \varepsilon$ para todo n e tal que $\pm M$ sejam pontos de continuidade de F. Então $F(M) - F(-M) = \lim_n F_n(M) - F_n(-M) \geqslant 1 - \varepsilon$, o que conclui a prova. $\square$

Como no caso de sequências de números determinísticos, podemos usar o corolário acima para obter outro que permitirá estabelecer convergência de uma sequência olhando para subsequências.

Corolário 10.39. *Seja $(X_n)_n$ uma sequência confinada de variáveis aleatórias. Suponha que exista uma variável aleatória X tal que $X_{n_k} \xrightarrow{\mathrm{d}} X$ para qualquer subsequência $(X_{n_k})_k$ que convirja em distribuição. Então $X_n \xrightarrow{\mathrm{d}} X$.*

Demonstração. Seja $g : \mathbb{R} \to \mathbb{R}$ contínua e limitada. Pelo Teorema de Helly-Bray, basta mostrar que a sequência numérica $(\mathbb{E}[g(X_n)])_n$ converge para $\mathbb{E}[g(X)]$. Para isso, basta mostrar que toda subsequência $(\mathbb{E}[g(X_{n_k})])_k$ tem uma subsubsequência $(\mathbb{E}[g(X_{n_{k_j}})])_j$ que converge para $\mathbb{E}[g(X)]$ (veja Teorema A.8). Seja $(n_k)_k$ uma tal subsequência. Pelo Corolário 10.38, existe uma subsubsequência $(n_{k_j})_j$ tal que $(X_{n_{k_j}})_j$ converge em distribuição para alguma variável aleatória. Por hipótese, essa variável aleatória é X, ou seja, $X_{n_{k_j}} \xrightarrow{\mathrm{d}} X$. Pelo Teorema de Helly-Bray, $(\mathbb{E}[g(X_{n_{k_j}})])_j \to \mathbb{E}[g(X)]$, como queríamos demonstrar. □

De volta à função característica, gostaríamos de um critério para mostrar que uma dada sequência de variáveis aleatórias está confinada.

Lema 10.40. *Seja X uma variável aleatória. Para todo $\delta > 0$, vale a estimativa $\mathbb{P}(|X| \geqslant \frac{2}{\delta}) \leqslant \frac{2}{\delta} \int_0^\delta \mathbb{E}[1 - \cos(tX)]\mathrm{d}t$.*

Demonstração. Comutando a esperança e a integral,

$$\begin{aligned}\frac{1}{\delta}\int_0^\delta \mathbb{E}[1-\cos(tX)]\mathrm{d}t = \mathbb{E}\left[\frac{1}{\delta}\int_0^\delta [1-\cos(tX)]\mathrm{d}t\right] &= \mathbb{E}\Big[1 - \tfrac{\operatorname{sen}(\delta X)}{\delta X}\Big] \\ &\geqslant \mathbb{E}\Big[\Big(1 - \tfrac{\operatorname{sen}(\delta X)}{\delta X}\Big)\mathbb{1}_{\{|X| \geqslant \frac{2}{\delta}\}}\Big] \geqslant \tfrac{1}{2}\mathbb{P}(|X| \geqslant \tfrac{2}{\delta}),\end{aligned}$$

pois $1 - \frac{\operatorname{sen} u}{u} \geqslant 0$ para todo $u \neq 0$ e $1 - \frac{\operatorname{sen} u}{u} \geqslant \frac{1}{2}$ se $|u| \geqslant 2$. Comutar a esperança e a integral está justificado pelo Teorema de Tonelli. □

A proposição abaixo, junto com o Teorema da Unicidade, permite usar funções características para identificar um limite em distribuição.

Proposição 10.41. *Se $X_n \xrightarrow{\mathrm{d}} X$, então $\varphi_{X_n}(t) \to \varphi_X(t)$ para todo $t \in \mathbb{R}$.*

Demonstração. Seja $t \in \mathbb{R}$. Como $\operatorname{sen}(tx)$ e $\cos(tx)$ são funções contínuas e limitadas de x, segue do Teorema de Helly-Bray que $\mathbb{E}[\cos(tX_n)] \to \mathbb{E}[\cos(tX)]$ e $\mathbb{E}[\operatorname{sen}(tX_n)] \to \mathbb{E}[\operatorname{sen}(tX)]$, donde $\varphi_{X_n}(t) \to \varphi_X(t)$. □

Agora estamos prontos para mostrar o Teorema da Continuidade de Lévy. Na verdade enunciaremos e provaremos abaixo uma versão um pouco mais forte que aquela enunciada na seção anterior. Ela nos diz que basta que a sequência de funções características $(\varphi_{X_n})_n$ convirja para uma função que seja contínua em $t = 0$, e isto já será suficiente para garantir que o limite das funções características também seja uma função característica. Esta última será a função característica de uma variável aleatória X que será o limite em distribuição da sequência $(X_n)_n$.

Teorema 10.42 (Teorema da Continuidade de Lévy)**.** *Sejam $(X_n)_{n\in\mathbb{N}}$ variáveis aleatórias e $(\varphi_{X_n})_{n\in\mathbb{N}}$ suas respectivas funções características. Se existe $\lim_n \varphi_{X_n}(t) = \varphi(t)$ para todo t real e φ é contínua em $t = 0$, então existe uma variável aleatória X, tal que $\varphi = \varphi_X$ e $X_n \xrightarrow{\mathrm{d}} X$.*

Demonstração. Do Lema 10.40 e do Teorema da Convergência Dominada,

$$\begin{aligned}\mathbb{P}(|X_n| \geqslant \tfrac{2}{\delta}) &\leqslant \tfrac{2}{\delta}\int_0^\delta [1-\cos(tX_n)]\mathrm{d}t \\ &= \tfrac{2}{\delta}\int_0^\delta [1-\Re\varphi_{X_n}(t)]\mathrm{d}t \to \tfrac{2}{\delta}\int_0^\delta [1-\Re\varphi(t)]\mathrm{d}t\end{aligned}$$

para todo $\delta > 0$. Seja $\varepsilon > 0$. Como φ é contínua em $t = 0$, podemos tomar $\delta_0 > 0$ tal que $|1 - \Re\varphi(t)| < \frac{\varepsilon}{2}$ para todo $t \in [0, \delta_0]$. Assim podemos tomar n_0 tal que $\mathbb{P}(|X_n| \geqslant 2\delta_0^{-1}) < \varepsilon$ para todo $n > n_0$. Para cada $n \leqslant n_0$, sempre existe $\delta_n > 0$ tal que $\mathbb{P}(|X_n| \geqslant 2\delta_n^{-1}) < \varepsilon$. Tomando $\delta = \min\{\delta_0, \delta_1, \ldots, \delta_{n_0}\}$, obtemos $\mathbb{P}(|X_n| \geqslant 2\delta^{-1}) < \varepsilon$ para todo $n \in \mathbb{N}$. Portanto, a sequência $(X_n)_n$ está confinada.

Pelo Corolário 10.38, existem uma variável aleatória X e uma subsequência $(X_{n_k})_k$ tais que $X_{n_k} \xrightarrow{\mathrm{d}} X$. Pela Proposição 10.41, $\varphi_X = \varphi$. Agora seja $X_{n'_k}$ outra subsequência que converge em distribuição para alguma X'. Novamente, pela Proposição 10.41, $\varphi_{X'} = \varphi$ e, pelo Teorema de Unicidade, $X' \sim X$. Pelo Corolário 10.39, $X_n \xrightarrow{\mathrm{d}} X$, concluindo a prova. □

Ficou faltando apenas demonstrar o Teorema de Unicidade.

Prova do Teorema de Unicidade. Basta mostrar que, dados $a, b \in \mathbb{R}$, vale $\mathbb{P}(a \leqslant X \leqslant b) = \mathbb{P}(a \leqslant Y \leqslant b)$, pois segue que $F_X = F_Y$ tomando-se $a \downarrow -\infty$. Para isso, vamos aproximar $\mathbb{1}_{[a,b]}$ por uma combinação de funções trigonométricas e usar a hipótese de que $\varphi_X = \varphi_Y$.

Sejam $-\infty < a < b < \infty$ e $\varepsilon > 0$. Tome $f^\varepsilon(x) : \mathbb{R} \to [0,1]$ contínua tal que $f(x) = 1$ para $x \in [a,b]$ e $f(x) = 0$ para $x \notin [a-\varepsilon, b+\varepsilon]$. Seja $n > |a|+|b|+\varepsilon$. Note que f^ε é uma função contínua em $[-n,n]$ e $f^\varepsilon(-n) = f^\varepsilon(n) = 0$.

Pela versão trigonométrica do Teorema de Weierstrass (Teorema A.16), f^ε pode ser aproximada em $[-n,n]$ por polinômios trigonométricos. Em particular, existem $m \in \mathbb{N}$ e $a_{-m}, \ldots, a_m \in \mathbb{C}$ tais que a função

$$f_n^\varepsilon(x) = \sum_{k=-m}^{m} a_k e^{i\frac{k\pi}{n}x}$$

assume valores reais e satisfaz

$$\sup_{x\in[-n,n]} |f_n^\varepsilon(x) - f^\varepsilon(x)| \leqslant \tfrac{1}{n} \quad \text{e} \quad \sup_{x\in\mathbb{R}} |f_n^\varepsilon(x)| \leqslant 2, \tag{10.43}$$

onde a segunda estimativa acima segue da primeira e do fato que f_n^ε é uma função periódica de período $2n$. Tomando-se a esperança,

$$\mathbb{E}[f_n^\varepsilon(X)] = \sum_{k=1}^{m} a_k \varphi_X(\tfrac{k\pi}{n}) = \sum_{k=1}^{m} a_k \varphi_Y(\tfrac{k\pi}{n}) = \mathbb{E}[f_n^\varepsilon(Y)]. \tag{10.44}$$

Observe que

$$\begin{aligned}
&\mathbb{P}(a \leqslant X \leqslant b) \leqslant \mathbb{E}[f^\varepsilon(X)] \leqslant \mathbb{E}[f_n^\varepsilon(X)] + \tfrac{1}{n} + 2\,\mathbb{P}(|X| \geqslant n) \quad \text{e} \\
&\mathbb{P}(a-\varepsilon \leqslant Y \leqslant b+\varepsilon) \geqslant \mathbb{E}[f^\varepsilon(Y)] \geqslant \mathbb{E}[f_n^\varepsilon(Y)] - \tfrac{1}{n} - 2\,\mathbb{P}(|Y| \geqslant n).
\end{aligned}$$

Em cada linha acima, a primeira desigualdade é devida à definição de f^ε e segunda segue das estimativas em (10.43). Substituindo a identidade (10.44), tomando $n \to \infty$ e depois $\varepsilon \downarrow 0$, chegamos a $\mathbb{P}(a \leqslant X \leqslant b) \leqslant \mathbb{P}(a \leqslant Y \leqslant b)$. De forma idêntica provamos a desigualdade oposta, donde segue que $F_X = F_Y$,

concluindo a prova. □

10.4 Fórmula de inversão

A prova do Teorema de Unicidade que demos na seção anterior é muito instrutiva. Ela está baseada no fato de que, se duas distribuições são indistinguíveis quando testadas em funções trigonométricas, então elas são as mesmas.

Entretanto, aquela prova não é auto-contida pois usa a versão trigonométrica do Teorema de Weierstrass. Daremos nesta seção uma prova auto-contida, que terá seus pontos altos como transformada de Fourier inversa e cálculo da integral de Dirichlet sem utilizar Análise Complexa.

Enunciamos e provamos agora a fórmula de inversão, que tem o Teorema de Unicidade como corolário. Se bem que nem o enunciado nem as técnicas de prova serão usadas no restante desse livro, sua leitura envolve um uso mais engenhoso da integral de Lebesgue, o que pode causar certa satisfação.

Teorema 10.45 (Fórmula de inversão)**.** *Sejam X uma variável aleatória e $a < b$ dois pontos de continuidade de F_X. Então*

$$F_X(b) - F_X(a) = \lim_{c \to +\infty} \frac{1}{2\pi} \int_a^b \int_{-c}^{+c} e^{-ixt} \varphi_X(t) \, \mathrm{d}t \, \mathrm{d}x.$$

Informalmente, se X tem densidade suave e com cauda leve,

$$\varphi_X(t) = \int_{-\infty}^{+\infty} e^{itx} f_X(x) \, \mathrm{d}x$$

cuja transformada de Fourier inversa[17] é dada por

$$f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-ixt} \varphi_X(t) \, \mathrm{d}t.$$

Integrando em $[a, b]$ obtemos

$$F_X(b) - F_X(a) = \int_a^b f_X(x)\,\mathrm{d}x = \frac{1}{2\pi}\int_a^b \int_{-\infty}^{+\infty} e^{-ixt}\varphi_X(t)\,\mathrm{d}t\,\mathrm{d}x,$$

explicando a fórmula de inversão. Havendo motivado a fórmula, passemos à demonstração, sem supor suavidade de nenhum tipo.

Demonstração. Sejam $-\infty < a < b < \infty$. Para cada $c > 0$, definimos

$$A_c = \frac{1}{2\pi}\int_a^b \int_{-c}^{+c} e^{-ixt}\varphi_X(t)\,\mathrm{d}t\,\mathrm{d}x.$$

Usando o Teorema de Fubini para inverter as integrais,

$$A_c = \frac{1}{2\pi}\mathbb{E}\Big[\int_{-c}^{+c}\int_a^b e^{-ixt}e^{itX}\,\mathrm{d}x\,\mathrm{d}t\Big],$$

que se justifica porque $\int_a^b \int_{-c}^{c} |e^{-itx}e^{itX}|\mathrm{d}t\mathrm{d}x = 2c(b-a) < \infty$.
A partir da expressão acima, definimos

$$g_c(z) = \int_{-c}^{+c}\int_a^b e^{-ixt}e^{itz}\,\mathrm{d}x\,\mathrm{d}t.$$

Integrando, expandindo, usando que o cosseno é par e substituindo, obtemos

$$g_c(z) = \int_{-c(z-a)}^{+c(z-a)} \frac{\operatorname{sen} v}{v}\,\mathrm{d}v - \int_{-c(z-b)}^{+c(z-b)} \frac{\operatorname{sen} v}{v}\,\mathrm{d}v.$$

Observe que a função $h(r, s) = \int_r^s \frac{\operatorname{sen} v}{v}\,\mathrm{d}v$ é limitada e tende a algum número positivo $\lambda > 0$ quando $s \to +\infty$ e $r \to -\infty$. Mais abaixo vamos descobrir o valor desse limite λ.

[17] A transformada de Fourier inversa não será usada neste livro, é mencionada apenas para explicar de onde veio a fórmula de inversão. Em Análise, o sinal de menos no expoente aparece na transformada e não na sua inversa, e usa-se $\sqrt{2\pi}$ em ambas ao invés de 2π na inversa.

De volta às funções g_c, separando em casos, obtemos

$$\lim_{c\to+\infty} g_c(z) = 2\lambda\Big[\mathbb{1}_{(a,b)}(z) + \tfrac{1}{2}\mathbb{1}_{\{a,b\}}(z)\Big].$$

Como g_c é limitada, pelo Teorema da Convergência Dominada,

$$\lim_{c\to+\infty} A_c = \lim_{c\to+\infty} \tfrac{1}{2\pi}\mathbb{E}[g_c(X)] = \tfrac{\lambda}{\pi}\mathbb{P}(a < X < b) + \tfrac{1}{2}\tfrac{\lambda}{\pi}\mathbb{P}(X \in \{a,b\}).$$

Repare que o enunciado do teorema é para pontos a e b onde F_X é contínua. Usando o fato de que $\int_{-c}^{+c} \frac{\operatorname{sen} v}{v}\mathrm{d}v \to \pi$, isso completa a prova do teorema.

Provaremos agora que $\lambda = \pi$, isto é, $\int_{-c}^{+c} \frac{\operatorname{sen} v}{v}\mathrm{d}v \to \pi$. Esta integral, condicionalmente convergente, é conhecida como integral de Dirichlet.

Consideramos o caso $X \sim \mathcal{N}(0,1)$, que, convenientemente, tem como função característica um múltiplo da sua densidade. Pela fórmula de inversão,

$$\begin{aligned}
\tfrac{\lambda}{\pi}\big[F_X(b) - F_X(a)\big] &= \lim_{c\to+\infty} \frac{1}{2\pi}\int_a^b \int_{-c}^{+c} e^{-ixt}\varphi_X(t)\,\mathrm{d}t\,\mathrm{d}x. \\
&= \lim_{c\to+\infty} \int_a^b \frac{1}{\sqrt{2\pi}}\Big(\int_{-c}^{+c} e^{-itx}\frac{e^{-t^2/2}}{\sqrt{2\pi}}\,\mathrm{d}t\Big)\,\mathrm{d}x \\
&= \int_a^b \frac{1}{\sqrt{2\pi}}\Big(\int_{-\infty}^{+\infty} e^{-itx}\frac{e^{-t^2/2}}{\sqrt{2\pi}}\,\mathrm{d}t\Big)\,\mathrm{d}x \\
&= \int_a^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\,\mathrm{d}x \\
&= F_X(b) - F_X(a),
\end{aligned}$$

donde concluímos que $\lambda = \pi$. Nas igualdades acima usamos o Teorema da Convergência Dominada (a densidade é integrável e $|e^{-itx}| = 1$), e o valor da integral correspondente à função característica da normal.

Essa prova de que $\int_{-c}^{+c} \frac{\operatorname{sen} v}{v}\mathrm{d}v \to \pi$ ilustra a técnica de aceitar um resultado parcial com constante indeterminada e descobrir a constante após aplicar o argumento a um caso concreto. Outra prova de $\int_{-c}^{+c} \frac{\operatorname{sen} v}{v}\mathrm{d}v \to \pi$, que

tampouco usa Análise Complexa, consiste em escrever

$$\begin{aligned}\int_0^{+c} \tfrac{\operatorname{sen} v}{v}\mathrm{d}v &= \int_0^{+c}\int_0^{+\infty} e^{-xv}\operatorname{sen} v\,\mathrm{d}x\mathrm{d}v = \int_0^{+\infty}\int_0^{+c} e^{-xv}\operatorname{sen} v\,\mathrm{d}v\mathrm{d}x \\ &= \int_0^{+\infty} \frac{1-e^{-cx}(\cos c + x\operatorname{sen} c)}{1+x^2}\mathrm{d}x \to \int_0^{+\infty}\frac{1}{1+x^2}\mathrm{d}x = \frac{\pi}{2}\end{aligned}$$

quando $c \to +\infty$. Comutar as integrais é justificado pelo Teorema de Fubini, pois $\int_0^c |\frac{\operatorname{sen} v}{v}|\mathrm{d}v \leqslant 1 + [\log c]^+ < \infty$, enquanto o limite é justificado pelo Teorema da Convergência Dominada pois o numerador é dominado por $1 + 1 + e^{-1}$ para todo $x \geqslant 0$ e $c \geqslant 1$.

Assim completamos a prova de que $\int_{-c}^{+c} \frac{\operatorname{sen} v}{v}\mathrm{d}v \to \pi$ e, com isso, provamos a fórmula de inversão. □

10.5 Exercícios

§10.1

1. Seja $X \sim \mathcal{N}(\mu, \sigma^2)$.

(a) Mostre que, se $\mu = 0$ e $\sigma = 1$, então $M_X(t) = e^{\frac{t^2}{2}}$.
(b) Mostre que $M_X(t) = e^{\frac{\sigma^2 t^2}{2} + \mu t}$.
(c) Use a função geradora de momentos para calcular $\mathbb{E}X$ e $\mathbb{V}X$.

Dica: $2tx - x^2 = t^2 - (x-t)^2$.

2. Seja $X \sim \mathcal{U}[a, b]$.

(a) Mostre que $M_X(t) = \frac{e^{tb} - e^{ta}}{t(b-a)}$.
(b) Use a função geradora de momentos para calcular $\mathbb{E}X$ e $\mathbb{V}X$.

3. Seja $X \sim \operatorname{Exp}(\lambda)$.

(a) Mostre que $M_X(t) = \frac{\lambda}{\lambda - t}$ para $t < \lambda$ e $+\infty$ para $t \geqslant \lambda$.
(b) Use a função geradora de momentos para calcular $\mathbb{E}X$ e $\mathbb{V}X$.

4. Seja Y uma variável aleatória absolutamente contínua com função de densidade dada por

$$f_Y(y) = \begin{cases} ye^{-y}, & y > 0, \\ 0, & \text{caso contrário.} \end{cases}$$

Ache a função geradora de momentos de Y e use-a para calcular $\mathbb{E}Y$ e $\mathbb{V}Y$.

5. Seja X uma variável aleatória. Mostre que o conjunto $\{t \in \mathbb{R} : M_X(t) < \infty\}$ é um intervalo que contém o ponto $t = 0$.

§10.2

6. Seja $X \sim \text{Exp}(\lambda)$, calcule $\varphi_X(t)$.

7. Sejam $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ e $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$ independentes. Utilize funções características para mostrar que $X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

8. Suponha que $\varphi_X(t) = e^{-|t|}$. Mostre que X não é integrável.

9. Mostre que uma variável aleatória X é simétrica, isto é, $X \sim -X$ se, e somente se, sua função caraterística é uma função real.

10. Dado $a > 0$, qual é a variável aleatória cuja função característica é $\cos(at)$? Qual a variável aleatória cuja função característica é $\cos^2 t$?

§10.3

11. Sejam X uma variável aleatória e φ sua respectiva função caraterística.

(a) Mostre que se existe $t_0 \neq 0$ tal que $|\varphi(t_0)| = 1$, então existe $a \in \mathbb{R}$ de modo que a distribuição de X esteja concentrada nos números da forma $a + \frac{2\pi n}{t_0}$, $n \in \mathbb{Z}$. Isto é, $\sum_{n\in\mathbb{Z}} \mathbb{P}(X = a + \frac{2\pi n}{t_0}) = 1$.

(b) Mostre que, se existe $\varepsilon > 0$ tal que $|\varphi(t)| = 1$ para todo $t \in (\varepsilon, +\varepsilon)$, então X é degenerada.

12. Sejam $(X_n)_n$ e $(\varphi_n)_n$ uma sequência de variáveis aleatórias e suas respectivas funções características. Mostre que $X_n \xrightarrow{\text{d}} 0$ se, e somente se, existe $\varepsilon > 0$ tal que $\varphi_n(t) \to 1$, quando $n \to \infty$, para todo $t \in (-\varepsilon, \varepsilon)$.

§10.4

13. Suponha que $\int_{\mathbb{R}} |\varphi_X(t)|\,\mathrm{d}t < \infty$. Prove que X é absolutamente contínua com densidade $f_X(x) = \frac{1}{2\pi}\int_{\mathbb{R}} e^{-itx}\varphi_X(t)\,\mathrm{d}t$.

14. Sejam X, Y independentes com distribuição $\mathrm{Exp}(1)$ e defina $Z = X - Y$.

(a) Mostre que $\varphi_Z(t) = \varphi_X(t)\varphi_Y(-t)$ para todo $t \in \mathbb{R}$.
(b) Encontre φ_Z.
(c) Encontre f_Z.
(d) Prove que $\frac{1}{2\pi}\int_{\mathbb{R}} e^{-itx}\frac{1}{1+t^2}\,\mathrm{d}t = \frac{e^{-|x|}}{2}$.

15. Suponha que $X \sim \mathrm{Cauchy}(0,1)$. Prove que $\varphi_X(t) = e^{-|t|}$.

Capítulo 11

Esperança Condicional

Em inúmeras situações queremos estudar o comportamento de uma dada variável aleatória X em termos de outra variável aleatória Y, seja porque dispomos de informação acerca de Y, seja porque é mais conveniente fazer cálculos primeiro tendo seu valor congelado e depois integrando sobre seus possíveis valores.

De forma mais abstrata, podemos estar interessados no comportamento de uma variável aleatória X tendo acesso a determinado tipo de informação, não necessariamente representada por outra variável aleatória. No caso mais geral, essa informação é representada pela σ-álgebra dos eventos cuja ocorrência podemos observar a priori. Começaremos por um caso mais simples.

11.1 Esperança condicional dada uma partição

Muitas vezes conseguimos dividir Ω em pedaços que podem ser estudados separadamente para depois ver-se o todo. Nesta seção vamos trabalhar com *partições finitas*, isto é, partições da forma $\mathcal{D} = \{D_1, D_2, \ldots, D_m\}$ para algum $m \in \mathbb{N}$.

Exemplo 11.1. Sejam $X_1, X_2, X_3, \ldots$ variáveis aleatórias assumindo valores

em $\{-1, 1\}$. O espaço Ω pode ser dividido em quatro eventos onde ambas X_1 e X_2 são constantes. $\triangle$

Recordemos a definição de esperança condicional de uma variável aleatória simples X dado um evento A, vista na Seção 5.4:

$$\mathbb{E}[X|A] = \sum_x x \cdot \mathbb{P}(X = x|A).$$

Definição 11.2 (Esperança condicional dada uma partição)**.** Sejam X uma variável aleatória simples e $\mathcal{D}$ uma partição finita de Ω. Definimos a *esperança condicional de X dado $\mathcal{D}$*, denotada por $\mathbb{E}[X|\mathcal{D}]$, como sendo a variável aleatória dada por

$$\mathbb{E}[X|\mathcal{D}](\omega) = \sum_k \mathbb{E}[X|D_k]\, \mathbb{1}_{D_k}(\omega).$$

Ou seja, para cada $D \in \mathcal{D}$, a variável aleatória $\mathbb{E}[X|\mathcal{D}]$ assume o valor $\mathbb{E}[X|D]$ quando D ocorre.

A esperança condicional $\mathbb{E}[X|\mathcal{D}]$ é a uma aproximação para X que depende apenas da informação relacionada à partição $\mathcal{D}$. Ela é grosseira o suficiente para atender à restrição de ser constante no eventos de $\mathcal{D}$, mas fina o suficiente para ser a melhor entre todas as aproximações sujeitas a essa restrição. Veja a Figura 11.1.

Exemplo 11.3. Um dado honesto é lançado. Seja X o valor exibido pelo dado

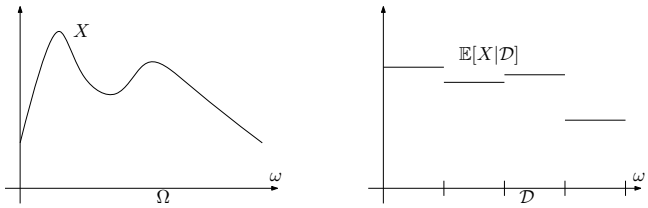


Figura 11.1. Ilustração da definição de esperança condicional.

e defina a partição $\mathcal{D} = \{\{X \text{ é par}\}, \{X \text{ é ímpar}\}\}$. Neste caso,

$$\mathbb{E}[X|\mathcal{D}](\omega) = \begin{cases} \mathbb{E}[X|X \text{ é par}], & \text{se } X(\omega) \text{ é par}, \\ \mathbb{E}[X|X \text{ é ímpar}], & \text{se } X(\omega) \text{ é ímpar}. \end{cases}$$

Assim,

$$\mathbb{E}[X|\mathcal{D}](\omega) = \begin{cases} 4, & \text{se } X(\omega) \text{ é par}, \\ 3, & \text{se } X(\omega) \text{ é ímpar}. \end{cases} \qquad \triangle$$

Proposição 11.4 (Propriedades da esperança condicional)**.** *Sejam X e Y variáveis aleatórias simples, $\mathcal{D}$ uma partição finita de Ω e $a, b \in \mathbb{R}$. Então valem as seguintes propriedades:*

(a) $\mathbb{E}[a \,|\, \mathcal{D}] = a$.
(b) *Se $X \leqslant Y$, então $\mathbb{E}[X|\mathcal{D}] \leqslant \mathbb{E}[Y|\mathcal{D}]$.*
(c) $\mathbb{E}[aX + bY|\mathcal{D}] = a\,\mathbb{E}[X|\mathcal{D}] + b\,\mathbb{E}[Y|\mathcal{D}]$.

Demonstração. A prova é baseada no fato de que essas mesmas propriedades valem quando condicionamos a um evento D fixo. Com efeito, $\mathbb{E}[a|\mathcal{D}] = \sum_k \mathbb{E}[a|D_k]\mathbb{1}_{D_k} = a\sum_k \mathbb{1}_{D_k} = a$, $\mathbb{E}[aX + bY|\mathcal{D}] = \sum_k \mathbb{E}[aX + bY|D_k]\mathbb{1}_{D_k} = a\sum_k \mathbb{E}[X|D_k]\mathbb{1}_{D_k} + b\sum_k \mathbb{E}[Y|D_k]\mathbb{1}_{D_k} = a\,\mathbb{E}[X|\mathcal{D}] + b\,\mathbb{E}[Y|\mathcal{D}]$ e, se $X \leqslant Y$, vale $\mathbb{E}[X|\mathcal{D}] = \sum_k \mathbb{E}[X|D_k]\mathbb{1}_{D_k} \leqslant \sum_k \mathbb{E}[Y|D_k]\mathbb{1}_{D_k} = \mathbb{E}[Y|\mathcal{D}]$. □

Teorema 11.5 (Esperança iterada)**.** *Sejam X uma variável aleatória simples e $\mathcal{D}$ uma partição finita. Então*

$$\mathbb{E}X = \mathbb{E}\big[\mathbb{E}[X|\mathcal{D}]\big].$$

Demonstração. Expandindo a definição de $\mathbb{E}[X|\mathcal{D}]$, obtemos

$$\begin{aligned} \mathbb{E}[\mathbb{E}[X|\mathcal{D}]] &= \mathbb{E}[\textstyle\sum_k \mathbb{E}[X|D_k]\,\mathbb{1}_{D_k}] = \sum_k \mathbb{E}[X|D_k]\,\mathbb{P}(D_k) \\ &= \textstyle\sum_k \mathbb{E}[X \cdot \mathbb{1}_{D_k}] = \mathbb{E}[X \cdot \sum_k \mathbb{1}_{D_k}] = \mathbb{E}X, \end{aligned}$$

sendo que a terceira igualdade segue da Proposição 5.53. □

Exemplo 11.6. No lançamento do dado considerado no Exemplo 11.3,

$$\mathbb{E}X = \mathbb{E}\big[\mathbb{E}[X|\mathcal{D}]\big] = \frac{1}{2}4 + \frac{1}{2}3 = \frac{7}{2}. \qquad \triangle$$

Definição 11.7. Seja $\mathcal{D} = \{D_1, \ldots, D_m\}$ uma partição finita e X uma variável aleatória simples. Dizemos que X é *$\mathcal{D}$-mensurável* se existem números $x_1, \ldots, x_m$, não necessariamente distintos, tais que

$$X = \sum_k x_k \mathbb{1}_{D_k}.$$

A equação acima diz que X é constante nos eventos de $\mathcal{D}$, o que também interpretamos como que a informação sobre $\mathcal{D}$ determina o valor de X.

Observe que $\mathbb{E}[X|\mathcal{D}]$ sempre é $\mathcal{D}$-mensurável.

O teorema a seguir diz que, se uma dada variável aleatória é $\mathcal{D}$-mensurável, então ela sai da esperança condicional como se fosse uma constante.

Teorema 11.8. *Sejam X e Y variáveis aleatórias simples e $\mathcal{D}$ uma partição finita. Se Y é $\mathcal{D}$-mensurável, então*

$$\mathbb{E}[XY|\mathcal{D}] = Y \cdot \mathbb{E}[X|\mathcal{D}]$$

e, em particular, $\mathbb{E}[Y|\mathcal{D}] = Y$.

Demonstração. Escrevendo $Y = \sum_j y_j \mathbb{1}_{D_j}$, para cada j fixado, vale a identidade

$$\mathbb{E}[XY|D_j] = \mathbb{E}[y_j X|D_j] = y_j \mathbb{E}[X|D_j],$$

donde $\mathbb{E}[XY|\mathcal{D}] = Y \cdot \mathbb{E}[X|\mathcal{D}]$ para todo $\omega \in D_j$. Como isso vale para todo j, vale a identidade para todo $\omega \in \Omega$. $\square$

Observação 11.9 (Melhor aproximação na média quadrática). Vejamos que $\mathbb{E}[X|\mathcal{D}]$ é a melhor aproximação $\mathcal{D}$-mensurável para X, no sentido de que, dentre todas as variáveis aleatórias Z que são $\mathcal{D}$-mensuráveis, é a que minimiza

o erro quadrático médio $\mathbb{E}|Z - X|^2$. Com efeito, tomando $W = \mathbb{E}[X|\mathcal{D}]$, mostraremos que $\mathbb{E}|X-W|^2 \leqslant \mathbb{E}|X-Z|^2$ para toda variável Z, $\mathcal{D}$-mensurável. Expandindo e usando o Teorema 11.8 duas vezes,

$$\begin{aligned}
&\mathbb{E}[(X-Z)^2|\mathcal{D}] - \mathbb{E}[(X-W)^2|\mathcal{D}] - \mathbb{E}[(Z-W)^2|\mathcal{D}] \\
&\quad = 2\mathbb{E}[(X-W)(W-Z)\,|\,\mathcal{D}] \\
&\quad = 2(W-Z)(\mathbb{E}[X|\mathcal{D}] - \mathbb{E}[W|\mathcal{D}]) \\
&\quad = 2(W-Z)(\mathbb{E}[X|\mathcal{D}] - W) = 0,
\end{aligned}$$

pois $W - Z$ e W são $\mathcal{D}$-mensuráveis. Tomando esperança na equação acima,

$$\mathbb{E}(X-Z)^2 = \mathbb{E}(X-W)^2 + \mathbb{E}(Z-W)^2 \geqslant \mathbb{E}(X-W)^2. \qquad \triangle$$

Observamos que a esperança condicional $\mathbb{E}[X|\mathcal{D}]$ é a *única* variável aleatória $\mathcal{D}$-mensurável tal que

$$\mathbb{E}[\mathbb{E}[X|\mathcal{D}]\mathbb{1}_B] = \mathbb{E}[X\mathbb{1}_B] \tag{11.10}$$

para todo $B \in \mathcal{D}$. A unicidade aqui é no sentido de que qualquer outra variável aleatória cumprindo essas duas condições tem que ser necessariamente igual a $\mathbb{E}[X|\mathcal{D}]$ exceto nos eventos de $\mathcal{D}$ que tenham probabilidade zero.

As principais propriedades da esperança condicional podem ser obtidas diretamente a partir desta caracterização de $\mathbb{E}[X|\mathcal{D}]$. Veremos como essa definição alternativa ajuda na prova do teorema abaixo.

Dadas duas partições $\mathcal{C}$ e $\mathcal{D}$, dizemos que $\mathcal{D}$ é *mais fina* que $\mathcal{C}$, denotado por $\mathcal{D} \succcurlyeq \mathcal{C}$, se para todo $D \in \mathcal{D}$ existe $C \in \mathcal{C}$ tal que $D \subseteq C$. Isso significa que $\mathcal{D}$ tem "mais informação" do que $\mathcal{C}$, pois os eventos de $\mathcal{D}$ formam uma subdivisão dos eventos de $\mathcal{C}$.

Teorema 11.11 (Esperança condicional iterada)**.** *Sejam X uma variável aleatória simples e $\mathcal{C}$ e $\mathcal{D}$ partições finitas de Ω. Se $\mathcal{D} \succcurlyeq \mathcal{C}$, então*

(1) $\mathbb{E}\big[\mathbb{E}[X\,|\,\mathcal{C}\,]|\mathcal{D}\,\big] = \mathbb{E}[\,X\,|\,\mathcal{C}\,]$ *quase certamente,*

(2) $\mathbb{E}\big[\mathbb{E}[X|\mathcal{D}]\,\big|\,\mathcal{C}\,\big] = \mathbb{E}[\,X\,|\,\mathcal{C}\,]$ *quase certamente.*

A propriedade acima é ilustrada na Figura 11.2.

Demonstração. Para clarificar a notação, denotamos $Y = \mathbb{E}[X|\mathcal{D}]$ e $Z = \mathbb{E}[X|\mathcal{C}]$. Para o item (1), como Z é $\mathcal{C}$-mensurável e $\mathcal{D} \succcurlyeq \mathcal{C}$, segue que Z é $\mathcal{D}$-mensurável, donde $\mathbb{E}[Z|\mathcal{D}] = Z$. Provemos agora o item (2). Seja $A \in \mathcal{C}$. Pela definição de Z, temos $\mathbb{E}[Z\mathbb{1}_A] = \mathbb{E}[X\mathbb{1}_A]$. Por outro lado, $A = B_1 \cup \cdots \cup B_k$, com $B_1, \ldots, B_k \in \mathcal{D}$ e, pela definição de Y, temos $\mathbb{E}[X\mathbb{1}_{B_j}] = \mathbb{E}[Y\mathbb{1}_{B_j}]$ para $j = 1, \ldots, k$. Somando sobre j, obtemos $\mathbb{E}[X\mathbb{1}_A] = \mathbb{E}[Y\mathbb{1}_A]$. Como Z é $\mathcal{C}$-mensurável e $\mathbb{E}[Z\mathbb{1}_A] = \mathbb{E}[Y\mathbb{1}_A]$ para todo $A \in \mathcal{C}$, concluímos que $Z = \mathbb{E}[Y|\mathcal{C}]$ q.c. pela observação acima. $\square$

11.2 Função de probabilidade condicional

Consideramos agora o caso em que a partição é dada pelos valores assumidos por uma outra variável aleatória.

Definição 11.12 (Partição induzida por uma variável aleatória)**.** Seja Y uma variável aleatória simples. Sejam $a_1, \ldots, a_m$ os distintos valores assumidos por Y e $D_j = \{Y = a_j\}$, de forma que $Y = \sum_{k=1}^{m} a_k \mathbb{1}_{D_k}$. Definimos a *partição induzida por* Y como $\mathcal{D}_Y = \{D_1, D_2, \ldots, D_m\}$.

Observe que Y sempre é $\mathcal{D}_Y$-mensurável. Ademais, se $X = g(Y)$ para alguma

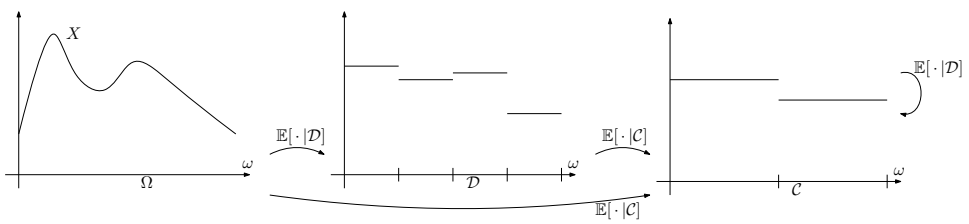


Figura 11.2. Diagrama ilustrando a esperança condicional iterada.

função $g : \mathbb{R} \to \mathbb{R}$, então X também é $\mathcal{D}_Y$-mensurável. Reciprocamente, se X é $\mathcal{D}_Y$-mensurável, então $X = g(Y)$ para alguma $g : \mathbb{R} \to \mathbb{R}$.

Definição 11.13 (Esperança condicional dada uma variável aleatória)**.** Sejam X e Y variáveis aleatórias simples. Definimos a *esperança condicional de X dado Y* como sendo a variável aleatória

$$\mathbb{E}[X|Y] = \mathbb{E}[X|\mathcal{D}_Y].$$

Ou seja, $\mathbb{E}[X|Y]$ assume o valor $\mathbb{E}[X|Y = y]$ no evento $\{Y = y\}$.

Neste contexto, a Observação 11.9 nos diz que $\mathbb{E}[X|Y]$ é, dentre todas a variáveis aleatórias Z que podem ser expressas como $Z = g(Y)$ para alguma $g : \mathbb{R} \to \mathbb{R}$, a que minimiza $\mathbb{E}(Z - X)^2$.

Na prática, podemos calcular $\mathbb{E}[X|Y]$ da seguinte maneira. Primeiro definimos a *função de probabilidade condicional de X dado Y* como

$$p_{X|Y}(x|y) = \frac{p_{X,Y}(x, y)}{p_Y(y)} \tag{11.14}$$

para todo y tal que $p_Y(y) > 0$. Nos pontos y tais que $p_Y(y) = 0$ podemos definir $p_{X|Y}(x|y) = p_X(x)$. Depois calculamos

$$\mathbb{E}[X|Y = y] = \sum_x x \cdot p_{X|Y}(x|y). \tag{11.15}$$

Exemplo 11.16. Seja (X, Y) um vetor aleatório com função de probabilidade conjunta dada por $p_{X,Y}(0, 0) = \frac{2}{5}$, $p_{X,Y}(0, 1) = \frac{1}{5}$, $p_{X,Y}(1, 0) = \frac{1}{10}$ e $p_{X,Y}(1, 1) = \frac{3}{10}$. Então, a função de probabilidade marginal de Y, p_Y, é dada por

$$p_Y(0) = \frac{1}{2} \quad \text{e} \quad p_Y(1) = \frac{1}{2},$$

o que por sua vez nos permite calcular a função de probabilidade condicional

de X dado Y, $p_{X|Y}(x|y)$, através de (11.14), obtendo

$$p_{X|Y}(0|0) = \frac{4}{5},\ p_{X|Y}(1|0) = \frac{1}{5},\ p_{X|Y}(0|1) = \frac{2}{5},\ p_{X|Y}(1|1) = \frac{3}{5},$$

isto é,

$$p_{X|Y}(0|y) = \frac{4-2y}{5} \quad \text{e} \quad p_{X|Y}(1|y) = \frac{1+2y}{5}.$$

Portanto, podemos calcular, via (11.15), que

$$\mathbb{E}[X|Y=y] = \frac{1+2y}{5}, \quad \text{para } y=0 \text{ ou } y=1,$$

logo $\mathbb{E}[X|Y] = \frac{1+2Y}{5}$ quase certamente. △

Exemplo 11.17. Sejam $X \sim \text{Binom}(n,p)$ e $Z \sim \text{Binom}(m,p)$ variáveis independentes e $Y = X + Z$. Conforme visto no Exemplo 4.20, $Y \sim \text{Binom}(n+m,p)$, logo, $p_Y(y) = \binom{m+n}{y} p^y (1-p)^{m+n-y}$. Assim,

$$\begin{aligned} p_{X|Y}(x|y) &= \frac{p_{X,Y}(x,y)}{p_Y(y)} = \frac{p_{X,Z}(x,y-x)}{p_Y(y)} = \frac{p_X(x)p_Z(y-x)}{p_Y(y)} \\ &= \frac{\binom{n}{x}p^x(1-p)^{n-x}\binom{m}{y-x}p^{y-x}(1-p)^{m-y+x}}{\binom{n+m}{y}p^y(1-p)^{n+m-y}} = \frac{\binom{n}{x}\binom{m}{y-x}}{\binom{n+m}{y}}, \end{aligned}$$

onde na terceira igualdade utilizamos o fato de X e Z serem independentes. Portanto, para todo $y = 0, \ldots, n+m$,

$$\begin{aligned} \mathbb{E}[X|Y=y] &= \sum_x x \frac{\binom{n}{x}\binom{m}{y-x}}{\binom{n+m}{y}} = \frac{n\sum_x \binom{n-1}{x-1}\binom{m}{y-x}}{\binom{n+m}{y}} \\ &= \frac{n\binom{n+m-1}{y-1}}{\binom{n+m}{y}} = \frac{ny}{n+m}. \end{aligned}$$

Logo, $\mathbb{E}[X|Y] = \frac{n}{m+n}Y$ quase certamente. △

Nos exemplos acima, $p_{X|Y}$ foi calculado a partir de $p_{X,Y}$ via (11.14). Isso é útil quando literalmente podemos observar Y e queremos atualizar nossas

expectativas com respeito à distribuição de X. Há também o caso em que $p_{X|Y}$, ao invés de calculado, é especificado ou deduzido por primeiros princípios, e serve para aplicar (11.15) entre outras ferramentas.

Exemplo 11.18. Um jogador lança um dado, e Y denota o número observado. Em seguida lança uma moeda honesta Y vezes, e X denota o número de coroas obtidas. Queremos calcular $\mathbb{E}[X|Y]$ e $\mathbb{E}X$. Para cada $y = 1, \ldots, 6$ e $x = 0, \ldots, y$, temos $p_{X|Y}(x|y) = \binom{y}{x}2^{-y}$. Calculamos então $\mathbb{E}[X|Y = y] = \sum_x x \cdot \binom{y}{x}2^{-y} = \frac{y}{2}$. Portanto, $\mathbb{E}[X|Y] = \frac{Y}{2}$ e, tomando a esperança iterada, $\mathbb{E}X = \mathbb{E}[\mathbb{E}[X|Y]] = \mathbb{E}[\frac{Y}{2}] = \frac{7}{4}$. △

A partir da função de probabilidade condicional de X dado Y, podemos também estudar a *distribuição condicional de X dado Y*, definida por

$$\mathbb{P}_{X|Y}(B|y) = \sum_{x \in B} p_{X|Y}(x|y) \tag{11.19}$$

para todo evento B.

Exemplo 11.20. Jogamos n moedas honestas, as que exibem cara permanecem como estão e as que exibem coroa são novamente lançadas. Sejam Y o número de coroas obtidas após a primeira rodada de lançamentos e X o número de coroas restantes após a segunda rodada de lançamentos. Neste caso, $p_{X|Y}(x|y) = \binom{y}{x}2^{-y}$. Sendo assim, a distribuição condicional $\mathbb{P}_{X|Y}(\cdot|y)$ corresponde à distribuição $\text{Binom}(y, \frac{1}{2})$. △

O comportamento conjunto de X e Y, ou de X isoladamente, pode ser estudado a partir dessa distribuição condicional, calculando-se a média sobre y. Mais precisamente,

$$\mathbb{P}(X \in B, Y \in C) = \sum_{y \in C} \mathbb{P}_{X|Y}(B|y)p_Y(y) \tag{11.21}$$

para quaisquer subconjuntos $B, C \subseteq \mathbb{R}$.

Exemplo 11.22. Sejam X e Y as variáveis aleatórias definidas no Exemplo 11.20. Observando que quase certamente as variáveis X e Y assumem

valores no conjunto $\{0,\dots,n\}$, podemos calcular a distribuição de X utilizando (11.21), com $B=\{k\}$ e $C=\{0,\dots,n\}$, isto é,

$$\begin{aligned}\mathbb{P}(X=k)=\mathbb{P}(X=k,Y\in\{0,\dots,n\})&=\sum_{y=0}^{n}\mathbb{P}_{X|Y}(\{k\}|y)p_Y(y)\\&=\sum_{y=0}^{n}\tbinom{y}{k}2^{-y}\tbinom{n}{y}2^{-n}=2^{-n}\sum_{y=0}^{n}\tbinom{n}{k}\tbinom{n-k}{n-y}2^{-y}\\&=2^{-n}\tbinom{n}{k}\sum_{j=0}^{n}\tbinom{n-k}{j}2^{j-n}=4^{-n}\tbinom{n}{k}\sum_{j=0}^{n-k}\tbinom{n-k}{j}2^{j}1^{n-k-j}\\&=4^{-n}\tbinom{n}{k}(2+1)^{n-k}=\tbinom{n}{k}(\tfrac{1}{4})^{k}(\tfrac{3}{4})^{n-k},\end{aligned}$$

onde utilizamos o Teorema Binomial. Portanto, $X\sim\text{Binom}(n,\frac{1}{4})$. △

A proposição abaixo diz que, se uma variável Y não nos dá informação alguma acerca de outra variável X, então a melhor aproximação para o valor de X sabendo-se o valor de Y nada mais é do que a própria esperança de X, não importando o valor de Y.

Proposição 11.23 (Variáveis independentes)**.** *Sejam X e Y variáveis aleatórias simples. Se X e Y são independentes, então $\mathbb{E}[X|Y]=\mathbb{E}X$.*

Demonstração. Imediato pois

$$\mathbb{E}[X|Y=y]=\sum_x x\cdot p_{X|Y}(x|y)=\sum_x x\cdot p_X(x)=\mathbb{E}X$$

para todo $y\in\mathbb{R}$. □

Exemplo 11.24. Sejam $X_1,\dots,X_m$ variáveis com a mesma esperança, e N uma variável aleatória assumindo valores em $\{1,\dots,m\}$ independente de $X_1,\dots,X_m$. Definimos

$$S_N=X_1+\cdots+X_N,$$

a soma dos N primeiros termos da sequência. Ou seja, S_N é a soma de uma quantidade aleatória de variáveis aleatórias. Mais formalmente, definimos $S_n = X_1 + \cdots + X_n$, para todo n, e então definimos $S_N = \sum_k S_k \mathbb{1}_{\{N=k\}}$. Vamos mostrar que

$$\mathbb{E}[S_N|N] = N \cdot \mathbb{E}X_1,$$

e, portanto, pelo Teorema 11.5,

$$\mathbb{E}[S_N] = \mathbb{E}N \cdot \mathbb{E}X_1,$$

isto é, o valor médio de uma soma aleatória é o valor médio do número de parcelas vezes o valor médio de cada parcela. Com efeito,

$$\begin{aligned}
\mathbb{E}[S_N|N] &= \sum_{k=1}^{m} \mathbb{E}[(X_1 + \cdots + X_k)\mathbb{1}_{\{N=k\}}|N] \\
&= \sum_{k=1}^{m} \mathbb{1}_{\{N=k\}} \mathbb{E}[(X_1 + \cdots + X_k)|N] \\
&= \sum_{k=1}^{m} \mathbb{1}_{\{N=k\}} \sum_{j=1}^{k} \mathbb{E}[X_j|N] \\
&= \sum_{k=1}^{m} \mathbb{1}_{\{N=k\}} \sum_{j=1}^{k} \mathbb{E}[X_j] \\
&= \mathbb{E}X_1 \cdot \sum_{k=1}^{m} k\mathbb{1}_{\{N=k\}} \\
&= N \cdot \mathbb{E}X_1.
\end{aligned}$$

Na segunda igualdade usamos o Teorema 11.8. Na quarta, usamos que X_j é independente de N, donde podemos aplicar a Proposição 11.23. Na quinta, usamos o fato de as variáveis $(X_j)_j$ terem a mesma esperança. $\triangle$

11.3 Densidade condicional

Dadas duas variáveis aleatórias simples X e Y, na seção anterior definimos a variável aleatória $\mathbb{E}[X \,|\, Y]$, que assume o valor $\mathbb{E}[X \,|\, Y = y]$ no evento $\{Y(\omega) = y\}$. Naquele contexto, o valor $\mathbb{E}[X \,|\, Y = y]$ para y tal que $\mathbb{P}(Y = y) > 0$ pode ser calculado a partir da distribuição condicional

$$\mathbb{P}_{X|Y}(B|y) = \mathbb{P}(X \in B \,|\, Y = y) = \frac{\mathbb{P}(X \in B, Y = y)}{\mathbb{P}(Y = y)}, \tag{11.25}$$

como feito nas Seções 3.4 e 5.4. Entretanto, quando Y é uma variável aleatória contínua, a fórmula (11.25) resulta na forma indeterminada $\frac{0}{0}$. Gostaríamos de poder definir $\mathbb{P}_{X|Y}(B|y)$ de forma tal que permita recuperar as propriedades vistas na seção anterior.

Nesta seção consideraremos o caso em que X e Y têm densidade conjunta. Faremos uma apresentação informal buscando motivar as definições e propriedades mais importantes. Uma justificativa mais rigorosa das propriedades aqui enunciadas será dada na Seção 11.6.

Por analogia a (11.14) definimos a *densidade condicional de X dado Y* por

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} \tag{11.26}$$

para todo y tal que $f_Y(y) > 0$. Para os pontos y onde $f_Y(y) = 0$, definimos, por convenção $f_{X|Y}(x|y) = f_X(x)$.

Observe-se que, se multiplicamos a equação acima por Δx, obtemos

$$\begin{aligned} f_{X|Y}(x|y)\,\Delta x &= \frac{f_{X,Y}(x,y)\Delta x \Delta y}{f_Y(y)\Delta y} \approx \frac{\mathbb{P}(x \leqslant X \leqslant x + \Delta x, y \leqslant Y \leqslant Y + \Delta y)}{\mathbb{P}(y \leqslant Y \leqslant y + \Delta y)} \\ &= \mathbb{P}(x \leqslant X \leqslant x + \Delta x | y \leqslant Y \leqslant y + \Delta y). \end{aligned}$$

Ou seja, coerentemente com a ideia de que $f_X(x)\,\Delta x$ representa a probabilidade de X estar em $[x, x+\Delta x]$, aqui $f_{X|Y}(x|y)\,\Delta x$ representa a probabilidade

condicional de tal evento dado que Y está em $[y, y + \Delta y]$.

A densidade condicional de X dado Y define uma distribuição parametrizada por y, à que chamamos *distribuição condicional de X dado Y*, dada por

$$\mathbb{P}_{X|Y}(B|y) = \int_B f_{X|Y}(x|y)\,\mathrm{d}x \tag{11.27}$$

para todo intervalo $B \subseteq \mathbb{R}$. Repare na semelhança com (11.19).

Assim como vimos no caso discreto, a distribuição conjunta de X e Y, ou de X isoladamente, pode ser calculada a partir dessa distribuição condicional, tomando-se a média sobre y, isto é,

$$\mathbb{P}(X \in B, Y \in C) = \int_C \mathbb{P}_{X|Y}(B|y) f_Y(y)\,\mathrm{d}y, \tag{11.28}$$

para quaisquer intervalos B e C, veja a semelhança com (11.21).

Observe que logramos obter expressões úteis envolvendo probabilidades condicionais, apesar de estarmos condicionando em um evento de medida zero. Essas expressões justificam-se mutuamente, pois

$$\iint_{B\times C} f_{X,Y}(x,y)\mathrm{d}x\mathrm{d}y = \int_C \Big(\int_B \tfrac{f_{X,Y}(x,y)}{f_Y(y)}\mathrm{d}x \Big) f_Y(y)\mathrm{d}y. \tag{11.29}$$

Exemplo 11.30. Sejam X e Y com densidade conjunta

$$f_{X,Y}(x,y) = \begin{cases} 6xy(2-x-y), & 0 < x < 1, 0 < y < 1, \\ 0, & \text{caso contrário.} \end{cases}$$

Vamos determinar a densidade condicional de X dado Y. Primeiramente,

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)\mathrm{d}x = \int_0^1 6xy(2-x-y)\mathrm{d}x = 4y - 3y^2$$

se $y \in (0,1)$ e 0 caso contrário. Para $y \in [0,1]$, calculamos

$$f_{X|Y}(x\,|\,y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{6x(2-x-y)}{4-3y}\mathbb{1}_{[0,1]}(x)$$

Para $y \notin [0,1]$, tomamos, por exemplo, $f_{X|Y}(x\,|\,y) = 6x(1-x)\mathbb{1}_{[0,1]}(x)$. $\triangle$

Exemplo 11.31. Sejam X e Y com densidade conjunta

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{2}ye^{-xy}, & 0 < x < \infty \ \text{ e } \ 0 < y < 2, \\ 0, & \text{caso contrário.} \end{cases}$$

Vamos determinar a densidade condicional de X dado Y. Temos

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)\mathrm{d}x = \frac{1}{2}\int_0^{\infty} ye^{-xy}\mathrm{d}x = \frac{1}{2}$$

para $0 < y < 2$. Logo $Y \sim \mathcal{U}[0,2]$. Assim, para $y \in (0,2)$ temos

$$f_{X|Y}(x\,|\,y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \begin{cases} ye^{-xy}, & x > 0, \\ 0, & x \leqslant 0. \end{cases}$$ $\triangle$

Uma situação muito comum na prática é quando são especificadas f_Y e $f_{X|Y}$. Neste caso, podemos estudar a variável X a partir de Y e da informação sobre como esta influencia aquela. Mais precisamente, podemos usar

$$f_{X,Y}(x,y) = f_{X|Y}(x|y)\, f_Y(y)$$

e

$$f_X(x) = \int_{-\infty}^{+\infty} f_{X|Y}(x|y) f_Y(y)\,\mathrm{d}y.$$

Exemplo 11.32. Sejam X e Y variáveis aleatórias tais que $Y \sim \mathcal{U}[0,2]$ e,

condicionado a que $Y = y$, X tem distribuição uniforme em $[0, y]$. Isto é,

$$f_{X|Y}(x \mid y) = \begin{cases} \frac{1}{y}, & 0 < x < y < 2, \\ 0, & \text{caso contrário.} \end{cases}$$

Sendo assim, podemos calcular as densidades conjunta e marginal utilizando as fórmulas acima, obtendo

$$f_{X,Y}(x, y) = f_{X|Y}(x|y)\, f_Y(y) = \begin{cases} \frac{1}{2y}, & 0 < x < y < 2, \\ 0, & \text{caso contrário.} \end{cases}$$

e

$$f_X(x) = \int_{-\infty}^{+\infty} f_{X|Y}(x|y) f_Y(y)\, \mathrm{d}y = \int_x^2 \frac{1}{2y}\frac{1}{2}\, \mathrm{d}y = -\frac{1}{2}\log\frac{x}{2}$$

para $x \in (0, 2]$ e zero caso contrário. △

Exemplo 11.33. Sejam X e Y variáveis aleatória, onde $Y \sim \text{Gama}(2, \lambda)$ e condicionado que $Y = y$, X tem distribuição $\mathcal{U}[0, y]$. Isto é, a densidade condicional $f_{X|Y}$ é especificada e vale $\frac{1}{y}\mathbb{1}_{[0,y]}(x)$. Sendo assim, podemos determinar a densidade de X como

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{+\infty} f_{X|Y}(x|y) f_Y(y)\, \mathrm{d}y = \mathbb{1}_{[0,+\infty)}(x) \int_x^{+\infty} \tfrac{1}{y}\lambda^2 y e^{-\lambda y}\, \mathrm{d}y \\ &= \lambda e^{-\lambda x}\mathbb{1}_{(0,+\infty)}(x). \end{aligned}$$

Isto é, $X \sim \text{Exp}(\lambda)$. △

A partir da distribuição condicional de X dado Y, podemos calcular a esperança condicional de X dado Y. Por analogia a (11.15), definimos

$$\mathbb{E}[X|Y = y] = \int_{-\infty}^{+\infty} x \cdot f_{X|Y}(x|y)\, \mathrm{d}x. \tag{11.34}$$

Exemplo 11.35. Se X e Y são as variáveis do Exemplo 11.32, podemos calcular

$$\mathbb{E}[X|Y=y] = \int_{-\infty}^{+\infty} x \cdot f_{X|Y}(x|y)\,\mathrm{d}x = \int_0^y \frac{x}{y}\,\mathrm{d}x = \frac{y}{2},$$

o que já era de se esperar dado que, condicionado ao evento $\{Y = y\}$, a variável X tem distribuição uniforme em $[0, y]$. △

Como na seção anterior, definimos $\mathbb{E}[X|Y]$ como a variável aleatória que toma valor $\mathbb{E}[X|Y = y]$ no evento $\{Y = y\}$. Observe que $\mathbb{E}[X|Y]$ é uma variável aleatória que pode ser expressa como uma função de Y.

Exemplo 11.36. No Exemplo 11.31, calculemos $\mathbb{E}\big[X^2\big|Y\big]$ e $\mathbb{E}\big[X^2\big|Y = 1\big]$.

Considerando a densidade condicional já calculada no Exemplo 11.31, temos

$$\mathbb{E}[X^2|Y=y] = \int_0^{+\infty} x^2 y e^{-xy}\mathrm{d}x = \frac{2}{y}\int_0^{+\infty} e^{-yx}\mathrm{d}x = \frac{2}{y^2}.$$

Logo,

$$\mathbb{E}\big[X^2\big|Y\big] = \frac{2}{Y^2} \quad \text{e} \quad \mathbb{E}\big[X^2\big|Y=1\big] = 2.$$ △

Como no Teorema 11.5, vale a propriedade da esperança iterada:

$$\mathbb{E}X = \mathbb{E}\big[\mathbb{E}[X|Y]\big]. \tag{11.37}$$

No contexto da seção anterior, esta fórmula toma a forma

$$\mathbb{E}X = \sum_y \mathbb{E}[X|Y=y]\,p_Y(y),$$

enquanto que no contexto desta seção temos,

$$\mathbb{E}X = \int_{-\infty}^{+\infty} \mathbb{E}[X|Y=y]\,f_Y(y)\,\mathrm{d}y.$$

Exemplo 11.38. Continuando os Exemplos 11.32 e 11.35,

$$\mathbb{E}X = \int_{-\infty}^{+\infty} \mathbb{E}[X|Y=y]\, f_Y(y)\, \mathrm{d}y = \int_0^2 \frac{y}{2}\,\frac{1}{2}\,\mathrm{d}y = \frac{1}{2}. \qquad \triangle$$

11.4 Esperança condicional dada uma σ-álgebra

Na Seção 11.1, mencionamos que partições mais finas que outras codificam uma situação em que se tem acesso a mais informação. Pensamos em "informação" como a coleção de eventos cuja ocorrência é acessível a um determinado observador. A forma mais geral de representar informação, quando tem-se acesso a infinitos eventos, é através de uma σ-álgebra. Sejam $(\Omega, \mathcal{F}, \mathbb{P})$ um espaço de probabilidade, X uma variável aleatória e $\mathcal{G}$ uma σ-álgebra mais grosseira que $\mathcal{F}$, isto é, $\mathcal{G} \subseteq \mathcal{F}$. Não há motivo algum para que X seja também mensurável com respeito à σ-álgebra $\mathcal{G}$. Em outras palavras, não há motivo para que a informação codificada por $\mathcal{G}$, que é mais grosseira que $\mathcal{F}$, seja suficiente para determinar o valor de X. Uma pergunta natural surge: qual seria a melhor variável aleatória $\mathcal{G}$-mensurável que poderia aproximar X em algum sentido? Reformulando a pergunta: qual a melhor aproximação para X quando temos acesso à informação codificada por $\mathcal{G}$? Nesta seção, daremos esta resposta. Trata-se de um conceito bastante abstrato, porém dos mais úteis e importantes em Probabilidade.

Teorema 11.39 (Esperança condicional dada uma σ-álgebra)**.** *Sejam $(\Omega, \mathcal{F}, \mathbb{P})$ um espaço de probabilidade, X uma variável aleatória estendida integrável ou não-negativa, e $\mathcal{G} \subseteq \mathcal{F}$ uma σ-álgebra. Então existe uma variável aleatória estendida Z que é $\mathcal{G}$-mensurável e satisfaz*

$$\int_A Z\,\mathrm{d}\mathbb{P} = \int_A X\,\mathrm{d}\mathbb{P} \quad \textit{para todo } A \in \mathcal{G}. \tag{11.40}$$

Dizemos que uma variável aleatória estendida com essas duas propriedades é uma esperança condicional de X dado $\mathcal{G}$*, e a denotamos por* $\mathbb{E}[X|\mathcal{G}]$.

A prova será dada na próxima seção.

Observação 11.41. A esperança condicional é única no seguinte sentido. Se Z e W são duas esperanças condicionais de X dado $\mathcal{G}$, então, pela Proposição 5.75, $Z = W$ q.c. Como a condição (11.40) é insensível ao que acontece em conjuntos de medida nula, somente podemos esperar unicidade nesse sentido. Por isso, toda afirmação a respeito de $\mathbb{E}[X|\mathcal{G}]$ virá com um quantificador de que vale quase certamente. $\triangle$

Observação 11.42. Se X é integrável, então $\mathbb{E}[X|\mathcal{G}]$ também é integrável e podemos supor que $\mathbb{E}[X|\mathcal{G}]$ é uma variável aleatória real (não assume valores infinitos). Se X é não-negativa, pelo Exercício 5.63 podemos assumir que $\mathbb{E}[X|\mathcal{G}]$ assume valores em $[0, +\infty]$. $\triangle$

Tomando $A = \Omega$ em (11.40), obtemos a propriedade da *esperança iterada*:

$$\mathbb{E}[\mathbb{E}[X|\mathcal{G}]] = \mathbb{E}X.$$

Outra propriedade da esperança condicional é que, se X é $\mathcal{G}$-mensurável, então $\mathbb{E}[X|\mathcal{G}] = X$ q.c.

Uma vantagem da definição de esperança condicional de uma variável aleatória estendida X dada uma σ-álgebra $\mathcal{G}$ é a sua generalidade, pois, como dissemos acima, σ-álgebras são a ferramenta ideal para codificar informação. Com efeito, os objetos definidos nas duas seções anteriores são casos particulares da definição abaixo, como será justificado na Seção 11.6. A esperança condicional dada uma partição também é um caso particular, o que segue de (11.10).

No restante desta seção, assumimos que $(\Omega, \mathcal{F}, \mathbb{P})$ é um espaço de probabilidade fixado.

Definição 11.43. Dadas duas variáveis aleatórias estendidas X e Y tais que X é integrável ou não-negativa, definimos a *esperança condicional de X dada Y* por

$$\mathbb{E}[X|Y] = \mathbb{E}[X|\sigma(Y)],$$

ou seja, é a esperança condicional dada a σ-álgebra gerada por Y.

Teorema 11.44 (Propriedades da esperança condicional)**.** *Sejam X e Y variáveis aleatórias integráveis, $\mathcal{G} \subseteq \mathcal{F}$ uma σ-álgebra e $a, b, c \in \mathbb{R}$. Então:*

(1) $\mathbb{E}[c|\mathcal{G}] = c$ *q.c.*
(2) *Se $X \leqslant Y$ q.c., então $\mathbb{E}[X|\mathcal{G}] \leqslant \mathbb{E}[Y|\mathcal{G}]$ q.c.*
(3) $\mathbb{E}[aX + bY|\mathcal{G}] = a\,\mathbb{E}[X|\mathcal{G}] + b\,\mathbb{E}[Y|\mathcal{G}]$ *q.c.*

Essas propriedades também valem para variáveis aleatórias estendidas não-negativas X e Y com constantes $a, b, c \in [0, +\infty]$.

Demonstração. Para o item (1), observe que a variável constante $X(\omega) = c$ é $\mathcal{G}$-mensurável e $\int_A X \,\mathrm{d}\mathbb{P} = \int_A c \,\mathrm{d}\mathbb{P}$ para todo $A \in \mathcal{G}$.

Para o item (3), observe que

$$\begin{aligned}\int_A (aX + bY)\,\mathrm{d}\mathbb{P} &= a\int_A X\,\mathrm{d}\mathbb{P} + b\int_A Y\,\mathrm{d}\mathbb{P} = a\int_A \mathbb{E}[X|\mathcal{G}]\,\mathrm{d}\mathbb{P} + b\int_A \mathbb{E}[Y|\mathcal{G}]\,\mathrm{d}\mathbb{P}\\ &= \int_A \Big(a\,\mathbb{E}[X|\mathcal{G}] + b\,\mathbb{E}[Y|\mathcal{G}]\Big)\mathrm{d}\mathbb{P},\end{aligned}$$

para todo $A \in \mathcal{G}$, ou seja, $\mathbb{E}[aX + bY|\mathcal{G}] = a\,\mathbb{E}[X|\mathcal{G}] + b\,\mathbb{E}[Y|\mathcal{G}]$ q.c.

Para provar o item (2), suponha que $X \leqslant Y$ q.c. Neste caso, podemos escrever $Y = X + Z$ q.c., onde Z é não-negativa. Pelo Exercício 5.63, $\mathbb{E}[Z|\mathcal{G}]$ é não-negativa q.c., e, pelo item (3), $\mathbb{E}[Y|\mathcal{G}] = \mathbb{E}[X|\mathcal{G}] + \mathbb{E}[Z|\mathcal{G}] \geqslant \mathbb{E}[X|\mathcal{G}]$ q.c., concluindo a prova. □

O teorema seguinte nos diz que, se o valor de uma variável aleatória estendida é determinado pela informação codificada pela σ-álgebra em questão, então ela sai da esperança condicional como se fosse uma constante. O Teorema 11.8 é um caso particular.

Teorema 11.45. *Se Y é $\mathcal{G}$-mensurável, $\mathbb{E}|X| < \infty$ e $\mathbb{E}|XY| < \infty$, então*

$$\mathbb{E}[XY|\mathcal{G}] = Y \cdot \mathbb{E}[X|\mathcal{G}] \text{ q.c.}$$

O mesmo vale se X e Y são não-negativas.

Demonstração. Consideramos primeiro o caso em que X e Y são não-negativas. Seja $A \in \mathcal{G}$. Tomando $Y_0 = \mathbb{1}_B$ para algum $B \in \mathcal{G}$,

$$\int_A XY_0 \, \mathrm{d}\mathbb{P} = \int_{A\cap B} X \, \mathrm{d}\mathbb{P} = \int_{A\cap B} \mathbb{E}[X|\mathcal{G}] \, \mathrm{d}\mathbb{P} = \int_A Y_0 \cdot \mathbb{E}[X|\mathcal{G}] \, \mathrm{d}\mathbb{P}.$$

Por linearidade, se Y_n é uma variável aleatória simples $\mathcal{G}$-mensurável, vale

$$\int_A XY_n \, \mathrm{d}\mathbb{P} = \int_A Y_n \cdot \mathbb{E}[X|\mathcal{G}] \, \mathrm{d}\mathbb{P}.$$

Tomando $0 \leqslant Y_n \uparrow Y$, pelo Teorema da Convergência Monótona obtemos

$$\int_A XY \, \mathrm{d}\mathbb{P} = \int_A Y \cdot \mathbb{E}[X|\mathcal{G}] \, \mathrm{d}\mathbb{P},$$

o que conclui a prova já que $Y \cdot \mathbb{E}[X|\mathcal{G}]$ é $\mathcal{G}$-mensurável.

Consideramos agora o caso em que X e XY são integráveis. Queremos mostrar que vale a identidade acima para todo $A \in \mathcal{G}$. Escrevendo $X = X^+ - X^-$ e $Y = Y^+ - Y^-$ e observando que $\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X^+|\mathcal{G}] - \mathbb{E}[X^-|\mathcal{G}]$, é suficiente mostrar que

$$\int_A X^\pm Y^\pm \, \mathrm{d}\mathbb{P} = \int_A Y^\pm \cdot \mathbb{E}[X^\pm|\mathcal{G}] \, \mathrm{d}\mathbb{P},$$

mas isso segue diretamente do caso anterior. □

O teorema seguinte generaliza o Teorema 11.11.

Teorema 11.46 (Esperança condicional iterada)**.** *Seja $\mathcal{H}$ uma σ-álgebra tal que $\mathcal{H} \subseteq \mathcal{G} \subseteq \mathcal{F}$, e X uma variável aleatória estendida integrável ou não-negativa. Então valem as seguintes identidades:*

(1) $\mathbb{E}\big[\mathbb{E}[X|\mathcal{H}]|\mathcal{G}\big] = \mathbb{E}\big[X|\mathcal{H}\big]$ *q.c.*
(2) $\mathbb{E}\big[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}\big] = \mathbb{E}\big[X|\mathcal{H}\big]$ *q.c.*

Uma interpretação visual do teorema acima no caso de σ-álgebras geradas por finitos eventos é dada na Figura 11.2.

Demonstração. Para clarificar, escrevemos $Y = \mathbb{E}[X|\mathcal{G}]$ e $Z = \mathbb{E}[X|\mathcal{H}]$. Para a primeira igualdade, basta observar que Z é $\mathcal{G}$-mensurável, donde $\mathbb{E}[Z|\mathcal{G}] = Z$ q.c. Provemos agora a segunda igualdade. Seja $A \in \mathcal{H}$. Pela definição de Z, temos $\int_A Z \, \mathrm{d}\mathbb{P} = \int_A X \, \mathrm{d}\mathbb{P}$. Por outro lado, como $A \in \mathcal{G}$, pela definição de Y temos $\int_A X \, \mathrm{d}\mathbb{P} = \int_A Y \, \mathrm{d}\mathbb{P}$. Como Z é $\mathcal{H}$-mensurável e $\int_A Z \, \mathrm{d}\mathbb{P} = \int_A Y \, \mathrm{d}\mathbb{P}$ para todo $A \in \mathcal{H}$, concluímos que $Z = \mathbb{E}[Y|\mathcal{H}]$ q.c. □

Definição 11.47. Dizemos que uma variável aleatória estendida X é *independente* da σ-álgebra $\mathcal{G}$ se $\{X \leqslant a\}$ e A são independentes para todos $a \in \mathbb{R}$ e $A \in \mathcal{G}$.

Proposição 11.48. *Se X é é uma variável aleatória estendida integrável ou não-negativa, e X é independente de $\mathcal{G}$, então $\mathbb{E}[X|\mathcal{G}] = \mathbb{E}X$ q.c.*

Demonstração. Pela independência de X e $\mathcal{G}$, temos, para todo $A \in \mathcal{G}$,

$$\int_A X \, \mathrm{d}\mathbb{P} = \mathbb{E}[X\mathbb{1}_A] = \mathbb{E}X \cdot \mathbb{E}[\mathbb{1}_A] = \int_A (\mathbb{E}X) \, \mathrm{d}\mathbb{P}$$

e, sendo constante, $\mathbb{E}X$ é $\mathcal{G}$-mensurável, o que conclui a prova. □

Os teoremas de convergência da integral, vistos na Seção 5.5, também têm seus análogos no contexto de esperança condicional. Nos três teoremas abaixo, $(X_n)_n$, X e Y denotam variáveis aleatórias definidas num mesmo espaço de probabilidade.

Teorema 11.49 (Convergência Monótona)**.** *Sejam $(X_n)_n$ e X variáveis aleatórias estendidas não-negativas tais que $0 \leqslant X_n \uparrow X$ q.c. Seja $\mathcal{G} \subseteq \mathcal{F}$ uma σ-álgebra. Então $\mathbb{E}[X_n|\mathcal{G}] \to \mathbb{E}[X|\mathcal{G}]$ q.c.*

Demonstração. Tome $Y = \limsup_n \mathbb{E}[X_n|\mathcal{G}]$, que é $\mathcal{G}$-mensurável. Como $0 \leqslant \mathbb{E}[X_n|\mathcal{G}] \leqslant \mathbb{E}[X_{n+1}|\mathcal{G}]$ q.c., segue que $0 \leqslant \mathbb{E}[X_n|\mathcal{G}] \uparrow Y$ q.c. Para $A \in \mathcal{G}$,

$$\int_A X_n \, \mathrm{d}\mathbb{P} = \int_A \mathbb{E}[X_n|\mathcal{G}] \, \mathrm{d}\mathbb{P}$$

e, aplicando o Teorema da Convergência Monótona em ambos os lados,

$$\int_A X \mathrm{d}\mathbb{P} = \int_A Y \mathrm{d}\mathbb{P},$$

donde $\mathbb{E}[X|\mathcal{G}] = Y$ q.c. □

Teorema 11.50 (Lema de Fatou)**.** *Sejam $(X_n)_n$ uma sequência de variáveis aleatórias estendidas não-negativas e uma σ-álgebra $\mathcal{G} \subseteq \mathcal{F}$. Então $\mathbb{E}[\liminf_n X_n|\mathcal{G}] \leqslant \liminf_n \mathbb{E}[X_n|\mathcal{G}]$ q.c.*

A demonstração é análoga à do Teorema 5.68, trocando-se $\int_\Omega \cdot \, \mathrm{d}\mu$ por $\mathbb{E}[\cdot|\mathcal{G}]$.

Teorema 11.51 (Convergência Dominada)**.** *Sejam $(X_n)_n$, X e Y variáveis aleatórias estendidas e uma σ-álgebra $\mathcal{G} \subseteq \mathcal{F}$. Se $X_n \to X$ q.c. e $|X_n| < Y$ q.c. para alguma Y integrável, então $\mathbb{E}[X_n|\mathcal{G}] \to \mathbb{E}[X|\mathcal{G}]$ q.c.*

A demonstração é análoga à do Teorema 5.69, trocando-se $\int_\Omega \cdot \, \mathrm{d}\mu$ por $\mathbb{E}[\cdot|\mathcal{G}]$.

Teorema 11.52 (Desigualdade de Jensen)**.** *Sejam $\mathcal{G} \subseteq \mathcal{F}$ uma σ-álgebra, I um intervalo aberto, $g : I \to \mathbb{R}$ uma função convexa, e X uma variável aleatória que assume valores em I. Suponha que X e $g(X)$ sejam integráveis. Então*

$$g(\mathbb{E}[X|\mathcal{G}]) \leqslant \mathbb{E}[g(X)|\mathcal{G}].$$

Demonstração. A prova é análoga à demonstração do Teorema 6.23, porém há que se contornar algumas complicações técnicas. Para cada $z \in I$ fixo, existe $c = c(z) \in \mathbb{R}$ tal que $g(x) \geqslant g(z) + c(z) \cdot (x - z)$ para todo $x \in I$. Como a função que leva z em $c(z)$ é não-decrescente, também é mensurável.

Suponhamos inicialmente que essa função c seja limitada. Tomando $Z = \mathbb{E}[X|\mathcal{G}]$, e observando que Z e $g(Z)$ são $\mathcal{G}$-mensuráveis, pelo Teorema 11.45

$$\begin{aligned}\mathbb{E}[g(X)|\mathcal{G}] \geqslant \mathbb{E}\big[g(Z) + c(Z) \cdot (X - Z)|\mathcal{G}\big] = \\ = g(Z) + c(Z) \cdot \mathbb{E}[X|\mathcal{G}] - c(Z) \cdot Z = g(Z),\end{aligned}$$

onde a primeira igualdade acima é devida ao Teorema 11.45.

Consideremos agora o caso geral. Somando uma constante a X, podemos supor que $0 \in I$. Subtraindo $c(0) \cdot x$, podemos supor que $g(x) \geqslant 0$ para todo $x \in I$. Tome $[a_n, b_n] \uparrow I$ com $a_n < 0 < b_n$. Para cada n fixo, definimos $g_n(x) = g(x)$ para $x \in [a_n, b_n]$, $g_n(x) = g(b_n) + c(b_n) \cdot (x - b_n)$ para $x \geqslant b_n$ e $g_n(x) = g(a_n) + c(a_n) \cdot (x - a_n)$ para $x \leqslant a_n$. Observe que as funções g_n são convexas, não-negativas, e satisfazem $g_n \uparrow g$. Ademais, suas respectivas c_n são limitadas a $[c(a_n), c(b_n)]$, donde $\mathbb{E}[g_n(X)|\mathcal{G}] \geqslant g_n(\mathbb{E}[X|\mathcal{G}])$. Como $g_n(\mathbb{E}[X|\mathcal{G}]) \uparrow g(\mathbb{E}[X|\mathcal{G}])$, pelo Teorema da Convergência Monótona, $\mathbb{E}[g_n(X)|\mathcal{G}] \uparrow \mathbb{E}[g(X)|\mathcal{G}]$, concluímos que $\mathbb{E}[g(X)|\mathcal{G}] \geqslant g(\mathbb{E}[X|\mathcal{G}])$. □

Teorema 11.53 (Contração em $\mathcal{L}^p$)**.** *Se $p \geqslant 1$ e $|X|^p$ é integrável, então*

$$\mathbb{E}\big[\, |\mathbb{E}[X|\mathcal{G}]|^p \,\big] \leqslant \mathbb{E}\big[\, |X|^p \,\big].$$

Demonstração. Primeiro, X é integrável pois $|x| \leqslant 1 + |x|^p$ para todo $x \in \mathbb{R}$. Pela desigualdade de Jensen, $|\mathbb{E}[X|\mathcal{G}]|^p \leqslant \mathbb{E}[|X|^p|\mathcal{G}]$ pois $g(x) = |x|^p$ é convexa. Tomando esperança iterada, segue a desigualdade desejada. □

11.5 Teorema de Radon-Nikodým

Esta seção é de natureza um pouco mais abstrata, onde mostraremos a existência da esperança condicional de uma variável aleatória dada uma σ-álgebra. Para isto faremos uso de uma das mais importantes ferramentas da Teoria da Medida, e que desempenha um papel fundamental na Teoria da Probabilidade, o Teorema de Radon-Nikodým, que será enunciado aqui e demonstrado no Apêndice D.5.

Sejam μ um medida definida em $(\Omega, \mathcal{F})$ e $f : \Omega \to [0, +\infty]$ uma função mensurável. Vimos na Seção 5.5.5 que

$$\nu(A) = \int_A f \, \mathrm{d}\mu$$

define uma nova medida em $(\Omega, \mathcal{F})$. Além disso, pela Proposição 5.66, $\nu(A) = 0$ para todo $A \in \mathcal{F}$ tal que $\mu(A) = 0$.

A próxima definição será crucial no papel de comparar medidas.

Definição 11.54 (Medidas absolutamente contínuas)**.** Sejam ν e μ medidas definidas em um mesmo espaço mensurável $(\Omega, \mathcal{F})$. Dizemos que a medida ν *é absolutamente contínua com respeito à medida* μ, o que denotamos por $\nu \ll \mu$, se $\nu(A) = 0$ para todo $A \in \mathcal{F}$ tal que $\mu(A) = 0$.

Gostaríamos de saber se a recíproca seria verdadeira, isto é, se $\nu \ll \mu$, então a medida ν poderia ser expressa como a integral de Lebesgue de alguma função $f : \Omega \to [0, +\infty]$ mensurável?

O próximo teorema nos fornece a resposta. O Teorema de Radon-Nikodým nos diz quando uma medida ν pode ser expressa em termos de uma outra medida μ, ponderada por uma função, que denotaremos por $\frac{\mathrm{d}\nu}{\mathrm{d}\mu}$ e chamaremos de derivada de Radon-Nikodým de ν com respeito a μ, conforme introduzido na Seção 5.5.5.

Teorema 11.55 (Teorema de Radon-Nikodým)**.** *Sejam ν e μ medidas σ-finitas definidas em um mesmo espaço mensurável $(\Omega, \mathcal{F})$. Então, $\nu \ll \mu$ se e somente se existe uma função mensurável $f : \Omega \to [0, +\infty)$ tal que*

$$\nu(A) = \int_A f \, \mathrm{d}\mu, \ \forall A \in \mathcal{F}.$$

A prova será dada no Apêndice D.5.

Antes de falar de esperança condicional, vejamos como o Teorema de Radon-Nikodým justifica algumas afirmações feitas no Capítulo 3.

Na Seção 3.5, demos a entender que, se X é uma variável aleatória que não tem densidade, então existe um conjunto $B \in \mathcal{B}$ tal que $m(B) = 0$ e $\mathbb{P}(X \in B) > 0$. Naquele ponto não tínhamos as ferramentas para provar essa propriedade, mas agora vemos que isso decorre do Teorema de Radon-Nikodým. Com efeito, a inexistência de tal conjunto B é equivalente a $\mathbb{P}_X \ll m$, o que por sua vez é equivalente à existência de uma densidade $\frac{\mathrm{d}\mathbb{P}_X}{\mathrm{d}m}$.

No final da Seção 3.4, dissemos que, dada uma variável X com densidade f_X, e dado um evento $A \in \mathcal{F}$ tal que $\mathbb{P}(A) > 0$, sempre existe a densidade condicional $f_{X|A}$. Isso pode ser demonstrado usando-se o Teorema de Radon-Nikodým. Com efeito, se X é absolutamente contínua, então para todo $B \in \mathcal{B}$ tal que $m(B) = 0$, temos $\mathbb{P}(X \in B|A) \leqslant \frac{\mathbb{P}(X \in B)}{\mathbb{P}(A)} = 0$, logo X também será absolutamente contínua sob a medida $\mathbb{P}(\,\cdot\,|A)$, o que implica a existência de $f_{X|A}$ tal que $\mathbb{P}(X \in B|A) = \int_B f_{X|A}\,\mathrm{d}x$ para todo $B \in \mathcal{B}$.

Concluímos esta seção com a prova do Teorema 11.39.

Demonstração do Teorema 11.39. Consideremos inicialmente o caso em que X é não-negativa e integrável. Definimos a medida ν no espaço mensurável $(\Omega, \mathcal{G})$ como

$$\nu(A) = \int_A X\,\mathrm{d}\mathbb{P} \quad \text{para cada } A \in \mathcal{G}.$$

Como $\nu(\Omega) = \mathbb{E}X < \infty$, temos que ν é σ-finita. No espaço de probabilidade $(\Omega, \mathcal{G}, \mathbb{P})$, temos que $\nu \ll \mathbb{P}$ (aqui cometemos um pequeno abuso de notação ao denotar também por $\mathbb{P}$ a restrição de $\mathbb{P}$ à σ-álgebra $\mathcal{G}$). Pelo Teorema de Radon-Nikodým, a derivada $\frac{\mathrm{d}\nu}{\mathrm{d}\mathbb{P}} : \Omega \to [0, +\infty)$ é $\mathcal{G}$-mensurável. Tome $Z = \frac{\mathrm{d}\nu}{\mathrm{d}\mathbb{P}}$. Como Z é $\mathcal{G}$-mensurável e satisfaz (11.40), isso conclui a prova.

Supomos agora que X seja integrável. Pelo caso anterior, existem variáveis aleatórias integráveis $\mathbb{E}[X^+|\mathcal{G}]$ e $\mathbb{E}[X^-|\mathcal{G}]$ tais que $\int_A \mathbb{E}[X^\pm|\mathcal{G}]\,\mathrm{d}\mathbb{P} = \int_A X^\pm\,\mathrm{d}\mathbb{P}$ para todo $A \in \mathcal{G}$. Definindo $Z = \mathbb{E}[X^+|\mathcal{G}] - \mathbb{E}[X^-|\mathcal{G}]$, obtemos,

$$\begin{aligned}\int_A Z\,\mathrm{d}\mathbb{P} &= \int_A \mathbb{E}[X^+|\mathcal{G}]\,\mathrm{d}\mathbb{P} - \int_A \mathbb{E}[X^-|\mathcal{G}]\,\mathrm{d}\mathbb{P} \\ &= \int_A X^+\,\mathrm{d}\mathbb{P} - \int_A X^-\,\mathrm{d}\mathbb{P} = \int_A X\,\mathrm{d}\mathbb{P}\end{aligned}$$

para todo $A \in \mathcal{G}$, o que está bem definido pois todas as variáveis aleatórias envolvidas são integráveis. Portanto Z cumpre as duas condições do teorema.

Por último, supomos que X seja não-negativa mas não necessariamente integrável. Tomamos uma sequência $0 \leqslant X_n \uparrow X$ onde X_n são variáveis aleatórias simples não-negativas. Pelo caso anterior, existe $\mathbb{E}[X_n|\mathcal{G}]$ para

todo n. Como na demonstração do Teorema 11.49, $\mathbb{E}[X_n|\mathcal{G}] \uparrow Z$ para alguma variável aleatória estendida Z que é $\mathcal{G}$-mensurável e satisfaz $\int_A X \mathrm{d}\mathbb{P} = \int_A Z \mathrm{d}\mathbb{P}$, para todo $A \in \mathcal{G}$. Ou seja, Z cumpre as duas condições do teorema, o que conclui esta demonstração. $\square$

11.6 Distribuição condicional regular

Na Seção 11.4 definimos $\mathbb{E}[X|Y]$ para quaisquer variáveis aleatórias X e Y com X integrável, mas não dissemos como calculá-la. A Seção 11.2 se restringe ao caso em que ambas as variáveis são discretas, enquanto a Seção 11.3 descreve o caso de variáveis com densidade conjunta porém sem fornecer demonstrações rigorosas das propriedades enunciadas. O objetivo agora é dar significado à noção de distribuição condicional de X dado Y no caso geral, unificando a abordagem das seções anteriores.

Definição 11.56. Sejam X e Y variáveis aleatórias definidas em um mesmo espaço de probabilidade. Uma *distribuição condicional regular de X dado Y* é qualquer função de $\mathcal{B} \times \mathbb{R}$ em $[0,1]$, que a cada $B \in \mathcal{B}$ e $y \in \mathbb{R}$ associa um número, denotado $\mathbb{P}_{X|Y}(B|y)$, satisfazendo:

(1) Para todo $y \in \mathbb{R}$ fixo, a função que leva $B \in \mathcal{B}$ em $\mathbb{P}_{X|Y}(B|y)$ é uma medida de probabilidade em $(\mathbb{R}, \mathcal{B})$;
(2) Para todo $B \in \mathcal{B}$ fixo, a função que leva $y \in \mathbb{R}$ em $\mathbb{P}_{X|Y}(B|y)$ é uma função mensurável;
(3) Para todos $B, C \in \mathcal{B}$, vale $\mathbb{P}(X \in B, Y \in C) = \int_C \mathbb{P}_{X|Y}(B|y)\, \mathbb{P}_Y(\mathrm{d}y)$.

Antes de prosseguir, é oportuno fazer algumas observações.

Primeiro, caso X e Y sejam discretas, a equação acima se reduz a (11.21), portanto a definição de distribuição condicional regular generaliza (11.19).

Segundo, quando Y é absolutamente contínua, a equação acima se reduz a (11.28) pela regra da cadeia. Se X e Y têm densidade conjunta, podemos deduzir, a partir de (11.29), que vale (11.28) para todos $B, C \in \mathcal{B}$, e portanto a definição (11.27) resulta em uma distribuição condicional regular.

Por último, a definição acima pode parecer bastante abstrata, e talvez inútil, pois mesmo sabendo que sempre existe uma distribuição condicional regular, isso não diz como encontrá-la. Mencionamos de passagem que uma forma explícita de se obter uma distribuição condicional regular seria a seguinte. Primeiro calculamos

$$F_{X|Y}(x\,|\,y) = \lim_{z\to x^+}\lim_{n\to\infty} \mathbb{P}\Big(X \leqslant z \,\Big|\, Y \in [y-\tfrac{1}{n}, y+\tfrac{1}{n}]\Big) \tag{11.57}$$

para os valores $y \in \mathbb{R}$ onde a expressão acima está bem definida e resulta em uma função de distribuição na variável x.[18] Depois definimos $\mathbb{P}_{X|Y}(\cdot|y)$ como sendo a única medida tal que $\mathbb{P}_{X|Y}\big((-\infty, x]\,\big|\,y\big) = F_{X|Y}(x|y)$ para todo x. Entretanto, essa forma não é a mais recomendada, nem do ponto de vista teórico nem do prático. Isso porque a fórmula (11.57) está na forma "diferencial" já que o limite em n nos dá uma "derivada" na variável y, enquanto a fórmula no item (3) está na forma integral, o que é bem mais robusto. Na prática, é melhor encontrar um candidato *ad hoc* para a distribuição condicional regular e verificar que ele satisfaz à definição (o limite acima pode nos ajudar a adivinhar quem deveria ser o candidato). Foi exatamente o que fizemos no parágrafo anterior, para justificar a fórmula (11.26)!

Os dois próximos teoremas serão demonstrados no Apêndice D.6.

Teorema 11.58. *Dadas duas variáveis aleatórias quaisquer X e Y, sempre existe uma distribuição condicional regular de X dado Y.*

Vamos definir $\mathbb{E}[X|Y=y]$ a partir de $\mathbb{P}_{X|Y}$, e depois usá-la para construir $\mathbb{E}[X|Y]$ explicitamente de forma a satisfazer às duas propriedades do Teorema 11.39. Ademais, o faremos de forma que seja coerente com as fórmulas (11.15) e (11.34).

[18] É importante tomar o limite primeiro em n e depois em z, caso contrário $F_{X|Y}(\cdot|y)$ pode não ser uma função de distribuição para nenhum $y \in \mathbb{R}$, como podemos ver tomando $X = Y \sim \mathcal{N}(0,1)$, caso em que teríamos "$F_{X|Y}(z|z) = \frac{1}{2}$" para todo $z \in \mathbb{R}$.

Teorema 11.59. *Sejam X e Y variáveis aleatórias e $\mathbb{P}_{X|Y}$ uma distribuição condicional regular. Então, para toda função mensurável $g : \mathbb{R}^2 \to [0, +\infty]$, vale*

$$\mathbb{E}[g(X,Y)] = \int_{\mathbb{R}} \Big(\int_{\mathbb{R}} g(x,y)\, \mathbb{P}_{X|Y}(\mathrm{d}x|y) \Big) \mathbb{P}_Y(\mathrm{d}y), \tag{11.60}$$

sendo que a integral interna fornece uma função mensurável de y.

Agora vamos supor que X é integrável ou não-negativa. Observamos que $\int_{\mathbb{R}} x\, \mathbb{P}_{X|Y}(\mathrm{d}x|y)$ está definida para $\mathbb{P}_Y$-quase todo y.[19] Definimos então

$$\mathbb{E}[X|Y=y] = \int_{\mathbb{R}} x\, \mathbb{P}_{X|Y}(\mathrm{d}x|y) \tag{11.61}$$

nos pontos y para os quais a integral está definida, e $\mathbb{E}[X|Y=y] = 0$ caso contrário. Pela regra da cadeia, a fórmula acima se reduz a (11.34) caso X e Y tenham densidade conjunta, ou (11.15) caso sejam discretas.

Definimos $\mathbb{E}[X|Y]$ como a variável aleatória que assume o valor $\mathbb{E}[X \mid Y = y]$ no evento $\{Y = y\}$, como havíamos feito nas Seções 11.2 e 11.3.

Proposição 11.62. *Se X é integrável ou não-negativa, então a variável aleatória $\mathbb{E}[X|Y]$, como definida acima, satisfaz às duas propriedades do Teorema 11.39.*

Demonstração. Seja $A \in \sigma(Y)$. Por definição, $A = \{Y \in C\}$ para algum $C \in \mathcal{B}$. Tomando $g(x,y) = x^+ \cdot \mathbb{1}_C(y)$, pelo Teorema 11.59,

$$\mathbb{E}[X^+ \mathbb{1}_A] = \int_{\mathbb{R}} \Big(\int_{\mathbb{R}} x^+\, \mathbb{P}_{X|Y}(\mathrm{d}x|y) \Big) \mathbb{1}_C(y)\, \mathbb{P}_Y(\mathrm{d}y).$$

Procedendo de modo idêntico com a parte negativa e observando que a

[19] Se X é não-negativa, $\mathbb{E}X^- = 0$, donde $\int_{\mathbb{R}} x^-\, \mathbb{P}_{X|Y}(\mathrm{d}x|y) = 0$ para $\mathbb{P}_Y$-quase todo y. Se X é integrável, $\mathbb{E}|X| < \infty$, donde $\int_{\mathbb{R}} |x|\, \mathbb{P}_{X|Y}(\mathrm{d}x|y) < \infty$ para $\mathbb{P}_Y$-quase todo y. Em ambos casos, $\int_{\mathbb{R}} x\, \mathbb{P}_{X|Y}(\mathrm{d}x|y)$ está definida para $\mathbb{P}_Y$-quase todo y.

diferença está bem definida para $\mathbb{P}_Y$-quase todo y, obtemos

$$\begin{aligned}\mathbb{E}[X\mathbb{1}_A] &= \int_{\mathbb{R}} \mathbb{1}_C(y)\,\mathbb{E}[X|Y=y]\,\mathbb{P}_Y(\mathrm{d}y)\\ &= \mathbb{E}\big[\mathbb{1}_C(Y)\cdot\mathbb{E}[X|Y]\big]\\ &= \mathbb{E}[\mathbb{E}[X|Y]\mathbb{1}_A],\end{aligned}$$

o que conclui a prova. □

Portanto, $\mathbb{E}[X|Y]$ goza de todas as propriedades enunciadas na Seção 11.4. Em particular, $\mathbb{E}X = \mathbb{E}\big[\mathbb{E}[X|Y]\big]$ e com isso justificamos também (11.37).

Veremos como se apresenta a distribuição condicional regular de X dado Y em alguns casos especiais, além dos casos quando ambas são discretas ou possuem densidade conjunta, que vimos nas Seções 11.2 e 11.3.

Caso em que Y é discreta

O caso em que Y é uma variável aleatória discreta generaliza a abordagem da Seção 11.1. Neste caso, não precisamos da teoria de distribuição condicional regular, e somos obrigados a tomar literalmente

$$\mathbb{P}_{X|Y}(B|y) = \frac{\mathbb{P}(X\in B, Y=y)}{\mathbb{P}(Y=y)}$$

para todo y tal que $\mathbb{P}(Y=y)>0$. Os valores y tais que $\mathbb{P}(Y=y)=0$ são irrelevantes, e para ter uma definição completa podemos tomar, por exemplo, $\mathbb{P}_{X|Y}(B|y) = \mathbb{P}_X(B)$.

Verifiquemos as condições da Definição 11.56. A condição (1) vale trivialmente. Defina $D = \{s : \mathbb{P}_Y(s) > 0\}$ é observe que D é enumerável. A condição (2) vale pois, para cada $B\in\mathcal{B}$ fixo, podemos expressar $\mathbb{P}_{X|Y}(B|y)$ como soma enumerável de funções mensuráveis $\sum_{s\in D}\mathbb{P}_{X|Y}(B|s)\mathbb{1}_{\{s\}}(y) + \mathbb{P}_X(B)\mathbb{1}_{D^c}(y)$. Já a condição (3) vale pois $\mathbb{P}(X\in B, Y\in C) = \sum_{y\in C}\mathbb{P}(X\in B, Y=y) = \sum_{y\in C}\mathbb{P}_{X|Y}(B|y)p_Y(y) = \int_C \mathbb{P}_{X|Y}(B|y)\mathbb{P}_Y(\mathrm{d}y)$.

Caso em que X e Y são independentes

Se X e Y são independentes, esse é o caso mais simples, pois o conhecimento de Y não afeta a variável X. Neste caso, podemos tomar

$$\mathbb{P}_{X|Y}(B|y) = \mathbb{P}_X(B).$$

Verifiquemos a Definição 11.56. As condições (1) e (2) valem trivialmente. A condição (3) vale pois $\int_C \mathbb{P}_{X|Y}(B|y)\mathbb{P}_Y(\mathrm{d}y) = \int_C \mathbb{P}_X(B)\mathbb{P}_Y(\mathrm{d}y) = \mathbb{P}_X(B)\int_C \mathbb{P}_Y(\mathrm{d}y) = \mathbb{P}(X \in B)\mathbb{P}(Y \in C) = \mathbb{P}(X \in B, Y \in C)$.

Caso de variável discreta com parâmetro contínuo desconhecido

Suponha que Y seja discreta, X seja absolutamente contínua, e que uma distribuição condicional regular $\mathbb{P}_{Y|X}$ seja conhecida. Seja $p_{Y|X}(y|x)$ uma função de probabilidade condicional associada.

Neste caso, uma distribuição condicional regular de X dado Y tem densidade dada por

$$f_{X|Y}(x|y) = \frac{p_{Y|X}(y|x)}{p_Y(y)} \cdot f_X(x)$$

se $p_Y(y) > 0$, e $f_{X|Y}(x|y) = f_X(x)$ caso contrário.

Mais precisamente, definimos $\mathbb{P}_{X|Y}(B|y) = \int_B f_{X|Y}(x|y)\,\mathrm{d}x$. Para verificar a condição (2), definimos o conjunto enumerável $D = \{s : p_Y(s) > 0\}$, e observamos que, para cada $B \in \mathcal{B}$ fixo, podemos expressar $\mathbb{P}_{X|Y}(B|y)$ como soma enumerável das funções mensuráveis

$$\sum_{s \in D} \frac{\int_B p_{Y|X}(y|x) f_X(x)\,\mathrm{d}x}{p_Y(y)} \mathbb{1}_{\{s\}}(y) + \mathbb{1}_{D^c}(y) \cdot \int_B f_X(x)\,\mathrm{d}x.$$

O numerador acima é uma função mensurável de y pelo Lema 5.82, pois é dado pela integral em x de uma função mensurável de x e y. Para verificar a condição (1), note que $\mathbb{P}_{X|Y}(B|y)$ é não-negativa por definição, e é σ-aditiva

em B como consequência da σ-aditividade da integral. Ademais,

$$\mathbb{P}_{X|Y}(\mathbb{R}|y) = \int_{\mathbb{R}} \frac{p_{Y|X}(y|x) f_X(x)}{p_Y(y)} \, \mathrm{d}x = \frac{\int_{\mathbb{R}} \mathbb{P}_{Y|X}(\{y\}|x)\mathbb{P}_X(\mathrm{d}x)}{p_Y(y)} = 1$$

se $p_Y(y) > 0$, e $\mathbb{P}_{X|Y}(\mathbb{R}|y) = \int_{\mathbb{R}} f_X(x)\,\mathrm{d}x = 1$ caso contrário. Finalmente, para verificar condição (3) desenvolvemos

$$\begin{aligned}
\int_C \mathbb{P}_{X|Y}(B|y)\mathbb{P}_Y(\mathrm{d}y) &= \sum_{y\in C} \Big(\int_B \frac{p_{Y|X}(y|x)}{p_Y(y)} \cdot f_X(x)\,\mathrm{d}x \Big) p_Y(y) \\
&= \sum_{y\in C} \int_B p_{Y|X}(y|x) \cdot f_X(x)\,\mathrm{d}x \\
&= \sum_{y\in C} \int_B \mathbb{P}_{Y|X}(\{y\} \mid x)\mathbb{P}_X(\mathrm{d}x) \\
&= \sum_{y\in C} \mathbb{P}(Y = y, X \in B) \\
&= \mathbb{P}(Y \in C, X \in B).
\end{aligned}$$

Exemplo 11.63 (Ensaios de Bernoulli com parâmetro dado por uma Beta). Sejam X e Y, variáveis aleatórias tais que $X \sim \mathrm{Beta}(a, b)$ e a distribuição condicional de Y dado que $X = x$ é $\mathrm{Binom}(n, x)$. Neste caso,

$$f_{X|Y}(x|y) = \frac{\binom{n}{y} x^y (1-x)^{n-y}}{p_Y(y)} f_X(x) = \frac{x^{y+a-1}(1-x)^{n-y+b-1}}{c(a,b,n,y)}$$

para todo $y = 0, \ldots, n$. Observamos também que $c = \int_0^1 x^{y+a-1}(1-x)^{n-y+b-1}\mathrm{d}x$, pois $f_{X|Y}(\cdot|y)$ é uma função de densidade. Portanto, a distribuição condicional de X dado que $Y = y$ é uma distribuição Beta de parâmetros $a + y$ e $b + (n - y)$. $\triangle$

Caso em que $\mathbb{P}_{X|Y}$ é especificado

Os Exemplos 11.18, 11.20, 11.32 e 11.33 ilustraram o caso em que $\mathbb{P}_{X|Y}$ é especificado, juntamente com $\mathbb{P}_Y$. Tal especificação deve satisfazer às condições (1) e (2), enquanto equação em (3) serve para determinar a distribuição conjunta $\mathbb{P}_{X,Y}$, cuja marginal serve para determinar $\mathbb{P}_X$, e a integral (11.61) serve para calcular $\mathbb{E}[X|Y]$. O seguinte exemplo não se enquadra nos contextos das Seções 11.2 ou 11.3.

Exemplo 11.64. Seja $Y \sim \mathcal{U}[0,1]$. Se $Y = y$, então uma moeda com probabilidade y de sair cara é lançada n vezes independentemente. Seja X a variável aleatória que representa o número de caras obtidas.

A distribuição condicional de X dado que $Y = y$ é $\text{Binom}(n, y)$. Portanto, $\mathbb{E}[X \mid Y = y] = ny$, ou seja, $\mathbb{E}[X \mid Y] = nY$, logo

$$\mathbb{E}\big[\mathbb{E}[X|Y]\big] = \mathbb{E}[nY] = \frac{n}{2}. \qquad \triangle$$

11.7 Exercícios

§11.1

1. Seja X uma variável aleatória simples definida em $(\Omega, \mathcal{F}, \mathbb{P})$ e $\mathcal{D}$ uma partição finita de Ω. Definimos a *variância condicional de X dada a partição $\mathcal{D}$* como:

$$\mathbb{V}[X|\mathcal{D}] = \mathbb{E}\big[(X - \mathbb{E}[X|\mathcal{D}])^2|\mathcal{D}\big].$$

Mostre que

$$\mathbb{V}[X|\mathcal{D}] = \mathbb{E}[X^2|\mathcal{D}] - (\mathbb{E}[X|\mathcal{D}])^2$$

e

$$\mathbb{V}X = \mathbb{E}\big[\mathbb{V}[X|\mathcal{D}]\big] + \mathbb{V}\big[\mathbb{E}[X|\mathcal{D}]\big].$$

2. Sejam X e Y variáveis aleatórias simples definidas em $(\Omega, \mathcal{F}, \mathbb{P})$ e $\mathcal{D}$ uma

partição. Mostre que

$$\mathbb{E}[\, X \cdot \mathbb{E}[Y|\mathcal{D}]\,] = \mathbb{E}[\, Y \cdot \mathbb{E}[X|\mathcal{D}]\,].$$

3. Sejam X e Y variáveis aleatórias simples definidas em $(\Omega, \mathcal{F}, \mathbb{P})$ e $\mathcal{D}$ uma partição finita de Ω. Mostre que, se

$$\mathbb{E}[Y^2 \,|\, \mathcal{D}] = X^2 \quad \text{e} \quad \mathbb{E}[Y|\mathcal{D}] = X,$$

então $\mathbb{P}(X = Y) = 1$. Dica: Desenvolva $\mathbb{E}(X - Y)^2$.

§11.2

4. Prove que, se $X_1, \ldots, X_m$ são simples e i.i.d., e N toma valores em $\{0, \ldots, m\}$ e é independente de $X_1, \ldots, X_m$, então

$$\mathbb{V}S_N = \mathbb{E}N \cdot \mathbb{V}X_1 + (\mathbb{E}X_1)^2 \mathbb{V}N.$$

5. Sejam X e Y variáveis aleatórias simples e i.i.d. Mostre que

$$\mathbb{E}[X|X+Y] = \mathbb{E}[Y|X+Y] = \frac{X+Y}{2}.$$

6. Dê um exemplo de variáveis aleatórias simples X e Y que não são independentes mas mesmo assim $\mathbb{E}[X|Y] = \mathbb{E}X$.

7. Dadas duas variáveis aleatórias simples X e Y, e $g : \mathbb{R} \to \mathbb{R}$, mostre que $\mathbb{E}[g(Y)\mathbb{E}[X|Y]] = \mathbb{E}[Xg(Y)]$.

§11.3

8. Um número não-negativo Y é escolhido com densidade $f_Y(y) = ye^{-y}$ para $y > 0$. Se $Y = y$, um número X é escolhido uniformemente no intervalo $[0, y]$.

(a) Descreva $f_{X|Y}$.
(b) Encontre $f_{X,Y}$.
(c) Encontre f_X.

(d) Encontre $f_{Y|X}$.
(e) Calcule $\mathbb{P}(X + Y \leqslant 2)$.

9. Para X e Y definidas no Exemplo 11.30, calcule $\mathbb{E}[X|Y]$.

10. Seja (X, Y) um vetor aleatório com densidade conjunta dada por

$$f_{X,Y}(x, y) = \begin{cases} \frac{e^{-y}}{y}, & \text{se } 0 < x < y, \\ 0, & \text{caso contrário.} \end{cases}$$

Calcule $\mathbb{E}[X^2|Y = y]$.

§11.4

11. Refaça os exercícios de §11.1 acima, supondo que X e Y têm segundo momento finito e usando uma σ-álgebra $\mathcal{G}$ no lugar de uma partição $\mathcal{D}$.

12.

(a) Suponha que $(X, Y) \sim (Y, X)$ e que X seja integrável. Mostre que

$$\mathbb{E}[X|X + Y] = \mathbb{E}[Y|X + Y].$$

(b) Sejam $X_1, \ldots, X_n$ i.i.d. e integráveis, e $S_n = \sum_{k=1}^{n} X_k$. Mostre que

$$\mathbb{E}[X_1|S_n] = \frac{S_n}{n}.$$

Dica: Expresse as esperanças como integrais de Lebesgue em $\mathbb{R}^2$ ou $\mathbb{R}^n$.

13. Mostre que, se X é integrável, então $\mathbb{E}[|\mathbb{E}[X|\mathcal{G}]|] \leqslant \mathbb{E}|X|$.

14. Sejam X e Y variáveis aleatórias com segundo momento finito. Mostre que $\mathbf{Cov}(X, \mathbb{E}[Y|X]) = \mathbf{Cov}(X, Y)$.

15. Sejam $X_1, X_2, \ldots$ variáveis aleatórias i.i.d. positivas. Suponha que X_1 e $\frac{1}{X_1}$ são integráveis, e defina $S_n = X_1 + \cdots + X_n$.

(a) Mostre que $\frac{1}{S_n}$ é integrável.

(b) Calcule $\mathbb{E}[\frac{X_1}{S_n}]$.
(c) Calcule $\mathbb{E}[\frac{S_m}{S_n}]$ para $1 \leqslant m \leqslant n$.

§11.5

16. Sejam λ, ν e μ medidas σ-finitas definidas no mesmo espaço mensurável $(\Omega, \mathcal{F})$. Mostre que

(a) Se $\lambda \ll \mu$ e $\nu \ll \mu$, então $(\lambda + \nu) \ll \mu$ e

$$\frac{\mathrm{d}(\lambda + \nu)}{\mathrm{d}\mu} = \frac{\mathrm{d}\lambda}{\mathrm{d}\mu} + \frac{\mathrm{d}\nu}{\mathrm{d}\mu}.$$

(b) Se $\nu \ll \mu$ e $c > 0$, então $(c\nu) \ll \mu$ e

$$\frac{\mathrm{d}(c\nu)}{\mathrm{d}\mu} = c\frac{\mathrm{d}\nu}{\mathrm{d}\mu}.$$

(c) Se $\lambda \ll \nu$ e $\nu \ll \mu$, então $\lambda \ll \mu$ e

$$\frac{\mathrm{d}\lambda}{\mathrm{d}\mu} = \frac{\mathrm{d}\lambda}{\mathrm{d}\nu} \cdot \frac{\mathrm{d}\nu}{\mathrm{d}\mu}.$$

17. Sejam μ e ν medidas definidas no mesmo espaço mensurável $(\Omega, \mathcal{F})$.

(a) Suponha que, para todo $\varepsilon > 0$, existe $\delta > 0$ tal que $\nu(A) < \varepsilon$ para todo $A \in \mathcal{F}$ com $\mu(A) < \delta$. Mostre que $\nu \ll \mu$.
(b) Mostre que a recíproca também é válida se $\nu(\Omega) < \infty$.
Dica: Suponha que a propriedade ε-δ não é válida, tome uma sequência adequada $(A_k)_k$ de subconjuntos, defina $B_n = \cup_{k \geqslant n} A_k$ e $A = \cap_n B_n$ para mostrar $\nu \not\ll \mu$.
(c) Mostre que a recíproca pode ser falsa sem a hipótese de que $\nu(\Omega) < \infty$.
Sugestão: Considere $\Omega = \mathbb{N}$, $\nu = \sum_n \delta_n$ e $\mu = \sum_n n^{-2}\delta_n$.

§11.6

18. Rita lança uma moeda, cuja probabilidade de sair cara é p, indefinidamente até obter a primeira cara; após isto ela irá receber uma quantia cuja

distribuição é exponencial com parâmetro igual ao número de lançamentos da moeda. Calcule a esperança da quantia recebida por Rita.

19. Joga-se um dado, depois uma moeda, depois o dado novamente e segue-se alternando entre o dado e a moeda. Quando se obtém cara na moeda, o jogo é imediatamente interrompido e conta-se o total Z de pontos obtidos nos lançamentos do dado. Calcule $\mathbb{E}Z$.

20. O número de passageiros que chegam ao ponto do ônibus 409 durante o intervalo de tempo $[0, t]$ tem distribuição de Poisson com parâmetro λt. O tempo T de chegada do próximo ônibus tem distribuição exponencial com parâmetro α. Mas precisamente, a distribuição condicional do número N de passageiros que chegam ao ponto antes do próximo ônibus dado que $T = t$ é Poisson(λt). Mostre que $N + 1 \sim \text{Geom}(\frac{\alpha}{\lambda+\alpha})$.

21. Suponha que $\mathbb{P}_{X|Y}(B|y) = \mathbb{P}_X(B)$ para todo $B \in \mathcal{B}$ e todo $y \in \mathbb{R}$. Mostre que X e Y são independentes.

22 (Princípio da substituição)**.** Seja $\mathbb{P}_{X|Y}$ uma distribuição condicional regular e $f : \mathbb{R}^2 \to \mathbb{R}$ uma função mensurável. Defina $Z = f(X, Y)$. Mostre que

$$\mathbb{P}_{Z|Y}(B|y) = \mathbb{P}_{X|Y}(\{x : f(x, y) \in B\}|y)$$

é uma distribuição condicional regular de Z dado Y.

Dica: Use $g(x, y) = \mathbb{1}_B(f(x, y)) \cdot \mathbb{1}_C(y)$.

23 (Princípio da substituição)**.** Seja $\mathbb{P}_{X|Y}$ uma distribuição condicional regular e $f : \mathbb{R}^2 \to [0, +\infty]$ uma função mensurável. Mostre que

$$\mathbb{E}[f(X, Y)|Y] = \int_{\mathbb{R}} f(x, Y)\, \mathbb{P}_{X|Y}(\mathrm{d}x|Y)$$

quase certamente. Dica: Use $g(x, y) = f(x, y) \cdot \mathbb{1}_C(y)$.

24. Sejam X e Y variáveis aleatórias independentes. Prove que

$$\mathbb{P}(X \leqslant Y) = \mathbb{E}[F_X(Y)].$$

Dica: Use o exercício anterior e esperança iterada.

25. Sejam X e Y variáveis aleatórias, onde X é discreta e Y tem função de densidade $f_Y(y)$. Seja $f_{Y|X}(y|x)$ uma densidade condicional regular de Y dado X, isto é, $f_{Y|X} : \mathbb{R}^2 \to \mathbb{R}$ é mensurável, e $\mathbb{P}_{Y|X}(B|x) = \int_B f_{Y|X}(y|x)\mathrm{d}y$ define uma distribuição condicional regular. Defina a função

$$p_{X|Y}(x|y) = \frac{f_{Y|X}(y|x)}{f_Y(y)} \cdot p_X(x)$$

para todo y tal que $f_Y(y) > 0$, e $p_{X|Y}(x|y) = p_X(x)$ se $f_Y(y) = 0$. Mostre que $\mathbb{P}_{X|Y}$, definida como $\mathbb{P}_{X|Y}(B|y) = \sum_{x \in B} p_{X|Y}(x|y)$, resulta em uma distribuição condicional regular.

Capítulo 12

Martingales

Uma pergunta que os autores ouvem frequentemente é: qual a melhor estratégia que posso adotar em um cassino? A resposta curta é: não apostar. O estudo de martingales explica, entre muitas outras coisas, como qualquer receita para lucrar em cassinos é, na verdade, uma receita para ir à falência.[20]

Considere um apostador que participa de um determinado jogo, e que a cada rodada ele vence com probabilidade $p \in [0, \frac{1}{2}]$. Em caso de vitória, ele ganha o montante apostado e, em caso de derrota, ele perde a mesma quantia. Se X_n representa o resultado obtido ($+1$ para vitória ou -1 para derrota) e C_n representa o valor apostado na n-ésima rodada, então $S_n = S_{n-1} + C_n X_n$ é o capital acumulado ao final da n-ésima rodada. Aqui C_n deve ser uma função determinística de $X_1, \ldots, X_{n-1}$ e representa a estratégia desse apostador. Efetivamente, o apostador tem todo o direito de utilizar a informação passada para decidir se e quanto vai apostar, mas deve decidi-lo sem saber qual será o resultado da próxima rodada. Analogamente, um investidor decide comprar uma ação sem saber se seu preço vai subir ou cair.

Observe que $S_{n+1} - S_n$ não precisa ser independente de $S_0, \ldots, S_n$. Sequências cujos incrementos não são independentes surgem naturalmente em várias situações, tanto teóricas quanto práticas. Alguns casos especiais têm sua

própria teoria, includindo as cadeias de Markov, que não são abordadas nesse livro, os processos estacionários, estudados no Capítulo 14, e os martingales, que são tema deste capítulo.

12.1 Definições e exemplos

Em um espaço de probabilidade $(\Omega, \mathcal{F}, \mathbb{P})$, uma *filtração* é uma sequência de σ-álgebras $(\mathcal{F}_n)_{n\in\mathbb{N}_0}$ contidas em $\mathcal{F}$ tais que $\mathcal{F}_n \subseteq \mathcal{F}_{n+1}$ para todo $n \in \mathbb{N}_0$. Uma sequência crescente de σ-álgebras representa um experimento em que tem-se acesso a mais e mais informação à medida que passa o tempo, representado por n. Lembramos que a ideia de "informação" é representada pela classe de eventos cuja ocorrência é acessível a um determinado observador. Por exemplo, imagine que uma moeda será lançada muitas vezes, e acabamos de observar o resultado do primeiro lançamento. Neste caso, particionamos Ω em dois eventos, e nossa informação até este momento é codificada pela σ-álgebra formada pelos quatro eventos: cara, coroa, o evento certo, e o evento impossível. Após o lançamento da segunda moeda, devemos particionar Ω em quatro eventos, e a σ-álgebra que codifica a informação observada terá $2^4 = 16$ eventos. Após o lançamento da terceira moeda, a σ-álgebra terá 2^{2^3} eventos.

Um *processo estocástico* é uma família indexada de variáveis aleatórias, o índice pode representar uma posição no espaço ou, mais frequentemente, um instante de tempo. Dizemos que o processo $(Z_n)_{n\in\mathbb{N}_0}$ é *adaptado à filtração* $(\mathcal{F}_n)_{n\in\mathbb{N}_0}$ se, para todo $n \in \mathbb{N}_0$, Z_n é $\mathcal{F}_n$-mensurável.

Em muitos casos, vamos querer que Z_{n+1} seja independente de $\mathcal{F}_n$, e será útil a seguinte definição. A *filtração natural do processo* $(Z_n)_{n\in\mathbb{N}_0}$ é a filtração

[20] Uma conhecida exceção é a do jogo de *blackjack* com poucos maços de baralho e regras de pagamento generosas (ambos raros hoje em dia), em que, após um período de lucro médio negativo, um apostador extremamente habilidoso poderia haver adquirido suficiente informação sobre as cartas restantes no baralho para começar a ter lucro médio ligeiramente positivo e, caso não fosse expulso em poucos minutos, poderia teoricamente compensar o período inicial e esperar algum lucro resultante positivo no final.

dada por $\mathcal{F}_n = \sigma(Z_0, Z_1, \ldots, Z_n)$. Em palavras, no instante n temos a informação dada por $Z_0, Z_1, \ldots, Z_n$ e nada mais. Essa é a menor filtração à qual o processo $(Z_n)_n$ é adaptado.

Observação 12.1. Se $Z_0, Z_1, Z_2, \ldots$ são independentes e $(\mathcal{F}_n)_{n\in\mathbb{N}_0}$ é a filtração natural dessa sequência, então Z_{n+1} é independente de $\mathcal{F}_n$ para todo $n \in \mathbb{N}_0$. Esse fato, intuitivamente óbvio, será formalizado no Corolário 13.10. $\triangle$

Todas as definições e enunciados deste capítulo pressupõem que há um espaço de probabilidade e uma filtração subjacentes, mesmo que não se lhes faça referência explícita. Em alguns exemplos pediremos explicitamente que a filtração seja aquela natural de algum processo, como forma de assegurar independência de certos eventos ocorridos no futuro.

Definição 12.2 (Martingale)**.** Dizemos que um processo adaptado $(M_n)_{n\in\mathbb{N}_0}$ é um *martingale* se, para todo $n \in \mathbb{N}_0$, M_n é integrável e

$$\mathbb{E}[M_{n+1}|\mathcal{F}_n] = M_n \tag{12.3}$$

quase certamente.

Note que a definição acima pressupõe um espaço $(\Omega, \mathcal{F}, \mathbb{P})$ e uma filtração $(\mathcal{F}_n)_n$ fixos, como dito logo antes. Quando for necessário especificar a filtração, diremos que $(M_n)_n$ é um *martingale com respeito à filtração* $(\mathcal{F}_n)_n$.

Da equação (12.3) surge a interpretação de que um martingale reflete os ganhos obtidos em um jogo justo: a esperança condicional do capital após a próxima rodada, dada a informação obtida até o presente momento, é igual ao capital acumulado até agora.

Exemplo 12.4. Seja $(X_n)_n$ uma sequência de variáveis aleatórias integráveis e independentes tais que $\mathbb{E}X_n = 0$. Considere a filtração natural desse processo. Defina

$$S_n = X_1 + \cdots + X_n.$$

Então $(S_n)_n$ é um martingale (subentende-se que o é com respeito à única filtração mencionada neste exemplo: a filtração natural de $(X_n)_n$). Veja que

S_n é $\mathcal{F}_n$-mensurável pois é dado pela soma de variáveis $\mathcal{F}_n$-mensuráveis, S_n é integrável pois é soma de variáveis integráveis, restando mostrar (12.3). Utilizando as propriedades da esperança condicional com respeito a σ-álgebras vistas no capítulo anterior, obtemos

$$\mathbb{E}[S_{n+1}|\mathcal{F}_n] = \mathbb{E}[X_{n+1}|\mathcal{F}_n] + \mathbb{E}[S_n|\mathcal{F}_n] = \mathbb{E}[X_{n+1}] + S_n = S_n.$$

As igualdades acima valem porque a esperança condicional é linear, S_n é $\mathcal{F}_n$-mensurável, e X_{n+1} é independente de $\mathcal{F}_n$. $\triangle$

No exemplo acima, assim como em várias passagens deste capítulo, usamos implicitamente a Observação 12.1.

Exemplo 12.5. Sejam $(X_n)_n$ variáveis aleatórias integráveis e independentes tais que $\mathbb{E}X_n = 1$ e considere a sua filtração natural. Defina

$$M_n = X_1 \times \cdots \times X_n.$$

Então $(M_n)_n$ é um martingale. Novamente, M_n é $\mathcal{F}_n$-mensurável pois é produto de variáveis $\mathcal{F}_n$-mensuráveis, e é integrável pois é produto de variáveis independentes e integráveis. Para mostrar (12.3), veja que

$$\mathbb{E}[M_{n+1}|\mathcal{F}_n] = \mathbb{E}[X_{n+1}M_n|\mathcal{F}_n] = M_n\mathbb{E}[X_{n+1}|\mathcal{F}_n] = M_n\mathbb{E}[X_{n+1}] = M_n,$$

onde as segunda igualdade vale porque M_n é $\mathcal{F}_n$-mensurável e a terceira porque X_{n+1} é independente de $\mathcal{F}_n$. $\triangle$

Exemplo 12.6. Sejam X uma variável aleatória integrável e $(\mathcal{F}_n)_n$ uma filtração qualquer. Defina $X_n = \mathbb{E}[X|\mathcal{F}_n]$ para todo $n \in \mathbb{N}$. Nesse caso, X_n é integrável e $\mathcal{F}_n$-mensurável pela definição de esperança condicional. Para verificar (12.3), veja que

$$\mathbb{E}[X_{n+1}|\mathcal{F}_n] = \mathbb{E}\Big[\mathbb{E}[X|\mathcal{F}_{n+1}]\Big|\mathcal{F}_n\Big] = \mathbb{E}[X|\mathcal{F}_n] = X_n.$$

A segunda igualdade acima segue da Esperança Condicional Iterada, observando que $\mathcal{F}_n \subseteq \mathcal{F}_{n+1}$. Portanto, $(X_n)_n$ é um martingale. $\triangle$

Assim como podemos interpretar um martingale como um jogo justo, a definição abaixo nos diz o que seria o análogo de um jogo favorável ou desfavorável ao apostador.

Definição 12.7 (Submartingale, Supermartingale)**.** Dizemos que um processo $(M_n)_n$ é um *submartingale* se $(M_n)_n$ é adaptado e, para todo $n \in \mathbb{N}_0$, M_n é integrável e

$$\mathbb{E}[M_{n+1}|\mathcal{F}_n] \geqslant M_n \tag{12.8}$$

quase certamente. Novamente, deixamos o espaço de probabilidade e a filtração implícitos. Dizemos que $(M_n)_n$ é um *supermartingale* se $(-M_n)_n$ é um submartingale, ou seja, $(M_n)_n$ é adaptado e vale a desigualdade oposta em (12.8). Observe que um martingale é um caso particular de submartingale.

Exemplo 12.9 (O passeio aleatório em $\mathbb{Z}$). Sejam $k \in \mathbb{Z}$ e $(X_n)_n$ uma sequência de variáveis aleatórias i.i.d. com distribuição comum $\mathbb{P}(X_n = +1) = 1 - \mathbb{P}(X_n = -1) = p$, e considere sua filtração natural. Defina $S_n = k + \sum_{j=1}^n X_j$ para todo $n \in \mathbb{N}_0$. A sequência $(S_n)_n$ é chamada de *passeio aleatório em* $\mathbb{Z}$. Assim como procedemos no Exemplo 12.4, S_n é $\mathcal{F}_n$-mensurável, integrável, e

$$\mathbb{E}[S_{n+1}|\mathcal{F}_n] = S_n + \mathbb{E}X_{n+1} = S_n + 2p - 1.$$

Portanto, $(S_n)_n$ é um martingale no caso simétrico $p = \frac{1}{2}$, um submartingale se $p \geqslant \frac{1}{2}$ e um supermartingale se $p \leqslant \frac{1}{2}$. $\triangle$

Observe que o exemplo acima é um caso particular do jogo mencionado no início deste capítulo, onde o valor apostado é igual a um em todas as rodadas.

Usaremos inúmeras vezes que $\mathbb{E}M_n \geqslant \mathbb{E}M_{n-1}$ se $(M_n)_n$ é um submartingale, $\mathbb{E}M_n \leqslant \mathbb{E}M_{n-1}$ se $(M_n)_n$ é um supermartingale, e $\mathbb{E}M_n = \mathbb{E}M_0$ se $(M_n)_n$ é um martingale. Para verificar que um processo adaptado $(M_n)_n$ é um submartingale, basta verificar que $\mathbb{E}[M_{n+1} - M_n|\mathcal{F}_n] \geqslant 0$ q.c., para todo n.

Salientamos que quase todos os enunciados que falam de submartingales têm seu análogo para supermartingales e vice-versa.

Proposição 12.10. *Sejam $(M_n)_n$ um martingale e $g : \mathbb{R} \to \mathbb{R}$ convexa tal que $g(M_n)$ é integrável para todo $n \in \mathbb{N}$. Então $(g(M_n))_n$ é um submartingale.*

Demonstração. Basta verificar (12.8), que segue de

$$\mathbb{E}[g(M_{n+1})|\mathcal{F}_n] \geqslant g(\mathbb{E}[M_{n+1}|\mathcal{F}_n]) = g(M_n),$$

onde utilizamos a Desigualdade de Jensen para Esperança Condicional. □

Exemplo 12.11. Se $(M_n)_n$, é um martingale, então os processos $(M_n^+)_n$, $(M_n^-)_n$, $(|M_n|)_n$ e $(M_n^2)_n$ são submartingales (com respeito à mesma filtração!). △

Proposição 12.12. *Sejam $(M_n)_n$ um submartingale e $g : \mathbb{R} \to \mathbb{R}$ convexa não-decrescente tal que $g(M_n)$ é integrável para todo $n \in \mathbb{N}$. Então $(g(M_n))_n$ é um submartingale.*

Demonstração. Como $\mathbb{E}[M_{n+1}|\mathcal{F}_n] \geqslant M_n$ e g é não-decrescente, obtemos

$$\mathbb{E}[g(M_{n+1})|\mathcal{F}_n] \geqslant g(\mathbb{E}[M_{n+1}|\mathcal{F}_n]) \geqslant g(M_n),$$

onde utilizamos a Desigualdade de Jensen para Esperança Condicional. □

Exemplo 12.13. Se $(M_n)_n$ é um submartingale e $a \in \mathbb{R}$, então $([M_n - a]^+)_n$ também o é, pois a função $g(x) = [x - a]^+$ é convexa e não-decrescente. △

12.2 Tempos de parada

Ao participar de rodadas sucessivas de um jogo de azar, ou mesmo ao investir dinheiro em um determinado ativo financeiro, surgem perguntas bastante naturais. Em que momento devemos parar de jogar? Em que momento devemos vender o ativo? A resposta óbvia seria quando obtivermos o maior lucro possível! A questão é que para sabermos que estamos passando por este momento, devemos ter informação acerca do futuro, pois se hoje é o

dia que obterei o maior lucro com meu ativo financeiro é porque amanhã ele se desvalorizará. Ou seja, há tempos aleatórios que somente conseguimos determinar se olharmos para o futuro. A seguinte definição caracteriza aqueles tempos aleatórios factíveis, isto é, aqueles que conseguimos determinar apenas com a informação contida até o presente.

Definição 12.14. Seja τ uma variável aleatória estendida que toma valores em $\mathbb{N}_0 \cup \{+\infty\}$. Dizemos que a variável τ é um *tempo de parada* com respeito à filtração $(\mathcal{F}_n)_{n\in\mathbb{N}_0}$ se $\{\tau \leqslant n\} \in \mathcal{F}_n$ para todo $n \in \mathbb{N}_0$.

Alguns livros definem tempo de parada pedindo que $\{\tau = n\} \in \mathcal{F}_n$ para todo $n \in \mathbb{N}_0$, o que é equivalente (verifique!). Se interpretamos que a σ-álgebra $\mathcal{F}_n$ carrega toda a informação disponível até o tempo n, temos que, τ ser um tempo de parada significa que o evento $\{\tau = n\}$ pode ser determinado com base na informação contida apenas até esse tempo n.

Exemplo 12.15. Sejam $(X_n)_n$ um processo adaptado e $B \in \mathcal{B}$. É frequente estarmos interessados no primeiro tempo de passagem pelo conjunto B (por exemplo, gostaríamos de vender determinado ativo financeiro quando ele atinge determinado preço pela primeira vez). Isto é, estamos interessados na variável aleatória

$$\tau_B = \inf\{n : X_n \in B\},$$

com a convenção $\inf \emptyset = +\infty$. Como

$$\{\tau_B \leqslant n\} = \cup_{k=0}^{n}\{X_k \in B\} \in \mathcal{F}_n$$

para todo $n \in \mathbb{N}_0$, segue que τ_B é um tempo de parada. $\triangle$

Dados um processo estocástico $(X_n)_{n\in\mathbb{N}_0}$ e um tempo aleatório τ assumindo valores em $\mathbb{N}_0$, definimos a variável X_τ que corresponde ao processo $(X_n)_n$ observado no tempo $n = \tau$. Definimos o *processo parado em* τ como $(X_{n\wedge\tau})_n$, onde $a \wedge b = \min\{a, b\}$, que geralmente representa um processo que para de evoluir quando uma determinada condição é cumprida.

Teorema 12.16 (Teorema do Martingale Parado)**.** *Sejam* $(X_n)_{n\in\mathbb{N}_0}$ *um*

submartingale e τ um tempo de parada, com respeito à mesma filtração. Então o processo parado $(X_{n\wedge\tau})_n$ também é um submartingale. Em particular, se $(X_n)_n$ é um martingale, então $(X_{n\wedge\tau})_n$ também é martingale.

Demonstração. Como

$$X_{n\wedge\tau} = X_n \cdot \mathbb{1}_{\{\tau\geqslant n\}} + \sum_{k=0}^{n-1} X_k \cdot \mathbb{1}_{\{\tau=k\}},$$

e todas as variáveis na equação acima são $\mathcal{F}_n$-mensuráveis, $X_{n\wedge\tau}$ é $\mathcal{F}_n$-mensurável, e como $|X_{n\wedge\tau}| \leqslant |X_1| + \cdots + |X_n|$, também é integrável. Para verificar (12.8), veja que

$$\begin{aligned}\mathbb{E}[X_{(n+1)\wedge\tau} - X_{n\wedge\tau}|\mathcal{F}_n] &= \mathbb{E}[(X_{n+1} - X_n) \cdot \mathbb{1}_{\{\tau>n\}}|\mathcal{F}_n] \\ &= \mathbb{1}_{\{\tau>n\}} \cdot \mathbb{E}[(X_{n+1} - X_n)|\mathcal{F}_n] \geqslant 0.\end{aligned}$$

Na segunda igualdade utilizamos o fato de que $\{\tau > n\} \in \mathcal{F}_n$, pois τ é tempo de parada. Na desigualdade usamos a propriedade de submartingale. □

Exemplo 12.17 (O segundo valete). Retiramos cartas do baralho ao acaso e sucessivamente, até que não reste nenhuma carta. Sabemos que, para cada $n = 1, \ldots, 52$, a probabilidade de que a n-ésima carta retirada seja um valete é igual a $\frac{1}{13}$. Uma propriedade talvez surpreendente é que a probabilidade de os dois primeiros valetes aparecerem juntos também é $\frac{1}{13}$!

Para justificar essa afirmação, definimos Y_n como a indicadora do evento em que a n-ésima carta retirada é um valete, e τ como o número de cartas retiradas até que saia um valete. Queremos mostrar que $\mathbb{E}[Y_{\tau+1}] = \frac{1}{13}$.

Tomamos $(\mathcal{F}_n)_n$ como a filtração natural de $(Y_n)_n$ e definimos X_n como a proporção de valetes restantes no baralho após a n-ésima retirada, isto é,

$$X_n = \frac{4 - Y_1 - \cdots - Y_n}{52 - n}.$$

Observe que, para cada n fixo, $\mathbb{E}[Y_{n+1}|\mathcal{F}_n] = X_n$, ou seja, a probabilidade

condicional de que a próxima carta seja um valete é igual à proporção de valetes restantes no baralho. Multiplicando por $\mathbb{1}_{\{\tau=n\}}$, somando sobre n e tomando a esperança, obtemos a identidade $\mathbb{E}[Y_{\tau+1}] = \mathbb{E}X_\tau$. Assim, queremos mostrar que $\mathbb{E}X_\tau = \frac{1}{13}$.

Afirmamos que $(X_n)_{n=0,\dots,49}$ é um martingale com respeito a $(\mathcal{F}_n)_n$, com $\mathcal{F}_0 = \{\emptyset, \Omega\}$ Com efeito,

$$X_{n+1} = \frac{(52-n)X_n - 1}{52-n-1} Y_{n+1} + \frac{(52-n)X_n}{52-n-1}(1 - Y_{n+1})$$

e, como X_n é $\mathcal{F}_n$-mensurável e $\mathbb{E}[Y_{n+1}|\mathcal{F}_n] = X_n$,

$$\mathbb{E}[X_{n+1}|\mathcal{F}_n] = \frac{(52-n)X_n - 1}{52-n-1} X_n + \frac{(52-n)X_n}{52-n-1}(1 - X_n) = X_n$$

e, portanto, $(X_n)_n$ é de fato um martingale. Pelo Teorema do Martingale Parado, $\mathbb{E}X_\tau = \mathbb{E}X_{49\wedge\tau} = \mathbb{E}X_0 = \frac{1}{13}$. △

Proposição 12.18. *Sejam $(X_n)_{n\in\mathbb{N}_0}$ um submartingale e τ um tempo de parada, com respeito à mesma filtração. Então*

$$\mathbb{E}[X_{n\wedge\tau}] \leqslant \mathbb{E}[X_n].$$

Demonstração. Basta expandir

$$\begin{aligned}
\mathbb{E}[X_n - X_{n\wedge\tau}] &= \mathbb{E}\Big[\sum_{k=\tau}^{n-1} X_{k+1} - X_k\Big] = \sum_{k=0}^{n-1} \mathbb{E}\big[(X_{k+1} - X_k)\mathbb{1}_{\{\tau\leqslant k\}}\big] \\
&= \sum_{k=0}^{n-1} \mathbb{E}\big[\mathbb{E}[(X_{k+1} - X_k)\mathbb{1}_{\{\tau\leqslant k\}}|\mathcal{F}_k]\big] \\
&= \sum_{k=0}^{n-1} \mathbb{E}\big[\mathbb{1}_{\{\tau\leqslant k\}} \cdot \mathbb{E}[X_{k+1} - X_k|\mathcal{F}_k]\big] \geqslant 0,
\end{aligned}$$

onde na desigualdade usamos a propriedade de submartingale. □

Essas propriedades de martingales parados permitem estender desigualdades envolvendo a posição final para desigualdades sobre toda a trajetória.

Teorema 12.19. *Seja $(S_n)_n$ um submartingale. Então, para cada $\lambda > 0$ valem*

$$\mathbb{P}\Big(\max_{j=1,\dots,n} S_j \geqslant \lambda\Big) \leqslant \frac{\mathbb{E}S_n^+}{\lambda}$$

e

$$\mathbb{P}\Big(\min_{j=1,\dots,n} S_j \leqslant -\lambda\Big) \leqslant \frac{\mathbb{E}S_n^+ - \mathbb{E}S_1}{\lambda}.$$

Demonstração. Seja $\tau = \inf\{j = 1, \dots, n : S_j \geqslant \lambda\} \wedge n$. Então τ é tempo de parada e $n \wedge \tau = \tau$. Pela Proposição 12.12, S_n^+ também é submartingale. Pela Proposição 12.18, $\mathbb{E}S_\tau^+ \leqslant \mathbb{E}S_n^+$, logo

$$\lambda\,\mathbb{P}(S_\tau \geqslant \lambda) \leqslant \mathbb{E}S_\tau^+ \leqslant \mathbb{E}S_n^+.$$

Analogamente, definindo $\tau = \inf\{j = 1, \dots, n : S_j \leqslant -\lambda\} \wedge n$, obtemos

$$\mathbb{E}S_1 \leqslant \mathbb{E}S_\tau = \mathbb{E}S_\tau^+ - \mathbb{E}S_\tau^- \leqslant \mathbb{E}S_n^+ - \lambda\,\mathbb{P}(S_\tau \leqslant -\lambda)$$

onde a primeira desigualdade é novamente dada pelo Teorema do Martingale Parado. □

Teorema 12.20 (Desigualdade de Doob-Kolmogorov)**.** *Seja $(S_n)_n$ um martingale. Então, para cada $\lambda > 0$ e $p \geqslant 1$ vale*

$$\mathbb{P}\Big(\max_{j=1,\dots,n} |S_j| \geqslant \lambda\Big) \leqslant \frac{\mathbb{E}|S_n|^p}{\lambda^p}.$$

Demonstração. Caso $\mathbb{E}|S_n|^p = +\infty$, a desigualdade vale trivialmente. Caso contrário, como a função $|\cdot|^p$ é convexa, $(|S_n|^p)_n$ é um submartingale e aplicamos o teorema anterior com λ^p no lugar de λ. □

Com algum esforço extra e usando a Desigualdade de Hölder (Apêndice D.7), é possível provar o seguinte.

Teorema 12.21 (Desigualdade de Doob em $\mathcal{L}^p$)**.** *Seja $(M_n)_n$ um martingale, ou um submartingale não-negativo. Então para todo $p > 1$ vale*

$$\mathbb{E}\Big[\max_{j=1,\dots,n} |M_j|^p\Big] \leqslant (\tfrac{p}{p-1})^p\, \mathbb{E}|M_n|^p.$$

A demonstração do teorema acima pode ser encontrada na Seção 14.11 de Williams (1991), e em Durrett (2019), Gut (2013) e Klenke (2014).

12.3 Amostragem opcional

Se $(X_n)_n$ é um martingale, então $\mathbb{E}X_n = \mathbb{E}X_0$ para todo $n \in \mathbb{N}$. Ou seja, se a sequência representa os ganhos em um jogo justo, o valor médio deve ser preservado ao longo das rodadas. Porém se τ é um tempo aleatório, pense por exemplo em τ como sendo o tempo em que o martingale atinge seu máximo, muito provavelmente $\mathbb{E}X_\tau \neq \mathbb{E}X_0$. Isso também é natural, pois neste caso precisamos de informação acerca do futuro para determinar quando estamos passando pelo tempo τ. Mas e se τ for um tempo de parada, será que poderíamos afirmar que $\mathbb{E}X_\tau = \mathbb{E}X_0$?

Exemplo 12.22 (O sistema de apostas chamado Martingale)*.* Seja $(Z_n)_n$ uma sequência de variáveis i.i.d. com $\mathbb{P}(Z_n = +1) = \mathbb{P}(Z_n = -1) = \frac{1}{2}$ e considere sua filtração natural. Tome $X_n = \sum_{k=1}^n 2^{k-1} Z_k$ para todo n. O martingale $(X_n)_n$ reflete o ganho acumulado de um jogador que aposta nos vermelhos, em um jogo de roleta justo,[21] e vai dobrando o valor apostado. Seja $\tau = \inf\{n : Z_n = +1\}$, isto é, τ é a rodada em que o jogador ganha pela primeira vez, recuperando todas as perdas anteriores e ainda tirando \$1 de lucro. Neste caso, $X_\tau = +1$ q.c. e $X_0 = 0$, logo $\mathbb{E}X_\tau \neq \mathbb{E}X_0$. Veja que essa estratégia exige que o jogador tenha um capital (ou crédito) infinito.[22] $\triangle$

Portanto, queremos saber sob quais hipóteses a propriedade de ser martingale ou supermartingale é preservada quando o processo é observado em um tempo

[21] Em um jogo de roleta justo, que não existe na prática, os vermelhos e pretos têm chance $\frac{1}{2}$, isto é, não existem as casas verdes 0 e 00.

de parada τ. O teorema a seguir nos fornece vários critérios que asseguram essa propriedade.

Teorema 12.23 (Teorema da Amostragem Opcional)**.** *Sejam $(X_n)_{n\in\mathbb{N}_0}$ um supermartingale e τ um tempo de parada finito. Então $\mathbb{E}X_\tau \leqslant \mathbb{E}X_0$ se pelo uma das condições abaixo for satisfeita para algum $K > 0$:*

(1) *$\tau \leqslant K$ q.c.;*
(2) *$|X_{n\wedge\tau}| \leqslant K$ q.c. para todo n;*
(3) *$\mathbb{E}\big[|X_{n+1} - X_n|\big|\mathcal{F}_n\big] \leqslant K$ q.c. para todo n e $\mathbb{E}\tau < \infty$;*
(4) *$X_n \geqslant 0$ q.c. para todo n.*

Em particular, se $(X_n)_n$ é um martingale e vale alguma das três primeiras condições acima, então $\mathbb{E}X_\tau = \mathbb{E}X_0$.

Demonstração. Usando o Teorema do Martingale Parado, $\mathbb{E}X_{n\wedge\tau} \leqslant \mathbb{E}X_0$. Supondo (1), basta tomar algum $n \geqslant K$ para ter $X_\tau = X_{n\wedge\tau}$. Para os outros casos, observe que, como τ é finito, $X_{n\wedge\tau} \to X_\tau$ q.c. Supondo (4), temos que $\mathbb{E}X_\tau \leqslant \liminf_n \mathbb{E}X_{n\wedge\tau} \leqslant \mathbb{E}X_0$ diretamente pelo Lema de Fatou.

Para (2) e (3), bastará mostrar que $\mathbb{E}X_{n\wedge\tau} \to \mathbb{E}X_\tau$. Usando o Teorema da Convergência Dominada, basta mostrar que $|X_{n\wedge\tau}| \leqslant Y$ com Y integrável. Para (2), tomamos $Y = K$ constante.

Para (3), definimos $Y_0 = |X_0|$ e $Y_n = Y_{n-1} + |X_{n\wedge\tau} - X_{(n-1)\wedge\tau}|$, observamos que $|X_{n\wedge\tau}| \leqslant Y_n \uparrow Y$, e estimamos

$$\begin{aligned}
\mathbb{E}Y_{n+1} - \mathbb{E}Y_n &= \mathbb{E}|X_{(n+1)\wedge\tau} - X_{n\wedge\tau}| \\
&= \mathbb{E}\big[\,|X_{n+1} - X_n| \cdot \mathbb{1}_{\{\tau>n\}}\big] \\
&= \mathbb{E}\Big[\mathbb{E}\big[\,|X_{n+1} - X_n| \cdot \mathbb{1}_{\{\tau>n\}}\big|\mathcal{F}_n\big]\Big] \\
&= \mathbb{E}\Big[\mathbb{1}_{\{\tau>n\}} \cdot \mathbb{E}\big[\,|X_{n+1} - X_n|\big|\mathcal{F}_n\big]\Big] \\
&\leqslant \mathbb{E}\big[K \cdot \mathbb{1}_{\{\tau>n\}}\big] = K \cdot \mathbb{P}(\tau > n).
\end{aligned}$$

[22]Há muitos relatos de apostadores que começaram com um grande capital e, depois de um longo período de lucro obtido com a aplicação sistemática dessa estratégia, terminaram perdendo todo o lucro e mais o capital que haviam levado consigo.

Somando em $n \in \mathbb{N}_0$, obtemos $\mathbb{E}Y \leqslant \mathbb{E}|X_0| + K\,\mathbb{E}\tau < \infty$. $\square$

Podemos ver que o Exemplo 12.22 viola cada uma das hipóteses acima:

- O tempo de parada τ não é limitado.
- O lucro (ou prejuízo!) acumulado $|X_{n\wedge\tau}|$ não é limitado.
- No evento $\{\tau > n\} \in \mathcal{F}_n$, temos que $|X_{n+1} - X_n| = 2^n$, portanto $\mathbb{E}\big[|X_{n+1} - X_n| \big| \mathcal{F}_n\big]$ não é limitado.
- O lucro acumulado X_n pode tomar valores negativos.

Exemplo 12.24 (Problema da ruína do apostador). Considere $(S_n)_{n\in\mathbb{N}_0}$ o passeio aleatório começando de $S_0 = x$. Seja $N \in \mathbb{N}$ fixo e suponha que $x \in \{0, \ldots, N\}$. Defina o tempo de parada $\tau = \inf\{n \in \mathbb{N}_0 : S_n = 0 \text{ ou } S_n = N\}$, e observe que $\tau < \infty$ q.c. (exercício!). Queremos calcular $\alpha(x) = \mathbb{P}(S_\tau = N)$, isto é, a probabilidade de que um passeio aleatório, começando em x, chegue a N antes de 0.

Podemos interpretar este exemplo como um jogador que dispõe de um capital inicial x e tem como adversário um outro jogador cujo capital inicial é de $N - x$. Em cada rodada o primeiro jogador ganha $+1$ com probabilidade p e -1 com probabilidade $1 - p$, até que um dos dois perca todo o capital.

Consideramos primeiro o caso $p = \frac{1}{2}$, em que $(S_n)_n$ é um martingale. Como $S_{n\wedge\tau}$ é limitado, pelo Teorema da Amostragem Opcional,

$$x = \mathbb{E}S_0 = \mathbb{E}S_\tau = 0 \cdot (1 - \alpha(x)) + N \cdot \alpha(x).$$

Portanto $\alpha(x) = \frac{x}{N}$ para todo $x \in \{0, \ldots, N\}$.

No caso $p \neq \frac{1}{2}$, observamos que a sequência definida por $M_n = (\frac{1-p}{p})^{S_n}$ é um martingale com respeito à mesma filtração. Tirar um martingale do chapéu é uma técnica muito comum em Probabilidade! Usando-se novamente Teorema da Amostragem Opcional, obtemos

$$(\tfrac{1-p}{p})^x = \mathbb{E}S_0 = \mathbb{E}S_\tau = (\tfrac{1-p}{p})^0 \cdot (1 - \alpha(x)) + (\tfrac{1-p}{p})^N \cdot \alpha(x),$$

e, resolvendo,

$$\alpha(x) = \frac{1 - (\frac{1-p}{p})^x}{1 - (\frac{1-p}{p})^N},$$

para todo $x \in \{0, \ldots, N\}$. △

O teorema abaixo é uma interessante aplicação do Teorema da Amostragem Opcional e generaliza o Exemplo 11.24, que diz que sob certas condições a média da soma de uma quantidade aleatória de parcelas aleatórias é a média das parcelas vezes o número médio de parcelas.

Teorema 12.25 (Identidade de Wald)**.** *Seja $(X_n)_{n\in\mathbb{N}_0}$ uma sequência de variáveis aleatórias independentes, integráveis e com a mesma esperança. Seja τ um tempo de parada integrável com respeito à filtração natural de $(X_n)_n$. Então*

$$\mathbb{E}[X_1 + \cdots + X_\tau] = \mathbb{E}\tau \cdot \mathbb{E}X_1.$$

Demonstração. Definindo $S_n = X_1 + \cdots + X_n - n\mathbb{E}X_1$, temos que $(S_n)_n$ é martingale e satisfaz

$$\mathbb{E}\Big[|S_{n+1} - S_n|\Big|\mathcal{F}_n\Big] \leqslant \mathbb{E}|X_{n+1} - \mathbb{E}X_1| \leqslant 2\mathbb{E}|X_1| < \infty.$$

Pelo Teorema da Amostragem Opcional, $\mathbb{E}S_\tau = \mathbb{E}S_0 = 0$, portanto

$$\mathbb{E}[X_1 + \cdots + X_\tau - \tau \cdot \mathbb{E}X_1] = 0,$$

o que conclui a prova. □

Exemplo 12.26 (Tempo médio de retorno do passeio aleatório)*.* No passeio aleatório simétrico, definia o tempo de parada $\tau = \inf\{n : S_n = 1\}$. Vamos mostrar que $\mathbb{E}\tau = +\infty$. Sabe-se que $\mathbb{P}(\tau < \infty) = 1$ (será mostrado na próxima seção, e se fosse falso teríamos $\mathbb{E}\tau = +\infty$ trivialmente). Se τ fosse integrável, teríamos $\mathbb{E}S_\tau = \mathbb{E}X_1 \cdot \mathbb{E}\tau = 0$ pela identidade de Wald. Como $S_\tau = 1$ q.c., podemos concluir que $\mathbb{E}\tau = +\infty$. △

12.4 Convergência quase certa de martingales

Nesta seção veremos que supermartingales cujo primeiro momento é limitado têm que convergir quase certamente. A razão por trás disso é a seguinte. Para que o processo não convirja, deve haver uma faixa $[a, b]$ tal que ele assuma valores abaixo de a e acima de b infinitas vezes. Pensemos que um supermartingale reflete o preço de um ativo que, na média, não sobe. Se o preço desse ativo atravessasse a faixa $[a, b]$ muitas vezes, um investidor poderia acumular lucro comprando uma ação cada vez que o preço está abaixo de a e vendendo essa ação cada vez que está acima de b. Entretanto, como essa estratégia também resulta em um supermartingale, esse investidor ainda tem lucro médio negativo, o que implica que o preço desse ativo deve ser cada vez mais disperso, no sentido de que $\mathbb{E}|X_n| \to +\infty$. Portanto, exceto nesse caso, o processo tem que convergir.

Teorema 12.27 (Teorema de Convergência de Martingales)**.** *Seja $(X_n)_n$ um supermartingale. Se $\sup_n \mathbb{E}|X_n| < \infty$, então existe uma variável aleatória integrável X_∞ tal que $X_n \to X_\infty$ quase certamente. Em particular, todo supermartingale não-negativo converge quase certamente.*

Demonstração. A maior parte da prova consiste em mostrar que

$$\mathbb{P}(X_n \leqslant a \text{ i.v. e } X_n \geqslant b \text{ i.v.}) = 0 \tag{12.28}$$

para todo par de números reais $a < b$.

Sejam $a < b$ números reais. Analisemos a estratégia de comprar em baixa e vender em alta. Para isso, definimos os tempos de parada:

$$\begin{aligned} \tau_1 &= \min\{n \geqslant 0 : X_n \leqslant a\}, & \tau_2 &= \min\{n > \tau_1 : X_n \geqslant b\} \\ \tau_{2k+1} &= \min\{n > \tau_{2k} : X_n \leqslant a\}, & \tau_{2k+2} &= \min\{n > \tau_{2k+1} : X_n \geqslant b\}, \\ &\vdots & &\vdots \end{aligned}$$

com a convenção de que $\min \emptyset = +\infty$. Uma interpretação é que os tempos τ_k

indicam os momentos de compra do ativo se k for ímpar ou venda se for par. Definimos

$$Z_n = \begin{cases} 1, & \text{se } \tau_k < n \leqslant \tau_{k+1} \text{ e } k \text{ é ímpar}, \\ 0, & \text{se } \tau_k < n \leqslant \tau_{k+1} \text{ e } k \text{ é par}, \end{cases}$$

com $\tau_0 = 0$. Observe que Z_n é $\mathcal{F}_{n-1}$-mensurável, pois

$$\{Z_n = 1\} = \bigcup_{k \text{ ímpar}} (\{\tau_k \leqslant n-1\} \cap \{\tau_{k+1} > n-1\}).$$

Essa variável indica se o investidor possui uma ação do ativo $(X_n)_n$ logo antes do instante n. O processo

$$W_n = \sum_{j=1}^{n} Z_j \cdot (X_j - X_{j-1})$$

descreve a variação do capital entre os instantes 0 e n. Definimos, para cada

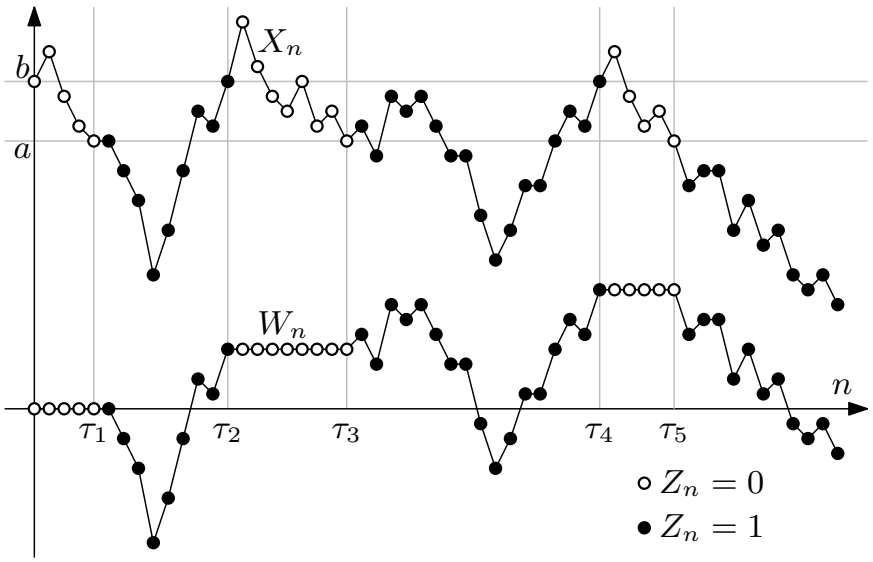


Figura 12.1. Argumento de travessias ascendentes completas.

$m \in \mathbb{N}$,

$$T_m = \max\{k : \tau_{2k} \leqslant m\},$$

que conta o numero de *travessias ascendentes completas* do processo $(X_n)_n$ sobre o intervalo $[a, b]$ até o tempo m. Observe que

$$W_n \geqslant (b-a)\, T_n - (X_n - a)^-,$$

pois $(b-a)\, T_n$ é uma cota inferior para o lucro obtido com as vendas em alta e $[X_n - a]^-$ é uma cota superior para a perda acumulada desde a última compra, veja Figura 12.1.

Por outro lado, $(W_n)_n$ também é supermartingale. Com efeito,

$$\begin{aligned}\mathbb{E}\big[W_{n+1} - W_n | \mathcal{F}_n\big] &= \mathbb{E}\big[Z_{n+1}(X_{n+1} - X_n) | \mathcal{F}_n\big] = \\ &= Z_{n+1} \cdot \mathbb{E}\big[X_{n+1} - X_n | \mathcal{F}_n\big] \leqslant 0,\end{aligned}$$

pois $Z_{n+1} \geqslant 0$ e $\mathbb{E}[X_{n+1} - X_n | \mathcal{F}_n] \leqslant 0$ q.c., já que $(X_n)_n$ é um supermartingale. Logo, $\mathbb{E}W_n \leqslant \mathbb{E}W_0 = 0$, donde concluímos que

$$\mathbb{E}T_n \leqslant \frac{\mathbb{E}(X_n - a)^-}{b-a}.$$

Agora definimos $T_\infty = \lim_n T_n$, que conta as travessias ascendentes completas do processo $(X_n)_n$ sobre o intervalo $[a, b]$, sem limite de tempo. Pelo Teorema da Convergência Monótona,

$$\mathbb{E}(T_\infty) = \lim_n \mathbb{E}T_n \leqslant \sup_n \frac{\mathbb{E}(X_n - a)^-}{b-a} \leqslant \sup_n \frac{\mathbb{E}|X_n| + |a|}{b-a},$$

que é finito por hipótese. Sendo integrável, T_∞ é finito quase certamente, o que prova (12.28).

Mostremos agora que $(X_n)_n$ converge quase certamente. No evento $\{\liminf_n X_n < \limsup_n X_n\}$, existem $a < b \in \mathbb{Q}$ tais que $X_n \leqslant a$ i.v. e

$X_n \geqslant b$ i.v. Como $\mathbb{Q}^2$ é enumerável, por subaditividade obtemos

$$\mathbb{P}(\liminf_n X_n < \limsup_n X_n) \leqslant \sum_{a,b\in\mathbb{Q}} \mathbb{P}(X_n \leqslant a \text{ i.v. e } X_n \geqslant b \text{ i.v.}) = 0.$$

Portanto, $X_n \overset{\text{q.c.}}{\to} X_\infty$ para alguma variável aleatória estendida X_∞. Pelo Lema de Fatou, $\mathbb{E}|X_\infty| \leqslant \liminf_n \mathbb{E}|X_n| \leqslant \sup_n \mathbb{E}|X_n| < \infty$. Ou seja, $|X_\infty|$ é integrável, logo quase certamente finito. $\square$

Exemplo 12.29 (Processo de ramificação). Seja $(X_{n,k})_{n,k\in\mathbb{N}}$ uma sequência duplamente indexada de variáveis aleatórias i.i.d. assumindo valores em $\mathbb{N}_0$. Definimos a sequência de variáveis aleatórias $(Z_n)_n$, por

$$Z_0 = 1 \quad \text{e} \quad Z_n = \sum_{k=1}^{Z_{n-1}} X_{n,k}, \quad \text{para todo } n \in \mathbb{N}.$$

Podemos interpretar o processo $(Z_n)_n$ como um modelo de crescimento populacional: cada indivíduo dá origem à próxima geração e logo desaparece, sendo que a prole gerada por cada indivíduo é independente dos demais e com a mesma distribuição. Deste modo, Z_n é o número de indivíduos da n-ésima geração. Gostaríamos de determinar se essa população se extingue ou não e, caso não se extingua, como cresce.

Para estudar o crescimento dessa população, vamos supor que as $X_{n,k}$ têm segundo momento finito, e denotar sua média por $\lambda = \mathbb{E}X_{1,1}$, a variância por $\sigma^2 = \mathbb{V}X_{1,1}$ e $p_j = \mathbb{P}(X_{1,1} = j)$. Defina $\mathcal{F}_n = \sigma((X_{j,k})_{k\in\mathbb{N},j=1,\dots,n})$ e

$$M_n = \lambda^{-n} Z_n$$

para todo $n \in \mathbb{N}$.

Afirmamos que $(M_n)_n$ é um martingale com respeito a $(\mathcal{F}_n)_n$. Com efeito, Z_n é mensurável com respeito a $\mathcal{F}_n$ e

$$\mathbb{E}[Z_{n+1}|\mathcal{F}_n] = \sum_{k\in\mathbb{N}} \mathbb{E}[X_{n+1,k}\mathbb{1}_{\{Z_n\geqslant k\}}|\mathcal{F}_n] = \sum_{k\in\mathbb{N}} \lambda\mathbb{1}_{\{Z_n\geqslant k\}} = \lambda Z_n,$$

onde na primeira igualdade usamos que as variáveis envolvidas são não-negativas e na segunda usamos que $\{Z_n \geqslant k\} \in \mathcal{F}_n$ e que $X_{n+1,k}$ é independente de $\mathcal{F}_n$. A independência entre $X_{n+1,k}$ e $\mathcal{F}_n$ é bastante óbvia, uma justificativa rigorosa será dada pelo Corolário 13.11.

Tomando a esperança na equação acima, Z_{n+1} é integrável, pois $\mathbb{E}Z_{n+1} = \lambda\,\mathbb{E}Z_n = \cdots = \lambda^{n+1} < \infty$. Substituindo, obtemos $\mathbb{E}[M_{n+1}|\mathcal{F}_n] = \lambda^{-n-1}\lambda Z_n = M_n$, provando que $(M_n)_n$ é um martingale.

Passemos então a estudar o comportamento de $(Z_n)_n$ em termos de λ.

O caso $0 \leqslant \lambda < 1$ é o mais fácil. Como $\mathbb{P}(Z_n > 0) = \mathbb{P}(Z_n \geqslant 1) \leqslant \mathbb{E}Z_n = \lambda^n$, temos que $Z_n \overset{\text{q.c.}}{\to} 0$ quando $n \to \infty$ pelo Lema de Borel-Cantelli. Ou seja, a população se extinguirá quase certamente.

Consideremos agora o caso $\lambda = 1$. Se $p_1 = 1$, então o processo $(Z_n)_n$ se torna trivial com $Z_n = 1$ para todo $n \in \mathbb{N}$ q.c. Supomos então que $p_1 < 1$, que neste caso implica $p_0 > 0$. Como $(Z_n)_n$ é um martingale não-negativo, pelo Teorema de Convergência de Martingales, existe uma variável aleatória Z_∞, tal que $Z_n \overset{\text{q.c.}}{\to} Z_\infty$. Como Z_n é sempre um número inteiro, $Z_n = Z_\infty$ para todo n grande, quase certamente. Por outro lado, como $p_0 > 0$, temos $\mathbb{P}(Z_n = k,\ \text{para todo } n \text{ grande}) = 0$ para todo $k \in \mathbb{N}$, pois $\mathbb{P}(Z_n = Z_{n+1} = \cdots = Z_{n+j} = k) \leqslant (1 - {p_0}^k)^j$ para todo j. Portanto, $Z_\infty = 0$ q.c., e a população se extinguirá nesse caso.

O caso $\lambda > 1$ exige mais ferramentas. Pelo Teorema de Convergência de Martingales, $M_n \overset{\text{q.c.}}{\to} M_\infty$ para alguma variável aleatória M_∞. Se $p_0 > 0$, a população tem chance de se extinguir logo nas primeiras gerações. Por outro lado, no evento $\{M_\infty > 0\}$, a população não apenas sobrevive como também cresce exponencialmente rápido, pois $Z_n \geqslant \frac{1}{2}\lambda^n M_\infty$ para todo n suficientemente grande. Logo, gostaríamos de mostrar que $\mathbb{P}(M_\infty > 0) > 0$, o que será feito na próxima seção. $\triangle$

Exemplo 12.30 (Recorrência do passeio aleatório simétrico). Sejam x e $y \in \mathbb{Z}$. Considere $(S_n)_n$ o passeio aleatório simétrico começando de $S_0 = x$ e defina

o tempo de parada $\tau = \inf\{n \geqslant 0 : S_n = y\}$. Afirmamos que

$$\mathbb{P}(\tau < \infty) = 1.$$

Podemos supor sem perda de generalidade que $y = 0$ e $x \geqslant 0$ (caso contrário consideramos $S_n - y$ ou $-S_n$ no lugar de S_n). Pelo Teorema do Martingale Parado, $(S_{n\wedge\tau})_n$ é um martingale. Como $S_{n\wedge\tau} \geqslant 0$, pelo Teorema de Convergência de Martingales, $(S_{n\wedge\tau})_n$ converge quase certamente. Porém, $(S_n)_n$ sempre dá saltos de ± 1, logo não pode convergir. Logo, a única forma de $(S_{n\wedge\tau})_n$ convergir é que $\tau < \infty$, donde concluímos que este evento é quase certo. Observamos que esta prova da recorrência tem a particularidade de não usar estimativas quantitativas de nenhum tipo. $\triangle$

O exemplo acima é talvez o tipo mais comum de aplicação desta teoria: uma vez que identificamos um martingale, observamos que ele deve convergir quase certamente e tiramos conclusões a partir disso. Abaixo damos exemplos de martingales e submartingales que não convergem.

Exemplo 12.31. Um exemplo trivial que ilustra a necessidade da hipótese $\sup_n \mathbb{E}|X_n| < \infty$ é o submartingale determinístico dado por $X_n = n$ para todo $n \in \mathbb{N}$, que obviamente não satisfaz tal hipótese e não converge. $\triangle$

Exemplo 12.32. Seja $(X_n)_n$ uma sequência independente tal que $\mathbb{P}(X_n = 4^n) = \mathbb{P}(X_n = -4^n) = \frac{1}{2}$, e tome $S_n = X_1 + \cdots + X_n$. Então $(S_n)_n$ é um martingale, $\limsup_n S_n = +\infty$ q.c. e $\liminf_n S_n = -\infty$ q.c. $\triangle$

Exemplo 12.33. Seja $(X_n)_n$ uma sequência independente tal que $\mathbb{P}(X_n = 1) = \frac{1}{2}$, $\mathbb{P}(X_n = -2^{n-1}) = 2^{-n}$ e $\mathbb{P}(X_n = 0) = \frac{1}{2} - 2^{-n}$. Tome $S_n = X_1 + \cdots + X_n$. Então $(S_n)_n$ é um martingale e $S_n \to +\infty$ q.c. $\triangle$

Exemplo 12.34. O passeio aleatório simétrico $(S_n)_n$ não converge, pois sempre dá saltos de ± 1. Podemos concluir pelo Teorema de Convergência de Martingales que $\sup_n \mathbb{E}|S_n| = +\infty$. Ademais, como $|S_n|$ é um submartingale, $\mathbb{E}|S_n|$ é não-decrescente em n, logo $\mathbb{E}|S_n| \to +\infty$. $\triangle$

12.5 Convergência de martingales em $\mathcal{L}^p$

Na seção anterior, estudamos condições para que um submartingale convirja quase certamente, o que não necessariamente implica convergência dos respectivos momentos, conforme podemos observar nos exemplos abaixo.

Exemplo 12.35. Seja $(X_n)_n$ uma sequência de variáveis aleatórias i.i.d. com $\mathbb{P}(X_n = 0) = \mathbb{P}(X_n = 2) = \frac{1}{2}$, e considere sua filtração natural. Defina $M_n = X_1 \times \cdots \times X_n$. Então $(M_n)_n$ é um martingale não-negativo e $M_n \to 0$ quase certamente. Observe que o martingale $(M_n)_n$ não converge em $\mathcal{L}^1$ pois $\mathbb{E}[\lim_n M_n] \neq \lim_n \mathbb{E}M_n$. △

O exemplo acima é uma reformulação do Exemplo 12.22. Há outros exemplos do mesmo fenômeno. No processo de ramificação crítico ($\lambda = 1$), vimos que $(Z_n)_n$ é um martingale com $\mathbb{E}Z_n = \mathbb{E}Z_0 = 1$, mas $Z_n \overset{\text{q.c.}}{\to} 0$. Também vimos no Exemplo 12.30 que o passeio aleatório simétrico começando em x e parado em y é um martingale que satisfaz $\mathbb{E}S_{n\wedge\tau} = x$ mas $S_{n\wedge\tau} \overset{\text{q.c.}}{\to} y \neq x$.

Gostaríamos de saber quando a esperança é preservada no limite. A definição a seguir terá um papel importante na convergência de martingales em $\mathcal{L}^1$.

Definição 12.36 (Integrabilidade uniforme)**.** Dizemos que uma sequência de variáveis aleatórias $(X_n)_n$ é *uniformemente integrável* se

$$\lim_{k\to\infty} \sup_n \int_{\{|X_n|\geqslant k\}} |X_n| \, \mathrm{d}\mathbb{P} = 0.$$

Teorema 12.37 (Teorema de Convergência de Vitali)**.** *Seja $(X_n)_n$ uma sequência uniformemente integrável. Então $\sup_n \mathbb{E}|X_n| < \infty$. Se, ademais, $X_n \overset{\text{q.c.}}{\to} X$, então X é integrável e $X_n \to X$ em $\mathcal{L}^1$.*

Demonstração. Para a primeira afirmação, basta tomar k tal que

$$\sup_n \int_{\{|X_n|\geqslant k\}} |X_n| \, \mathrm{d}\mathbb{P} \leqslant 1,$$

pois isso nos dá $\sup_n \mathbb{E}|X_n| \leqslant k+1$ para todo n.

Para a segunda, suponha que $X_n \overset{\text{q.c.}}{\to} X$. Pelo Lema de Fatou, $\mathbb{E}|X| \leqslant \liminf_n \mathbb{E}|X_n| \leqslant \sup_n \mathbb{E}|X_n| < \infty$, logo X é integrável.

Agora seja $\varepsilon > 0$. Tome k tal que $\int_{\{|X|>k\}} |X| \,\mathrm{d}\mathbb{P} < \varepsilon$ e $\int_{\{|X_n|>k\}} |X_n| \,\mathrm{d}\mathbb{P} < \varepsilon$ para todo $n \in \mathbb{N}$. Defina $g(x) = x\mathbb{1}_{[-k,k]}(x) + k\mathbb{1}_{(k,+\infty)}(x) - k\mathbb{1}_{(-\infty,-k)}(x)$. Estimaremos

$$\mathbb{E}|X_n - X| \leqslant \mathbb{E}|g(X_n) - g(X)| + \mathbb{E}|g(X_n) - X_n| + \mathbb{E}|g(X) - X|.$$

Como g é contínua, $g(X_n) \overset{\text{q.c.}}{\to} g(X)$ e, pelo Teorema da Convergência Dominada, $\mathbb{E}|g(X_n) - g(X)| \to 0$. Como $|g(x) - x| < |x|\mathbb{1}_{|x|>k}$, temos que $\mathbb{E}|g(X_n) - X_n| \leqslant \varepsilon$ e $\mathbb{E}|g(X) - X| \leqslant \varepsilon$. Portanto, $\limsup_n \mathbb{E}|X_n - X| \leqslant 2\varepsilon$, para todo ε, concluindo a prova. □

Observação 12.38. Integrabilidade uniforme implica $\sup_n \mathbb{E}|X_n| < \infty$ mas não vale a recíproca. Por exemplo, se $\mathbb{P}(X_n = n) = 1 - \mathbb{P}(X_n = 0) = \frac{1}{n}$, então $\mathbb{E}|X_n| = 1$ mas essa sequência não é uniformemente integrável. △

Teorema 12.39. *Seja $(X_n)_{n\in\mathbb{N}_0}$ um submartingale uniformemente integrável. Então, existe uma variável aleatória X_∞, tal que $X_n \to X_\infty$ quase certamente e em $\mathcal{L}^1$. Além disso, $\mathbb{E}[X_\infty|\mathcal{F}_k] \geqslant X_k$ q.c. para todo $k \in \mathbb{N}_0$. Em particular, se $(X_n)_n$ é um martingale uniformemente integrável, $\mathbb{E}[X_\infty|\mathcal{F}_k] = X_k$ q.c.*

Demonstração. Pelo Teorema de Convergência de Vitali, $\sup_n \mathbb{E}|X_n| < \infty$. Logo, pelo Teorema da Convergência de Martingales, existe uma variável aleatória estendida X_∞ tal que $X_n \overset{\text{q.c.}}{\to} X_\infty$. Pelo Teorema de Convergência de Vitali novamente, X_∞ é integrável e $X_n \to X_\infty$ em $\mathcal{L}^1$. Finalmente, seja $A \in \mathcal{F}_k$. Como $X_n \to X_\infty$ em $\mathcal{L}^1$ e $\mathbb{E}[X_n|\mathcal{F}_k] \geqslant X_k$ para $n \geqslant k$, temos

$$\int_A \mathbb{E}[X_\infty|\mathcal{F}_k] \,\mathrm{d}\mathbb{P} = \int_A X_\infty \,\mathrm{d}\mathbb{P} = \lim_n \int_A X_n \,\mathrm{d}\mathbb{P} \geqslant \int_A X_k \,\mathrm{d}\mathbb{P}.$$

Como ambos $\mathbb{E}[X_\infty|\mathcal{F}_k]$ e X_k são $\mathcal{F}_k$-mensuráveis e a desigualdade vale para todo $A \in \mathcal{F}_k$, pelo Exercício 5.63 concluímos que $\mathbb{E}[X_\infty|\mathcal{F}_k] \geqslant X_k$ q.c. □

Podemos inspecionar o martingale definido no Exemplo 12.35 e verificar diretamente que ele converge certamente mas não é uniformemente integrável nem converge em $\mathcal{L}^1$. Abaixo, discutimos mais alguns exemplos.

Exemplo 12.40. Seja M_n o passeio aleatório simétrico começando de $x = 1$ e parado em $y = 0$, como definido no Exemplo 12.30. Observe que $\sup_n \mathbb{E}|M_n| = 1$ pois $M_n \geqslant 0$ q.c. e $\mathbb{E}M_n = 1$ para todo n, donde pudemos concluir que $(M_n)_n$ converge q.c. Entretanto, $M_n \overset{\text{q.c.}}{\to} 0$, logo não pode valer convergência em $\mathcal{L}^1$ e esse martingale não é uniformemente integrável. $\triangle$

Exemplo 12.41. O processo de ramificação crítico satisfaz $\sup_n \mathbb{E}|Z_n| < \infty$ pois $Z_n \geqslant 0$ q.c. e $\mathbb{E}Z_n = 1$ para todo n. Isso foi usado para mostrar que $Z_n \overset{\text{q.c.}}{\to} 0$. Novamente, não pode valer convergência em $\mathcal{L}^1$ e esse martingale não é uniformemente integrável. $\triangle$

As proposições seguintes são critérios úteis para verificar a integrabilidade uniforme.

Proposição 12.42. *Seja $(X_n)_n$ uma sequência de variáveis aleatórias. Se $\sup_n \mathbb{E}|X_n|^p < \infty$ para algum $p > 1$, então $(X_n)_n$ é uniformemente integrável.*

Demonstração. Basta observar que

$$\sup_n \int_{\{|X_n| \geqslant k\}} |X_n| \, \mathrm{d}\mathbb{P} \leqslant \sup_n \int_{\{|X_n| \geqslant k\}} \frac{|X_n|^p}{k^{p-1}} \, \mathrm{d}\mathbb{P} \leqslant \frac{\sup_n \mathbb{E}|X_n|^p}{k^{p-1}} \to 0$$

quando $k \to \infty$, pois $\sup_n \mathbb{E}|X_n|^p < \infty$. $\square$

Proposição 12.43. *Seja $(X_n)_n$ um martingale tal que $X_0 = 0$ q.c. Suponha que $\mathbb{E}(X_n - X_{n-1})^2 < \infty$ para todo n. Então*

$$\mathbb{E}X_n^2 = \sum_{k=1}^{n} \mathbb{E}(X_k - X_{k-1})^2 \text{ para todo } n.$$

Em particular, se $(M_n)_n$ é um martingale tal que $\sum_{k=1}^{\infty} \mathbb{E}(M_k - M_{k-1})^2 < \infty$, então $(M_n)_n$ é uniformemente integrável, logo converge q.c. e em $\mathcal{L}^1$.

Demonstração. Mostraremos a primeira parte por indução em n. Expandindo os todos termos, vemos que

$$X_{n+1}^2 = X_n^2 + (X_{n+1} - X_n)^2 + 2\, X_n(X_{n+1} - X_n).$$

Observe que $X_n(X_{n+1} - X_n)$ é integrável, pois $|xy - x^2| \leqslant y^2 + 2x^2$ para todos x e y em $\mathbb{R}$. Como $(X_n)_n$ é um martingale,

$$\mathbb{E}[X_n(X_{n+1} - X_n)|\mathcal{F}_n] = X_n\, \mathbb{E}[X_{n+1} - X_n|\mathcal{F}_n] = 0 \text{ q.c.}$$

Tomando a esperança na equação acima, obtemos $\mathbb{E}[X_n(X_{n+1} - X_n)] = 0$, o que termina a prova por indução.

Para a segunda parte, tome $X_n = M_n - M_0$, observe que $\sup_n \mathbb{E}X_n^2 = \sum_n \mathbb{E}(M_n - M_{n-1})^2 < \infty$. Pela Proposição 12.42, $(X_n)_n$ é uniformemente integrável. Pelo Teorema 12.39, $(X_n)_n$ converge q.c. e em $\mathcal{L}^1$ para alguma X_∞. Portanto, $\mathbb{E}|M_n - M_0 - X_\infty| \to 0$, donde concluímos que $(M_n)_n$ converge q.c. e em $\mathcal{L}^1$ para $X_\infty + M_0$. □

Exemplo 12.44 (Processo de ramificação supercrítico). Continuando o Exemplo 12.29, mostraremos que quando $\lambda > 1$ há probabilidade positiva de a população crescer exponencialmente ao invés de se extinguir. Como $(M_n)_n$ é um martingale não-negativo, sabemos que converge quase certamente para alguma variável aleatória M_∞. Basta então mostrar que $\mathbb{P}(M_\infty > 0) > 0$.

Para isso, vamos estimar $\mathbb{E}(M_n - M_{n-1})^2$. Observe que

$$\begin{aligned}(M_n - M_{n-1})^2 &= \lambda^{-2n}\,(Z_n - \lambda Z_{n-1})^2 \\ &= \lambda^{-2n}\Big[\sum_k \mathbb{1}_{\{Z_{n-1}=k\}}\Big(\sum_{j=1}^k (X_{n,j} - \lambda)\Big)^2\Big].\end{aligned}$$

Tomando a esperança e usando independência entre Z_{n-1} e $(X_{n,j})_j$,

$$\mathbb{E}(M_n - M_{n-1})^2 = \lambda^{-2n}\,\Big[\sum_k \mathbb{P}(Z_{n-1} = k)\cdot \mathbb{E}\Big(\sum_{j=1}^k (X_{n,j} - \lambda)\Big)^2\Big]$$

$$
\begin{aligned}
&= \lambda^{-2n} \Big[\sum_k \mathbb{P}(Z_{n-1} = k) \cdot k\, \sigma^2 \Big] \\
&= \lambda^{-2n}\, \sigma^2\, \mathbb{E}[Z_{n-1}] = \sigma^2\, \lambda^{-n-1}.
\end{aligned}
$$

Logo, $\sum_n \mathbb{E}(M_n - M_{n-1})^2 < \infty$. Pela Proposição 12.43, $(M_n)_n$ converge em $\mathcal{L}^1$. Portanto, $\mathbb{E}M_\infty = \mathbb{E}M_1 = 1$ e $\mathbb{P}(M_\infty > 0) > 0$. $\triangle$

O teorema abaixo diz que uma variável aleatória pode ser aproximada por sua esperança condicional dada uma σ-álgebra de uma filtração qualquer.

Teorema 12.45. *Sejam $(\mathcal{F}_n)_n$ uma filtração e Z uma variável aleatória integrável. Defina $\mathcal{F}_\infty = \sigma(\cup_n \mathcal{F}_n)$. Então*

$$
\mathbb{E}[Z|\mathcal{F}_n] \to \mathbb{E}[Z|\mathcal{F}_\infty] \text{ q.c. e em } \mathcal{L}^1.
$$

Demonstração. Para todo $n \in \mathbb{N}$, definimos $X_n = \mathbb{E}[Z|\mathcal{F}_n]$, que é um martingale com respeito à filtração $(\mathcal{F}_n)_n$ conforme vimos no Exemplo 12.6. Mostraremos inicialmente que $(X_n)_n$ é uniformemente integrável. Seja $\varepsilon > 0$. Como Z é integrável, podemos tomar β de modo que $\int_{\{|Z|>\beta\}} |Z| \, \mathrm{d}\mathbb{P} < \frac{\varepsilon}{2}$. Sendo assim,

$$
\begin{aligned}
\int_{\{|X_n|>k\}} |X_n| \, \mathrm{d}\mathbb{P} &\leqslant \int_{\{|X_n|>k\}} \mathbb{E}[|Z||\mathcal{F}_n] \, \mathrm{d}\mathbb{P} = \int_{\{|X_n|>k\}} |Z| \, \mathrm{d}\mathbb{P} \\
&= \int_{\{|X_n|>k, |Z|\leqslant\beta\}} |Z| \, \mathrm{d}\mathbb{P} + \int_{\{|X_n|>k, |Z|>\beta\}} |Z| \, \mathrm{d}\mathbb{P} \\
&\leqslant \beta \mathbb{P}(|X_n| \geqslant k) + \int_{\{|Z|>\beta\}} |Z| \, \mathrm{d}\mathbb{P} \\
&\leqslant \frac{\beta \mathbb{E}|X_n|}{k} + \frac{\varepsilon}{2} \leqslant \frac{\beta \mathbb{E}|Z|}{k} + \frac{\varepsilon}{2} \leqslant \varepsilon,
\end{aligned}
$$

onde na última desigualdade, tomamos $k > \frac{2\beta\mathbb{E}|Z|}{\varepsilon}$. Isto mostra a integrabilidade uniforme de $(X_n)_n$. Portanto, pelo Teorema 12.39, $X_n = \mathbb{E}[Z|\mathcal{F}_n] \to X_\infty$ q.c. e em $\mathcal{L}^1$.

Resta mostrar que $\mathbb{E}[Z|\mathcal{F}_\infty] = X_\infty$. Primeiro, podemos tomar $X_\infty = \lim_n X_n \cdot \mathbb{1}_{\{(X_n)_n \text{ converge}\}}$, de forma que X_∞ é $\mathcal{F}_\infty$-mensurável pelo Lema 3.50.

Observe que para todos $A \in \mathcal{F}_n$ e $m \geqslant n$, pela definição de X_m temos que

$$\int_A X_m \, \mathrm{d}\mathbb{P} = \int_A Z \, \mathrm{d}\mathbb{P}.$$

Como $X_m \overset{\mathcal{L}^1}{\to} X_\infty$, podemos tomar o limite na integral, obtendo

$$\int_A X_\infty \, \mathrm{d}\mathbb{P} = \lim_m \int_A X_m \, \mathrm{d}\mathbb{P} = \int_A Z \, \mathrm{d}\mathbb{P} \quad \text{para todo } A \in \cup_{n\in\mathbb{N}} \mathcal{F}_n.$$

Lembrando que Z e X_∞ são integráveis, podemos definir para todo $A \in \mathcal{F}_\infty$,

$$\mu(A) = \int_A (Z^+ + X_\infty^-) \, \mathrm{d}\mathbb{P} \quad \text{e} \quad \nu(A) = \int_A (X_\infty^+ + Z^-) \, \mathrm{d}\mathbb{P},$$

obtendo duas medidas finitas μ e ν definidas em $(\Omega, \mathcal{F}_\infty)$, que coincidem na classe $\cup_{n\in\mathbb{N}} \mathcal{F}_n$. Como essa classe é um π-sistema, pelo Teorema 3.37 (unicidade de medidas), as medidas coincidem na σ-álgebra $\sigma(\cup_n \mathcal{F}_n) = \mathcal{F}_\infty$. Assim,

$$\int_A X_\infty \, \mathrm{d}\mathbb{P} = \int_A Z \, \mathrm{d}\mathbb{P} \quad \text{para todo } A \in \mathcal{F}_\infty.$$

Como X_∞ é $\mathcal{F}_\infty$-mensurável, segue que $X_\infty = \mathbb{E}[Z|\mathcal{F}_\infty]$ q.c. □

Se um martingale satisfaz $\sup_n \mathbb{E}|X_n|^p < \infty$ para algum $p > 1$, podemos combinar a Proposição 12.42 com o Teorema 12.39 para concluir que $X_n \to X_\infty$ q.c. e em $\mathcal{L}^1$. Entretanto, nesse caso podemos concluir algo bem mais forte, conforme enunciado abaixo.

Teorema 12.46 (Convergência em $\mathcal{L}^p$)**.** *Sejam $p > 1$ e $(X_n)_{n\in\mathbb{N}_0}$ um martingale, ou um submartingale não-negativo, e suponha que $\sup_n \mathbb{E}|X_n|^p < \infty$. Então, existe uma variável aleatória X_∞, tal que $X_n \to X_\infty$ quase certamente e em $\mathcal{L}^p$.*

Demonstração. Defina $Y = \sup_n |X_n| = \lim_n \max_{j=1,\dots,n} |X_j|$. Usando o Teorema 12.21 e o Teorema da Convergência Monótona, temos $\mathbb{E}Y^p \leqslant (\frac{p}{p-1})^p \sup_n \mathbb{E}|X_n|^p < \infty$. Já sabemos que $X_n \overset{\text{q.c.}}{\to} X_\infty$ pelo parágrafo acima.

Como $|X_n| \leqslant Y$ para todo n, podemos aplicar o Teorema da Convergência Dominada em $\mathcal{L}^p$, concluindo a prova. $\square$

12.6 Decomposição de Doob

Nesta seção compreenderemos um pouco mais da estrutura dos submartingales, pois estes sempre poderão ser escritos como a soma de um martingale e um processo com propriedades especiais.

Definição 12.47. Dizemos que o processo $(A_n)_{n\in\mathbb{N}_0}$ é *previsível com respeito à filtração* $(\mathcal{F}_n)_n$ se A_n é $\mathcal{F}_{n-1}$-mensurável para todo $n \in \mathbb{N}$ e $A_0 = 0$.

O processo $(C_n)_n$ definido no preâmbulo deste capítulo é um exemplo de processo previsível. A variável C_n representa quanto o jogador decide apostar na n-ésima rodada do jogo, esta é sempre determinada sabendo-se o resultado das $n-1$ primeiras rodadas.

Teorema 12.48 (Teorema de Decomposição de Doob)**.** *Sejam $(\mathcal{F}_n)_{n\in\mathbb{N}_0}$ uma filtração e $(X_n)_{n\in\mathbb{N}_0}$ um processo adaptado tal que $\mathbb{E}|X_n| < \infty$ para todo $n \in \mathbb{N}_0$. Então existem um martingale $(M_n)_n$ e um processo previsível $(A_n)_n$ tais que $X_n = M_n + A_n$ para todo $n \in \mathbb{N}_0$. Além disso, tal decomposição é única no sentido de que qualquer outra decomposição em martingale e processo previsível é igual a esta q.c.*

Demonstração. Defina $A_0 = 0$, $M_0 = X_0$,

$$M_{n+1} = M_n + X_{n+1} - \mathbb{E}[X_{n+1}|\mathcal{F}_n],$$

e

$$A_{n+1} = A_n + \mathbb{E}[X_{n+1}|\mathcal{F}_n] - X_n$$

para todo $n \in \mathbb{N}_0$. Por indução, verificamos que $X_n = M_n + A_n$ somando as duas equações, e também que A_n é $\mathcal{F}_{n-1}$-mensurável e M_n é $\mathcal{F}_n$-mensurável diretamente das fórmulas. Ademais, tomando a esperança condicional com

respeito a $\mathcal{F}_n$, vemos que $\mathbb{E}[M_{n+1}|\mathcal{F}_n] = M_n$. Logo, $(M_n)_n$ é um martingale. Isso mostra a existência da decomposição.

Para a unicidade, suponha que haja outra decomposição $X_n = M'_n + A'_n$ com $(M'_n)_n$ martingale e $(A'_n)_n$ processo previsível. Neste caso,

$$A'_{n+1} - A'_n = A_{n+1} + M_{n+1} - M'_{n+1} - A_n - M_n + M'_n.$$

Tomando a esperança condicional com respeito a $\mathcal{F}_n$, obtemos

$$\begin{aligned} A'_{n+1} - A'_n &= \mathbb{E}[A'_{n+1} - A'_n|\mathcal{F}_n] \\ &= \mathbb{E}[A_{n+1} + M_{n+1} - M'_{n+1} - A_n - M_n + M'_n|\mathcal{F}_n] \\ &= \mathbb{E}[A_{n+1} - A_n|\mathcal{F}_n] = A_{n+1} - A_n \quad \text{q.c.}, \end{aligned}$$

onde na primeira e terceira igualdades estamos utilizando que $(A'_n)_n$ e $(A_n)_n$ são previsíveis e na segunda que $(M'_n)_n$ e $(M_n)_n$ são martingales. Como $A_0 = A'_0 = 0$, concluímos que $A_n = A'_n$ q.c. e, por conseguinte, $M_n = M'_n$ q.c. para todo $n \in \mathbb{N}$. □

A sequência $(A_n)_n$ que aparece na Decomposição de Doob acima é denominada *compensador do processo* $(X_n)_n$. Observando que $A_{n+1} - A_n = \mathbb{E}[X_{n+1} - X_n|\mathcal{F}_n]$, temos que o processo previsível $(A_n)_n$ é q.c. não-decrescente se, e somente se, o processo $(X_n)_n$ é um submartingale.

Um caso particular muito importante é quando $(X_n)_n$ é um martingale com segundo momento finito e consideramos o submartingale $(X_n^2)_n$. Neste caso, a decomposição de Doob do submartingale é denotada $X_n^2 = M_n + \langle X\rangle_n$, e o compensador $\langle X\rangle_n$ é denominado *variação quadrática do martingale* $(X_n)_n$. Pela Proposição 12.43, a variação quadrática satisfaz

$$\mathbb{E}[\langle X\rangle_n] = \mathbb{E}X_n^2 - \mathbb{E}X_0^2 = \sum_{k=1}^{n} \mathbb{E}(X_k - X_{k-1})^2.$$

Exemplo 12.49. Seja $(S_n)_n$ o passeio aleatório simétrico (quando $p = \frac{1}{2}$),

conforme definido no Exemplo 12.9, de forma que $(S_n)_n$ é um martingale com respeito à filtração $(\mathcal{F}_n)_n$ induzida pelas parcelas $(X_n)_n$. Pela equação acima, variação quadrática de $(S_n)_n$ é dada por $\langle S\rangle_n = n$ para todo $n \in \mathbb{N}$. Isso também pode ser verificado diretamente, expandindo

$$A_{n+1} - A_n = \mathbb{E}[S_{n+1}^2|\mathcal{F}_n] - S_n^2 = S_n^2 + 2S_n\,\mathbb{E}[X_{n+1}|\mathcal{F}_n] + \mathbb{E}[X_{n+1}^2|\mathcal{F}_n] - S_n^2 = 1,$$

de forma que o compensador $(A_n)_n$ é dado por $A_n = n$ para todo n. $\triangle$

12.7 Exercícios

§ 12.1

1. Seja $(S_n)_n$ o passeio aleatório em $\mathbb{Z}$ e $(\mathcal{F})_n$ a filtração definidos no Exemplo 12.9. Defina $M_n = (\frac{1-p}{p})^{S_n}$ para todo $n \in \mathbb{N}$. Mostre que $(M_n)_n$ é um martingale.

2. Seja X uma variável aleatória ilimitada que toma valores em $\mathbb{N}$. Defina as taxas $\lambda_k = \mathbb{P}(X > k|X \geqslant k)$ e suponha que $\lambda_k > 0$ para todo k. Defina $M_0 = 1$ e

$$M_n = \begin{cases} \frac{1}{\lambda_1 \cdots \lambda_n}, & X > n, \\ 0, & X \leqslant n. \end{cases}$$

Mostre que $(M_n)_n$ é um martingale com respeito à sua própria filtração natural.

§ 12.2

3. Se τ_1 e τ_2 são tempos de parada com respeito à mesma filtração $(\mathcal{F}_n)_n$, então podemos afirmar que as variáveis abaixo são tempos de parada? Prove ou dê contra-exemplo.

(a) $\tau_1 + \tau_2$.
(b) $|\tau_1 - \tau_2|$.
(c) $\max\{\tau_1, \tau_2\}$.

(d) $\tau_1 \wedge \tau_2$.
(e) $\tau_1 \wedge k$ para $k \in \mathbb{N}$.

4. Sejam $(X_n)_n$ um processo adaptado e $B \in \mathcal{B}$. Definimos o tempo de última passagem em B como $\sigma_B = \sup\{n : X_n \in B\}$. Dê um exemplo ilustrando que σ_B não é tempo de parada.

5. Seja τ um tempo de parada com respeito à filtração $(\mathcal{F}_n)_{n\in\mathbb{N}_0}$. Defina

$$\mathcal{F}_\tau = \{A \in \mathcal{F} : A \cap \{\tau \leqslant n\} \in \mathcal{F}_n \text{ para todo } n \in \mathbb{N}_0\}.$$

Mostre que:

(a) $\mathcal{F}_\tau$ é uma σ-álgebra.
(b) τ é $\mathcal{F}_\tau$-mensurável.
(c) Se $(X_n)_n$ é um processo adaptado à filtração $(\mathcal{F}_n)_n$, então X_τ é $\mathcal{F}_\tau$-mensurável.
(d) Se τ_1 e τ_2 são tempos de parada tais que $\tau_1 \leqslant \tau_2$ para todo ω, então $\mathcal{F}_{\tau_1} \subseteq \mathcal{F}_{\tau_2}$.

§12.3

6 (Critério de integrabilidade para tempos de parada)**.** Seja τ um tempo de parada com respeito à filtração $(\mathcal{F}_n)_{n\in\mathbb{N}_0}$. Suponha que existem $C \in \mathbb{N}$ e $\varepsilon > 0$ satisfazendo

$$\mathbb{E}[\mathbb{1}_{\{\tau\leqslant n+C\}}|\mathcal{F}_n] \geqslant \varepsilon \quad \text{q.c., para todo } n \in \mathbb{N}_0.$$

Mostre por indução em k que

$$\mathbb{P}(\tau > kC) \leqslant (1-\varepsilon)^k$$

para todo $k \in \mathbb{N}$ e, em particular, $\mathbb{E}\tau < \infty$.

7. No problema da ruína do apostador, mostre que $\mathbb{E}\tau < \infty$.

8 (Um problema de otimização estocástica)**.** Uma companhia deseja recrutar um novo funcionário para uma vaga de emprego. São N candidatos para a

única vaga; eles são entrevistados um de cada vez e após o término de cada entrevista o candidato é contratado, ignorando-se os não entrevistados, ou ele é dispensado definitivamente e entrevista-se o próximo. Suponha que os atributos dos candidatos sejam dados por uma sequência $(X_n)_{n=1}^N$ de variáveis aleatórias i.i.d. com distribuição $\mathcal{U}[0,1]$. Seja $(\mathcal{F}_n)_{n=1}^N$ sua filtração natural. Vamos mostrar que existe um tempo de parada que maximiza $\mathbb{E}X_\tau$ dentre todos os possíveis tempos de parada. Com este objetivo, defina indutivamente a sequência $(\alpha_n)_{n=1}^N$ como $\alpha_N = 0$ e $\alpha_{n-1} = \frac{1+\alpha_n^2}{2}$ para $n = 1, \ldots, n-1$. Defina o tempo de parada $\tau^* = \inf\{n : X_n \geqslant \alpha_n\}$, e seja τ um outro tempo de parada qualquer.

(a) Mostre que $\mathbb{E}[\max\{U, \alpha\}] = \frac{1+\alpha^2}{2}$, onde $U \sim \mathcal{U}[0,1]$ e $\alpha \in [0,1]$.

(b) Defina $Y_n = \max\{X_{n\wedge\tau}, \alpha_n\}$. Mostre que

$$\mathbb{E}[Y_n \cdot \mathbb{1}_{\{\tau \leqslant n-1\}} | \mathcal{F}_{n-1}] \leqslant Y_{n-1} \cdot \mathbb{1}_{\{\tau \leqslant n-1\}}.$$

(c) Mostre que

$$\mathbb{E}[Y_n \cdot \mathbb{1}_{\{\tau \geqslant n\}} | \mathcal{F}_{n-1}] \leqslant Y_{n-1} \cdot \mathbb{1}_{\{\tau \geqslant n\}}$$

e conclua que $(Y_n)_{n=1}^N$ é um supermartingale.

(d) Defina $Y_n^* = \max\{X_{n\wedge\tau^*}, \alpha_n\}$. Mostre que $(Y_n^*)_{n=1}^N$ é um martingale.

(e) Mostre que $\mathbb{E}Y_{\tau^*}^* \geqslant \mathbb{E}Y_\tau$.

(f) Conclua que $\mathbb{E}X_{\tau^*} \geqslant \mathbb{E}X_\tau$, finalizando o exercício.

9 (Generalizando o problema anterior)**.** Seja $(X_n)_{n=1}^N$ um processo adaptado à filtração $(\mathcal{F}_n)_{n=1}^N$ e suponha que as X_n sejam integráveis. Vamos mostrar que existe um tempo de parada τ^* tal que $\mathbb{E}X_{\tau^*} \geqslant \mathbb{E}X_\tau$ para todo tempo de parada τ. Defina $Y_N = X_N$ e $Y_n = \max\{X_n, \mathbb{E}[Y_{n+1}|\mathcal{F}_n]\}$ para todo $n = 1, \ldots, N-1$, que representa o preço justo para entrar no jogo no instante n (pois o jogador pode escolher entre ficar com X_n ou esperar o instante $n+1$). Tome $\tau^* = \min\{n : Y_n = X_n\}$ e seja τ um tempo de parada qualquer.

(a) Verifique que $(Y_n)_{n=1}^N$ é um supermartingale.

(b) Mostre que $\mathbb{E}X_\tau \leqslant \mathbb{E}Y_\tau \leqslant \mathbb{E}Y_1$.

(c) Defina $Z_n = Y_{n\wedge\tau^*}$. Verifique que $Z_{n+1} \cdot \mathbb{1}_{\{\tau^* \leqslant n\}} = Z_n \cdot \mathbb{1}_{\{\tau^* \leqslant n\}}$ q.c.

(d) Verifique que $\mathbb{E}[Z_{n+1}|\mathcal{F}_n] \cdot \mathbb{1}_{\{\tau^*>n\}} = Z_n \cdot \mathbb{1}_{\{\tau^*>n\}}$ q.c.
(e) Conclua que $(Z_n)_{n=1}^N$ é um martingale.
(f) Mostre que $\mathbb{E}X_{\tau^*} = \mathbb{E}Y_{\tau^*} = \mathbb{E}Y_1$, concluindo o exercício.

§ 12.4

10. Seja $(X_n)_n$ uma sequência i.i.d. assumindo valores ± 1 com probabilidade $\frac{1}{2}$ cada. Defina $Z_1 = \frac{1}{3}$ e $Z_{n+1} = Z_n + (1 - Z_n)Z_n X_n$ para todo $n \in \mathbb{N}$. Mostre que $Z_n \xrightarrow{\mathrm{d}} \mathrm{Bernoulli}(\frac{1}{3})$.

§ 12.5

11. Mostre que a posição do passeio aleatório $(S_n)_n$ não é uniformemente integrável.

12. Mostre que o processo de ramificação com $\lambda > 1$ tem probabilidade positiva de sobreviver, sem supor que $\sigma^2 < \infty$.

13. Mostre que o martingale definido no Exercício 2 converge quase certamente mas não converge em $\mathcal{L}^1$.

§ 12.6

14. Sejam $(X_n)_{n\in\mathbb{N}_0}$ processo adaptado e $(C_n)_{n\in\mathbb{N}_0}$ processo previsível com respeito à filtração $(\mathcal{F}_n)_n$. Suponha que existe $K > 0$ tal que $\mathbb{P}(|C_n| \leqslant K) = 1$ para todo n. Defina $M_n = C_0 X_0 + \sum_{k=1}^n C_k(X_k - X_{k-1})$ para todo $n \in \mathbb{N}$.

(a) Mostre que se $(X_n)_n$ é um martingale, então $(M_n)_n$ também o é.
(b) Mostre que se $(X_n)_n$ é um submartingale e $C_n \geqslant 0$ para todo $n \in \mathbb{N}$, então $(M_n)_n$ é um submartingale.

O processo $(M_n)_n$ é chamado *transformação de* $(X_n)_n$ *por* $(C_n)_n$, denotado por $(C \bullet X)_n$. Ele aparece implicitamente no preâmbulo deste capítulo e na prova do Teorema de Convergência de Martingales.

15. Sejam $(X_n)_{n\in\mathbb{N}_0}$ e $(Y_n)_{n\in\mathbb{N}_0}$ martingales com respeito à mesma filtração $(\mathcal{F}_n)_{n\in\mathbb{N}_0}$ tais que $\mathbb{E}X_n^2 < \infty$ e $\mathbb{E}Y_n^2 < \infty$ para todo $n \in \mathbb{N}$. Defina o processo

$\langle X, Y\rangle$ como

$$\langle X, Y\rangle_n = \frac{1}{4}[\langle X+Y\rangle_n - \langle X-Y\rangle_n].$$

Mostre que $(X_nY_n - \langle X, Y\rangle_n)_{n\in\mathbb{N}_0}$ é um martingale.

16. Sejam $(X_n)_n$ e $(Y_n)_n$ sequencias de variáveis aleatórias independentes com $\mathbb{E}X_n^2 < \infty$, $\mathbb{E}Y_n^2 < \infty$, $\mathbb{E}X_n = \mathbb{E}Y_n = 0$, $S_n = X_1 + \cdots + X_n$ e $T_n = Y_1 + \cdots + Y_n$ para todo $n \in \mathbb{N}$. Mostre que $\langle S, T\rangle_n = \sum_{k=1}^n \mathbf{Cov}(X_k, Y_k)$.

Capítulo 13

Leis 0-1 e Séries Aleatórias

De modo geral, dizemos que vale uma lei 0-1 quando podemos estabelecer que um dado evento, ou todos os eventos de uma dada classe, têm probabilidade zero ou um. Já nos deparamos com duas situações com esta característica ao longo deste texto, a Proposição 2.16 (um evento é independente de si mesmo se, e somente se, sua probabilidade é zero ou um) e o Corolário 7.18 (dada uma sequência enumerável de eventos independentes, a ocorrência de infinitos deles é um evento de probabilidade zero ou um). Leis 0-1 são muito úteis em Probabilidade, pois excluem a possibilidade de valores de probabilidade no interior do intervalo $[0, 1]$, e, dessa forma, podemos concluir que a probabilidade é 1 verificando que não é 0, e vice-versa. Estudaremos duas das principais leis desse tipo: a Lei 0-1 de Kolmogorov e a Lei 0-1 de Hewitt-Savage.

Uma consequência da Lei 0-1 de Kolmogorov é que uma série de números aleatórios independentes será convergente com probabilidade zero ou um. Na última seção deste capítulo daremos condições necessárias e suficientes para que essa probabilidade seja um.

Os enunciados e demonstrações dos principais teoremas deste capítulo dependem de alguns conceitos técnicos dos quais ainda não dispomos.

Começaremos com uma seção cujo objetivo é introduzir definições e ferramentas que serão usadas nas seções e capítulos seguintes.

13.1 Álgebras e espaços de sequências

Começamos pelo espaço de sequências infinitas. Denotamos o conjunto de todas as sequências de números reais por

$$\mathbb{R}^{\mathbb{N}} = \{\boldsymbol{x} = (x_1, x_2, x_3, \dots) : x_n \in \mathbb{R} \text{ para todo } n \in \mathbb{N}\}.$$

Dizemos que um conjunto $A \subseteq \mathbb{R}^{\mathbb{N}}$ é um conjunto *cilíndrico* se

$$A = \{\boldsymbol{x} = (x_n)_n \in \mathbb{R}^{\mathbb{N}} : (x_1, x_2 \dots, x_k) \in B\}$$

para algum $k \in \mathbb{N}$ e $B \in \mathcal{B}(\mathbb{R}^k)$.

Definição 13.1 (Conjuntos borelianos em $\mathbb{R}^{\mathbb{N}}$)**.** Definimos a *σ-álgebra de Borel em* $\mathbb{R}^{\mathbb{N}}$, denotada por $\mathcal{B}(\mathbb{R}^{\mathbb{N}})$, como a σ-álgebra gerada pela classe dos conjuntos cilíndricos.

Observe que a projeção de $\mathbb{R}^{\mathbb{N}}$ em $\mathbb{R}$ que leva $\boldsymbol{x}$ em sua n-ésima coordenada x_n é uma função mensurável em $\mathbb{R}^{\mathbb{N}}$, pois $\{\boldsymbol{x} : x_n \in B\}$ é um conjunto cilíndrico para todo $B \in \mathcal{B}(\mathbb{R})$.

Definição 13.2. Dado um espaço amostral Ω e uma família de funções $(f_\alpha)_{\alpha\in\Lambda}$ de Ω em $\mathbb{R}$, dizemos que $A \subseteq \Omega$ é *cilíndrico com respeito a* $(f_\alpha)_{\alpha\in\Lambda}$ se existem $k \in \mathbb{N}$, $\alpha_1, \dots, \alpha_k \in \Lambda$ e $B \in \mathcal{B}(\mathbb{R}^k)$ tais que $A = \{\omega \in \Omega : (f_{\alpha_1}(\omega), \dots, f_{\alpha_k}(\omega)) \in B\}$. Definimos $\sigma\big((f_\alpha)_{\alpha\in\Lambda}\big)$ como a σ-álgebra gerada pelos conjuntos cilíndricos com respeito a $(f_\alpha)_{\alpha\in\Lambda}$. A σ-álgebra $\sigma\big((f_\alpha)_{\alpha\in\Lambda}\big)$ é a menor σ-álgebra em Ω para a qual todas as f_α são mensuráveis.

Em geral, a classe dos conjuntos cilíndricos não forma uma σ-álgebra. Por isso, consideramos uma estrutura mais simples.

Definição 13.3 (Álgebra)**.** Uma classe $\mathcal{A}$ de subconjuntos de Ω é chamada uma *álgebra em* Ω se

(1) $\Omega \in \mathcal{A}$,
(2) $A^c \in \mathcal{A}$ para todo $A \in \mathcal{A}$,
(3) $A \cup B \in \mathcal{A}$ para todos $A, B \in \mathcal{A}$.

É muito útil poder trabalhar com álgebras por pelo menos dois motivos: os elementos de uma álgebra costumam ser muito mais simples de descrever e estudar, e algumas classes importantes de conjuntos são álgebras mas não são σ-álgebras. Por exemplo, a classe dos conjuntos cilíndricos com respeito a $(f_\alpha)_{\alpha\in\Lambda}$ é uma álgebra mas não é uma σ-álgebra, assim como a classe dos conjuntos cilíndricos de $\mathbb{R}^{\mathbb{N}}$.

Mas o que significa poder trabalhar com álgebras? Significa que conjuntos de uma σ-álgebra podem ser aproximados, em um sentido bastante forte, por conjuntos de uma álgebra, como veremos a seguir.

Definição 13.4 (Diferença simétrica)**.** Dados dois conjuntos A e B, a sua *diferença simétrica* $A\triangle B$ é definida como $A\triangle B = (A \setminus B) \cup (B \setminus A)$.

Observe que a diferença simétrica satisfaz um análogo da desigualdade triangular: $A\triangle C \subseteq (A\triangle B) \cup (B\triangle C)$, e também satisfaz $A^c\triangle B^c = A\triangle B$ e $(\cup_n A_n)\triangle(\cup_n B_n) \subseteq \cup_n(A_n\triangle B_n)$ (verifique essas três propriedades!).

Teorema 13.5 (Aproximação por álgebras)**.** *Sejam* $(\Omega, \mathcal{F}, \mathbb{P})$ *um espaço de probabilidade e* $\mathcal{A} \subseteq \mathcal{F}$ *uma álgebra. Então, dados quaisquer* $B \in \sigma(\mathcal{A})$ *e* $\varepsilon > 0$*, existe* $A \in \mathcal{A}$ *tal que* $\mathbb{P}(A\triangle B) < \varepsilon$.

Demonstração. Seja $\mathcal{C}$ a classe dos conjuntos que satisfazem à afirmação do teorema: $\mathcal{C} = \{B \in \mathcal{F} : \forall\varepsilon > 0,\ \exists A \in \mathcal{A} \text{ com } \mathbb{P}(A\triangle B) < \varepsilon\}$. Observe que $\mathcal{A} \subseteq \mathcal{C}$, pois $\mathbb{P}(A\triangle A) = \mathbb{P}(\emptyset) = 0$. Logo, basta mostrar que $\mathcal{C}$ é uma σ-álgebra. É imediato que $\emptyset \in \mathcal{C}$, pois $\mathbb{P}(\emptyset\triangle\emptyset) = 0$. Se $C \in \mathcal{C}$, então dado $\varepsilon > 0$, existe $A \in \mathcal{A}$ tal que $\mathbb{P}(A\triangle C) < \varepsilon$; como $A^c\triangle C^c = A\triangle C$ e $A^c \in \mathcal{A}$, segue que $C^c \in \mathcal{C}$. Para concluir que $\mathcal{C}$ é uma σ-álgebra, falta

mostrar que $\mathcal{C}$ é fechado por uniões enumeráveis. Sejam $C_1, \ldots, C_n, \cdots \in \mathcal{C}$ e $\varepsilon > 0$. Defina $C = \cup_{n=1}^{+\infty} C_n$ e tome $n \in \mathbb{N}$ tal que $\mathbb{P}(C \setminus (\cup_{k=1}^{n} C_k)) < \frac{\varepsilon}{2}$. Tome $A_1, A_2, \ldots, A_n \in \mathcal{A}$ tais que $\mathbb{P}(A_k \triangle C_k) < \frac{\varepsilon}{2n}$, para todo $k = 1, \ldots, n$. Observando que

$$(\cup_{k=1}^{n} A_k) \triangle (\cup_{k=1}^{+\infty} C_k) \subseteq (\cup_{k=1}^{n} (A_k \triangle C_k)) \cup (C \setminus (\cup_{k=1}^{n} C_k)),$$

por subaditividade obtemos

$$\mathbb{P}\Big((\cup_{k=1}^{n} A_k) \triangle (\cup_{k=1}^{+\infty} C_k)\Big) \leqslant \sum_{k=1}^{n} \mathbb{P}(A_k \triangle C_k) + \mathbb{P}(C \setminus (\cup_{k=1}^{n} C_k)) < \varepsilon.$$

Portanto, $C \in \mathcal{C}$, o que completa a prova de que $\mathcal{C}$ é uma σ-álgebra. $\square$

O uso mais frequente do teorema acima é quando queremos aproximar um boreliano de $\mathbb{R}^{\mathbb{N}}$ por conjuntos cilíndricos.

13.2 Independência de σ-álgebras

Nesta seção, o espaço de probabilidade $(\Omega, \mathcal{F}, \mathbb{P})$ está fixado e todas as σ-álgebras mencionadas estarão contidas em $\mathcal{F}$.

Dizemos que duas σ-álgebras $\mathcal{F}_1$ e $\mathcal{F}_2$ são *independentes* se $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$ para todos $A \in \mathcal{F}_1$ e $B \in \mathcal{F}_2$. Recordando-nos da Definição 11.47, uma variável aleatória X é independente de uma σ-álgebra $\mathcal{G}$ se, e somente se, $\sigma(X)$ é independente de $\mathcal{G}$. Ademais, duas variáveis aleatórias X e Y são independentes se, e somente se, $\sigma(X)$ e $\sigma(Y)$ são independentes. Esse conceito pode ser estendido para uma família de σ-álgebras.

Definição 13.6. Uma família de σ-álgebras $(\mathcal{F}_\alpha)_{\alpha \in \Lambda}$ é *independente* se, para todo $k \in \mathbb{N}$, todos $\alpha_1, \ldots, \alpha_k \in \Lambda$ distintos, e todos $A_1 \in \mathcal{F}_{\alpha_1}, \ldots, A_k \in \mathcal{F}_{\alpha_k}$,

$$\mathbb{P}(A_1 \cap \cdots \cap A_k) = \mathbb{P}(A_1) \cdots \mathbb{P}(A_k).$$

Vale observar que uma família de eventos é independente (conforme definido no Capítulo 2) se, e somente se, a família das σ-álgebras geradas respectivamente por cada um desses eventos é independente. Ademais, uma família de variáveis aleatórias é independente (conforme definido no Capítulo 4 se, e somente se, a família das σ-álgebras geradas por essas variáveis aleatórias é independente. Portanto, independência de σ-álgebras generaliza os conceitos de independência de eventos e de variáveis aleatórias.

A definição acima também é útil para classes de eventos que não são necessariamente σ-álgebras. Dizemos que uma família de classes de eventos $(\mathcal{C}_\alpha)_{\alpha\in\Lambda}$ é *independente* se satisfaz a condição análoga à da definição acima.

Lema 13.7. *Seja $(\mathcal{C}_\alpha)_{\alpha\in\Lambda}$ uma família independente de classes de eventos. Suponha que $\mathcal{C}_\alpha$ seja um π-sistema para todo $\alpha \in \Lambda$. Defina $\mathcal{F}_\alpha = \sigma(\mathcal{C}_\alpha)$ para todo $\alpha \in \Lambda$. Então a família $(\mathcal{F}_\alpha)_{\alpha\in\Lambda}$ é independente.*

A prova será dada no Apêndice D.1.

Observação 13.8. O lema acima pode falhar se não pedimos que as classes $\mathcal{C}_\alpha$ sejam π-sistemas. Considere $\Omega = \{1, 2, 3, 4\}$, e $\mathbb{P}$ a medida de probabilidade uniforme. Então $\{1, 2\}$ é independente de $\{1, 3\}$ e de $\{1, 4\}$, mas não é independente de $\{1, 3\}\cap\{1, 4\}$. Ou seja, a classe $\mathcal{C}_1 = \{\{1, 2\}\}$ é independente da classe $\mathcal{C}_2 = \{\{1, 3\}, \{1, 4\}\}$ mas não é independente da σ-álgebra gerada por $\mathcal{C}_2$ (esse contra-exemplo é uma reedição do Exemplo 2.19). $\triangle$

Proposição 13.9. *Seja $(\mathcal{F}_\alpha)_{\alpha\in\Lambda}$ uma família independente de σ-álgebras. Suponha que $\Lambda = \cup_{n\in\mathbb{N}}\Lambda_n$ e que essa união seja disjunta. Defina $\mathcal{G}_n = \sigma(\cup_{\alpha\in\Lambda_n}\mathcal{F}_\alpha)$ para $n \in \mathbb{N}$. Então $(\mathcal{G}_n)_{n\in\mathbb{N}}$ é uma família independente.*

Demonstração. Para cada $n \in \mathbb{N}$, considere a classe $\mathcal{C}_n$ dos eventos da forma $A_1 \cap \cdots \cap A_k$, onde $A_j \in \mathcal{F}_{\alpha_j}$ para todo $j = 1, \ldots, k$, com $\alpha_1, \ldots, \alpha_k \in \Lambda_n$ distintos, para qualquer $k \in \mathbb{N}$. Observe que $\mathcal{C}_n$ é um π-sistema que gera $\mathcal{G}_n$. Observe também que família $(\mathcal{C}_n)_n$ é independente. Pelo lema acima, $(\mathcal{G}_n)_{n\in\mathbb{N}}$ é uma família independente. $\square$

Na proposição acima, se $\Lambda_n = \emptyset$, então $\mathcal{G}_n = \{\emptyset, \Omega\}$ é independente de qualquer classe de eventos.

A proposição acima tem as seguintes consequências, talvez óbvias.

Corolário 13.10. *Sejam $X_1, \dots X_n, Y_1, \dots, Y_m$ variáveis aleatórias independentes. Então $\sigma(X_1, \dots, X_n)$ é independente de $\sigma(Y_1, \dots, Y_m)$.*

Corolário 13.11. *Seja $(X_\alpha)_{\alpha\in\Lambda}$ uma família de variáveis aleatórias independentes. Suponha que $\Lambda = \Lambda_1 \cup \Lambda_2$ e que a união seja disjunta. Então $\sigma((X_\alpha)_{\alpha\in\Lambda_1})$ e $\sigma((X_\alpha)_{\alpha\in\Lambda_2})$ são σ-álgebras independentes.*

Corolário 13.12. *Sejam $(Z_{j,k})_{(j,k)\in\mathbb{N}^2}$ variáveis aleatórias independentes e $g_j : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}$ funções mensuráveis. Defina $U_j = g_j((Z_{j,k})_{k\in\mathbb{N}})$. Então a família $(U_j)_j$ é independente.*

13.3 Lei 0-1 de Kolmogorov

Nesta seção vamos estudar uma importante classe de eventos, chamados de eventos caudais, e a famosa Lei 0-1 de Kolmogorov associada a tais eventos.

Sejam $(\Omega, \mathcal{F}, \mathbb{P})$ um espaço de probabilidade e $\boldsymbol{X} = (X_n)_n$ uma sequência de variáveis aleatórias. Definimos as seguintes σ-álgebras contidas em $\mathcal{F}$:

$$\mathcal{F}_n^\infty = \sigma((X_j)_{j\geqslant n}), \quad \mathscr{X} = \cap_n \mathcal{F}_n^\infty.$$

Em palavras, os eventos da σ-álgebra $\mathcal{F}_{n+1}^\infty$ são aqueles cuja ocorrência, apesar de ser determinada por $\boldsymbol{X}$, não depende dos valores das n primeiras variáveis aleatórias $X_1, \dots, X_n$, enquanto os eventos $A \in \mathscr{X}$ são aqueles cuja ocorrência não depende das k primeiras variáveis aleatórias $X_1, \dots, X_k$, qualquer que seja $k \in \mathbb{N}$, isto é, A não depende do conhecimento de nenhuma quantidade finita das variáveis aleatórias $(X_k)_k$, apenas do comportamento remoto desta sequência.

Definição 13.13. Dizemos que um evento A é *caudal* com respeito à

sequência $\boldsymbol{X}$ se $A \in \mathscr{X}$. Chamamos $\mathscr{X}$ de *σ-álgebra dos eventos caudais* com respeito à sequência $\boldsymbol{X}$, ou simplesmente *σ-álgebra caudal*.

Exemplo 13.14. São exemplos de eventos caudais com respeito a $\boldsymbol{X}$:

(a) $\{\frac{X_1+\cdots+X_n}{n} \text{ converge}\}$;
(b) $\{X_n \in B \ i.v.\}$ para um $B \in \mathcal{B}$ dado;
(c) $\{\limsup_n \frac{X_1+\cdots+X_n}{n} < C\}$ para um $C \in \mathbb{R}$ dado;
(d) $\{\sum_{n=1}^{\infty} X_n \text{ converge}\}$.

Mostraremos apenas o primeiro dos itens acima e deixamos os outros como exercício. Primeiramente, o evento está em $\mathcal{F}_1^{\infty}$ pelo Lema 3.50. Para todo $k \in \mathbb{N}$, podemos escrever

$$\frac{X_1 + \cdots + X_n}{n} = \frac{X_1 + \cdots + X_k}{n} + \frac{X_{k+1} + \cdots + X_n}{n}.$$

Como $\frac{X_1+\cdots+X_k}{n} \to 0$ quando $n \to \infty$, temos que a convergência de $\frac{X_1+\cdots+X_n}{n}$ é equivalente à convergência de $\frac{X_{k+1}+\cdots+X_n}{n}$, e este evento está em $\mathcal{F}_k^{\infty}$. Como isso vale para todo k, o evento está em $\cap_k \mathcal{F}_k^{\infty}$, ou seja, $\{\frac{X_1+\cdots+X_n}{n} \text{ converge}\}$ é um evento caudal com respeito à sequência $\boldsymbol{X}$. $\triangle$

Exemplo 13.15. Dados $B \in \mathcal{B}$ e $C \in \mathbb{R}$, em geral os eventos abaixo não são caudais com respeito a $\boldsymbol{X}$:

(a) $\{X_1 + \cdots + X_n \in B \ i.v.\}$;
(b) $\{\limsup_n X_1 + \cdots + X_n < C\}$. $\triangle$

A primeira das leis 0-1 que vamos estudar, conhecida como a Lei 0-1 de Kolmogorov, tem um enunciado bem simples. Se as variáveis aleatórias $(X_n)_n$ são independentes, então todo evento caudal tem probabilidade 0 ou 1.

Teorema 13.16 (Lei 0-1 de Kolmogorov)**.** *Se $(X_n)_n$ é uma sequência de variáveis aleatórias independentes, então $\mathbb{P}(A) = 0$ ou 1 para todo $A \in \mathscr{X}$.*

Demonstração. Seja $A \in \mathscr{X}$. No espírito da Proposição 2.16, vamos mostrar que A é independente de si mesmo. Para isso vamos considerar uma aproximação para A que é independente de A.

Como os eventos cilíndricos com respeito à sequência $\boldsymbol{X}$ formam uma álgebra, pelo Teorema 13.5, existem eventos $A_n \in \sigma(X_1, \dots, X_n)$ tais que

$$\mathbb{P}(A \triangle A_n) \to 0$$

quando $n \to \infty$. Daí pode-se verificar que

$$\mathbb{P}(A_n) \to \mathbb{P}(A) \quad \text{e} \quad \mathbb{P}(A \cap A_n) \to \mathbb{P}(A),$$

o que deixamos como um instrutivo exercício sobre aproximações (ao final deste capítulo).

Por outro lado, como $A \in \mathscr{X}$, segue que $A \in \sigma(X_{n+1}, X_{n+2}, \dots)$ para todo $n \in \mathbb{N}$. Logo, pelo Corolário 13.11, A é independente de A_n, ou seja,

$$\mathbb{P}(A \cap A_n) = \mathbb{P}(A) \cdot \mathbb{P}(A_n).$$

Tomando o limite em n, obtemos

$$\mathbb{P}(A) = \mathbb{P}(A) \cdot \mathbb{P}(A),$$

ou seja, $\mathbb{P}(A) = 0$ ou 1. □

Corolário 13.17. *Se Z é uma variável aleatória $\mathscr{X}$-mensurável, então Z é uma variável aleatória degenerada.*

Demonstração. Por hipótese, $F_Z(z) = \mathbb{P}(Z \leqslant z) = 0$ ou 1. Tomando $C = \sup\{z \in \mathbb{R} : F_Z(z) = 0\}$, obtemos $\mathbb{P}(Z = C) = 1$ q.c. □

13.4 Lei 0-1 de Hewitt-Savage

No Exemplo 13.15 foram exibidos alguns exemplos de eventos que não são caudais. Todavia, se a sequência $(X_n)_n$ associada a tais eventos for i.i.d., tais eventos satisfarão a outra importante lei 0-1, a Lei 0-1 de Hewitt-Savage.

Dizemos que uma função $\pi : \mathbb{N} \to \mathbb{N}$ é uma *permutação finita* se π é uma bijeção e $\pi(n) \neq n$ para no máximo finitos valores de $n \in \mathbb{N}$. Dada uma sequência $\boldsymbol{x} \in \mathbb{R}^{\mathbb{N}}$, definimos a sequência permutada $\pi(\boldsymbol{x}) = (x_{\pi(n)})_n$.

Definição 13.18 (Evento simétrico)**.** Seja $(\Omega, \mathcal{F}, \mathbb{P})$ um espaço de probabilidade e $\boldsymbol{X} = (X_n)_n$ uma sequência de variáveis aleatórias. Dizemos que um conjunto $B \in \mathcal{B}(\mathbb{R}^{\mathbb{N}})$ é *simétrico* se $\pi^{-1}(B) = B$ para toda permutação finita π. Dizemos que um evento $A \in \mathcal{F}$ é *simétrico com respeito à sequência* $\boldsymbol{X}$ se $A = \{\omega : X(\omega) \in B\}$ para algum conjunto simétrico $B \in \mathcal{B}(\mathbb{R}^{\mathbb{N}})$.

Teorema 13.19 (Lei 0-1 de Hewitt-Savage)**.** *Sejam $(X_n)_n$ uma sequência de variáveis aleatórias independentes e identicamente distribuídas, e A um evento simétrico com respeito à sequência $(X_n)_n$. Então $\mathbb{P}(A) = 0$ ou 1.*

Demonstração. Novamente, a ideia é mostrar que A é independente de si mesmo. De forma um pouco mais elaborada do que fizemos na Lei 0-1 de Kolmogorov, desta vez vamos considerar diferentes aproximações para A que são independentes entre si.

Tome $B \in \mathcal{B}(\mathbb{R}^{\mathbb{N}})$ simétrico tal que $A = \{\omega : \boldsymbol{X}(\omega) \in B\}$. Vamos mostrar que $\mathbb{P}_{\boldsymbol{X}}(B) = 0$ ou 1. Pelo Teorema 13.5, existem $C_n \in \mathcal{B}(\mathbb{R}^n)$ tais que

$$\mathbb{P}_{\boldsymbol{X}}(B \triangle B_n) \to 0,$$

onde $B_n = \{\boldsymbol{x} \in \mathbb{R}^{\mathbb{N}} : (x_1, \ldots, x_n) \in C_n\}$.

Vamos agora definir outra aproximação independente desta. Para cada $n \in \mathbb{N}$, seja $\pi_n : \mathbb{N} \to \mathbb{N}$ a permutação finita que troca de lugar os blocos $(1, \ldots, n)$ e $(n+1, \ldots, 2n)$, isto é,

$$\pi_n(j) = \begin{cases} j + n, & \text{se } 1 \leqslant j \leqslant n, \\ j - n, & \text{se } n + 1 \leqslant j \leqslant 2n, \\ j, & \text{se } j > 2n, \end{cases}$$

e defina

$$\tilde{B}_n = \pi_n^{-1}(B_n) = \{\boldsymbol{x} \in \mathbb{R}^{\mathbb{N}} : (x_{n+1}, \ldots, x_{2n}) \in C_n\}.$$

Vejamos primeiro que $\tilde{B}_n$ aproxima B. Como a sequência $\boldsymbol{X}$ é i.i.d., temos

$$\mathbb{P}_{\boldsymbol{X}} = \mathbb{P}_{\pi_n(\boldsymbol{X})}$$

para cada $n \in \mathbb{N}$. Sendo assim,

$$\mathbb{P}_{\boldsymbol{X}}(B \triangle B_n) = \mathbb{P}_{\pi_n(\boldsymbol{X})}(B \triangle B_n) = \mathbb{P}_{\boldsymbol{X}}(\pi_n^{-1}(B) \triangle \pi_n^{-1}(B_n)) = \mathbb{P}_{\boldsymbol{X}}(B \triangle \tilde{B}_n).$$

A segunda igualdade é devida ao fato que $\pi_n^{-1}(B \triangle B_n) = \pi_n^{-1}(B) \triangle \pi_n^{-1}(B_n)$, que é válida pois $\pi_n : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}^{\mathbb{N}}$ é uma bijeção. Na última igualdade usamos que B é simétrico. Logo,

$$\mathbb{P}_{\boldsymbol{X}}(B \triangle \tilde{B}_n) = \mathbb{P}_{\boldsymbol{X}}(B \triangle B_n) \to 0.$$

Daí pode-se verificar que

$$\mathbb{P}_{\boldsymbol{X}}(B_n \cap \tilde{B}_n) \to \mathbb{P}_{\boldsymbol{X}}(B), \quad \mathbb{P}_{\boldsymbol{X}}(B_n) \to \mathbb{P}_{\boldsymbol{X}}(B) \quad \text{e} \quad \mathbb{P}_{\boldsymbol{X}}(\tilde{B}_n) \to \mathbb{P}_{\boldsymbol{X}}(B),$$

o que novamente deixamos como exercício sobre aproximações.

Por outro lado, como $(X_1, \ldots, X_n)$ é independente de $(X_{n+1}, \ldots, X_{2n})$, pelo Corolário 13.10 temos que $\mathbb{P}_{\boldsymbol{X}}(B_n \cap \tilde{B}_n) = \mathbb{P}_{\boldsymbol{X}}(B_n) \cdot \mathbb{P}_{\boldsymbol{X}}(\tilde{B}_n)$ e, tomando o limite em n, obtemos

$$\mathbb{P}_{\boldsymbol{X}}(B) = \mathbb{P}_{\boldsymbol{X}}(B) \cdot \mathbb{P}_{\boldsymbol{X}}(B).$$

Portanto, $\mathbb{P}(A) = 0$ ou 1, o que conclui esta demonstração. □

Corolário 13.20. *Sejam $(X_n)_n$ uma sequência de variáveis aleatórias i.i.d. e $S_n = X_1 + \cdots + X_n$ para todo $n \in \mathbb{N}$. Então, $\mathbb{P}(A) = 0$ ou 1 para todo evento A caudal com respeito à sequência $(S_n)_n$.*

Demonstração. Em geral, o evento A não é caudal com respeito à sequência

$(X_n)_n$. Porém, observe que se A é caudal com respeito à sequência $(S_n)_n$, então A é um evento simétrico com respeito à sequência $(X_n)_n$. $\square$

Exemplo 13.21 (De volta ao Exemplo 13.15). São exemplos de eventos simétricos com respeito à sequência $(X_n)_n$:

(a) $\{X_1 + \cdots + X_n \in B \ i.v.\}$ para todo $B \in \mathcal{B}$;
(b) $\{\limsup_n X_1 + \cdots + X_n < C\}$ para todo $C \in \mathbb{R}$.

Portanto, se a sequência $(X_n)_n$ for i.i.d., vale a Lei 0-1 de Hewitt-Savage para os eventos acima. $\triangle$

Contra-exemplo 13.22. Seja $(X_n)_n$ uma sequência de variáveis aleatórias independentes com distribuição $\mathbb{P}(X_n = 0) = 1 - \mathbb{P}(X_n = 1) = 2^{-n}$ para todo $n \in \mathbb{N}$. Observe que o evento $A = \{\lim_n (X_1 \times \cdots \times X_n) = 1\}$ é simétrico com respeito à sequência $(X_n)_n$. Porém, $\mathbb{P}(A) = \prod_{n=1}^{\infty}(1 - 2^{-n}) \in (0, 1)$. Assim vemos que hipótese de as variáveis $(X_n)_n$ terem a mesma distribuição é crucial para a validade da Lei 0-1 de Hewitt-Savage, apesar de desnecessária para a Lei 0-1 de Kolmogorov. $\triangle$

13.5 Convergência de séries aleatórias

Seja $(X_n)_n$ uma sequência de variáveis aleatórias independentes. Conforme vimos na seção anterior o evento $\{\omega : \sum_{n=1}^{\infty} X_n(\omega) \text{ converge}\}$ é caudal, logo pela Lei 0-1 de Kolmogorov, sua probabilidade é zero ou um. Isto é, $\sum_{n=1}^{\infty} X_n$ converge quase certamente ou diverge quase certamente. Nesta seção estudaremos critérios que nos permitem determinar qual dois casos acima ocorre quase certamente.

Exemplo 13.23. Seja $(X_n)_n$ uma sequência de variáveis aleatórias independentes, com distribuição $\mathbb{P}(X_n = 1) = 1 - \mathbb{P}(X_n = 0) = 2^{-n}$ para todo $n \in \mathbb{N}$. Pelo Lema de Borel-Cantelli, $\mathbb{P}(X_n \neq 0 \text{ i.v}) = 0$ e portanto $\sum_{n=1}^{\infty} X_n$ converge quase certamente. $\triangle$

No exemplo acima, foi relativamente fácil concluir que a série converge quase certamente. Suponha agora que $(Z_n)_n$ é i.i.d. com distribuição $\mathbb{P}(Z_n = 1) =$

$\mathbb{P}(Z_n = -1) = \frac{1}{2}$ e considere a série harmônica com sinal aleatório $\sum_{n=1}^{\infty} \frac{Z_n}{n}$. Esta série converge ou diverge quase certamente? Não podemos proceder como no exemplo acima, e precisamos de critérios mais efetivos.

Inicialmente, estabeleçamos um critério para a convergência quase certa de uma sequência de variáveis aleatórias.

Lema 13.24. *Uma sequência $(X_n)_{n\in\mathbb{N}}$ de variáveis aleatórias converge quase certamente se, e somente se, $\lim_n \mathbb{P}(\sup_{k\in\mathbb{N}} |X_{n+k} - X_n| > \varepsilon) = 0$, para todo $\varepsilon > 0$.*

Demonstração. Observamos que uma sequência $(x_n)_n$ de números reais converge se, e somente se, $\lim_n(\sup_{k\geqslant n} x_k - \inf_{k\geqslant n} x_k) = 0$. Logo, $(X_n)_n$ converge q.c. se, e somente se, $\sup_{k\geqslant n} X_k - \inf_{k\geqslant n} X_k \overset{\text{q.c.}}{\to} 0$. Seja $Z_n = \sup_{k\geqslant n} X_k - \inf_{k\geqslant n} X_k$. Como $(Z_n)_n$ é uma sequência monótona, segue das Proposições 7.25 e 7.32 que $Z_n \overset{\text{q.c.}}{\to} 0$ se, e somente se, $Z_n \overset{\mathbb{P}}{\to} 0$. Mas como $\sup_{k\in\mathbb{N}} |X_{n+k} - X_n| \leqslant \sup_{k\geqslant n} X_k - \inf_{k\geqslant n} X_k \leqslant 2\sup_{k\in\mathbb{N}} |X_{n+k} - X_n|$, concluímos que $Z_n \overset{\mathbb{P}}{\to} 0$ se, e somente se, $\sup_{k\in\mathbb{N}} |X_{n+k} - X_n| \overset{\mathbb{P}}{\to} 0$, o que prova o lema. □

A seguir provaremos dois lemas. O primeiro é uma interessante desigualdade envolvendo o segundo momento da soma de variáveis aleatórias, pois vai na direção oposta à de Tchebyshev; enquanto o segundo lema é uma versão mais forte da Desigualdade de Tchebyshev e é interessante por si só.

Lema 13.25. *Sejam $c \in \mathbb{R}$ e $(X_n)_n$ uma sequência de variáveis aleatórias independentes tais que $\mathbb{E}X_n = 0$ e $\mathbb{P}(|X_n| \leqslant c) = 1$ para todo $n \in \mathbb{N}$. Então, para todo $\varepsilon > 0$, vale*

$$\mathbb{P}(S_n^2 \geqslant \varepsilon\, \mathbb{V}S_n) \geqslant \frac{(1-\varepsilon)^2}{3 + \frac{c^2}{\mathbb{V}S_n}},$$

onde $S_n = \sum_{k=1}^{n} X_k$.

Demonstração. Pela Desigualdade de Paley-Zygmund,

$$\mathbb{P}(S_n^2 \geqslant \varepsilon\, \mathbb{E}S_n^2) \geqslant (1-\varepsilon)^2 \frac{(\mathbb{E}S_n^2)^2}{\mathbb{E}S_n^4}.$$

Para estimar o quociente, expandimos como na demonstração do Teorema 8.5 e majoramos

$$\begin{aligned}\mathbb{E}S_n^4 &= \sum_{j=1}^n \mathbb{E}X_j^4 + 3\sum_{j=1}^n\sum_{k=1}^n \mathbb{1}_{k\neq j}\mathbb{E}X_j^2 \cdot \mathbb{E}X_k^2 \\ &\leqslant c^2 \sum_{j=1}^n \mathbb{E}X_j^2 + 3\Big(\sum_{j=1}^n \mathbb{V}X_j\Big)^2 = (3+\tfrac{c^2}{\mathbb{V}S_n})(\mathbb{V}S_n)^2,\end{aligned}$$

o que conclui a prova. □

Lema 13.26 (Desigualdade Maximal de Kolmogorov)**.** *Sejam $X_1, \ldots, X_n$ variáveis aleatórias independentes com média zero e segundo momento finito. Defina $S_k = \sum_{j=1}^k X_j$. Então, para todo $\varepsilon > 0$,*

$$\mathbb{P}\Big(\max_{1\leqslant k\leqslant n} |S_k| \geqslant \varepsilon\Big) \leqslant \frac{\mathbb{V}S_n}{\varepsilon^2}.$$

Demonstração. Defina $A_k = \{|S_1| < \varepsilon, \ldots, |S_{k-1}| < \varepsilon, |S_k| \geqslant \varepsilon\}$. Observe que $\sum_k \mathbb{1}_{A_k} = \mathbb{1}_{\{\max_{1\leqslant k\leqslant n}|S_k|\geqslant\varepsilon\}}$. Expandindo $(S_k + (S_n - S_k))^2$,

$$\begin{aligned}\mathbb{E}S_n^2 &\geqslant \textstyle\sum_{k=1}^n \mathbb{E}[S_n^2 \mathbb{1}_{A_k}] \\ &= \textstyle\sum_{k=1}^n (\mathbb{E}[S_k^2\mathbb{1}_{A_k}] + 2\cdot\mathbb{E}[S_k\mathbb{1}_{A_k}]\cdot\mathbb{E}[S_n - S_k] + \mathbb{E}[(S_n-S_k)^2\mathbb{1}_{A_k}]) \\ &\geqslant \textstyle\sum_{k=1}^n \mathbb{E}[\varepsilon^2\mathbb{1}_{A_k}] = \varepsilon^2\cdot\mathbb{P}(\max_{1\leqslant k\leqslant n}|S_k|\geqslant\varepsilon),\end{aligned}$$

onde utilizamos que $S_n - S_k$ tem média zero e é independente de $S_k\mathbb{1}_{A_k}$. □

O lema acima também segue diretamente da Desigualdade de Doob-Kolmogorov, provada no Capítulo 12.

Teorema 13.27 (Teorema de Uma Série)**.** *Seja $(X_n)_{n\in\mathbb{N}}$ uma sequência de variáveis aleatórias independentes com $\mathbb{E}X_n = 0$ para todo $n \in \mathbb{N}$. Então*

(a) *Se $\sum_{n=1}^{\infty} \mathbb{E}X_n^2 < \infty$, então $\mathbb{P}(\sum_{n=1}^{\infty} X_n$ converge$) = 1$.*
(b) *Se $\mathbb{P}(\sum_{n=1}^{\infty} X_n$ converge$) = 1$ e existe $c > 0$ tal que $\mathbb{P}(|X_n| \leqslant c) = 1$ para todo n, então $\sum_{n=1}^{\infty} \mathbb{E}X_n^2 < \infty$.*

Demonstração. Queremos estudar a convergência de $S_n = X_1 + \cdots + X_n$.

Começamos pelo item (a). Pela Desigualdade Maximal de Kolmogorov,

$$\begin{aligned}\mathbb{P}\Big(\sup_{k\in\mathbb{N}} |S_{n+k} - S_n| > \varepsilon\Big) &= \lim_m \mathbb{P}\Big(\max_{k\leqslant m} |S_{n+k} - S_n| > \varepsilon\Big) \\ &\leqslant \lim_m \frac{\sum_{k=n}^{n+m} \mathbb{E}X_k^2}{\varepsilon^2} = \frac{\sum_{k=n}^{\infty} \mathbb{E}X_k^2}{\varepsilon^2} \to 0\end{aligned}$$

para todo $\varepsilon > 0$. Pelo Lema 13.24, S_n converge quase certamente.

Passamos ao item (b). Suponha que $\mathbb{V}S_n \to +\infty$ e que existe $c > 0$ tal que $|X_n| \leqslant c$ q.c. para todo n. Pelos Lema de Fatou e Lema 13.25,

$$\mathbb{P}(S_n^2 \geqslant \tfrac{1}{2}\mathbb{E}S_n^2 \text{ i.v.}) \geqslant \limsup_n \mathbb{P}(S_n^2 \geqslant \tfrac{1}{2}\mathbb{E}S_n^2) \geqslant \lim_n \tfrac{1}{4}\Big(3 + \tfrac{c^2}{\mathbb{V}S_n}\Big)^{-1} = \tfrac{1}{12},$$

donde concluímos que $\mathbb{P}(S_n \text{ diverge}) > 0$. □

Exemplo 13.28 (p-série aleatória)*.* Seja $(Z_n)_n$ uma sequência de variáveis aleatórias i.i.d. com distribuição $\mathbb{P}(Z_n = 1) = \mathbb{P}(Z_n = -1) = \frac{1}{2}$ para todo $n \in \mathbb{N}$. Considere a p-série aleatória $\sum_{n=1}^{\infty} \frac{Z_n}{n^p}$ com $p > 0$. Como $\mathbb{P}(|X_n| \leqslant 1) = 1$ e $\mathbb{E}[\frac{Z_n}{n^p}] = 0$ para todo $n \in \mathbb{N}$, segue do Teorema de Uma Série que $\sum_{n=1}^{\infty} \frac{Z_n}{n^p}$ converge quase certamente se, e somente se, $\sum_{n=1}^{\infty} \frac{1}{n^{2p}}$ converge. Portanto, $\sum_{n=1}^{\infty} \frac{Z_n}{n^p}$ converge quase certamente se $p > \frac{1}{2}$ e diverge quase certamente se $0 < p \leqslant \frac{1}{2}$. Em particular, a série harmônica com sinal aleatório, mencionada no início desta seção, converge quase certamente. △

O próximo teorema é um refinamento do anterior, que retira a hipótese de $\mathbb{E}X_n = 0$. Em troca, teremos que verificar a convergência de duas séries de números reais ao invés de uma.

Teorema 13.29 (Teorema das Duas Séries)**.** *Seja $(X_n)_{n\in\mathbb{N}}$ uma sequência de variáveis aleatórias independentes. Valem as seguintes proposições:*

(a) *Se $\sum_{n=1}^{\infty}\mathbb{E}X_n$ e $\sum_{n=1}^{\infty}\mathbb{V}X_n$ convergem, então $\mathbb{P}(\sum_{n=1}^{\infty}X_n$ converge$)=1$.*

(b) *Se $\mathbb{P}(\sum_{n=1}^{\infty}X_n$ converge$)=1$ e existe $c>0$ tal que $\mathbb{P}(|X_n|\leqslant c)=1$ para todo $n\in\mathbb{N}$, então $\sum_{n=1}^{\infty}\mathbb{E}X_n$ e $\sum_{n=1}^{\infty}\mathbb{V}X_n$ convergem.*

Demonstração. Começamos pelo item (a). Como $\sum_{n=1}^{\infty}\mathbb{V}X_n=\sum_{n=1}^{\infty}\mathbb{E}(X_n-\mathbb{E}X_n)^2$ converge, pelo Teorema de Uma Série, $\sum_{n=1}^{\infty}(X_n-\mathbb{E}X_n)$ converge q.c. Por hipótese, $\sum_{n=1}^{\infty}\mathbb{E}X_n$ também converge, logo $\sum_{n=1}^{\infty}X_n$ converge q.c.

Passamos ao item (b). Seja $(Y_n)_n$ uma sequência de variáveis aleatórias independentes entre si e independentes de $(X_n)_n$, tais que $Y_n\sim X_n$ para todo $n\in\mathbb{N}$ (para construir tal sequência é suficiente tomar o produto de $(\Omega,\mathcal{F},\mathbb{P})$ consigo mesmo). Se $\sum_{n=1}^{\infty}X_n$ converge q.c., então $\sum_{n=1}^{\infty}Y_n$ converge q.c. e consequentemente, $\sum_{n=1}^{\infty}(X_n-Y_n)$ também converge quase certamente.

Por hipótese, $\mathbb{P}(|X_n|\leqslant c)=1$ para todo $n\in\mathbb{N}$, logo $\mathbb{P}(|X_n-Y_n|\leqslant 2c)=1$. Como $\mathbb{E}[X_n-Y_n]=0$, pelo item (b) do Teorema de Uma Série, $\sum_{n=1}^{\infty}\mathbb{E}[X_n-Y_n]^2$ converge. Por outro lado, como X_n e Y_n são independentes,

$$\mathbb{E}(X_n-Y_n)^2=\mathbb{V}[X_n-Y_n]=2\mathbb{V}X_n.$$

Logo, $\sum_{n=1}^{\infty}\mathbb{V}X_n$ converge e, utilizando o item (a) do Teorema de Uma Série, $\sum_{n=1}^{\infty}(X_n-\mathbb{E}X_n)$ converge q.c. Como $\sum_{n=1}^{\infty}X_n$ converge q.c., concluímos que $\sum_{n=1}^{\infty}\mathbb{E}X_n$ converge. □

O próximo teorema nos fornece uma condição equivalente à convergência quase certa de uma série de variáveis independentes. Em compensação, ganhamos uma terceira série de números reais para analisar.

Dada uma variável X aleatória e $c>0$, definimos a variável truncada

$$X^c(\omega)=\begin{cases}X(\omega), & \text{se } |X(\omega)|\leqslant c,\\ 0, & \text{se } |X(\omega)|>c.\end{cases}$$

Teorema 13.30 (Teorema das Três Séries)**.** *Sejam $(X_n)_n$ uma sequência de variáveis aleatórias independentes e $(X_n^c)_n$ a sequência truncada em $c > 0$.*

(a) *Se existe $c > 0$ tal que as três séries $\sum_{n=1}^{\infty} \mathbb{E}X_n^c$, $\sum_{n=1}^{\infty} \mathbb{V}X_n^c$ e $\sum_{n=1}^{\infty} \mathbb{P}(|X_n| > c)$ convergem, então $\mathbb{P}(\sum_{n=1}^{\infty} X_n$ converge$) = 1$.*

(b) *Reciprocamente, se $\mathbb{P}(\sum_{n=1}^{\infty} X_n$ converge$) = 1$, então as três séries $\sum_{n=1}^{\infty} \mathbb{E}X_n^c$, $\sum_{n=1}^{\infty} \mathbb{V}X_n^c$ e $\sum_{n=1}^{\infty} \mathbb{P}(|X_n| > c)$ convergem para todo $c > 0$.*

Demonstração. Começamos pelo item (a). Pelo Teorema das Duas Séries, $\sum_{n=1}^{\infty} X_n^c$ converge q.c. Por outro lado, pelo Lema de Borel-Cantelli,

$$\mathbb{P}(X_n \neq X_n^c \text{ i.v.}) = \mathbb{P}(|X_n| > c \text{ i.v.}) = 0.$$

Isto é, quase certamente, X_n e X_n^c diferem para, no máximo, finitos valores do índice n. Como $\sum_{n=1}^{\infty} X_n^c$ converge quase certamente, segue que $\sum_{n=1}^{\infty} X_n$ também converge quase certamente.

Passamos ao item (b). Seja $c > 0$. Como $\sum_{n=1}^{\infty} X_n$ converge quase certamente, $X_n \overset{\text{q.c.}}{\to} 0$. Pela Proposição 7.16, $\mathbb{P}(|X_n| > c \text{ i.v.}) = 0$, logo $\sum_{n=1}^{\infty} X_n^c$ converge q.c. Logo, pelo Teorema das Duas Séries, $\sum_{n=1}^{\infty} \mathbb{E}X_n^c$ e $\sum_{n=1}^{\infty} \mathbb{V}X_n^c$ convergem. Ademais, como as variáveis $(X_n)_n$ são independentes, o Lema de Borel-Cantelli assegura que $\sum_{n=1}^{\infty} \mathbb{P}(|X_n| > c)$ converge. □

13.6 Exercícios

§13.1

1. Sejam $\{\mathcal{F}_n\}_n$ σ-álgebras em Ω tais que $\mathcal{F}_k \subseteq \mathcal{F}_{k+1}$ para todo k. Prove que $\mathcal{A} = \cup_n \mathcal{F}_n$ é uma álgebra. Prove que essa propriedade pode falhar sem a hipótese de que $\mathcal{F}_k \subseteq \mathcal{F}_{k+1}$.

2. Sejam $A, B \in \mathcal{F}$ eventos e sejam $(A_n)_n$ e $(B_n)_n$ sequências de eventos tais que $\mathbb{P}(A_n \triangle A) \to 0$ e $\mathbb{P}(B_n \triangle B) \to 0$. Prove que:

(a) $\mathbb{P}(A_n) \to \mathbb{P}(A)$ e $\mathbb{P}(B_n) \to \mathbb{P}(B)$.

(b) $\mathbb{P}(A_n \cap B_n) \to \mathbb{P}(A \cap B)$.
(c) $\mathbb{P}(A \cap A_n) \to \mathbb{P}(A)$.
(d) Se $A = B$, então $\mathbb{P}(A_n \triangle B_n) \to 0$.

§13.3

3. Prove que os eventos do Exemplo 13.14 são caudais com respeito à sequência $(X_n)_n$.

4. Dado $0 < \alpha < 1$, dê um exemplo de uma sequência $(X_n)_n$ de variáveis aleatórias independentes tais que $\mathbb{P}(X_n = X_1 \text{ i.v.}) = \alpha$. Dê também um exemplo em que as variáveis sejam i.i.d.

5. Dê um exemplo de uma sequência de variáveis aleatórias $(X_n)_n$ e um evento A, caudal com respeito à sequência $(X_n)_n$, tais que $\mathbb{P}(A) \in (0, 1)$.

6. Forneça uma nova prova da Lei 0-1 de Kolmogorov utilizando o Teorema 12.45 e o Corolário 13.11.

§13.4

7. Prove que os eventos do Exemplo 13.15 são simétricos com respeito à sequência $(X_n)_n$.

8. Sejam $(X_n)_n$ uma sequência de variáveis aleatórias e $S_n = \sum_{k=1}^{n} X_k$ para todo $n \in \mathbb{N}$. Mostre que se A é um evento caudal com respeito à sequência $(S_n)_n$, então A é um evento simétrico com respeito à sequência $(X_n)_n$.

9. Seja $\mathcal{S}$ a classe de todos os eventos simétricos com respeito às variáveis aleatórias $(X_n)_n$. A classe $\mathcal{S}$ é uma σ-álgebra? Justifique.

10. Sejam B, B_n e $\tilde{B}_n$ tais que $\mathbb{P}(B \triangle B_n) \to 0$ e $\mathbb{P}(B \triangle \tilde{B}_n) \to 0$.

(a) Mostre que $B \cup B_n \cup \tilde{B}_n \subseteq (B \cap B_n \cap \tilde{B}_n) \cup (B \triangle B_n) \cup (B \triangle \tilde{B}_n)$.
(b) Mostre que $\mathbb{P}(B_n \cap \tilde{B}_n) \to \mathbb{P}(B)$ e $\mathbb{P}(B_n) \to \mathbb{P}(B)$.

§13.5

11. Seja $(X_n)_n$ uma sequência de variáveis aleatórias independentes. Determine se $\sum_{n=1}^{\infty} X_n$ converge ou diverge quase certamente nos casos abaixo. Quando a resposta for converge quase certamente, determine o valor da soma se possível.

(a) $\mathbb{P}(X_n = 2^{-n}) = \mathbb{P}(X_n = 0) = \frac{1}{2}$ para todo $n \in \mathbb{N}$.
(b) $\mathbb{P}(X_n = 1) = 1 - \mathbb{P}(X_n = 0) = \frac{1}{n}$ para todo $n \in \mathbb{N}$.
(c) $\mathbb{P}(|X_n| \leqslant \frac{1}{n^2}) \geqslant 1 - \frac{1}{n^2}$ para todo $n \in \mathbb{N}$.
(d) $\mathbb{P}(X_n = \frac{1}{2^n}) = \mathbb{P}(X_n = -\frac{1}{2^n}) = \frac{1}{2}$ para todo $n \in \mathbb{N}$.

12. Seja $(Z_n)_n$ uma sequência de variáveis aleatórias independentes, com $\mathbb{E}Z_n = 0$ para todo $n \in \mathbb{N}$.

(a) Mostre que $\sum_{n=1}^{\infty} \mathbb{E}\big[\frac{Z_n^2}{1+|Z_n|}\big]$ converge se, e somente se,

$$\sum_{n=1}^{\infty} \mathbb{E}\Big[Z_n^2 \mathbb{1}_{\{|Z_n| \leqslant 1\}} + |Z_n| \mathbb{1}_{\{|Z_n| > 1\}}\Big] \text{ converge.}$$

(b) Mostre que, se $\sum_{n=1}^{\infty} \mathbb{E}\big[\frac{Z_n^2}{1+|Z_n|}\big]$ converge, então $\sum_{n=1}^{\infty} \mathbb{E}Z_n^c$ e $\sum_{n=1}^{\infty} \mathbb{V}Z_n^c$ convergem quando truncadas em $c = 1$.
(c) Utilize o Teorema das Três Séries para concluir que, se $\sum_{n=1}^{\infty} \mathbb{E}\big[\frac{Z_n^2}{1+|Z_n|}\big]$ converge, então $\sum_{n=1}^{\infty} Z_n$ converge quase certamente.

13. Dê um exemplo de uma sequência $(X_n)_n$ de variáveis aleatórias independentes e uma constante $c > 0$ tais que $\sum_{n=1}^{\infty} \mathbb{E}X_n^c$ e $\sum_{n=1}^{\infty} \mathbb{V}X_n^c$ convirjam, e $\sum_{n=1}^{\infty} \mathbb{P}(|X_n| > c)$ divirja.

14. Dê um exemplo de uma sequência $(X_n)_n$ de variáveis aleatórias independentes e uma constante $c > 0$ tais que $\sum_{n=1}^{\infty} \mathbb{P}(|X_n| > c)$ e $\sum_{n=1}^{\infty} \mathbb{V}X_n^c$ convirjam, e $\sum_{n=1}^{\infty} \mathbb{E}X_n^c$ divirja.

15. Seja $(X_n)_n$ uma sequência de variáveis aleatórias independentes com média zero. Suponha que existe $c > 0$ tal que $\mathbb{P}(|X_n| \leqslant c) = 1$ para todo n, e que $\mathbb{E}S_n^2 \to +\infty$, onde $S_n = \sum_{k=1}^{n} X_k$. Mostre que $\limsup_n \frac{S_n^2}{\mathbb{E}S_n^2} \geqslant 1$ q.c.

Capítulo 14

Teoria Ergódica

Este capítulo explora os conceitos mais básicos da Teoria Ergódica, que são fundamentais para o estudo de Probabilidade e Processos Estocásticos em geral. A maioria dos usos da Teoria Ergódica no estudo de Processos Estocásticos consiste no fato de que certos eventos definidos em termos de processos satisfazendo determinadas propriedades deverão ter probabilidade 0 ou 1. Além disso, fazendo-se observações ao longo das trajetórias de certos processos, o valor médio dessas observações tende a se estabilizar, o que significa uma nova forma de Lei dos Grandes Números.

14.1 Transformações que preservam medida

Nesta seção introduzimos conceitos básicos que serão usados no restante deste capítulo, e apresentaremos o Teorema de Recorrência de Poincaré.

Definição 14.1 (Sequência aleatória e operador deslocamento)**.** No espaço das sequências de números reais, definimos o *operador deslocamento* $\theta : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}^{\mathbb{N}}$, dado por $\theta(x_1, x_2, x_3 \dots) = (x_2, x_3, x_4 \dots)$. Se $\boldsymbol{X} = (X_1, X_2, \dots)$ é uma sequência aleatória, $\theta\boldsymbol{X}$ é a sequência deslocada $\theta\boldsymbol{X} = (X_2, X_3, \dots)$.

Definição 14.2 (Sequência estacionária)**.** A sequência $\boldsymbol{X} = (X_1, X_2, \dots)$ é

dita *estacionária* se $\theta \boldsymbol{X} \sim \boldsymbol{X}$. Isto é, $\boldsymbol{X}$ é estacionária se

$$\mathbb{P}(\boldsymbol{X} \in B) = \mathbb{P}(\theta \boldsymbol{X} \in B)$$

para todo $B \in \mathcal{B}(\mathbb{R}^{\mathbb{N}})$, onde a σ-álgebra $\mathcal{B}(\mathbb{R}^{\mathbb{N}})$ foi definida na Seção 13.1.

Vejamos alguns exemplos de sequências estacionárias, com algumas observações preliminares sobre o comportamento de sua média observada.

Exemplo 14.3. Toda sequência $(X_n)_n$ i.i.d. é estacionária. Se X_1 é integrável, pela Lei Forte dos Grandes Números, $n^{-1}(X_1 + \cdots + X_n) \overset{\text{q.c.}}{\to} \mathbb{E}X_1$, ou seja, a média observada converge a um valor determinístico. $\triangle$

Exemplo 14.4 (Sequência estacionária com correlações não-nulas). Sejam $(Z_n)_n$ variáveis aleatórias i.i.d. com distribuição $\mathcal{U}[-1,1]$ e $Y \sim \mathcal{U}[-1,1]$ independente de $(Z_n)_n$. Defina $X_n = Z_n + Y$. Essa sequência é estacionária mas $\mathbf{Cov}(X_j, X_k) = \frac{1}{3}$ para todos $j, k \in \mathbb{N}$ distintos. Este é um exemplo de sequência que tem "muita memória". Veja que $n^{-1}(X_1 + \cdots + X_n) \overset{\text{q.c.}}{\to} Y$, ou seja, o limite é aleatório. $\triangle$

Exemplo 14.5 (Sequência estacionária com correlações não-nulas). Defina $\boldsymbol{X}$ de forma que $\boldsymbol{X} = (0,1,0,1,0,1,\dots)$ e $\boldsymbol{X} = (1,0,1,0,1,0,\dots)$ com probabilidade $\frac{1}{2}$ cada. Essa sequência é estacionária mas $\mathbf{Cov}(X_j, X_k) = \frac{1}{4}(-1)^{k-j}$ para todos $j, k \in \mathbb{N}$. Veja que, apesar de que essa sequência tem "muita memória", $n^{-1}(X_1 + \cdots + X_n) \overset{\text{q.c.}}{\to} \frac{1}{2} = \mathbb{E}X_1$, que é determinístico. $\triangle$

Observação 14.6. Uma forma de produzir várias sequências estacionárias (ou de verificar que uma dada sequência é estacionária) é a seguinte. Dada uma sequência estacionária $\boldsymbol{X}$ e uma função mensurável $f : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}^{\mathbb{N}}$ tal que $\theta \circ f = f \circ \theta$, a sequência aleatória $f(\boldsymbol{X})$ é estacionária. Por exemplo, se $(X_n)_n$ é estacionária, então a sequência $(Z_n)_n$ dada por $Z_n = X_{n+1} + X_{n+2}^2 X_{n+3}$ é estacionária. $\triangle$

O estudo de sequências estacionárias pode servir-se da Teoria Ergódica a partir da conexão que agora passamos a descrever.

Seja $(\Omega, \mathcal{F}, \mathbb{P})$ um espaço de probabilidade que modela um certo experimento aleatório. Seja $\boldsymbol{X}$ uma sequência de variáveis aleatórias que são observadas

ao longo desse experimento. Podemos ver $\boldsymbol{X}$ como uma função definida em Ω, que toma valores no espaço $E = \mathbb{R}^{\mathbb{N}}$, mensurável com respeito às σ-álgebras $\mathcal{F}$ em Ω e $\mathcal{E} = \mathcal{B}(\mathbb{R}^{\mathbb{N}})$ em E. (Alguns livros dizem que $\boldsymbol{X}$ nada mais é do que uma variável aleatória que toma valores em E ao invés de $\mathbb{R}$, mas é mais comum dizer que $\boldsymbol{X}$ é um "elemento aleatório" e reservar o termo "variável aleatória" para o caso em que tomam-se valores em $\mathbb{R}$.)

Voltemos a falar de $\boldsymbol{X}$. Definindo a medida $\mu = \mathbb{P}_{\boldsymbol{X}}$ em $(E, \mathcal{E})$, temos que $(E, \mathcal{E}, \mu)$ é um espaço de probabilidade. Chamemos de T o operador de deslocamento θ nesse espaço. Observe que $T : E \to E$ é um operador $\mathcal{E}$-mensurável. Além disso, $\boldsymbol{X}$ é uma sequência estacionária se, e somente se, $\mu(T^{-1}(A)) = \mu(A)$ para todo $A \in \mathcal{E}$.

Neste capítulo, vamos estudar operadores determinísticos T definidos em espaços de probabilidade $(E, \mathcal{E}, \mu)$, sem perder de vista que tudo o que fizermos com essa notação e terminologia terá suas consequências no estudo de sequências de variáveis aleatórias.

Definição 14.7 (Transformação que preserva medida)**.** Dado um espaço de probabilidade $(E, \mathcal{E}, \mu)$ e um operador mensurável $T : E \to E$, dizemos que T é uma *transformação que preserva medida* se $T_*\mu = \mu$, isto é, $\mu(T^{-1}(A)) = \mu(A)$ para todo $A \in \mathcal{E}$.

Exemplo 14.8. Sejam $E = \{1, \ldots, n\}$, $\mathcal{E} = \mathcal{P}(E)$ e $\mu(A) = \frac{\#A}{n}$, então toda bijeção $T : E \to E$ é uma transformação que preserva medida. $\triangle$

Observamos que, se T preserva a medida μ, então segue do Teorema 5.65 que

$$\int_E f \, \mathrm{d}\mu = \int_E (f \circ T) \, \mathrm{d}\mu$$

para toda $f : \Omega \to [0, +\infty]$ mensurável.

A proposição abaixo segue do Teorema 3.37 (unicidade de medidas).

Proposição 14.9. *Sejam $(E, \mathcal{E}, \mu)$ um espaço de probabilidade e $T : E \to E$ um operador mensurável. Se $\mu(T^{-1}(A)) = \mu(A)$ para todo A em algum π-sistema que gera $\mathcal{E}$, então T é uma transformação que preserva medida.*

Exemplo 14.10 (Rotação do círculo). Sejam $E = [0,1)$, $\mathcal{E} = \mathcal{B}([0,1))$ e μ a medida de Lebesgue em $[0,1)$. Dado $\lambda \in E$, defina $T(x) = x + \lambda - \lfloor x + \lambda \rfloor \in E$. Se interpretamos o espaço E como sendo um círculo ("colando" os extremos do intervalo), essa transformação é equivalente a uma rotação de $2\pi\lambda$ radianos. Observe que $\mu(T^{-1}[a,b)) = \mu([a,b))$ para todo $[a,b) \subseteq [0,1)$. Pela Proposição 14.9, podemos concluir que T preserva medida. $\triangle$

Exemplo 14.11. Sejam $E = [0,1)$, $\mathcal{E} = \mathcal{B}([0,1))$ e μ a medida de Lebesgue em $[0,1)$. Fixe $n \in \mathbb{N}$, e defina $S(x) = nx - \lfloor nx \rfloor \in E$. Novamente, $\mu(S^{-1}[a,b)) = \mu([a,b))$ para todo $[a,b) \subseteq [0,1)$. Pela Proposição 14.9, podemos concluir que S é uma transformação que preserva medida. $\triangle$

Em geral estaremos interessados em estudar o comportamento da trajetória definida através de iterações sucessivas de uma transformação que preserva medida a partir de um ponto do espaço E. De modo mais preciso, dados $x \in E$ e T uma transformação que preserva medida, definimos a *órbita de x* como a sequência $(x, T(x), T^2(x), \dots)$. O teorema a seguir nos diz que, dado qualquer conjunto $A \in \mathcal{E}$, a órbita de quase todo ponto de A deve regressar ao conjunto A infinitas vezes.

Teorema 14.12 (Teorema de Recorrência de Poincaré)**.** *Sejam $(E, \mathcal{E}, \mu)$ um espaço de probabilidade e $T : E \to E$ uma transformação que preserva medida. Dado $A \in \mathcal{E}$, tomando $A' = \{x \in A : T^n(x) \in A \text{ i.v.}\}$, temos $\mu(A \setminus A') = 0$.*

Demonstração. Para cada $n \in \mathbb{N}$, seja

$$C_n = \{x \in A : T^k(x) \notin A,\ \forall k \geqslant n\}.$$

Neste caso podemos escrever $A' = A \setminus \cup_{n=1}^{\infty} C_n$. Como

$$C_n = A \setminus \cup_{k=n}^{\infty} T^{-k}(A)$$

e T é mensurável, $C_n \in \mathcal{E}$ para todo $n \in \mathbb{N}$ e, por conseguinte, $A' \in \mathcal{E}$. Logo, basta mostrar que $\mu(C_n) = 0$ para todo $n \in \mathbb{N}$. Como μ é uma medida finita

e T preserva a medida μ,

$$\begin{aligned}
\mu(C_n) &= \mu(A \setminus \cup_{k=n}^{\infty} T^{-k}(A)) \\
&\leqslant \mu\big(\big(\cup_{k=0}^{\infty} T^{-k}(A)\big) \setminus \cup_{k=n}^{\infty} T^{-k}(A)\big) \\
&= \mu(\cup_{k=0}^{\infty} T^{-k}(A)) - \mu(\cup_{k=n}^{\infty} T^{-k}(A)) \\
&= \mu(\cup_{k=0}^{\infty} T^{-k}(A)) - \mu(T^{-n}(\cup_{k=0}^{\infty} T^{-k}(A))) \\
&= \mu(\cup_{k=0}^{\infty} T^{-k}(A)) - \mu(\cup_{k=0}^{\infty} T^{-k}(A)) = 0,
\end{aligned}$$

concluindo a prova. □

Observe que o fato de μ ser uma medida finita foi crucial nessa demonstração, pois nesse caso $\mu(B \setminus D) = \mu(B) - \mu(D)$ sempre que $D \subseteq B$. Intuitivamente, como o espaço E tem medida finita, as órbitas que começam em A não têm para onde fugir, restando-lhes apenas a alternativa de retornar ao conjunto A. Como contra-exemplo, considere $(\mathbb{R}, \mathcal{B}, m)$, $T(x) = x + 2$, $A = [0, 1]$. Veja que T é uma translação, logo preserva a medida de Lebesgue, mas a órbita de todo ponto do intervalo $[0, 1]$ foge para a direita sem nunca retornar.

Em termos de sequências estacionárias, o Teorema de Recorrência de Poincaré diz que, se uma dada condição é observada em alguma região de índices de um processo estocástico estacionário, então essa mesma condição será cumprida em infinitas outras regiões de índices.

14.2 Teorema Ergódico de Birkhoff

Sejam T uma transformação que preserva medida e $A \in \mathcal{E}$. O Teorema de Recorrência de Poincaré nos diz que a órbita de quase todo ponto $x \in A$ retornará ao conjunto A infinitas vezes. É natural perguntarmos com que frequência ocorrem tais visitas. Mais precisamente, será que a frequência relativa

$$\frac{\#\{k \in \{0, \ldots, n-1\} : T^k(x) \in A\}}{n}$$

com que a órbita de x visita A durante seus n primeiros passos tem algum comportamento assintótico quando $n \to \infty$?

A pergunta acima pode ser colocada em um contexto mais geral. Seja $f : E \to \mathbb{R}$ uma função mensurável, isto é, um observável que depende do estado x da nossa dinâmica. O valor médio dos valores observados de f ao longo dos n primeiros passos da órbita de x é dado pelo quociente

$$\frac{1}{n}\sum_{k=0}^{n-1} f(T^k(x)).$$

Há algum comportamento de longo prazo para o quociente acima? Ele converge? Se sim, qual o limite? Converge em que sentido? Poderia convergir para o valor médio de f com respeito a μ?[23]

É intuitivo que este comportamento de longo prazo deva depender de quanto a transformação T revolve o espaço E.

Definição 14.13. Dizemos que um conjunto $A \in \mathcal{E}$ *é T-invariante* se $T^{-1}(A) = A$. Denotamos por $\mathcal{I}_T = \{A \in \mathcal{E} : T^{-1}(A) = A\}$ a classe dos conjuntos T-invariantes. Dizemos que uma função mensurável $f : E \to \mathbb{R}$ é *T-invariante* se $f = f \circ T$.

Em palavras, a classe $\mathcal{I}_T$ é formada pelas regiões do espaço E que não se mesclam com seu complementar quando aplicamos T.

Exercício 14.14. Mostre que $\mathcal{I}_T$ é uma σ-álgebra. Mostre que f é T-invariante se, e somente se, for $\mathcal{I}_T$-mensurável. $\triangle$

Daqui em diante vamos usar $\mathbf{E}$ para denotar a integral com respeito a μ. Com isso, poderemos usar a conveniente notação de esperança condicional.

[23] Questões como estas precederam à abordagem matemática e têm suas origens na Física-Estatística, notadamente no desenvolvimento da Teoria dos Gases com Boltzmann e outros. Um grande avanço, do ponto de vista matemático, ocorreu no ano de 1931 quando Birkhoff e von Neumann anunciaram, de modo independente, suas versões do Teorema Ergódico.

Teorema 14.15 (Teorema Ergódico de Birkhoff)**.** *Sejam T uma transformação que preserva medida no espaço de probabilidade $(E, \mathcal{E}, \mu)$ e $f : E \to \mathbb{R}$ uma função mensurável e integrável. Então,*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f \circ T^k = \mathbf{E}[f|\mathcal{I}_T] \qquad q.t.p.$$

Antes da demostração do Teorema de Birkhoff, precisaremos do lema abaixo

Lema 14.16. *Seja $T : E \to E$ transformação que preserva medida e $g : E \to \mathbb{R}$ uma função mensurável e integrável. Então*

$$\int_A g \, \mathrm{d}\mu \geqslant 0,$$

onde

$$A = \Big\{ x \in E : \limsup_n \frac{1}{n} \sum_{k=0}^{n-1} g(T^k(x)) > 0 \Big\}.$$

Demonstração. Defina $S_n = \sum_{j=0}^{n-1} g \circ T^j$ e $G_n = \max\{S_1, \ldots, S_n\}$. Observe que $(G_n)_n$ é não-decrescente, que $G_n \circ T + g = \max\{S_2, \ldots, S_{n+1}\}$ e que

$$G_{n+1} = \max\{S_1, \ldots, S_{n+1}\} = \max\{g, G_n \circ T + g\} = g + [G_n \circ T]^+.$$

Subtraindo,

$$G_{n+1} - G_n \circ T = g + [G_n \circ T]^-.$$

Agora, observe que, se $x \in A$, então existe $k \in \mathbb{N}$ tal que $G_k(T(x)) \geqslant 0$ e, neste caso, $G_{n+1}(x) - G_n(T(x)) = g(x)$ para todo $n \geqslant k$. Portanto,

$$G_{n+1}(x) - G_n(T(x)) \to g(x) \quad \text{para todo } x \in A.$$

Tomando o limite dentro da integral, obtemos

$$0 \leqslant \int_A (G_{n+1} - G_n) \mathrm{d}\mu = \int_A (G_{n+1} - G_n \circ T) \mathrm{d}\mu \to \int_A g \, \mathrm{d}\mu,$$

que é o que queríamos demonstrar. Na igualdade usamos que T preserva medida e $A \in \mathcal{I}_T$. Para justificar o limite dentro da integral, observamos que $|G_{n+1} - G_n \circ T| \leqslant |g| + |g \circ T|$, que é integrável por hipótese. $\square$

Demonstração do Teorema Ergódico de Birkhoff. Seja $\varepsilon > 0$ e tome

$$g = f - \mathbf{E}[f|\mathcal{I}_T] - \varepsilon.$$

Como $f = \mathbf{E}[f|\mathcal{I}_T] + \varepsilon + g$ e os dois primeiros termos são T-invariantes,

$$\frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k = \mathbf{E}[f|\mathcal{I}_T] + \varepsilon + \frac{1}{n}\sum_{k=0}^{n-1} g \circ T^k \qquad \text{q.t.p.}$$

para todo $n \in \mathbb{N}$. Definindo o conjunto A como no Lema 14.16, temos que

$$0 \leqslant \int_A (f - \mathbf{E}[f|\mathcal{I}_T] - \varepsilon)\,\mathrm{d}\mu = -\varepsilon\mu(A),$$

donde $\mu(A) = 0$, e, portanto,

$$\limsup_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k \leqslant \mathbf{E}[f|\mathcal{I}_T] + \varepsilon \qquad \text{q.t.p.}$$

Aplicando o mesmo argumento à função $g = -f + \mathbf{E}[f|\mathcal{I}_T] - \varepsilon$, obtemos

$$\liminf_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k \geqslant \mathbf{E}[f|\mathcal{I}_T] - \varepsilon \qquad \text{q.t.p.}$$

e como ε é arbitrário, isso conclui a prova. $\square$

Corolário 14.17. *Se T é uma transformação que preserva a medida μ, então, para todos $A, B \in \mathcal{E}$,*

$$\lim_n \frac{1}{n}\sum_{k=0}^{n-1} \mu\Big(T^{-k}(A) \cap B\Big) = \mathbf{E}\big[\mathbb{1}_B \cdot \mathbf{E}[\mathbb{1}_A|\mathcal{I}_T]\big].$$

Demonstração. Denotando $f = \mathbb{1}_A$, temos $\mu(T^{-k}(A) \cap B) = \mathbf{E}[\mathbb{1}_B \cdot f \circ T^k]$. Assim,

$$\frac{1}{n}\sum_{k=0}^{n-1} \mu\Big(T^{-k}(A) \cap B\Big) = \mathbf{E}\Big[\mathbb{1}_B \cdot \frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k\Big] \to \mathbf{E}\Big[\mathbb{1}_B \cdot \mathbf{E}[\mathbb{1}_A | \mathcal{I}_T]\Big].$$

A convergência vale pelos Teorema Ergódico de Birkhoff e Teorema da Convergência Dominada, já que $|\mathbb{1}_B(x)| \leqslant 1$ e $|\frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k(x)| \leqslant 1$ para todo $x \in E$. $\square$

14.3 Transformações ergódicas e misturadoras

Voltemos à questão da frequência com que as órbitas visitam um dado conjunto. Seja $(E, \mathcal{E}, \mu)$ um espaço de probabilidade e T uma transformação que preserva medida. Dado $A \in \mathcal{E}$, o Teorema Ergódico nos diz que a frequência de visitas a A, dada por

$$\lim_n \frac{\#\{k \in \{0, \ldots, n-1\} : T^k(x) \in A\}}{n},$$

está bem definida. Queremos saber qual propriedade a transformação T deve ter de modo a assegurar que o limite seja simplesmente $\mu(A)$ para quase todo ponto inicial x. A resposta é que a transformação T deve revolver todo o espaço E. A próxima definição dá uma forma precisa a essa ideia.

Definição 14.18 (Transformação ergódica)**.** Uma transformação T que preserva medida é dita *ergódica* se $\mu(A) = 0$ ou 1 para todo $A \in \mathcal{I}_T$.

Em outras palavras, a transformação T é ergódica se não há forma de separar hermeticamente o espaço em regiões A e A^c, ambas com medida positiva, de forma que cada órbita de T fique restrita a uma dessas duas regiões.

A proposição abaixo responde à pergunta colocada no início desta seção.

Proposição 14.19. *Seja T uma transformação que preserva medida. São equivalentes:*

(*i*) *A transformação T é ergódica;*
(*ii*) *Toda $f : E \to \mathbb{R}$ mensurável e T-invariante é constante q.t.p.;*
(*iii*) *Para toda $f : E \to \mathbb{R}$ mensurável e integrável, $\mathbf{E}[f|\mathcal{I}_T] = \mathbf{E}f$ q.t.p.;*
(*iv*) *Para toda $f : E \to \mathbb{R}$ mensurável e limitada, $\mathbf{E}[f|\mathcal{I}_T] = \mathbf{E}f$ q.t.p.;*
(*v*) *Para todos $A, B \in \mathcal{E}$, vale $\frac{1}{n}\sum_{k=0}^{n-1} \mu(T^{-k}(A) \cap B) \to \mu(A) \cdot \mu(B)$.*

Demonstração. A estratégia será mostrar a seguinte cadeia de implicações: (*i*) $\Rightarrow$ (*ii*) $\Rightarrow$ (*iii*) $\Rightarrow$ (*iv*) $\Rightarrow$ (*v*) $\Rightarrow$ (*i*). A demonstração da primeira implicação é quase idêntica à prova do Corolário 13.17 e será omitida.

Assuma (*ii*) e seja f integrável. Pelo Exercício 14.14, $\mathbf{E}[f|\mathcal{I}_T]$ é T-invariante. Por (*ii*), $\mathbf{E}[f|\mathcal{I}_T] = c$ q.t.p. Integrando ambos os lados, concluímos que $c = \mathbf{E}f$, o que prova (*iii*).

A implicação (*iii*) $\Rightarrow$ (*iv*) é trivial.

Dados quaisquer $A, B \in \mathcal{E}$ e assumindo o item (*iv*), $\mathbf{E}[\mathbb{1}_A|\mathcal{I}_T] = \mu(A)$ q.t.p. e, pelo Corolário 14.17,

$$\lim_n \frac{1}{n}\sum_{k=0}^{n-1} \mu(T^{-k}(A) \cap B) = \mathbf{E}[\mathbb{1}_B \cdot \mathbb{E}[\mathbb{1}_A|\mathcal{I}_T]] = \mu(A) \cdot \mu(B).$$

Isso mostra o item (*v*).

Finalmente, seja $A \in \mathcal{I}_T$ e tome $B = A$. Assumindo o item (*v*),

$$\mu(A) = \lim_n \frac{1}{n}\sum_{k=0}^{n-1} \mu(T^{-k}(A) \cap A) = \mu(A) \cdot \mu(A),$$

logo, $\mu(A) = 0$ ou 1, ou seja, T é ergódica. □

Combinando a proposição acima com o Teorema Ergódico de Birkhoff, obtemos o seguinte corolário.

Corolário 14.20. *Sejam T uma transformação ergódica e $f : E \to \mathbb{R}$ uma função mensurável e integrável. Então*

$$\lim_n \frac{1}{n} \sum_{k=0}^{n-1} f \circ T^k = \int_E f \, \mathrm{d}\mu \quad \mu\text{-}q.t.p.$$

Em palavras, o corolário acima nos diz que a "média temporal" (tomada ao longo dos sucessivos passos da órbita) do observável f converge para a sua "média espacial" (calculada sobre o espaço E).

Exemplo 14.21 (Rotação do círculo). O operador de rotação no círculo $T : [0, 1) \to [0, 1)$, definido no Exemplo 14.10, é ergódico se, e somente se, seu parâmetro λ for irracional. Na Seção 14.4 daremos uma prova curta que usa séries de Fourier e uma mais maçante baseada em aproximações. △

Consideraremos agora um tipo especial de transformação que preserva medida. São transformações cuja dinâmica não apenas revolve o espaço mas também misturam todas as suas regiões.

Definição 14.22 (Transformação misturadora)**.** Uma transformação que preserva medida $T : E \to E$ é dita *misturadora* se

$$\lim_n \mu(T^{-n}(A) \cap B) = \mu(A) \cdot \mu(B)$$

para todos $A, B \in \mathcal{E}$.

A definição acima nos diz que os eventos $T^{-n}(A)$ e B se tornam praticamente independentes quando n cresce. Isto é bem mais forte do que o comportamento médio das órbitas não depender da condição inicial.

Proposição 14.23. *Toda transformação misturadora é ergódica.*

Demonstração. Sejam T uma transformação misturadora e $A \in \mathcal{I}_T$. Tomando $B = A$ na Definição 14.22, temos

$$\mu(A) \cdot \mu(A) = \lim_n \mu(T^{-n}(A) \cap A) = \mu(A),$$

pois $T^{-n}(A) \cap A = A$, logo $\mu(A) = 0$ ou 1, isto é, T é ergódica. $\square$

Um exemplo de transformação misturadora será dado no final desta seção, dentro do contexto de sequências estacionárias. A transformação do Exemplo 14.11 também é misturadora.[24] Vejamos alguns exemplos de transformações ergódicas que não são misturadoras.

Contra-exemplo 14.24. Tome $E = \{0, 1\}$, $\mu(\{0\}) = \mu(\{1\}) = \frac{1}{2}$, e $T(x) = 1-x$. Neste caso, $\mathcal{I}_T = \{\emptyset, \{0, 1\}\}$, logo T é ergódica. Tomando $A = B = \{0\}$, vemos que $\mu(A) \cdot \mu(B) = \frac{1}{4}$, enquanto $\mu(T^{-n}(A) \cap B) = 0$ ou $\frac{1}{2}$ dependendo da paridade de n. Logo, T não pode ser misturadora. Este exemplo é equivalente ao Exemplo 14.5. $\triangle$

Contra-exemplo 14.25 (Rotação do círculo). Outro exemplo de transformação ergódica que não é misturadora é a rotação irracional no círculo $[0, 1)$, vista no Exemplo 14.21. Daremos mais detalhes na Seção 14.4. $\triangle$

Sequências estacionárias

Seja $\boldsymbol{X} = (X_1, X_2, \dots)$ uma sequência estacionária definida no espaço de probabilidade $(\Omega, \mathcal{F}, \mathbb{P})$. Como visto na Seção 14.1, há uma correspondência entre sequências estacionárias definidas no contexto de um experimento aleatório e sequências da forma $(f \circ T^k)_{k \in \mathbb{N}_0}$, onde T é uma transformação que preserva medida. Sendo assim, todas as definições e teoremas obtidos podem ser reformulados em termos de sequências estacionárias.

Sejam $\boldsymbol{X}$ sequência estacionária e $A \in \mathcal{F}$ um evento da forma $A = \{\omega : \boldsymbol{X}(\omega) \in B\}$, para algum $B \in \mathcal{B}(\mathbb{R}^{\mathbb{N}})$ dado. Dizemos que A é invariante por translações de $\boldsymbol{X}$ se $A = \{\theta \boldsymbol{X}(\omega) \in B\}$. Denotaremos por $\mathcal{I}_{\boldsymbol{X}}$ a σ-álgebra dos eventos invariantes por translações de $\boldsymbol{X}$. Dizemos que a sequência estacionária $\boldsymbol{X}$ é *ergódica*, se $\mathbb{P}(A) = 0$ ou 1 para todo $A \in \mathcal{I}_{\boldsymbol{X}}$. Isso é equivalente à propriedade de o operador de deslocamento θ ser ergódico no espaço de probabilidade induzido $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^{\mathbb{N}}), \mathbb{P}_{\boldsymbol{X}})$.

[24] De fato, esse exemplo é equivalente a aplicar o operador deslocamento em uma sequência i.i.d., no espírito do Lema 4.28, mas não entraremos em detalhes.

Enunciamos agora uma consequência imediata do Teorema Ergódico de Birkhoff.

Teorema 14.26 (Teorema Ergódico para sequências estacionárias)**.** *Seja* $\boldsymbol{X} = (X_1, X_2, \dots)$ *uma sequência estacionária com* X_1 *integrável. Então,*

$$\lim_n \frac{1}{n} \sum_{k=1}^{n} X_k = \mathbb{E}[X_1 | \mathcal{I}_{\boldsymbol{X}}]$$

q.t.p. Se $\boldsymbol{X}$ *for uma sequência ergódica, então* $\mathbb{E}[X_1|\mathcal{I}_{\boldsymbol{X}}] = \mathbb{E}X_1$ *q.t.p.*

Passemos ao exemplo mais forte de sequência ergódica.

Proposição 14.27. *Se* $\boldsymbol{X}$ *é uma sequência i.i.d., então é ergódica.*

Demonstração. A ideia é bem semelhante à prova da Lei 0-1 de Kolmogorov. Dado qualquer $A \in \mathcal{I}_{\boldsymbol{X}}$, por definição existe $B \in \mathcal{B}(\mathbb{R}^{\mathbb{N}})$ tal que $A = \{(X_n, X_{n+1}, \dots) \in B\}$ para todo $n \in \mathbb{N}$. Pelo Teorema 13.5 (aproximação por álgebras), existe uma sequência de eventos $B_n \in \mathcal{B}(\mathbb{R}^n)$ tais que

$$\mathbb{P}(A \triangle A_n) \to 0,$$

onde $A_n = \{(X_1, \dots, X_n) \in B_n\}$. Logo,

$$\mathbb{P}(A_n) \to \mathbb{P}(A) \quad \text{e} \quad \mathbb{P}(A \cap A_n) \to \mathbb{P}(A).$$

Seja $n \in \mathbb{N}$ fixo. Como $A \in \mathcal{I}_{\boldsymbol{X}}$, temos $A = \{(X_{n+1}, X_{n+2}, \dots) \in B\}$, logo A e A_n são independentes, isto é,

$$\mathbb{P}(A \cap A_n) = \mathbb{P}(A) \cdot \mathbb{P}(A_n).$$

Tomando o limite em n, $\mathbb{P}(A) = \mathbb{P}(A) \cdot \mathbb{P}(A)$, logo $\mathbb{P}(A) = 0$ ou 1. $\square$

Combinando os enunciados acima, temos o seguinte corolário.

Corolário 14.28 (Lei Forte dos Grandes Números)**.** *Seja* $\boldsymbol{X} = (X_1, X_2, \dots)$ *uma sequência de variáveis aleatórias i.i.d. com* X_1 *integrável. Então*

$$\frac{1}{n}\sum_{k=1}^{n} X_k \overset{\text{q.c.}}{\to} \mathbb{E}X_1.$$

Dizemos que a sequência estacionária $\boldsymbol{X}$ é *misturadora* se

$$\mathbb{P}(\theta^n \boldsymbol{X} \in A,\ \boldsymbol{X} \in B) \to \mathbb{P}(\boldsymbol{X} \in A)\mathbb{P}(\boldsymbol{X} \in B)$$

para todos $A, B \in \mathcal{B}(\mathbb{R}^{\mathbb{N}})$. Isso é equivalente à propriedade de o operador de deslocamento θ ser misturador no espaço de probabilidade induzido $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^{\mathbb{N}}), \mathbb{P}_{\boldsymbol{X}})$.

Exemplo 14.29. A sequência definida no Exemplo 14.5 é ergódica mas não é misturadora. △

Exemplo 14.30 (Deslocamento de sequências i.i.d.). Se as variáveis aleatórias $X_1, X_2, \dots$ são i.i.d., então a sequência $\boldsymbol{X} = (X_1, X_2, \dots)$ é misturadora.

Tome $\mu = \mathbb{P}_X$ em $\mathbb{R}^{\mathbb{N}}$. Sejam $A, B \in \mathcal{B}(\mathbb{R}^{\mathbb{N}})$ e seja $\varepsilon > 0$. Pelo Teorema 13.5 (aproximação por álgebras), existem $n \in \mathbb{N}$ e eventos $\tilde{A}, \tilde{B} \in \mathcal{B}(\mathbb{R}^{\mathbb{N}})$ que dependem apenas das n primeiras coordenadas, tais que

$$\mu(A \triangle \tilde{A}) < \varepsilon \quad \text{e} \quad \mu(B \triangle \tilde{B}) < \varepsilon. \tag{14.31}$$

Seja $k \geqslant n$. Observe que

$$\mu(\theta^{-k}(A) \triangle \theta^{-k}(\tilde{A})) = \mu(\theta^{-k}(A \triangle \tilde{A})) = \mu(A \triangle \tilde{A}) < \varepsilon,$$

onde a primeira igualdade é devida ao fato de que pré-imagem comuta com diferença simétrica e a segunda se deve ao fato de que θ preserva medida.

Por outro lado, como $(X_1, \dots, X_n)$ é independente de $(X_{k+1}, \dots, X_{k+n})$,

$$\mu(\theta^{-k}(\tilde{A}) \cap \tilde{B}) = \mu(\theta^{-k}(\tilde{A})) \cdot \mu(\tilde{B}) = \mu(\tilde{A}) \cdot \mu(\tilde{B}).$$

Combinando essa identidade com (14.31), segue que

$$|\mu(\theta^{-k}(A) \cap B) - \mu(A)\mu(B)| \leqslant 4\varepsilon$$

para todo $k \geqslant n$. Como ε é arbitrário, $\mu(\theta^{-k}(A) \cap B) \to \mu(A)\mu(B)$. $\triangle$

Eventos e funções quase invariantes

Em muitas situações práticas, temos uma transformação ergódica T e gostaríamos de concluir que um dado evento A satisfaz $\mu(A) = 0$ ou 1, ou que uma dada função f é constante q.t.p. Para isso, basta que A ou f sejam invariantes, porém essa condição é muito forte. Vejamos outra mais fácil de verificar.

Definição 14.32 (Quase invariante)**.** Dizemos que um evento $A \in \mathcal{E}$ é *quase T-invariante* se $\mu(A \triangle T^{-1}(A)) = 0$. Denotamos por $\mathcal{I}_T^*$ a classe dos conjuntos quase T-invariantes. Dizemos que uma função mensurável $f : E \to \mathbb{R}$ é *quase T-invariante* se $f = f \circ T$ q.t.p.

Exercício 14.33. Mostre que $\mathcal{I}_T^*$ é uma σ-álgebra. Mostre que f é quase T-invariante se, e somente se for $\mathcal{I}_T^*$-mensurável. $\triangle$

As proposições abaixo estabelecem as relações entre conjuntos T-invariantes, quase T-invariantes e transformações ergódicas.

Proposição 14.34. *Seja T é uma transformação que preserva medida. Para todo $A \in \mathcal{I}_T^*$, existe $B \in \mathcal{I}_T$ tal que $\mu(A \triangle B) = 0$.*

Demonstração. Seja $A \in \mathcal{I}_T^*$. Tome $B = \{x : T^n(x) \in A \text{ i.v.}\}$. Temos que $B \in \mathcal{I}_T$, pois $T^{-1}(B) = \{x : T^{n+1}(x) \in A \text{ i.v.}\} = B$. Além disso, observe que $A \triangle B \subseteq \cup_{k=0}^{\infty}(T^{-k}(A) \triangle T^{-k-1}(A))$. Por outro lado, como T preserva medida e $A \in \mathcal{I}_T^*$,

$$\mu(T^{-k}(A) \triangle T^{-k-1}(A)) = \mu(T^{-k}(A \triangle T^{-1}(A))) = \mu(A \triangle T^{-1}(A)) = 0$$

para todo $k \in \mathbb{N}_0$, o que conclui a prova. $\square$

Proposição 14.35. *Se T é ergódica, então $\mu(A) = 0$ ou 1 para todo A quase T-invariante e toda f quase T-invariante é constante q.t.p.*

Demonstração. Dado $A \in \mathcal{I}_T^*$, pela Proposição 14.34, existe $B \in \mathcal{I}_T$ com $\mu(A \triangle B) = 0$, logo $\mu(A) = \mu(B) = 0$ ou 1, pois T é ergódica. A prova da afirmação sobre f quase T-invariante é idêntica à do Corolário 13.17 e será omitida. □

14.4 Rotação do círculo

Aqui justificaremos os Exemplos 14.21 e 14.25. Começamos pelo caso mais simples.

Rotação racional do círculo não é ergódica

Seja $\lambda = \frac{p}{q} \in \mathbb{Q}$ com $p \in \mathbb{Z}$ e $q \in \mathbb{N}$. Considerando o evento

$$A = \cup_{k=1}^{q} [\tfrac{2k-1}{2q}, \tfrac{2k}{2q}],$$

podemos ver que A é T-invariante e $\mu(A) = \frac{1}{2}$, logo T não é ergódica.

Rotação irracional do círculo é ergódica

Vamos supor daqui em diante que λ é irracional.

Provaremos inicialmente que T é ergódica usando séries de Fourier. Uma prova auto-contida porém mais tediosa será dada mais abaixo.

Seja $f : [0, 1) \to \mathbb{R}$ uma função mensurável limitada, periódica de período 1 e T-invariante. Da Análise de Fourier, sabemos que existe a única sequência

$(a_n)_{n\in\mathbb{Z}}$ tal que $g_k \to f$ em $\mathcal{L}^2$ quando $k \to \infty$, onde

$$g_k(x) = \sum_{n=-k}^{k} a_n e^{2\pi i n x}.$$

Substituindo x por $T(x)$ na equação acima,

$$g_k(T(x)) = g_k(x+\lambda) = \sum_{n=-k}^{k} (e^{2\pi i n\lambda} a_n) e^{2\pi i n x}.$$

Por outro lado, como $g_k \to f$ em $\mathcal{L}^2$ e T preserva medida, segue que $g_k \circ T \to f \circ T$ em $\mathcal{L}^2$. E como $f \circ T = f$, temos que $g_k \circ T \to f$ em $\mathcal{L}^2$.

Assim, pela unicidade da sequência $(a_n)_{n\in\mathbb{Z}}$, concluímos que

$$a_n(1 - e^{2\pi i n\lambda}) = 0, \text{ para todo } n \in \mathbb{Z}.$$

Entretanto, como λ é irracional, $e^{2\pi i n\lambda} \neq 1$, logo $a_n = 0$, para todo $n \neq 0$. Portanto $f(x) = a_0$ para quase todo ponto x. Logo, f é constante q.t.p. Pela Proposição 14.19, T é ergódica.

Rotação irracional do círculo não é misturadora

Vejamos agora que T não é misturadora.

Observamos inicialmente que, para todos $x \in [0,1)$ e $I \subseteq [0,1)$ intervalo aberto não-vazio, $T^n(x) \in I$ para infinitos valores de n, isto é, a órbita de x é densa em $[0,1)$. A demonstração fica como exercício.

Agora tome $A = B = [0, \frac{1}{2})$. Como consequência do fato de que as órbitas são densas, temos $\mu(T^{-n}(A) \cap B) < \frac{1}{8}$ para infinitos valores de n, mas $\mu(A) \cdot \mu(B) = \frac{1}{4}$, logo T não pode ser misturadora.

Rotação irracional do círculo é ergódica. Demostração alternativa

Faremos aqui uma prova alternativa de que T é ergódica. Ela será mais longa. Todavia, não será necessário o uso de Análise de Fourier.

Seja $A \in \mathcal{I}_T$. Faremos uma sucessão de aproximações para concluir que

$$\int_0^1 |\mathbb{1}_A(x) - \mu(A)| \mathrm{d}x = 0$$

donde segue que $\mathbb{1}_A(x) = \mu(A)$ q.t.p. e, portanto, $\mu(A) = 0$ ou 1.

Seja $\varepsilon > 0$. Mostraremos primeiro que existe $\delta > 0$ e uma função $f : \mathbb{R} \to [0, 1]$ periódica de período 1 tal que

$$\int_0^1 |\mathbb{1}_A(x) - f(x)| \mathrm{d}x < 2\varepsilon \tag{14.36}$$

e tal que

$$|f(x) - f(y)| < \varepsilon \tag{14.37}$$

para todos $x, y \in \mathbb{R}$ tais que $|x - y| < \delta$.

Com efeito, Pelo Teorema 13.5 (aproximação por álgebras), existe um boreliano $B \subseteq [0, 1)$ da forma $B = \cup_{j=1}^k [a_j, b_j)$ e tal que $\mu(A \triangle B) < \varepsilon$. Agora defina $f(x) = 0$ para $x \in [0, 1) \setminus B$ e $f(x) = 1$ para x tal que $a_j + \frac{\varepsilon}{2k} \leqslant x \leqslant b_j - \frac{\varepsilon}{2k}$ para algum j. Definimos f nos intervalos restantes interpolando linearmente. Isso define f em $[0, 1)$, depois definimos f para todo $x \in \mathbb{R}$ de forma periódica. Como $\int_0^1 |f(x) - \mathbb{1}_B(x)| \mathrm{d}x < \varepsilon$, vale (14.36). Ademais, tomando $\delta = \frac{\varepsilon^2}{2k}$, também vale (14.37).

Agora, usando (14.36), mostraremos que

$$\int_0^1 |f \circ T^n - f| \, \mathrm{d}\mu < 4\varepsilon. \tag{14.38}$$

Com efeito,

$$\begin{aligned}\int_0^1 |f \circ T^n - f| \mathrm{d}\mu &\leqslant \int_0^1 |f \circ T^n - \mathbb{1}_A \circ T^n| \mathrm{d}\mu + \int_0^1 |f - \mathbb{1}_A \circ T^n| \mathrm{d}\mu \\ &= 2 \int_0^1 |f - \mathbb{1}_A| \mathrm{d}\mu < 4\varepsilon,\end{aligned}$$

onde a igualdade vale porque T preserva medida e A é T-invariante.

Finalmente, mostraremos que,

$$\int_0^1 |f(x) - \alpha| \, \mathrm{d}x < 5\varepsilon, \tag{14.39}$$

onde $\alpha = \int_0^1 f \, \mathrm{d}\mu$. Para isso, vamos mostrar que, para todo $y \in [0, 1)$,

$$\int_0^1 |f(x + y) - f(x)| \mathrm{d}x < 5\varepsilon. \tag{14.40}$$

Seja $y \in [0, 1)$. Como as órbitas de T são densas em $[0, 1)$, existe $n \in \mathbb{N}$ tal que $|T^n(0) - y| < \delta$. Daí decorre que $|f(x + y) - f(T^n(x))| < \varepsilon$ para todo $x \in [0, 1)$. Combinando essa desigualdade com (14.38), obtemos (14.40). Invertendo as integrais, obtemos

$$\begin{aligned}\int_0^1 |\alpha - f(x)| \, \mathrm{d}x &= \int_0^1 \left| \int_0^1 [f(x + y) - f(x)] \mathrm{d}y \right| \mathrm{d}x \\ &\leqslant \int_0^1 \left(\int_0^1 \left| f(x + y) - f(x) \right| \mathrm{d}y \right) \mathrm{d}x \\ &= \int_0^1 \left(\int_0^1 \left| f(x + y) - f(x) \right| \mathrm{d}x \right) \mathrm{d}y < 5\varepsilon,\end{aligned}$$

onde usamos o Teorema de Tonelli para inverter as integrais, provando (14.39).

Para concluir, observe que $|\alpha - \mu(A)| < 2\varepsilon$ por (14.36) e, combinando com (14.36) e (14.39), obtemos

$$\int_0^1 |\mathbb{1}_A - \mu(A)| \, \mathrm{d}\mu \leqslant \int_0^1 (|\mathbb{1}_A - f| + |f - \alpha| + |\alpha - \mu(A)|) \, \mathrm{d}\mu \leqslant 9\varepsilon.$$

Como ε é arbitrário, $\int_0^1 |\mathbb{1}_A - \mu(A)| \, \mathrm{d}\mu = 0$, o que termina a prova.

14.5 Teorema Ergódico de von Neumann

Recordemos que estamos trabalhando em um espaço de probabilidade $(E, \mathcal{E}, \mu)$ e denotando a integral por $\mathbf{E}$. Fixado $p \geqslant 1$, definimos

$$\|f\|_p = \left(\mathbf{E}|f|^p\right)^{\frac{1}{p}}$$

e o conjunto $\mathcal{L}^p$ das funções $f : \Omega \to \mathbb{R}$ mensuráveis tais que $\|f\|_p < \infty$. Provaremos no Apêndice D.7 que vale a desigualdade triangular

$$\|f + g\|_p \leqslant \|f\|_p + \|g\|_p$$

para todas $f, g \in \mathcal{L}^p$. Observe também que $\|cf\|_p = |c| \cdot \|f\|_p$ para todo $c \in \mathbb{R}$, logo $\mathcal{L}^p$ e um espaço vetorial real.[25]

O teorema abaixo complementa o Teorema Ergódico de Birkhoff, mostrando que também vale a convergência em $\mathcal{L}^p$ sob certas condições.

Teorema 14.41 (Teorema Ergódico de von Neumann)**.** *Sejam T uma transformação que preserva medida no espaço de probabilidade $(E, \mathcal{E}, \mu)$ e $f \in \mathcal{L}^p$ com $p \geqslant 1$. Então $\|\mathbf{E}[f \,|\, \mathcal{I}_T]\|_p \leqslant \|f\|_p$ e*

$$\frac{1}{n} \sum_{k=0}^{n-1} f \circ T^k \overset{\mathcal{L}^p}{\to} \mathbf{E}[f|\mathcal{I}_T]. \tag{14.42}$$

Demonstração. Segue direto do Teorema 11.53 que $\|\mathbf{E}[f \,|\, \mathcal{I}_T]\|_p \leqslant \|f\|_p$.

[25] Observamos que $\|\cdot\|_p$ não é exatamente uma norma, nem mesmo restrita ao espaço $\mathcal{L}^p$, pois $\|f\|_p = 0$ não implica necessariamente que f seja a função constante igual a zero, apenas que $f(x) = 0$ para μ-quase todo x. Em Análise Funcional, costuma-se definir $\mathcal{L}^p$ em termos de classes de equivalência de funções que coincidem μ-q.t.p., mas não nos preocuparemos com isso e usaremos a notação $\|\cdot\|_p$ mesmo sabendo que não é de fato uma norma.

Para a demonstração da convergência em $\mathcal{L}^p$, consideramos inicialmente o caso em que f é limitada q.t.p., isto é, existe constante $c > 0$ tal que $f \leqslant c$ q.t.p. Combinando o Teorema Ergódico de Birkhoff com o Teorema da Convergência Dominada em $\mathcal{L}^p$, podemos concluir que vale (14.42).

Consideramos agora o caso geral. Dado $\varepsilon > 0$, como $|f|^p$ é integrável, usando truncamento e o Teorema da Convergência Dominada, podemos obter g limitada tal que $\|f - g\|_p < \varepsilon$. Procedemos a estimar

$$\begin{aligned}
\left\| \mathbf{E}[f \,|\, \mathcal{I}_T] - \frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k \right\|_p &\leqslant \left\| \mathbf{E}[f \,|\, \mathcal{I}_T] - \mathbf{E}[g \,|\, \mathcal{I}_T] \right\|_p + \\
&\quad + \left\| \mathbf{E}[g \,|\, \mathcal{I}_T] - \frac{1}{n}\sum_{k=0}^{n-1} g \circ T^k \right\|_p + \left\| \frac{1}{n}\sum_{k=0}^{n-1} (f - g) \circ T^k \right\|_p \\
&\leqslant \left\| \mathbf{E}[f - g \,|\, \mathcal{I}_T] \right\|_p + \left\| \mathbf{E}[g \,|\, \mathcal{I}_T] - \frac{1}{n}\sum_{k=0}^{n-1} g \circ T^k \right\|_p + \|f - g\|_p \\
&\leqslant 2\|f - g\|_p + \left\| \mathbf{E}[g \,|\, \mathcal{I}_T] - \frac{1}{n}\sum_{k=0}^{n-1} g \circ T^k \right\|_p \to 2\|f - g\|_p < 2\varepsilon.
\end{aligned}$$

Na primeira desigualdade, utilizamos a desigualdade triangular. Na segunda, a desigualdade triangular combinada com o fato de que T preserva medida. Na terceira, novamente o Teorema 11.53. O limite vale pelo caso anterior, pois g é limitada q.t.p. Como a desigualdade acima é válida para todo $\varepsilon > 0$, concluímos que (14.42) vale para toda $f \in \mathcal{L}^p$. □

14.6 Exercícios

§14.1

1. Sejam U variável aleatória com distribuição uniforme em $[0, 1)$ e $(X_n)_n$ a sequência de variáveis aleatórias que denota a representação decimal de U (caso haja mais de uma, considere aquela que termina com infinitos zeros);

isto é, $U = \sum_{n=1}^{\infty} \frac{X_n}{10^n}$. Dados $l \in \mathbb{N}$ e $\boldsymbol{k} = (k_1, \dots, k_l) \in \{0, 1, \dots, 9\}^l$ uma sequência de l dígitos fixada, defina a sequência $(Z_n^{\boldsymbol{k}})$, onde $Z_n^{\boldsymbol{k}} = \mathbb{1}_{\{X_n = k_1, \dots, X_{n+l-1} = k_l\}}$. Mostre que a sequência $(Z_n^{\boldsymbol{k}})$ é estacionária.

2. Seja $(X_1, X_2, \dots)$ uma sequência estacionária com $\mathbb{E}X_1^2 < \infty$. Mostre que $\mathbb{E}X_n$ não depende de n e a covariância entre duas variáveis da sequência é da forma $\mathbf{Cov}(X_n, X_k) = g(n-k)$ para alguma função $g : \mathbb{R} \to \mathbb{R}$.

3 (Transformação de Gauss)**.** Considere $T : [0,1] \to [0,1]$ dada por

$$T(x) = \begin{cases} \frac{1}{x} - \lfloor \frac{1}{x} \rfloor, & \text{se } x \neq 0, \\ 0, & \text{se } x = 0. \end{cases}$$

Mostre que T é uma transformação que preserva a medida μ dada por

$$\mu(A) = \frac{1}{\log 2} \int_A \frac{1}{1+x} \mathrm{d}x, \qquad A \in \mathcal{B}([0,1]).$$

4. Seja $p > 1$ e considere $T : [0,1] \to [0,1]$ dada por $T(x) = x^p$. Seja $f : [0,1] \to [0, +\infty)$ uma função mensurável com $\int_0^1 f(x)\,\mathrm{d}x = 1$, e defina a medida μ por

$$\mu(A) = \int_A f(x)\,\mathrm{d}x, \qquad A \in \mathcal{B}([0,1]).$$

Mostre que T não preserva a medida μ.

5. Seja $(X_1, X_2, \dots)$ uma sequência estacionária. Mostre que, quase certamente, se $X_1 \neq 0$ então $\sum_n |X_n| = +\infty$.

§14.2

6. Sejam $E = \{1, \dots, m\}$, $\mathcal{E} = \mathcal{P}(E)$, $\mu(A) = \frac{\#A}{m}$ e $T : E \to E$ uma bijeção. Seja $f : E \to \mathbb{R}$ é uma função qualquer. Descreva o limite da sequência

$$\frac{1}{n} \sum_{k=0}^{n-1} f(T^k(x))$$

como uma função de x.

7. Dê um exemplo de uma sequência estacionária $(X_1, X_2, \dots)$ com X_1 integrável tal que

$$\mathbb{P}\left(\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} X_k = \mathbb{E}X_1\right) = 0.$$

§14.3

8. Estabeleça condições para que a transformação T definida no Exercício 6 seja ergódica.

9. Seja S a transformação definida no Exemplo 14.11.

(a) Mostre que S é uma transformação ergódica

Dica: Pense nos pontos de $[0, 1)$ escritos na base n.

(b) Utilize o Teorema Ergódico para fornecer uma prova alternativa que quase todo número em $[0, 1)$ é normal (conforme definido na Seção 8.4).

10. Sejam $n, k \in \mathbb{N}$, $E = \{1, \dots, n\}$, $\mathcal{E} = \mathcal{P}(E)$ e $\mu(A) = \frac{\#A}{n}$. Defina $T_k : E \to E$ por $T(x) = x + k$ se $x + k < n$ e $x + k - n$ caso contrário.

(a) Mostre que T preserva a medida μ.

(b) Mostre que T é ergódica se, e somente se, n e k são primos entre si.

(c) Nos casos em que T for ergódica, podemos afirmar que T é misturadora?

11. Mostre que uma transformação que preserva medida T é misturadora se, e somente se,

$$\lim_{n} \mathbf{Cov}(f \circ T^n, g) = 0$$

para todas f e g funções mensuráveis.

12. Prove o Teorema 14.26.

13. Sejam $\boldsymbol{X} = (X_1, X_2, \dots)$ uma sequência de variáveis aleatórias e $f : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}$ uma função $\mathcal{B}(\mathbb{R}^{\mathbb{N}})$-mensurável. Defina a sequência $\boldsymbol{Z} = (Z_1, Z_2, \dots)$, onde $Z_n = f(X_n, X_{n+1}, \dots)$. Mostre que:

(a) Se $\boldsymbol{X}$ é estacionário, então $\boldsymbol{Z}$ é estacionário.

(b) $\mathcal{I}_{\boldsymbol{Z}} \subseteq \mathcal{I}_{\boldsymbol{X}}$.

(c) Se $\boldsymbol{X}$ é ergódico, então $\boldsymbol{Z}$ é ergódico.

(d) Se $\boldsymbol{X}$ é estacionário e Z_1 é integrável, então

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} Z_k = \mathbb{E}[Z_1|\mathcal{I}_{\boldsymbol{X}}] = \mathbb{E}[Z_1|\mathcal{I}_{\boldsymbol{Z}}] \quad \text{q.c.}$$

(e) Se $\boldsymbol{X}$ é ergódico e Z_1 é integrável, então $\mathbb{E}[Z_1|\mathcal{I}_{\boldsymbol{X}}] = \mathbb{E}[Z_1]$ q.c.

14. Seja T uma função que preserva medida e $A \in \mathcal{E}$ tal que $\mu(T^{-1}(A)\backslash A) = 0$. Mostre que A é quase T-invariante.

§14.5

15. Mostre que uma sequência estacionária $(X_1, X_2, \dots)$ é ergódica se, e somente se

$$\frac{1}{n} \sum_{j=1}^{n} \mathbb{1}_B(X_j, \dots, X_{j+k-1}) \overset{\text{q.c.}}{\to} \mathbb{P}((X_1, \dots, X_k) \in B)$$

para todos $k \in \mathbb{N}$ e $B \in \mathcal{B}(\mathbb{R}^k)$.

Capítulo 15

Grandes Desvios

Seja $(X_n)_n$ uma sequência de variáveis aleatórias i.i.d. com média μ e variância σ^2. Escrevendo $S_n = X_1 + \cdots + X_n$, a Lei dos Grandes Números nos diz que

$$\frac{S_n}{n} \overset{\text{q.c.}}{\to} \mu,$$

e o Teorema do Limite Central mostra que

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \overset{\text{d}}{\to} \mathcal{N}.$$

Em palavras, a Lei dos Grandes Números mostra que a média observada $\frac{S_n}{n}$ converge para μ, enquanto o Teorema do Limite Central nos diz que os desvios de S_n em torno de sua média $n\mu$ são tipicamente da ordem de $\sqrt{n}$. Neste capítulo estudamos

$$\mathbb{P}\Big(\Big|\frac{S_n}{n} - \mu\Big| \geqslant \varepsilon n\Big)$$

para n grande, isto é, consideramos o evento, muito improvável, de que a diferença entre S_n e sua média seja da ordem de n. Um desvio dessa ordem é chamado um *grande desvio*. Mais precisamente, vamos estudar a

Desigualdade de Concentração de Chernoff e o Princípio dos Grande Desvios de Cramér, que nos dizem como se comporta a probabilidade de tais desvios.

Notação. O termo $o(b_n)$ denota alguma função $g(n)$ satisfazendo $\frac{g(n)}{b_n} \to 0$ quando $n \to \infty$. Cada vez que aparece, pode denotar uma função diferente.

15.1 Desigualdade de concentração

Seja X uma variável aleatória integrável com média μ. Sejam $(X_n)_n$ independentes e com a mesma distribuição de X, e tome $S_n = X_1 + \cdots + X_n$.

A Lei dos Grandes Números de Tchebyshev para este caso foi provada da seguinte forma: dado qualquer $a \in (\mu, +\infty)$,

$$\mathbb{P}\Big(\tfrac{S_n}{n} \geqslant a\Big) \leqslant \mathbb{P}\Big((\tfrac{S_n}{n} - \mu)^2 \geqslant (a-\mu)^2\Big) \leqslant \tfrac{1}{(a-\mu)^2}\mathbb{E}\Big(\tfrac{S_n}{n} - \mu\Big)^2 = \tfrac{\mathbb{V}X}{(a-\mu)^2} \cdot \frac{1}{n}$$

A desigualdade acima diz que, quando $\mathbb{E}X^2 < \infty$, a probabilidade de $\frac{S_n}{n}$ diferir de μ por mais que uma quantidade fixa $a - \mu$ decai pelo menos tão rápido quanto $\frac{1}{n}$. Na prova da Lei dos Grandes Números de Cantelli, vimos que, quando $\mathbb{E}X^4 < \infty$, esta probabilidade decai pelo menos tão rápido quanto $\frac{1}{n^2}$. Em geral, se $\mathbb{E}X^{2k} < \infty$ ela decai pelo menos tão rápido quanto $\frac{1}{n^k}$.

Tentaremos agora obter estimativas melhores usando momentos de e^{tX} ao invés de X^{2k}. Para todos $t \geqslant 0$ e $a \in (\mu, +\infty)$,

$$\mathbb{P}\Big(\tfrac{S_n}{n} \geqslant a\Big) \leqslant \mathbb{P}\Big[e^{tS_n} \geqslant e^{atn}\Big] \leqslant \tfrac{1}{e^{atn}}\mathbb{E}[e^{tS_n}] = e^{-atn}M(t)^n = e^{-[at - \log M(t)]n},$$

onde $M(t) = \mathbb{E}[e^{tX}]$ é a função geradora de momentos de X; observe que na penúltima igualdade utilizamos que as variáveis (X_n) são i.i.d com distribuição comum X. De modo análogo, para $a \in (-\infty, \mu)$ e $t \leqslant 0$, obtemos

$$\mathbb{P}\Big(\tfrac{S_n}{n} \leqslant a\Big) \leqslant e^{-[at - \log M(t)]n}. \tag{15.1}$$

Portanto, se mostrarmos que a expressão entre colchetes é positiva para algum t, teremos estabelecido que essa probabilidade de fato decai pelo menos exponencialmente rápido. Sabemos que a função geradora de momentos é finita em um intervalo que contém o ponto $t = 0$. Denotaremos os extremos desse intervalo por $\mathcal{D}_M^- \leqslant 0$ e $\mathcal{D}_M^+ \geqslant 0$.

Teorema 15.2 (Desigualdade de Concentração de Chernoff)**.** *Sejam $(X_n)_n$ variáveis aleatórias i.i.d. com distribuição comum X, média μ, função geradora de momentos $M(t)$ e $S_n = X_1 + \cdots + X_n$. Se $\mathcal{D}_M^+ > 0$, então, para qualquer $a > \mu$, existe $t > 0$ tal que $[at - \log M(t)] > 0$. Como*

$$\mathbb{P}\Big(\tfrac{S_n}{n} \geqslant a\Big) \leqslant e^{-[at-\log M(t)]n},$$

segue que essa probabilidade decai pelo menos exponencialmente rápido em n. Analogamente, se $\mathcal{D}_M^- < 0$ e $a < \mu$, então $[at - \log M(t)] > 0$ para algum $t < 0$ e a estimativa em (15.1) decai, pelo menos, exponencialmente rápido.

Demonstração. Suponha que $\mathcal{D}_M^+ > 0$ e seja $a > \mu$. Consideramos inicialmente o caso em que $\mathcal{D}_M^- < 0$. Pela Proposição 10.7, $(\log M)'(0) = M'(0) = \mu < a$ e, por conseguinte, $[at - \log M(t)] > 0$ para todo $t > 0$ suficientemente pequeno, provando o teorema neste caso.

Consideramos agora o caso geral. Para reduzir ao caso em que $\mathcal{D}_M^- < 0$, vamos truncar as variáveis X_n. Para cada n e k, defina $X_{n;k} = X_n \mathbb{1}_{\{X_n \geqslant -k\}}$. Pelo Teorema da Convergência Dominada, $\lim_k \mathbb{E}X_{n;k} = \mu$ e $\lim_k (as - \log \mathbb{E}[e^{sX_{n;k}}]) = as - \log M(s)$ para todo $s \in [0, \mathcal{D}_M^+)$. Tome k tal que $\mathbb{E}X_{n;k} < a$. Pelo caso anterior, podemos tomar $t > 0$ tal que $(at - \log \mathbb{E}[e^{tX_{n;k}}]) > 0$. Como $X_n \leqslant X_{n;k}$, vale $[at - \log M(t)] > 0$, concluindo a prova do teorema.

A segunda parte do teorema, com $\mathcal{D}_M^- < 0$ e $a < \mu$, segue da primeira parte trocando-se os sinais de a, μ, t e X. $\square$

15.2 Princípio dos Grandes Desvios

A Desigualdade de Concentração de Chernoff diz que a probabilidade de um grande desvio decai, pelo menos, exponencialmente rápido. Porém, em muitos casos, vale algo muito mais forte: a velocidade de decaimento é exatamente exponencial, com uma taxa que pode ser descrita quase explicitamente.

Definição 15.3 (Função taxa)**.** Seja X uma variável aleatória. Definimos a *função taxa* I associada à distribuição de X, como a função $I : \mathbb{R} \to [0, +\infty]$ dada por

$$I(a) = \sup_{t \in \mathbb{R}} \big[at - \log M(t)\big], \tag{15.4}$$

onde $M(t)$ é a função geradora de momentos da variável X.

Podemos pensar na função taxa como uma tentativa de obter a melhor estimativa possível a partir de (15.1). A razão pela qual a função I merece esse nome é que, uma vez que maximizamos $[at - \log M(t)]$ sobre todo t, a desigualdade (15.1) deixa de ser apenas mais uma cota superior, sendo de fato a melhor cota superior possível. A maximização em (15.4) é conhecida como a *transformada de Legendre*, isto é, a função taxa é a transformada de Legendre do logaritmo da função geradora de momentos.

Dado $A \subseteq \mathbb{R}$, para descrever a maneira mais fácil (ou menos difícil) de $\frac{S_n}{n}$ estar em A, vamos denotar

$$I(A) = \inf_{a \in A} I(a).$$

Se A for um intervalo da reta, denotaremos por A° o seu interior (o intervalo excluindo as extremidades de A) e por $\bar{A}$ o seu fecho (o intervalo incluindo as extremidades de A caso sejam números reais), respectivamente.

Teorema 15.5 (Princípio dos Grandes Desvios de Cramér)**.** *Sejam $(X_n)_n$ variáveis aleatórias i.i.d., com distribuição comum X, e J um intervalo de*

$\mathbb{R}$. *Se* $I(J) < \infty$, *então*

$$e^{-I(J^\circ)\cdot n+o(n)} \leqslant \mathbb{P}\Big(\tfrac{S_n}{n} \in J\Big) \leqslant e^{-I(\bar{J})\cdot n+o(n)},$$

onde I *é a função taxa da variável* X. *Em particular, quando* $I(J^\circ) = I(\bar{J})$, *temos a taxa exata de decaimento exponencial para estas probabilidades:*

$$\mathbb{P}\Big(\tfrac{S_n}{n} \in J\Big) = e^{-I(J)\cdot n+o(n)}.$$

Caso $I(\bar{J}) = +\infty$, *vale*

$$\mathbb{P}\Big(\tfrac{S_n}{n} \in J\Big) \leqslant e^{-cn+o(n)}$$

para todo $c \in \mathbb{R}$.

A demonstração do teorema acima será dada na próxima seção. Antes disso, vamos discutir a relação entre a função geradora de momentos, M, e sua transformada de Legendre, a função taxa I.

Vejamos como encontrar $I(a)$ graficamente e algebricamente. No caso de o supremo em (15.4) ser atingido em $t = y$ para algum y no interior do intervalo onde M é finita, a derivada de $at - \log M(t)$ se anula em $t = y$, donde

$$a = \frac{M'(y)}{M(y)}. \tag{15.6}$$

Às vezes é possível expressar y em termos de a, e assim calcular I por

$$I(a) = a \cdot y - \log M(y), \quad y = y(a).$$

Esse processo de encontrar y tal que $(\log M)'(y) = a$ e expressar $I(a) = ay - \log M(y)$ está ilustrado na Figura 15.1 e nos exemplos abaixo.

Reciprocamente, se existe y tal que $(\log M)'(y) = a$, então o supremo em (15.4) é atingido em $t = y$, pois, como mostraremos mais abaixo, $\log M$ é uma função convexa.

Exemplo 15.7. Se $X \sim \mathcal{N}(\mu, \sigma^2)$, temos

$$\log M(t) = \frac{\sigma^2 t^2}{2} + t\mu,$$

assim

$$a = (\log M)'(y) = \sigma^2 y + \mu, \qquad y = \frac{a-\mu}{\sigma^2},$$

portanto,

$$I(a) = \frac{a(a-\mu)}{\sigma^2} - \left[\frac{(a-\mu)^2}{2\sigma^2} + \frac{\mu(a-\mu)}{\sigma^2}\right] = \frac{(a-\mu)^2}{2\sigma^2}, \quad a \in \mathbb{R}. \qquad \triangle$$

Exemplo 15.8. Se $X \sim \text{Poisson}(\lambda)$, temos

$$\log M(t) = \lambda(e^t - 1).$$

Analisando o gráfico de $\log M(t)$, observamos que o supremo de $at - \log M(t)$ é atingido em algum $y \in \mathbb{R}$ somente se $a > 0$. Neste caso,

$$a = (\log M)'(y) = \lambda e^y, \qquad y = \log \tfrac{a}{\lambda},$$

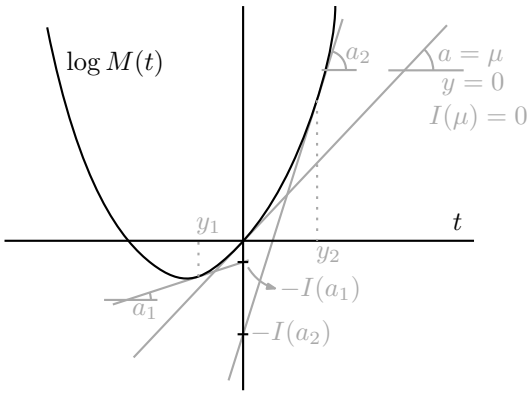


Figura 15.1. Obtenção da função taxa a partir da função $\log M_X$.

e

$$I(a) = ay - \log M(y) = a \log \tfrac{a}{\lambda} - a + \lambda.$$

Se $a = 0$, o supremo vale λ, pois $\log M(t)$ tende a $-\lambda$ quando $t \to -\infty$. Se $a < 0$, o supremo vale $+\infty$ pois $at - \log M(t) \geqslant at - \lambda$ para todo $t < 0$.

Em resumo,

$$I(a) = \begin{cases} +\infty, & a < 0, \\ \lambda, & a = 0, \\ a \log \frac{a}{\lambda} - a + \lambda, & a > 0. \end{cases}$$

Os casos $a = 0$ e $a < 0$ refletem o fato de que variáveis de Poisson nunca tomam valores negativos, e tomam o valor zero com probabilidade $e^{-\lambda}$. △

Exercício 15.9. Calcule a função taxa da variável X nos seguintes casos:

(a) $X \sim \text{Bernoulli}(p)$.
(b) $X \sim \text{Exp}(\lambda)$.
(c) $X = \mu$ q.c. △

Como dito anteriormente, $\log M$ é uma função convexa, a função I também goza da mesma propriedade.

Proposição 15.10. *As funções I e $\log M$ são convexas.*

Demonstração. Consideremos inicialmente a função I. Sejam $0 < \alpha < 1$ e $\beta = 1 - \alpha$. Para a_1 e $a_2 \in \mathbb{R}$,

$$\begin{aligned} I(\alpha a_1 + \beta a_2) &= \sup_{t \in \mathbb{R}} \big[(\alpha a_1 + \beta a_2)t - (\alpha + \beta) \log M(t)\big] \\ &= \sup_{t \in \mathbb{R}} \big[\alpha(a_1 t - \log M(t)) + \beta(a_2 t - \log M(t))\big] \\ &\leqslant \sup_{t \in \mathbb{R}} \big[\alpha(a_1 t - \log M(t))\big] + \sup_{t \in \mathbb{R}} \big[\beta(a_2 t - \log M(t))\big] \\ &= \alpha I(a_1) + \beta I(a_2), \end{aligned}$$

o que mostra que a função taxa I é convexa. Passamos agora à convexidade de

$\log M$. Sejam t_1 e $t_2 \in \mathbb{R}$. Usando a Desigualdade de Hölder (Apêndice D.7),

$$\begin{aligned}
\log M(\alpha t_1 + \beta t_2) &= \log \mathbb{E}\Big[e^{\alpha t_1 X} \cdot e^{\beta t_2 X}\Big] \\
&\leqslant \log\left\{\left(\mathbb{E}\left[\left(e^{\alpha t_1 X}\right)^{\frac{1}{\alpha}}\right]\right)^{\alpha}\left(\mathbb{E}\left[\left(e^{\beta t_2 X}\right)^{\frac{1}{\beta}}\right]\right)^{\beta}\right\} \\
&= \alpha \log \mathbb{E}[e^{t_1 X}] + \beta \log \mathbb{E}[e^{t_2 X}] \\
&= \alpha \log M(t_1) + \beta \log M(t_2),
\end{aligned}$$

o que conclui a prova. □

15.3 Demonstração do Teorema de Cramér

Dividiremos a prova do Teorema de Cramér em duas partes distintas, uma dedicada à prova da cota inferior e outra dedicada à cota superior.

Prova da cota inferior

Vamos primeiro mostrar que, para qualquer $a \in \mathbb{R}$ tal que $I(A) < \infty$ e qualquer $\delta > 0$, vale

$$\mathbb{P}\Big(\tfrac{S_n}{n} \in [a-\delta, a+\delta]\Big) \geqslant e^{-I(a)\cdot n + o(n)}. \tag{15.11}$$

Essa estimativa vale sob as hipóteses do Teorema 15.5. No entanto, vamos também supor que o supremo em (15.4) é atingido para algum $y \in \mathbb{R}$. De fato, vamos supor que tal y está no interior do intervalo onde M é finita, donde vale (15.6). Abandonar essa hipótese requer passos técnicos e complicados que vão além dos nossos objetivos.

Usando (15.6), pode-se adaptar a prova da Proposição 10.7 para obter

$$\frac{\mathbb{E}[Xe^{yX}]}{\mathbb{E}[e^{yX}]} = a.$$

A principal observação é que a expressão do lado esquerdo corresponde à esperança de uma variável aleatória Y cuja distribuição é obtida a partir da distribuição de X, distorcida por um fator da forma $g(x) = e^{yx}$. Ou seja, para uma variável aleatória Y cuja distribuição é dada por

$$\mathbb{P}(Y \in B) = \tfrac{\mathbb{E}[\mathbb{1}_B(X)e^{yX}]}{\mathbb{E}[e^{yX}]} = \mathbb{E}\Big[\mathbb{1}_B(X)\tfrac{e^{yX}}{M(y)}\Big],$$

temos $\mathbb{E}Y = a$ pela regra da cadeia, pois a derivada de Radon-Nikodým entre as duas distribuições é dada por

$$\frac{\mathrm{d}\mathbb{P}_Y}{\mathrm{d}\mathbb{P}_X}(x) = \frac{1}{c}e^{yx},$$

onde $c = \mathbb{E}[e^{yX}]$. Portanto, para variáveis $Y_1, Y_2, \ldots$ i.i.d. distribuídas como esta versão distorcida de X, a ocorrência de $\frac{Y_1+\cdots+Y_n}{n} \approx a$ não é um evento raro.

A prova então consiste em controlar a razão entre as probabilidades de $(X_1, \ldots, X_n)$ e $(Y_1, \ldots, Y_n)$ tomarem valores em um subconjunto de $\mathbb{R}^n$ que é típico para este último vetor, de forma tal que tal razão não fique menor que $e^{-I(a)\cdot n - o(n)}$.

Seja $\delta > 0$. Fixe $\varepsilon \in (0, \delta]$, e defina o conjunto

$$B_n^\varepsilon = \big\{(z_1, \ldots, z_n) : \big|\tfrac{z_1+\cdots+z_n}{n} - a\big| \leqslant \varepsilon\big\} \subseteq \mathbb{R}^n.$$

Então

$$\begin{aligned}
\mathbb{P}\Big(\tfrac{S_n}{n} \in [a-\varepsilon, a+\varepsilon]\Big) &= \mathbb{E}[\mathbb{1}_{B_n^\varepsilon}(X_1, \ldots, X_n)] \\
&= \mathbb{E}\Big[\tfrac{M(y)^n}{e^{y(X_1+\cdots+X_n)}}\,\mathbb{1}_{B_n^\varepsilon}(X_1, \ldots, X_n)\tfrac{e^{yX_1}}{M(y)}\cdots\tfrac{e^{yX_n}}{M(y)}\Big] \\
&\geqslant \mathbb{E}\Big[\frac{M(y)^n}{e^{ayn+|y|\varepsilon n}}\mathbb{1}_{B_n^\varepsilon}(X_1, \ldots, X_n)\tfrac{e^{yX_1}}{M(y)}\cdots\tfrac{e^{yX_n}}{M(y)}\Big] \\
&= e^{-[ay-\log M(y)+|y|\varepsilon]\cdot n}\cdot\mathbb{P}\big((Y_1, \ldots, Y_n) \in B_n^\varepsilon\big) \\
&= e^{-[ay-\log M(y)+|y|\varepsilon]\cdot n}\cdot\mathbb{P}\Big(\tfrac{Y_1+\cdots+Y_n}{n} \in [a-\varepsilon, a+\varepsilon]\Big).
\end{aligned}$$

Esta última probabilidade converge para um pela Lei dos Grandes Números, e portanto

$$\mathbb{P}\Big(\Big|\tfrac{S_n}{n} - a\Big| \leqslant \delta\Big) \geqslant \mathbb{P}\Big(\Big|\tfrac{S_n}{n} - a\Big| \leqslant \varepsilon\Big) \geqslant e^{-I(a)\cdot n - 2|y|\varepsilon\cdot n}$$

para todo n suficientemente grande. Como $\varepsilon \in (0, \delta]$ é arbitrário, isso prova (15.11).

Agora seja $\varepsilon > 0$ novamente. Tome $a \in J^\circ$ tal que $I(a) \leqslant I(J^\circ) + \varepsilon$. Tome $\delta > 0$ tal que $[a - \delta, a + \delta] \subseteq J$. Por (15.11),

$$\mathbb{P}\Big(\tfrac{S_n}{n} \in J\Big) \geqslant \mathbb{P}\Big(\tfrac{S_n}{n} \in [a - \delta, a + \delta]\Big) \geqslant e^{-I(a)\cdot n + o(n)} \geqslant e^{-I(J^\circ)\cdot n - \varepsilon n + o(n)}.$$

Como a desigualdade acima é válida para todo $\varepsilon > 0$, temos

$$\mathbb{P}\Big(\tfrac{S_n}{n} \in J\Big) \geqslant e^{-I(J^\circ)\cdot n + o(n)},$$

o que conclui a prova da cota inferior no Teorema de Cramér.

Prova da cota superior

Vejamos como a cota superior no Teorema 15.5 é uma consequência imediata do Teorema 15.2. Começamos com propriedades de monotonicidade da função taxa.

Proposição 15.12. *A função taxa I é não-crescente em $(-\infty, \mu]$ e não-decrescente em $[\mu, +\infty)$. Além disso, $I(\mu) = 0$,*

$$I(a) = \sup_{t \geqslant 0}[at - \log M(t)] \text{ para } a \geqslant \mu \text{ e } I(a) = \sup_{t \leqslant 0}[at - \log M(t)] \text{ para } a \leqslant \mu.$$

Demonstração. Tomando $t = 0$, temos $[at - \log M(t)] = 0$, logo $I(a) \geqslant 0$ para todo $a \in \mathbb{R}$. Agora, pela desigualdade de Jensen, $M(t) = \mathbb{E}[e^{tX}] \geqslant e^{\mathbb{E}tX} = e^{t\mu}$, donde

$$\mu t - \log M(t) \leqslant 0.$$

Isso implica que $I(\mu) = 0$. Isso também implica que, para $a > \mu$ e $t < 0$, $at - \log M(t) < 0$, assim $I(a) = \sup_{t \geqslant 0}[at - \log M(t)]$. Analogamente, para $a < \mu$, temos $I(a) = \sup_{t \leqslant 0}[at - \log M(t)]$.

Para provar monotonicidade em $[\mu, +\infty)$, vejamos que, se $\mu < c < a$,

$$I(a) = \sup_{t \geqslant 0}[at - \log M(t)] \geqslant \sup_{t \geqslant 0}[ct - \log M(t)] = I(c) \geqslant 0 = I(\mu).$$

A prova da monotonicidade da função taxa em $(-\infty, \mu]$ se faz de modo análogo. $\square$

Demonstração da cota superior no Teorema 15.5. Escrevemos $\bar{J} = [c, a] \subseteq \mathbb{R}$. Se $c \leqslant \mu \leqslant a$, $I(\bar{J}) = 0$ e não há nada a provar. Podemos então assumir que $a < \mu$, pois o caso $c > \mu$ é análogo. Seja $\varepsilon > 0$. Pela Proposição 15.12, temos

$$I(\bar{J}) = I(a) = \sup_{t \leqslant 0}[at - \log M(t)],$$

observe que a Proposição 15.12 foi utilizada em ambas as igualdades acima, na primeira, a monotonicidade, na segunda, que o supremo é atingido em $t \leqslant 0$. Logo, dado $\varepsilon > 0$, podemos tomar $t \leqslant 0$ tal que $[at - \log M(t)] \geqslant I(\bar{J}) - \varepsilon$. Agora, usando a estimativa (15.1) obtemos

$$\mathbb{P}\Big(\tfrac{S_n}{n} \in J\Big) \leqslant \mathbb{P}\Big(\tfrac{S_n}{n} \leqslant a\Big) \leqslant e^{-I(\bar{J}) \cdot n + \varepsilon n}.$$

Como ε é arbitrário, $\mathbb{P}\Big(\frac{S_n}{n} \in J\Big) \leqslant e^{-I(\bar{J}) \cdot n + o(n)}$, concluindo a prova. $\square$

Apêndice A

Preliminares

Neste apêndice vamos listar alguns conceitos preliminares que serão utilizados ao longo do livro. Na Seção A.1, listamos um conjunto de conceitos do Cálculo que suporemos como pré-requisito para todo o livro. Se alguma parte dessa seção parece muito difícil ou misteriosa, isso indica que uma base mais sólida de Cálculo é necessária antes que se comece a ler este livro.

Na Seção A.2, faremos uma exposição auto-contida de um tópico que não é visto em qualquer curso de Cálculo: a expansão de Taylor. Ter familiaridade com esse tópico não é pré-requisito, ele pode ser aprendido com uma breve leitura da própria Seção A.2 e somente será usado nas Seções 9.2, 9.3, e 10.2.

Na Seção A.3, listamos um conjunto de conceitos de Análise Real que suporemos como pré-requisito para as partes mais avançadas do livro. O leitor que já está familiarizado com esses conceitos poderá ver as generalizações de $\limsup$ e $\liminf$ na Seção C.2, e com isso cobrir todo o livro.

A.1 Cálculo

Nesta seção, listamos de forma muito sucinta um conjunto de conceitos do Cálculo que suporemos como pré-requisito para todo o livro.

Sequências e séries

Dada uma sequência $(x_n)_n$ de números reais e $L \in \mathbb{R}$, dizemos que $x_n \to L$ quando $n \to \infty$, também denotado por $\lim_{n\to\infty} x_n = L$ se, para todos $L' < L$ e $L'' > L$ existe n_0 tal que $L' \leqslant x_n \leqslant L''$ para todo $n \geqslant n_0$. Se for óbvio que o índice em questão é n e o ponto é ∞, podemos escrever apenas $x_n \to L$ omitindo o "quando $n \to \infty$". Também podemos omitir o "∞", escrevendo apenas $\lim_n x_n = L$. De forma similar, dizemos que $x_n \to +\infty$ se, para todo $L' < \infty$, existe n_0 tal que $x_n \geqslant L'$ para todo $n \geqslant n_0$. A definição de $x_n \to -\infty$ é análoga com $L'' > -\infty$. Se não for o caso que $x_n \to L$, escrevemos $x_n \not\to L$. Dizemos que $(x_n)_n$ *converge* se $\lim_n x_n$ existe e é finito. Caso contrário, ou seja, se $\lim_n x_n$ não existe ou é $\pm\infty$, dizemos que a sequência $(x_n)_n$ *diverge*.

Escrevemos $x_n \uparrow a$ ou $x_n \downarrow a$ para indicar que $(x_n)_n$ é não-decrescente, ou não-crescente, e $x_n \to a$. De forma análoga, escrevemos $x_n \uparrow +\infty$ ou $x_n \downarrow -\infty$ para limites infinitos. Salientamos que toda sequência monótona tem um limite, possivelmente infinito.

O limite satisfaz às seguintes propriedades. Se $x_n \to a \in \mathbb{R}$ e $y_n \to b \in \mathbb{R}$, então $x_n + cy_n \to a + bc$ para todo $c \in \mathbb{R}$. Se $a_n \leqslant b_n \leqslant c_n$ para todo n, $a_n \to L$ e $c_n \to L$, então $b_n \to L$, onde L pode ser um número real ou $\pm\infty$. Usando-se bisseção e essas propriedades, pode-se provar o seguinte.

Teorema A.1. *Sejam $A, B \subseteq \mathbb{R}$ não-vazios tais que $a \leqslant b$ para todo $a \in A$ e $b \in B$. Então existe $c \in \mathbb{R}$ tal que $a \leqslant c \leqslant b$ para todos $a \in A$ e $b \in B$.*

Dada uma sequência $(x_n)_n$ de números reais, definimos a *série*

$$\sum_{n=1}^{\infty} x_n = \lim_n \sum_{k=1}^{n} x_k$$

como o limite das somas parciais, caso exista. Se ademais o limite for finito, diremos que *a série converge*. Caso contrário, dizemos que *a série diverge*. Observe que a soma de números não-negativos sempre está bem definida, podendo ser finita ou infinita.

Se $\sum_{n=1}^{\infty} x_n$ converge, então $x_n \to 0$ e $\sum_{k=n}^{\infty} x_k \to 0$.

A *série geométrica* $\sum_{n=1}^{\infty} q^n$ converge se $-1 < q < 1$, e neste caso ela converge para $q/(1-q)$. A chamada *p-série*, dada por $\sum_{n=1}^{\infty} n^{-p}$, converge se, e somente se, $p > 1$. Veremos a justificativa mais adiante. O caso particular com $p = 1$ se chama *série harmônica*, dada por $\sum_{n=1}^{\infty} \frac{1}{n}$, que diverge.

Dizemos que os termos x_n são *somáveis*, e que a respectiva série é *absolutamente convergente*, se $\sum_{n=1}^{\infty} |x_n| < \infty$. Nesse caso, $\sum_{n=1}^{\infty} x_n$ também converge. Se $|x_n| \leqslant y_n \, \forall n$ e $\sum_{n=1}^{\infty} y_n < \infty$, então os x_n são somáveis.

Uma série é chamada *série alternada* se os sinais dos termos são alternados entre positivo e negativo. Se, ademais, $|x_n| \downarrow 0$, então a série alternada $\sum_{n=1}^{\infty} x_n$ converge.

Se uma série é convergente mas não é absolutamente convergente, dizemos que ela é *condicionalmente convergente*. Um exemplo é a chamada *série harmônica alternada*, dada por $\sum_{n=1}^{\infty} \frac{(-1)^n}{n}$; essa série converge pelo parágrafo anterior, e ela é condicionalmente convergente pois $\sum_{n=1}^{\infty} n^{-1} = +\infty$.

Limites de funções

Sejam $f : I \to \mathbb{R}$, onde $I \subseteq R$ é um intervalo aberto, e $a \in I$. Dizemos que $\lim_{x \to a} f(x) = L$ se, para toda sequência $(x_n)_n$ em I tal que $x_n \neq a$ para todo n e $x_n \to a$, vale $f(x_n) \to L$. Dizemos que $\lim_{x \to a^+} f(x) = L$ ou $f(a+) = L$ se para toda sequência $(x_n)_n$ em I tal que $x_n > a$ para todo n e $x_n \to a$, vale $f(x_n) \to L$. Definição análoga vale para $\lim_{x \to a^-} f(x) = f(a-) = L$.

Observação A.2. Na definição de $\lim_{x \to a^+} f(x) = L$, podemos nos restringir às sequências $(x_n)_n$ tais que $x_n > a$ para todo n e que $x_n \downarrow a$, isto é, podemos supor que $(x_n)_n$ é monótona. $\triangle$

De maneira similar, dizemos que $\lim_{x \to +\infty} f(x) = L$ se $f(x_n) \to L$ para qualquer sequência $(x_n)_n$ tal que $x_n \uparrow +\infty$, e tomamos definição análoga para $\lim_{x \to -\infty} f(x) = L$.

Dizemos que f *é contínua em* a se $f(a) = \lim_{x \to a} f(x)$. Dizemos que

f é contínua à direita em $a \in \mathbb{R}$ se $f(a) = \lim_{x \to a^+} f(x)$, análogo para continuidade à esquerda. Dizemos que *f é contínua* se f é contínua em todo ponto de I. Dizemos que *f é contínua por partes* se, em cada intervalo finito, f tem no máximo uma quantidade finita de pontos onde é descontínua e, nos pontos onde f é descontínua, f tem ambos os limites laterais definidos e finitos. Dizemos que f é uma *função-degrau* se f é contínua por partes e, em todo intervalo finito, f assume apenas finitos valores.

Integral de Riemann

Seja $f : [a, b] \to \mathbb{R}$ e $L \in \mathbb{R}$. Dizemos que a *integral de Riemann* de f em $[a, b]$ é igual a L, o que denotamos por

$$\int_a^b f(x)\,\mathrm{d}x = L,$$

se, para todos $L' < L$ e $L'' > L$, existe $\varepsilon > 0$ tal que, para todo $k \in \mathbb{N}$, para todos $x_0 < x_1 < \cdots < x_k$ com $x_0 = a$, $x_k = b$ e $x_j - x_{j-1} < \varepsilon$ para cada $j = 1, \ldots, k$, vale que

$$L' \leqslant \sum_{j=1}^{k} f(y_j)(x_j - x_{j-1}) \leqslant L'', \quad \text{para todos } y_j \in [x_{j-1}, x_j].$$

O somatório na expressão acima é chamado *soma de Riemann*. Dizemos que f é *Riemann-integrável* se existe $L \in \mathbb{R}$ tal que $\int_a^b f(x)\,\mathrm{d}x = L$.

Teorema A.3. *Se $f : [a, b] \to \mathbb{R}$ é contínua, então f é Riemann-integrável.*

Teorema A.4. *Se $f : [a, b] \to \mathbb{R}$ é monótona, então f é Riemann-integrável.*

Teorema A.5. *Se $f : [a, b] \to \mathbb{R}$ é uma função Riemann-integrável e $g : [a, b] \to \mathbb{R}$ é uma função contínua por partes, então fg é Riemann-integrável.*

Integrais impróprias e iteradas

Seja $f : \mathbb{R} \to [0, +\infty)$ uma função não-negativa. Definimos as *integrais impróprias*

$$\begin{aligned}
\int_a^{+\infty} f(x)\,\mathrm{d}x &= \lim_{b\to+\infty} \int_a^b f(x)\,\mathrm{d}x, \\
\int_{-\infty}^b f(x)\,\mathrm{d}x &= \lim_{a\to-\infty} \int_a^b f(x)\,\mathrm{d}x, \\
\int_{-\infty}^{+\infty} f(x)\,\mathrm{d}x &= \int_0^{+\infty} f(x)\,\mathrm{d}x + \int_{-\infty}^0 f(x)\,\mathrm{d}x.
\end{aligned}$$

Dizemos que a integral imprópria é *convergente* se o limite correspondente existe e é finito.

As integrais impróprias podem ajudar a determinar se as respectivas séries são convergentes e vice-versa. Se $f : [1, +\infty) \to [0, +\infty)$ é não-crescente e definimos $a_n = f(n)$, então a série $\sum_{n=1}^{+\infty} a_n$ converge se, e somente se, a integral $\int_1^{+\infty} f(x)\mathrm{d}x$ converge. Como $\int_1^{+\infty} x^{-p}\mathrm{d}x$ converge se, e somente se, $p > 1$, justificamos o critério de convergência da p-série $\sum_{n=1}^{\infty} n^{-p}$ enunciado anteriormente.

Seja $f : \mathbb{R}^n \to [0, +\infty)$ uma função não-negativa. Dizemos que f é *Riemann-integrável* se, para qualquer escolha de $-\infty < a_j < b_j < \infty$, a integral iterada

$$\int_{a_n}^{b_n} \cdots \int_{a_1}^{b_1} f(x_1, \ldots, x_n)\,\mathrm{d}x_1 \cdots \mathrm{d}x_n$$

está definida, é finita, e não depende da ordem de integração.

A.2 Expansão de Taylor

Nem todo curso de Cálculo inclui a expansão de Taylor. Faremos aqui uma breve exposição para preencher esta lacuna. Polinômios de Taylor serão usados em algumas passagens dos Capítulos 9 e 10 e no Apêndice B.

Teorema A.6 (Expansão de Taylor)**.** *Sejam $f : (a, b) \to \mathbb{R}$ e $z \in (a, b)$. Se f for k vezes derivável no ponto z, então*

$$f(z+h) = f(z) + f'(z)h + \tfrac{1}{2}f''(z)h^2 + \cdots + \tfrac{1}{k!}f^{(k)}(z)h^k + r(h),$$

para todo h tal que $z + h \in (a, b)$, onde a função resto r é tal que

$$\frac{r(h)}{h^k} \to 0$$

quando $h \to 0$. Se $f^{(k+1)}$ existe e é contínua em (a, b), então

$$r(h) = \tfrac{f^{(k+1)}(x)}{(k+1)!}h^{k+1}$$

para algum x entre z e $z + h$.

Para lembrar-se da fórmula, basta ignorar o resto e impor que o valor da função e das k primeiras derivadas de ambos lados da equação coincidam no ponto $h = 0$. A última fórmula se chama *resto de Lagrange*.

Demonstração. Veja que

$$r(h) = f(z+h) - \left[f(z) + f'(z) \cdot h + \tfrac{1}{2}f''(z)h^2 + \cdots + \tfrac{1}{k!}f^{(k)}(z)h^k\right]$$

satisfaz

$$r(0) = r'(0) = r''(0) = \cdots = r^{(k)}(0) = 0. \tag{A.7}$$

Aplicando a regra de L'Hôpital $k - 1$ vezes, obtemos

$$\lim_{h\to 0} \frac{r(h)}{h^k} = \frac{1}{k!} \lim_{h\to 0} \frac{r^{(k-1)}(h)}{h} = \frac{1}{k!}r^{(k)}(0) = 0,$$

provando a primeira parte. Para a segunda parte, consideramos o caso $h > 0$, pois $h < 0$ é idêntico e $h = 0$ é trivial. Sejam M' e M o mínimo e o máximo de $r^{(k+1)}$ em $[0, h]$. Pelo Teorema do Valor Extremo, M' e M são atingidos por $r^{(k+1)}$ em pontos de $[0, h]$. Usando (A.7) e integrando $k + 1$ vezes, pela

cota $M' \leqslant r^{(k+1)}(s) \leqslant M$ obtemos $M' \leqslant \frac{(k+1)!}{h^{k+1}} r(h) \leqslant M$. Pelo Teorema do Valor Intermediário, existe $t \in [0, h]$ tal que $h^{(k+1)}(t) = \frac{(k+1)!}{h^{k+1}} r(h)$. Como $f^{(k+1)}(z+t) = h^{(k+1)}(t)$, isso conclui a prova. $\square$

A.3 Análise Real

Este apêndice lista os principais tópicos de um curso introdutório de Análise Real que serão necessários para acessar as partes mais avançadas deste livro, como por exemplo as Seções 1.4 e 7.4, e portanto todas aquelas que dependem direta ou indiretamente delas, incluindo os Capítulos 12, 13 e 14.

Teorema A.8 (Subsubsequências)**.** *Seja* $(x_n)_n$ *e* $L \in \mathbb{R}$. *Então* $\lim_n x_n = L$ *se, e somente se, toda subsequência* $(x_{n_k})_k$ *possui uma subsubsequência* $(x_{n_{k_j}})_j$ *tal que* $\lim_j x_{n_{k_j}} = L$. *A mesma equivalência vale quando* $L = \pm\infty$.

Dado um conjunto $A \subseteq \mathbb{R}$, dizemos que A é *aberto* se, para todo $x \in A$, existe $\varepsilon > 0$ tal que $(x - \varepsilon, x + \varepsilon) \subseteq A$. Dizemos que A é *fechado* se A^c é aberto. Dizemos que A é *denso* em $\mathbb{R}$ se $A \cap B \neq \emptyset$ para todo aberto $B \subseteq \mathbb{R}$.

Teorema A.9. *Se* $A \subseteq \mathbb{R}$ *é um conjunto aberto, então existem intervalos abertos* $I_1, I_2, I_3, \ldots$ *tais que* $A = \cup_k I_k$.

Dado um conjunto $A \subseteq \mathbb{R}^n$, dizemos que A é *aberto* se, para todo $\boldsymbol{x} \in A$, existe $\varepsilon > 0$ tal que $\{\boldsymbol{y} : \|\boldsymbol{y} - \boldsymbol{x}\| < \varepsilon\} \subseteq A$.

Teorema A.10. *Uma função* $f : \mathbb{R}^n \to \mathbb{R}$ *é contínua se, e somente se,* $f^{-1}(A)$ *é aberto para todo* $A \subseteq \mathbb{R}$ *aberto.*

Teorema A.11. *Se* $A \subseteq \mathbb{R}^n$ *é um conjunto aberto, então existem paralelepípedos abertos* $I_1, I_2, I_3, \ldots$ *tais que* $A = \cup_k I_k$. *Um paralelepípedo aberto é um conjunto da forma* $I = (a_1, b_1) \times \cdots \times (a_n, b_n) \subseteq \mathbb{R}^n$.

Recordemos a definição de conjunto enumerável dada na página 24. Mais formalmente, um conjunto qualquer J é enumerável se, e somente se, J é finito ou existe uma função $f : \mathbb{N} \to J$ sobrejetiva.

Teorema A.12. *O conjunto $\mathbb{R}$ não é enumerável.*

Corolário A.13. *Todo intervalo não-degenerado é não-enumerável*

Dizemos que um conjunto $K \subseteq \mathbb{R}$ é *compacto* se K é fechado e limitado.

Observação A.14. Em alguns livros há uma definição mais abstrata de conjuntos compactos, e o Teorema de Heine-Borel diz que a definição mais abstrata é equivalente à definição acima. $\triangle$

Teorema A.15. *Se $\mathbb{R} \supseteq K_1 \supseteq K_2 \supseteq K_3 \supseteq \cdots$ são compactos e $K_n \neq \emptyset$ para todo $n \in \mathbb{N}$ então $\cap_n K_n \neq \emptyset$.*

Teorema A.16 (Weierstrass)**.** *Seja $g : \mathbb{R} \to \mathbb{R}$ uma função contínua e periódica, de período 2π. Então, para todo $\varepsilon > 0$, existem $m \in \mathbb{N}$ e $a_{-m}, \ldots, a_m \in \mathbb{C}$ tais que*

$$\Big| \sum_{k=-m}^{m} a_k e^{ikx} - g(x) \Big| \leqslant \varepsilon$$

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

Dado $J \subseteq \mathbb{R}$ e $f : J \to \mathbb{R}$, dizemos que f é *uniformemente contínua* se, para todo $\varepsilon > 0$, existe $\delta > 0$ tal que $|f(x) - f(y)| < \varepsilon$ para todos $x, y \in J$ tais que $|x - y| < \delta$. A diferença entre f ser contínua e uniformemente contínua é que δ não pode depender de x.

Teorema A.17. *Se $f : K \to \mathbb{R}$ é contínua e $K \subseteq \mathbb{R}$ é compacto, então f é uniformemente contínua.*

Apêndice B

Fórmula de Stirling

Este apêndice tem como objetivo demonstrar a importante aproximação abaixo.

Teorema B.1 (Fórmula de Stirling)**.** *Quando* $n \to \infty$*,*

$$\frac{n!}{n^n\, e^{-n}\, \sqrt{2\pi n}} \to 1.$$

Obtenção da fórmula e demonstração

Para entender como surge essa a fórmula, observe que

$$\log n! = \log 1 + \log 2 + \cdots + \log n = \sum_{k=1}^{n} \log k$$

é uma aproximação para

$$\int_0^n \log x \,\mathrm{d}x = n \log n - n = \log(n^n e^{-n}).$$

Seguindo argumentos de aproximação, poderíamos mostrar que

$$\log n! = n \log n - n + r(n), \quad \text{onde } \tfrac{r(n)}{n} \to 0.$$

Isto já seria o suficiente em muitas aplicações, mas queremos obter uma aproximação mais fina. De fato, queremos aproximar assintoticamente $n!$ e não apenas $\log n!$.

Admitindo uma correção polinomial, busquemos uma aproximação da forma

$$\lambda n^n e^{-n} n^\alpha$$

com $\lambda > 0$ e $\alpha \in \mathbb{R}$. Veremos agora que tal aproximação é de fato válida com $\alpha = \frac{1}{2}$ e $\lambda > 0$ desconhecido. Posteriormente, descobriremos que $\lambda = \sqrt{2\pi}$.

Escrevendo

$$c_n = \log\left(\frac{n^n e^{-n} n^\alpha}{n!}\right),$$

temos

$$c_{n+1} - c_n = \log\Big((n+1)\Big(\tfrac{n+1}{n}\Big)^n \tfrac{e^{-n-1}}{e^{-n}} \Big(\tfrac{n+1}{n}\Big)^\alpha \tfrac{n!}{(n+1)!}\Big) \tag{B.2}$$

$$= \Big[n \log(1 + \tfrac{1}{n}) - 1\Big] + \alpha \log(1 + \tfrac{1}{n}). \tag{B.3}$$

Fazendo a expansão de Taylor de $\log(1+x)$ para $x \in [0,1]$ temos

$$\log(1+x) = x - \frac{x^2}{2} + r(x),$$

onde $r(x)$ é igual a $\frac{2}{(1+z)^3}\frac{x^3}{6}$ para algum $z \in [0,x]$ e satisfaz $0 \leqslant r(x) \leqslant \frac{x^3}{3}$.

Continuando o desenvolvimento de $c_{n+1} - c_n$, temos

$$\begin{aligned} c_{n+1} - c_n &= \Big[n\big(\tfrac{1}{n} - \tfrac{1}{2n^2} + r(\tfrac{1}{n})\big) - 1\Big] + \alpha\Big(\tfrac{1}{n} - \tfrac{1}{2n^2} + r(\tfrac{1}{n})\Big) \\ &= \tfrac{\alpha}{n} - \tfrac{1}{2n} + n\, r(\tfrac{1}{n}) - \tfrac{\alpha}{2n^2} + \alpha\, r(\tfrac{1}{n}). \end{aligned}$$

Para anular os termos de ordem $\frac{1}{n}$, vamos escolher $\alpha = \frac{1}{2}$. Assim sendo,

$$c_{n+1} - c_n = n\, r(\tfrac{1}{n}) + \tfrac{1}{2}\, r(\tfrac{1}{n}) - \tfrac{1}{4n^2}.$$

Combinando a identidade acima com expansão de Taylor, obtemos

$$|c_{n+1} - c_n| \leqslant \tfrac{1}{2n^2},$$

que é somável. Logo $c_n \to c$ para algum $c \in \mathbb{R}$ e, portanto,

$$\frac{n!}{n^n e^{-n}\sqrt{n}} \to e^{-c} = \lambda$$

para algum $\lambda > 0$. Resta mostrar que essa constante é dada por $\lambda = \sqrt{2\pi}$.

Cálculo da constante

Vamos provar que $\lambda = \sqrt{2\pi}$ de duas formas diferentes.

Usando a demonstração do Teorema de De Moivre-Laplace A primeira prova supõe que o leitor viu a demonstração do Teorema de De Moivre-Laplace na Seção 9.2. Observe inicialmente que a demonstração do Teorema de De Moivre-Laplace funciona assumindo a fórmula de Stirling com uma constante desconhecida λ no lugar de $\sqrt{2\pi}$.

Pela Desigualdade de Tchebyshev,

$$1 - \tfrac{1}{m^2} \leqslant \mathbb{P}\Big(-m \leqslant \tfrac{S_n - np}{\sqrt{npq}} \leqslant +m\Big) \leqslant 1.$$

Fazendo $n \to \infty$, temos pelo Teorema de De Moivre-Laplace que

$$1 - \tfrac{1}{m^2} \leqslant \frac{1}{\lambda}\int_{-m}^{m} e^{-\frac{x^2}{2}}\, \mathrm{d}x \leqslant 1.$$

Fazendo agora $m \to \infty$ obtemos

$$\frac{1}{\lambda}\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}}\,\mathrm{d}x = 1,$$

logo $\lambda = \sqrt{2\pi}$.

Usando o produto de Wallis O produto de Wallis é dado pela seguinte identidade

$$\frac{\pi}{2} = \prod_{n=1}^{\infty}\left(\frac{2n}{2n-1}\cdot\frac{2n}{2n+1}\right) = \lim_{n\to\infty}\frac{2}{1}\cdot\frac{2}{3}\cdot\frac{4}{3}\cdot\frac{4}{5}\cdot\frac{6}{5}\cdot\frac{6}{7}\cdots\frac{2n}{2n-1}\cdot\frac{2n}{2n+1},$$

que será demonstrada mais abaixo.

Tomando a raiz quadrada e usando que $\frac{2n}{2n+1} \to 1$ obtemos

$$\sqrt{\frac{\pi}{2}} = \lim_{n\to\infty}\frac{2\cdot 4\cdot 6\cdots(2n-2)}{3\cdot 5\cdot 7\cdots(2n-1)}\cdot\sqrt{2n}.$$

Multiplicando numerador e denominador da fração acima pelo produto dos n primeiros números pares chegamos a

$$\begin{aligned}\sqrt{\frac{\pi}{2}} &= \lim_{n\to\infty}\frac{2\cdot 2\cdot 4\cdot 4\cdot 6\cdot 6\cdots(2n-2)\cdot(2n-2)}{2\cdot 3\cdot 4\cdot 5\cdot 6\cdot 7\cdots(2n-2)\cdot(2n-1)}\cdot\frac{2n\cdot 2n}{2n}\cdot\frac{\sqrt{2n}}{2n}\\ &= \lim_{n\to\infty}\frac{2^{2n}(1^2\cdot 2^2\cdot 3^2\cdots n^2)}{(2n)!}\cdot\frac{1}{\sqrt{2n}} = \lim_{n\to\infty}\frac{2^{2n}(n!)^2}{(2n)!\,\sqrt{2n}}.\end{aligned}$$

Finalmente, substituindo na fórmula de Stirling chegamos a

$$\sqrt{\frac{\pi}{2}} = \lim_{n\to\infty}\frac{2^{2n}n^{2n}e^{-2n}\lambda^2 n}{(2n)^{2n}e^{-2n}\sqrt{2\lambda^2 n}\sqrt{2n}} = \sqrt{\frac{\lambda^2}{4}},$$

e portanto $\lambda = \sqrt{2\pi}$.

Demonstração do produto de Wallis Daremos a demonstração sob a forma de exercício. Seja

$$I_n = \int_0^{\pi/2} \operatorname{sen}^n x \, \mathrm{d}x, \qquad n \geqslant 0.$$

(a) Mostre que para todo $n \geqslant 2$ vale

$$I_n = \frac{n-1}{n} I_{n-2}.$$

Sugestão: Integrando $\operatorname{sen}^n x = \operatorname{sen}^{n-1} x \cdot \operatorname{sen} x$ por partes, mostre que $\int_0^{\pi/2} \operatorname{sen}^n x \, \mathrm{d}x = (n-1) \int_0^{\pi/2} (\operatorname{sen}^{n-2} x)(\cos^2 x) \, \mathrm{d}x = (n-1)[I_{n-2} - I_n]$.

(b) Verifique que para todo $n \geqslant 1$ vale

$$\frac{I_{2n-2}}{I_{2n-1}} = \frac{2n}{2n-1} \cdot \frac{2n}{2n+1} \cdot \frac{I_{2n}}{I_{2n+1}}.$$

(c) Verifique que $I_0 = \frac{\pi}{2}$ e $I_1 = 1$.

(d) Mostre por indução que para todo $n \geqslant 0$ vale

$$\frac{\pi}{2} = \frac{2}{1} \cdot \frac{2}{3} \cdot \frac{4}{3} \cdot \frac{4}{5} \cdot \frac{6}{5} \cdot \frac{6}{7} \cdots \frac{2n}{2n-1} \cdot \frac{2n}{2n+1} \cdot \frac{I_{2n}}{I_{2n+1}}.$$

(e) Mostre que $\frac{2n}{2n+1} = \frac{I_{2n+1}}{I_{2n-1}} \leqslant \frac{I_{2n+1}}{I_{2n}} \leqslant 1$, e portanto $\frac{I_{2n}}{I_{2n+1}} \to 1$, concluindo a prova do produto de Wallis.

Apêndice C

A Reta Real e o Infinito

O objetivo deste apêndice é tornar certos conceitos mais gerais e robustos. Alguns dos enunciados e argumentos vistos na Seção 5.2 em diante se tornam mais sucintos se permitirmos que determinadas quantidades sejam infinitas. Com esse objetivo introduziremos a noção de reta estendida na primeira parte deste apêndice.

Na segunda parte, definiremos os conceitos de supremo e ínfimo que generalizam as noções de máximo e mínimo, além dos conceitos de limites superior e inferior de uma sequência de elementos da reta estendida. Essa segunda parte pressupõe familiaridade com os tópicos do Apêndice A.3.

C.1 Reta estendida

Definimos o conjunto dos números reais estendidos, denotado por $[-\infty, +\infty]$, agregando-se os símbolos $+\infty$ e $-\infty$ ao conjunto $\mathbb{R}$. Salientamos que os símbolos $+\infty$ e $-\infty$ não são números reais. Às vezes escrevemos ∞ no lugar de $+\infty$.

Adotaremos as seguintes convenções:

- Declaramos que $-\infty < +\infty$ e $-\infty < c < +\infty$ para todo $c \in \mathbb{R}$.
- Deixamos sem definir $(+\infty) + (-\infty)$, ao que chamamos "$\infty - \infty$". Temos que cuidar que tudo o que escrevamos não contenha $\infty - \infty$.
- Também deixamos sem definir $\frac{z}{\pm\infty}$, qualquer que seja $z \in [-\infty, +\infty]$.
- Definimos $0 \cdot (\pm\infty) = 0$.
  No Cálculo, $0 \cdot (\pm\infty)$ era considerada uma forma indeterminada, mas em Teoria da Medida e Teoria da Probabilidade, essa definição é muito conveniente e faz sentido. Por exemplo, a área de um retângulo de largura zero e altura infinita é igual a zero.
- Definimos $(+\infty)+(+\infty) = +\infty$, $(-\infty)+(-\infty) = -\infty$, $(+\infty)\cdot(\pm\infty) = \pm\infty$, $(-\infty) \cdot (\pm\infty) = \mp\infty$, $-(-\infty) = +\infty$, $|\pm\infty| = +\infty$. Ademais, $c + (\pm\infty) = \pm\infty$ para todo $c \in \mathbb{R}$ e $a \cdot (\pm\infty) = \pm\infty$ para todo $a > 0$.

Estas convenções fazem com que a soma e produto em $[-\infty, +\infty]$ sejam operações comutativas, distributivas e associativas sempre que bem definidas, ou seja, exceto pela restrição do $\infty - \infty$.

Observe que algumas propriedades antes válidas para números reais agora exigem maior escrutínio. Mais precisamente, para $x, y, z \in [-\infty, +\infty]$, a relação $x + z = y + z$ não necessariamente implica que $x = y$, pois não podemos subtrair z de ambos os lados de uma igualdade antes de verificar que $|z| < \infty$. Da mesma forma, $x < y$ não implica que $x + z < y + z$.

Com a definição $0 \cdot (\pm\infty) = 0$ ao invés da forma indeterminada, não podemos dizer que $\lim_n(a_n b_n) = (\lim_n a_n)(\lim_n b_n)$ mesmo que os limites existam em $[-\infty, +\infty]$, pois a igualdade pode falhar em exemplos como $a_n = n$ e $b_n = \frac{1}{n}$. Entretanto, vale essa igualdade se as sequências são não-negativas e não-decrescentes.

Apesar do problema acima, limites funcionam bem com a soma, como em

$$\lim_{n\to\infty}(2 + n^2) = \lim_{n\to\infty} 2 + \lim_{n\to\infty} n^2 = 2 + (+\infty) = +\infty.$$

O mais importante é lembrar que sempre teremos problema se encontrarmos $\infty - \infty$.

C.2 Supremo e limite superior

A cada subconjunto $A \subseteq [-\infty, +\infty]$ podemos associar um elemento $\sup(A) \in [-\infty, +\infty]$, chamado *supremo de* A, dado pela menor cota superior possível para A. De modo análogo definimos o *ínfimo de* A, denotado por $\inf(A)$, como a maior das cotas inferiores de A (observe que $+\infty$ e $-\infty$ são sempre cotas superior e inferior, respectivamente, para qualquer $A \subseteq [-\infty, +\infty]$). Com essas definições, temos por exemplo que $\sup(\mathbb{N}) = +\infty$, $\inf(\mathbb{N}) = 1$, $\sup(\emptyset) = -\infty$, $\inf(\emptyset) = +\infty$ e $\sup(\mathbb{R}) = \sup([-\infty, +\infty]) = +\infty$. Se $A \subseteq B \subseteq [-\infty, +\infty]$, então $\sup(A) \leqslant \sup(B)$ e $\inf(A) \geqslant \inf(B)$.

Dada uma sequência $(x_n)_n$ de elementos de $[-\infty, +\infty]$, podemos sempre definir seu *limite superior* e seu *limite inferior*, denotados por $\limsup_n x_n$ e $\liminf_n x_n$, como

$$\limsup_n x_n = \inf_n \sup_{k \geqslant n} x_k \quad \text{e} \quad \liminf_n x_n = \sup_n \inf_{k \geqslant n} x_k.$$

Sempre vale $\liminf_n x_n \leqslant \limsup_n x_n$ e, quando há igualdade, dizemos que a sequência tem um limite dado por

$$\lim_n x_n = \liminf_n x_n = \limsup_n x_n.$$

Quando os números da sequência são todos reais, essa definição de $\lim_n x_n$ coincide com a definição dada no Apêndice A.1.

Uma propriedade útil é que $\limsup_n (x_n + y_n) \leqslant \limsup_n x_n + \limsup_n y_n$ desde que as somas não resultem em $\infty - \infty$. Ademais, se $x_n \leqslant y_n$ para todo n, vale $\limsup_n x_n \leqslant \limsup_n y_n$ e $\liminf_n x_n \leqslant \liminf_n y_n$. Além disso, $\liminf_n (-x_n) = -\limsup_n x_n$.

O uso de sup e lim sup é muito robusto porque nem todo conjunto tem um elemento maximal, e nem toda sequência tem um limite, mas todo conjunto tem um supremo e toda sequência tem um limite superior. Uma técnica poderosa para mostrar que $x_n \to L$ é a seguinte. Verificamos que, para

qualquer $\varepsilon > 0$ dado, $\limsup_n x_n \leqslant L + \varepsilon$ e $\liminf_n x_n \geqslant L - \varepsilon$, donde concluímos que, sendo ε arbitrário, $\limsup_n x_n = \liminf_n x_n = L$. A grande vantagem é que podemos estimar o limite superior comparando com outras sequências que sim têm limite, mesmo sem saber *a priori* que a sequência $(x_n)_n$ o tem.

Somas de números em $[0, +\infty]$ estão sempre bem definidas, através de uma fórmula bem simples. Se Λ denota um conjunto de índices e $x_\alpha \in [0, +\infty]$ para todo $\alpha \in \Lambda$, definimos

$$\sum_{\alpha \in \Lambda} x_\alpha = \sup_{\substack{A \subseteq \Lambda \\ A \text{ finito}}} \sum_{\alpha \in A} x_\alpha. \tag{C.1}$$

Apesar de que o conjunto Λ pode ser um conjunto não-enumerável, observamos que a soma acima poderá ser finita somente se, no máximo, uma quantidade enumerável dos termos forem positivos (exercício). Observe que, neste caso, $\sum_{n \in \mathbb{N}} x_n$ estende a definição de $\sum_{n=1}^{\infty} x_n$ dada no Apêndice A.1, ao permitir que x_n tome valor $+\infty$.

Observamos que $\sum_{\alpha \in \Lambda}(x_\alpha + y_\alpha) = \sum_{\alpha \in \Lambda} x_\alpha + \sum_{\alpha \in \Lambda} y_\alpha$ como consequência da definição acima (deixamos os detalhes como exercício). Em particular, se $x_\alpha \geqslant y_\alpha \geqslant 0$ para todo α, temos $\sum_{\alpha \in \Lambda} x_\alpha \geqslant \sum_{\alpha \in \Lambda} y_\alpha \geqslant 0$.

Teorema C.2. *Dada uma sequência $x_{j,k} \in [0, +\infty]$ duplamente indexada,*

$$\sum_{j=1}^{\infty} \sum_{k=1}^{\infty} x_{j,k} = \sum_{k=1}^{\infty} \sum_{j=1}^{\infty} x_{j,k}.$$

Demonstração. Mostraremos que cada soma domina a outra. Seja $a < \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} x_{j,k}$. Tome m tal que $\sum_{j=1}^{m} \sum_{k=1}^{\infty} x_{j,k} > a$. Logo, $\sum_{k=1}^{\infty} \sum_{j=1}^{m} x_{j,k} > a$, o que implica que $\sum_{k=1}^{\infty} \sum_{j=1}^{\infty} x_{j,k} > a$. Como isso vale para todo $a < \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} x_{j,k}$, concluímos que $\sum_{k=1}^{\infty} \sum_{j=1}^{\infty} x_{j,k} \geqslant \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} x_{j,k}$. A outra desigualdade é provada de forma idêntica. $\square$

Apêndice D

Elementos de Teoria da Medida

O presente apêndice é mais extenso e detalhado que os demais. Seu propósito é fornecer os elementos de Teoria da Medida necessários para compreender os tópicos mais avançados deste livro. As demostrações de alguns teoremas mais sofisticados podem ser ignoradas em uma primeira leitura.

D.1 Teorema π-λ de Dynkin

Nesta seção estudaremos o Teorema π-λ de Dynkin e o usaremos para provar o Teorema 3.37 (unicidade de medidas) e o Lema 13.7.

Antes de introduzir os conceitos usados na sua prova, vejamos de onde eles surgem. Suponha que queiramos verificar se duas medidas *finitas* μ e ν em um espaço mensurável $(\Omega, \mathcal{F})$ coincidem em uma grande classe de conjuntos, como $\mathcal{F}$. Suponha também que $\mu(\Omega) = \nu(\Omega)$. Em quantos conjuntos A precisamos testar que $\mu(A) = \nu(A)$? Testar todos $A \in \mathcal{F}$ seria demais. Havendo testado para uma sequência de conjuntos disjuntos $(A_n)_n$, a propriedade valerá para $(A_n^c)_n$ e também para $\cup_n A_n$, por propriedades básicas das medidas finitas.

Entretanto, havendo testado para um par de conjuntos $A, B \in \mathcal{F}$, não há garantia de que as duas medidas coincidam em $A \cap B$. Queremos mostrar que a classe

$$\mathcal{D} = \{A \in \mathcal{F} : \mu(A) = \nu(A)\} \tag{D.1}$$

contém muitos conjuntos. Essa classe já tem uma estrutura proveniente da forma como medidas finitas tratam complementos e uniões disjuntas, que formalizamos com a seguinte definição.

Definição D.2 (λ-sistema)**.** Seja Ω um espaço amostral. Uma classe $\mathcal{D}$ de subconjuntos de Ω é um *λ-sistema de* Ω, se $\mathcal{D}$ contém Ω e é fechada por complementos e uniões enumeráveis disjuntas,[26] isto é,

(1) $\Omega \in \mathcal{D}$,
(2) Se $A \in \mathcal{D}$, então $A^c \in \mathcal{D}$,
(3) Se $\{A_k\}_{k=1}^{\infty} \subseteq \mathcal{D}$, $A_k \cap A_j = \emptyset$ para todos $k \neq j$, então $\bigcup_{k=1}^{\infty} A_k \in \mathcal{D}$.

Formalizamos a discussão anterior em forma de proposição.

Proposição D.3. *Se μ e ν são medidas em um espaço mensurável $(\Omega, \mathcal{F})$ tais que $\mu(\Omega) = \nu(\Omega) < \infty$, então a classe $\mathcal{D}$ definida em (D.1) é um λ-sistema.*

Demonstração. Seja $A \in \mathcal{D}$. Temos $\mu(A^c) = \mu(\Omega) - \mu(A) = \nu(\Omega) - \nu(A) = \nu(A^c)$, logo $A^c \in \mathcal{D}$. Agora considere uma coleção enumerável $\{A_n\}_n \subseteq \mathcal{D}$ de conjuntos disjuntos. Então $\mu(\cup_n A_n) = \sum_n \mu(A_n) = \sum_n \nu(A_n) = \nu(\cup_n A_n)$, logo $\cup_n A_n \in \mathcal{D}$. $\square$

Agora observamos que, à estrutura da classe definida em (D.1) lhe falta um aspecto: ser fechada por interseções. Essa é a razão da definição de π-sistemas dada na Seção 3.6.

Lema D.4. *Um λ-sistema fechado por interseções é uma σ-álgebra.*

[26] Salientamos que alguns livros usam uma definição diferente de λ-sistema. É relativamente fácil verificar que as duas as definições são equivalentes.

Demonstração. Seja $\mathcal{F}$ um λ-sistema fechado por interseções. Como $\mathcal{F}$ é não-vazio e fechado por complementos, basta mostrar que é fechado por uniões enumeráveis. Seja $\{A_n\}_n \subseteq \mathcal{F}$. Defina $B_1 = A_1$ e $B_n = A_1^c \cap \cdots \cap A_{n-1}^c \cap A_n$ para todo $n \geqslant 2$. Observe que $B_n \in \mathcal{F}$ para todo n, pois $\mathcal{F}$ é fechado por interseções e complemento, e que $\cup_n A_n = \cup_n B_n$, sendo esta última uma união disjunta. Como $\mathcal{F}$ é fechado por uniões enumeráveis disjuntas, $\cup_n A_n = \cup_n B_n \in \mathcal{F}$. Isso conclui a prova. $\square$

Os conceitos de π-sistemas e λ-sistemas são úteis quando, por um lado, a classe $\mathcal{D}$ (grande e complicada) de conjuntos satisfazendo determinada propriedade formam naturalmente um λ-sistema e, por outro lado, podemos encontrar uma *subclasse* $\mathcal{C}$ (menor e mais simples) que forma um π-sistema e gera a σ-álgebra desejada. Este é o caso do Teorema 3.37 (unicidade de medidas). Para aplicar esses conceitos, usamos uma poderosa ferramenta.

Teorema D.5 (Teorema π-λ de Dynkin)**.** *Seja Ω um espaço amostral. Suponha que $\mathcal{C}$ é um π-sistema de subconjuntos de Ω e $\mathcal{D}$ é um λ-sistema em Ω. Se $\mathcal{D}$ contém $\mathcal{C}$, então $\mathcal{D}$ contém $\sigma(\mathcal{C})$.*

Demonstração. Definimos $\mathcal{G}$ como o menor λ-sistema[27] que contém o π-sistema $\mathcal{C}$, e vamos mostrar que $\mathcal{G}$ também é um π-sistema.

A ideia principal é considerar, para cada $B \in \mathcal{G}$, a classe

$$\mathcal{F}_B = \{A \in \mathcal{G} : A \cap B \in \mathcal{G}\}.$$

Podemos ver que $\mathcal{F}_B$ é um λ-sistema usando $A^c \cap B = ((A \cap B) \cup B^c)^c$.

Agora, seja $B \in \mathcal{C}$. Como $\mathcal{C}$ é um π-sistema, $\mathcal{C} \subseteq \mathcal{F}_B$. Mas $\mathcal{G}$ é o menor λ-sistema que contém $\mathcal{C}$, logo $\mathcal{F}_B = \mathcal{G}$.

Finalmente, seja $D \in \mathcal{G}$. Dado $B \in \mathcal{C}$, temos $D \in \mathcal{F}_B$, o que é equivalente a $B \in \mathcal{F}_D$. Portanto, $\mathcal{C} \subseteq \mathcal{F}_D$. Mas $\mathcal{G}$ é o menor λ-sistema que contém $\mathcal{C}$, donde $\mathcal{F}_D = \mathcal{G}$. Como isso vale para todo $D \in \mathcal{G}$, concluímos que $\mathcal{G}$ é um

[27] Isso é definido da mesma forma que a menor σ-álgebra que contém uma dada classe.

π-sistema. Pelo Lema D.4, $\mathcal{G}$ é uma σ-álgebra, logo $\mathcal{C} \subseteq \sigma(\mathcal{C}) \subseteq \mathcal{G} \subseteq \mathcal{D}$, provando o teorema. $\square$

Demonstração do Teorema 3.37 (unicidade de medidas). Suponha momentaneamente que $\mu(\Omega) = \nu(\Omega) < \infty$. Defina a classe $\mathcal{D}$ como em (D.1). Pela Proposição D.3, $\mathcal{D}$ é um λ-sistema. Pelo Teorema π-λ, $\sigma(\mathcal{C}) \subseteq \mathcal{D}$, o que quer dizer precisamente que $\mu = \nu$ em $\sigma(\mathcal{C})$, como queríamos demonstrar.

Agora abandonamos a suposição inicial, e supomos apenas que $\mu(A_n) = \nu(A_n) < \infty$ para alguma sequência $A_n \uparrow \Omega$ de conjuntos em $\mathcal{C}$, tal como enunciado. Para cada $n \in \mathbb{N}$, defina as medidas μ_n e ν_n por $\mu_n(A) = \mu(A \cap A_n)$ e $\nu_n(A) = \nu(A \cap A_n)$. Com essa definição, $\mu_n(\Omega) = \mu(A_n) = \nu(A_n) = \nu_n(\Omega) < \infty$. Repare que, para cada $C \in \mathcal{C}$, temos $C \cap A_n \in \mathcal{C}$ porque $\mathcal{C}$ é um π-sistema. Logo, $\mu_n(C) = \mu(C \cap A_n) = \nu(C \cap A_n) = \nu_n(C)$ para todo $C \in \mathcal{C}$.

Portanto, para cada n, as medidas μ_n e ν_n estão no caso anterior e, para todo $A \in \sigma(\mathcal{C})$, temos $\mu_n(A) = \nu_n(A)$. Usando continuidade por baixo de μ e ν, temos

$$\mu(A) = \lim_n \mu(A \cap A_n) = \lim_n \mu_n(A) = \lim_n \nu_n(A) = \lim_n \nu(A \cap A_n) = \nu(A),$$

o que conclui a prova. $\square$

Demonstraçãodo Lema 13.7. Sem perda de generalidade, podemos supor que $\Omega \in \mathcal{C}_\alpha$ para todo $\alpha \in \Lambda$. Consideramos inicialmente o caso em que $\Lambda = \{1, \ldots, n\}$ para algum $n \in \mathbb{N}$. Fixe $B_2 \in \mathcal{C}_2, \ldots, \mathcal{C}_n \in \mathcal{C}_n$. Defina

$$\mathcal{D} = \{A \in \mathcal{F}_1 : \mathbb{P}(A \cap B_2 \cap \cdots \cap B_n) = \mathbb{P}(A)\mathbb{P}(B_2) \cdots \mathbb{P}(B_n)\}.$$

Observe que $\mathcal{D}$ é um λ-sistema (exercício!). Como, por hipótese, $\mathcal{C}_1 \subseteq \mathcal{D}$, pelo Teorema π-λ, concluímos que $\mathcal{D} = \mathcal{F}_1$. Ou seja, $\mathbb{P}(A \cap B_2 \cap \cdots \cap B_n) = \mathbb{P}(A)\mathbb{P}(B_2) \cdots \mathbb{P}(B_n)$ para todo $A \in \mathcal{F}_1$. Como $B_2 \in \mathcal{C}_2, \ldots, \mathcal{C}_n \in \mathcal{C}_n$ são arbitrários, segue que $\mathcal{F}_1, \mathcal{C}_2, \ldots, \mathcal{C}_n$ são classes independentes. Por um argumento idêntico, $\mathcal{F}_1, \mathcal{F}_2, \mathcal{C}_3, \ldots, \mathcal{C}_n$ são classes independentes.

Aplicando o mesmo argumento n vezes, concluímos que $\mathcal{F}_1 \ldots, \mathcal{F}_n$ são classes independentes.

Consideramos agora o caso geral. Sejam $k \in \mathbb{N}$, $\alpha_1, \ldots, \alpha_k \in \Lambda$ distintos, e $A_1 \in \mathcal{F}_{\alpha_1}, \ldots, A_k \in \mathcal{F}_{\alpha_k}$. Como $\mathcal{C}_{\alpha_1}, \ldots, \mathcal{C}_{\alpha_k}$ são independentes por hipótese, pelo caso anterior $\mathcal{F}_{\alpha_1}, \ldots, \mathcal{F}_{\alpha_k}$ são independentes, logo $\mathbb{P}(A_1, \ldots, A_k) = \mathbb{P}(A_1) \cdots \mathbb{P}(A_k)$, provando o lema. $\square$

D.2 Teorema de Extensão de Carathéodory

Nesta seção estudaremos o Teorema de Extensão de Carathéodory e o usaremos para provar o Teorema 1.51 (existência da medida de Lebesgue).

Recordemo-nos da definição de álgebra vista na Seção 13.1. Dado um espaço amostral Ω e uma álgebra $\mathcal{A}$, dizemos que a função $\mu : \mathcal{A} \to [0, +\infty]$ é uma *medida finitamente aditiva* se $\mu(\emptyset) = 0$ e $\mu(A \cup B) = \mu(A) + \mu(B)$ para todo par $A, B \in \mathcal{A}$ disjuntos. Note que medidas finitamente aditivas podem não ser medidas. A teoria das medidas finitamente aditivas é bastante limitada, pois elas podem não ter a propriedade de serem contínuas por baixo. Felizmente, se uma medida finitamente aditiva está definida em uma álgebra, é suficiente pedir que μ seja "σ-aditiva sempre que possível" para garantir que ela não apresente anomalias. Dado um espaço amostral Ω e uma álgebra $\mathcal{A}$, dizemos que uma medida finitamente aditiva $\mu : \mathcal{A} \to [0, +\infty]$ é uma *pré-medida* se for σ-aditiva, isto é, $\mu(\bigcup_n A_n) = \sum_n \mu(A_n)$ para toda família enumerável de conjuntos disjuntos $A_n \in \mathcal{A}$ cuja união A esteja em $\mathcal{A}$.

Teorema D.6 (Teorema de Extensão de Carathéodory)**.** *Seja $\mathcal{A}$ uma álgebra em Ω e μ uma pré-medida em $(\Omega, \mathcal{A})$. Então existe uma medida $\bar{\mu}$ em $(\Omega, \sigma(\mathcal{A}))$ tal que $\bar{\mu}(A) = \mu(A)$ para todo $A \in \mathcal{A}$. Ademais, se $\mu(A_n) < \infty$ para alguma sequência $A_n \uparrow \Omega$ de conjuntos em $\mathcal{A}$, então há uma única medida $\bar{\mu}$ com essa propriedade.*

A demonstração será dada mais abaixo.

Agora fixe $\Omega = (0,1]$ e seja $\mathcal{E} = \{(a,b] : 0 \leqslant a \leqslant b \leqslant 1\}$, a classe dos intervalos contidos em $(0,1]$, abertos à esquerda e fechados à direita. Definimos $\mathcal{A}$ como a classe de todos os conjuntos que podem ser escritos como união disjunta e finita de elementos de $\mathcal{E}$. Observe que $\mathcal{A}$ é uma álgebra em $(0,1]$.

Vamos definir uma pré-medida μ em $\mathcal{A}$ da seguinte maneira. Dado $A = (a_1, b_1] \cup \cdots \cup (a_n, b_n]$ com $a_1 \leqslant b_1 \leqslant a_2 \leqslant b_2 \leqslant \ldots \leqslant a_n \leqslant b_n$, definimos

$$\mu(A) = \mu\big((a_1, b_1] \cup \cdots \cup (a_n, b_n]\big) = \sum_{j=1}^{n} b_j - a_j.$$

Afirmamos que a fórmula acima está bem definida ainda que A possa ser escrito de muitas formas diferentes como união finita disjunta de elementos de $\mathcal{E}$. Omitimos a tediosa prova deste fato, mas enfatizamos uma consequência importante, que a função $\mu : \mathcal{A} \to [0, +\infty]$ é finitamente aditiva.

Lema D.7. *A função $\mu : \mathcal{A} \to [0, +\infty]$ é uma pré-medida.*

Demonstração. É suficiente mostrar que, se $A_n \downarrow \emptyset$ e $A_n \in \mathcal{A}$ para todo n, então $\mu(A_n) \to 0$. Com efeito, dados D_n conjuntos disjuntos em $\mathcal{A}$ tais que $\cup_n D_n = D \in \mathcal{A}$, podemos definir $A_n = D \setminus \cup_{j=1}^{n} D_j$ de forma que $A_n \in \mathcal{A}$, $A_n \downarrow \emptyset$ e $\mu(D) = \mu(A_n) + \sum_{j=1}^{n} \mu(D_j) \to \sum_{j=1}^{\infty} \mu(D_j)$. Vamos mostrar por contraposição. Seja $\{A_n\}_n \subseteq \mathcal{A}$ uma sequência decrescente com $\mu(A_n) \geqslant 2\varepsilon$ para todo n. Vamos provar que $\cap_n A_n \neq \emptyset$. Para cada $k \in \mathbb{N}$, podemos tomar $B_k \in \mathcal{A}$ tal que $\overline{B_k} \subseteq A_k$ e $\mu(A_k \setminus B_k) \leqslant \varepsilon 2^{-k}$, basta fazer os intervalos A_k um pouco mais curtos do lado esquerdo. Pela aditividade de μ em $\mathcal{A}$, temos $\mu(\cup_{k=1}^{n}(A_k \setminus B_k)) \leqslant \varepsilon$ para cada n. Por outro lado, $A_n \setminus \cap_{k=1}^{n} B_k = \cup_{k=1}^{n}(A_n \setminus B_k) \subseteq \cup_{k=1}^{n} A_k \setminus B_k$, logo $\mu(A_n \setminus \cap_{k=1}^{n} B_k) \leqslant \varepsilon$, e como $\mu(A_n) \geqslant 2\varepsilon$, temos $\mu(\cap_{k=1}^{n} B_k) \geqslant \varepsilon$. Definimos então $K_n = \cap_{k=1}^{n} \overline{B_k} \neq \emptyset$. Como os K_n são conjuntos compactos, $\cap_{n=1}^{\infty} K_n \neq \emptyset$ pelo Teorema A.15. Portanto, $\cap_{n=1}^{\infty} A_n \supseteq \cap_{n=1}^{\infty} \overline{B_n} = \cap_{n=1}^{\infty} K_n \neq \emptyset$, concluindo a prova. $\square$

Combinando o lema acima com o Teorema de Extensão de Carathéodory, podemos estender μ a $\mathcal{B}((0,1])$. Munidos da existência de tal medida, podemos

finalmente justificar a existência da medida de Lebesgue em $\mathbb{R}$.

Demonstração do Teorema 1.51. Para todo $k \in \mathbb{Z}$, podemos definir a medida μ_k em $(\mathbb{R}, \mathcal{B})$ por $\mu_k(A) = \mu\big((A-k) \cap (0,1]\big)$, onde μ é a medida em $(0,1]$ obtida acima, e finalmente $m(A) = \sum_{k\in\mathbb{Z}} \mu_k(A)$. Observe que m é uma medida. Observe também que $m\big((a,b]\big) = b - a < \infty$ para todo $a \leqslant b \in \mathbb{R}$. Unicidade segue do Teorema 3.37 (unicidade de medidas): consideramos novamente o π-sistema $\mathcal{C} = \{(a,b] : a \leqslant b \in \mathbb{R}\}$, observamos que $\sigma(\mathcal{C}) = \mathcal{B}$, e que $(-n,n]$ são conjuntos em $\mathcal{C}$ de medida finita cuja união é $\mathbb{R}$. $\square$

No resto desta seção, vamos provar o Teorema de Extensão de Carathéodory.

Para cada subconjunto E de Ω, consideramos a coleção $\mathcal{D}_E$ de todas as sequências $(A_n)_n$ de conjuntos em $\mathcal{A}$ que cobrem E:

$$\mathcal{D}_E = \Big\{(A_n)_n : A_n \in \mathcal{A} \text{ para todo } n \text{ e } E \subseteq \bigcup\nolimits_n A_n\Big\}.$$

Definição D.8 (Medida exterior)**.** Seja μ uma pré-medida em $(\Omega, \mathcal{A})$. Definimos a *medida exterior* $\mu^* : \mathcal{P}(\Omega) \to [0, +\infty]$ como

$$\mu^*(E) = \inf_{(A_n)_n \in \mathcal{D}_E} \sum_n \mu(A_n).$$

Podemos pensar na medida exterior como uma forma de medir o conjunto visto de fora. Observe que ela está definida para qualquer subconjunto de Ω e não apenas para os elementos de $\mathcal{A}$.

Lema D.9 (Propriedades da medida exterior)**.** *A medida exterior μ^* satisfaz:*

(1) *Para todos $E \subseteq F \subseteq \Omega$, $\mu^*(E) \leqslant \mu^*(F)$,*
(2) *Para $(E_n)_{n\in\mathbb{N}}$ subconjuntos de Ω, vale $\mu^*(\cup_n E_n) \leqslant \sum_n \mu^*(E_n)$,*
(3) *Para $A \in \mathcal{A}$, vale $\mu^*(A) = \mu(A)$.*

Demonstração. Se $E \subseteq F$, então $\mathcal{D}_F \subseteq \mathcal{D}_E$, o que implica $\mu^*(E) \leqslant \mu^*(F)$.

Provaremos agora que $\mu^*(\cup_n E_n) \leqslant \sum_n \mu^*(E_n)$. Seja $\varepsilon > 0$. Para cada $n \in \mathbb{N}$, pela definição de $\mu^*(E_n)$ podemos tomar $(A_{n,k})_k$ em $\mathcal{A}$ tais que $E_n \subseteq \cup_k A_{n,k}$ e $\sum_k \mu(A_{n,k}) \leqslant \mu^*(E_n) + \varepsilon 2^{-n}$. Como $\cup_n E_n \subseteq \cup_{n,k} A_{n,k}$, pela definição de $\mu^*(\cup_n E_n)$ temos que $\mu^*(\cup_n E_n) \leqslant \sum_{n,k} \mu(A_{n,k}) = \sum_n \sum_k \mu(A_{n,k}) \leqslant \sum_n (\mu^*(E_n) + \varepsilon 2^{-n}) = \varepsilon + \sum_n \mu^*(E_n)$. Como ε é arbitrário, a desigualdade segue.

Finalmente provaremos que $\mu^*(A) = \mu(A)$ para $A \in \mathcal{A}$. Que $\mu^*(A) \leqslant \mu(A)$ segue imediatamente tomando-se $(A, \emptyset, \emptyset, \emptyset \dots) \in \mathcal{D}_A$. Mostraremos que $\mu^*(A) \geqslant \mu(A)$ usando a hipótese de que μ é σ-aditiva em $\mathcal{A}$. Seja $(A_n)_n \in \mathcal{D}_A$. Tome $B_n = A \cap A_n \cap (\cap_{k=1}^{n-1} A_k^c)$, de forma que eles formam uma partição de A. Logo, $\mu(A) = \sum_n \mu(B_n) \leqslant \sum_n \mu(A_n)$. Como isso vale para qualquer $(A_n)_n \in \mathcal{D}_A$, temos $\mu(A) \leqslant \mu^*(A)$, concluindo a prova. $\square$

Informalmente, os conjuntos problemáticos são maiores quando vistos por fora do que por dentro. Tentaremos excluir esse tipo de problema considerando conjuntos $A \subseteq \Omega$ tais que, para todo $E \subseteq \Omega$, vale que

$$\mu^*(E) = \mu^*(E \cap A) + \mu^*(E \cap A^c).$$

Os conjuntos A satisfazendo a essa condição são chamados *μ^*-mensuráveis*, e $\mathcal{F}^*$ denota a coleção desses conjuntos; isto é,

$$\mathcal{F}^* = \{A \subseteq \Omega : \mu^*(E) = \mu^*(E \cap A) + \mu^*(E \cap A^c) \text{ para todo } E \subseteq \Omega\}.$$

A segunda propriedade listada no lema acima implica que

$$\mu^*(E) \leqslant \mu^*(E \cap A) + \mu^*(E \cap A^c),$$

de forma que os conjuntos problemáticos são aqueles $A \subseteq \Omega$ para os quais a desigualdade acima é estrita para algum $E \subseteq \Omega$.

Proposição D.10 (Carathéodory)**.** *A classe $\mathcal{F}^*$ de conjuntos μ^*-mensuráveis é uma σ-álgebra em Ω que contém $\mathcal{A}$, e a restrição de μ^* a $\mathcal{F}^*$ é uma medida.*

Demonstração. Faremos uso extensivo do Lema D.9. Vamos provar primeiro que $\mathcal{A} \subseteq \mathcal{F}^*$. Sejam $A \in \mathcal{A}$, $E \subseteq \Omega$ e $\varepsilon > 0$. Pela definição de μ^* existem $A_1, A_2, \dots \in \mathcal{A}$ com $E \subseteq \cup_n A_n$ e $\sum_n \mu(A_n) \leqslant \mu^*(E) + \varepsilon$. Logo,

$$\begin{aligned}\mu^*(E) &\leqslant \mu^*(E \cap A) + \mu^*(E \cap A^c) \leqslant \sum_n \mu^*(A_n \cap A) + \sum_n \mu^*(A_n \cap A^c) \\ &= \sum_n \mu(A_n \cap A) + \sum_n \mu(A_n \cap A^c) = \sum_n \mu(A_n) \leqslant \mu^*(E) + \varepsilon,\end{aligned}$$

onde na segunda desigualdade usamos subaditividade de μ^* duas vezes e na primeira igualdade, o fato de que $\mu^* = \mu$ em $\mathcal{A}$. Como ε é arbitrário, temos que $A \in \mathcal{F}^*$, logo $\mathcal{A} \subseteq \mathcal{F}^*$.

Agora provaremos que $\mathcal{F}^*$ é uma σ-álgebra e μ^* é σ-aditiva em $\mathcal{F}^*$.

Passo 1. A classe $\mathcal{F}^*$ é uma álgebra.

Trivialmente, $\Omega \in \mathcal{F}^*$ e $A^c \in \mathcal{F}^*$ para todo $A \in \mathcal{F}^*$. Sejam $A, B \in \mathcal{F}^*$, usando a subaditividade de μ^* obtemos

$$\begin{aligned}\mu^*(E) &\leqslant \mu^*(E \cap (A \cap B)) + \mu^*(E \cap (A \cap B)^c) \leqslant \mu^*(E \cap A \cap B) \\ &\quad + \mu^*(E \cap A \cap B^c) + \mu^*(E \cap A^c \cap B) + \mu^*(E \cap A^c \cap B^c) \\ &= \mu^*(E \cap A) + \mu^*(E \cap A^c) = \mu^*(E),\end{aligned}$$

logo $A \cap B \in \mathcal{F}^*$; concluindo que $\mathcal{F}^*$ é uma álgebra.

Passo 2. Sejam $A_1, \dots, A_n \in \mathcal{F}^*$ disjuntos. Mostraremos que

$$\mu^*\big(E \cap (\cup_{k=1}^n A_k)\big) = \sum_{k=1}^n \mu^*(E \cap A_k)$$

para todo $E \subseteq \Omega$. Com efeito, como $A_1 \in \mathcal{F}^*$, temos que

$$\begin{aligned}\mu^*\big(E \cap (\cup_{k=1}^n A_k)\big) &= \mu^*\big(E \cap (\cup_{k=1}^n A_k) \cap A_1\big) + \mu^*\big(E \cap (\cup_{k=1}^n A_k) \cap A_1^c\big) \\ &= \mu^*\big(E \cap A_1\big) + \mu^*\big(E \cap (\cup_{k=2}^n A_k)\big),\end{aligned}$$

onde a última igualdade é devida ao fato que A_1 e $(\cup_{k=2}^n A_k)$ são disjuntos.

Utilizando o mesmo argumento para A_2, depois A_3, e assim por diante até A_n, obtemos a identidade desejada.

Passo 3. Sejam $A_1, A_2, \cdots \in \mathcal{F}^*$ disjuntos. Mostraremos que $\cup_n A_n \in \mathcal{F}^*$ e $\mu^*(\cup_n A_n) = \sum_n \mu^*(A_n)$.

Para isso, vamos mostrar que

$$\mu^*(E) \leqslant \mu^*(E \cap G) + \mu^*(E \cap G^c) \leqslant \sum_{k=1}^{\infty} \mu^*(E \cap A_k) + \mu^*(E \cap G^c) \leqslant \mu^*(E),$$

onde $G = \cup_n A_n$.

As duas primeiras desigualdades seguem da subaditividade de μ^*. Para mostrar a última desigualdade, definindo $G_n = \cup_{k=1}^n A_k$, temos

$$\begin{aligned}\mu^*(E) = \mu^*(E \cap G_n) + \mu^*(E \cap G_n^c) &= \sum_{k=1}^{n} \mu^*(E \cap A_k) + \mu^*(E \cap G_n^c) \\ &\geqslant \sum_{k=1}^{n} \mu^*(E \cap A_k) + \mu^*(E \cap G^c) \to \sum_{k=1}^{\infty} \mu^*(E \cap A_k) + \mu^*(E \cap G^c),\end{aligned}$$

onde a primeira igualdade segue do Passo 1, pois $G_n \in \mathcal{F}^*$, e a segunda igualdade segue do Passo 2. Isso conclui a demonstração do Passo 3.

Como $\mathcal{F}^*$ é uma álgebra fechada por uniões enumeráveis de conjuntos disjuntos, é um λ-sistema fechado por interseções, logo, pelo Lema D.4, $\mathcal{F}^*$ é uma σ-álgebra. Como μ^* é σ-aditiva em $\mathcal{F}^*$, isso conclui a prova da proposição. □

Demonstração do Teorema de Extensão de Carathéodory. Observe que, pela Proposição D.10, $\mathcal{A} \subseteq \mathcal{F}^*$, logo $\sigma(\mathcal{A}) \subseteq \mathcal{F}^*$. Ademais, μ^*, restrita a $\sigma(\mathcal{A})$, é σ-aditiva, logo é uma medida. Além disso, μ^*, restrita a $\mathcal{A}$, coincide com μ. Isso prova a parte de existência. Para unicidade, suponha que μ^* seja σ-finita, de forma que exista uma coleção enumerável $\{A_n\}_n$ em $\sigma(\mathcal{A})$ tal que $\mu^*(A_n) < \infty$ para todo n, e $\cup_n A_n = \Omega$. Da Definição D.8, para cada n existe uma coleção enumerável $\{A_{n,k}\}_k$ tal que $\cup_k A_{n,k} \supseteq A_n$ e $\mu^*(A_{n,k}) < \infty$.

Reindexando essa coleção enumerável $\{A_{n,k}\}_{n,k}$ como $\{B_j\}_{j\in\mathbb{N}}$ e definindo $C_\ell = B_1 \cup \cdots \cup B_\ell$, temos que $C_\ell \in \mathcal{A}$, $\mu(C_\ell) < \infty$ e $C_\ell \uparrow \Omega$. Observando-se que $\mathcal{A}$ é um π-sistema, segue do Teorema 3.37 (unicidade de medidas) que μ^* é a única medida em $\sigma(\mathcal{A})$ que coincide com μ em $\mathcal{A}$, concluindo essa demonstração. □

D.3 Operações com funções mensuráveis

Esta seção é uma continuação da Seção 3.7 e consiste em fornecer as demonstrações omitidas dos últimos lemas. Estas demonstrações não são particularmente difíceis, mas podem ser omitidas sem prejuízo à compreensão de outras seções do livro.

Demonstração do Lema 3.48. Como $\mathbb{Q}$ é denso em $\mathbb{R}$, para cada $\omega \in \Omega$, $f(\omega) < g(\omega)$ se, e somente se, existe $r \in \mathbb{Q}$ tal que $f(\omega) < r < g(\omega)$. Portanto, podemos expressar $\{\omega : f(\omega) < g(\omega)\}$ como a união de $\{\omega : f(\omega) < r < g(\omega)\}$ sobre todo $r \in \mathbb{Q}$. Então, podemos escrever

$$\{f < g\} = \bigcup_{r\in\mathbb{Q}} (\{f < r < g\}) = \bigcup_{r\in\mathbb{Q}} (\{f < r\} \cap \{r < g\}).$$

Como $\mathbb{Q}$ é enumerável, concluímos $\{f < g\} \in \mathcal{F}$. Além disso,

$$\begin{aligned}
\{f \neq g\} &= \{f < g\} \cup \{f > g\} \in \mathcal{F}, \\
\{f = g\} &= \{f \neq g\}^c \in \mathcal{F}, \\
\{f \leqslant g\} &= \{f < g\} \cup \{f = g\} \in \mathcal{F},
\end{aligned}$$

o que prova o lema. □

Demonstração do Lema 3.49. Começamos pela segunda parte. Suponha que a soma $f(\omega) + g(\omega)$ esteja bem definida para todo ω. Seja $b \in \mathbb{R}$. A função $b - g$ é mensurável, pois $\{b - g < c\} = \{g > b - c\} \in \mathcal{F}$ para todo $c \in \mathbb{R}$. Pelo

Lema 3.48, $\{f+g<b\}=\{f<b-g\}\in\mathcal{F}$, provando que $f+g$ é mensurável.

Agora mostraremos que a função fg é mensurável. Suponha inicialmente que f e g tomem valores em $\mathbb{R}$. Como $x\mapsto x^2$ é contínua, pelas afirmações demonstradas acima, $f+g$, $f-g$, $(f+g)^2$ e $(f-g)^2$ são mensuráveis e, como

$$fg=\frac{(f+g)^2-(f-g)^2}{4},$$

fg também é mensurável. Consideremos finalmente o caso geral, em que f e g tomam valores em $[-\infty,+\infty]$. Seja $A=\{\omega : f(\omega)\in\mathbb{R} \text{ e } g(\omega)\in\mathbb{R}\}$. Note que $A\in\mathcal{F}$. Defina $f_1=f\mathbb{1}_A$, $f_2=f\mathbb{1}_{A^c}$, $g_1=g\mathbb{1}_A$, $g_2=g\mathbb{1}_{A^c}$. Observe que f_2g_2 assume valores em $\{-\infty,0,+\infty\}$ e é mensurável. Por outro lado, pelo caso anterior f_1g_1 é mensurável. Como $fg=f_1g_1+f_2g_2$, concluímos que fg é mensurável. □

Segue do Lema 3.49 que, se $f:\Omega\to[-\infty,+\infty]$ é uma função mensurável e $a\in[-\infty,+\infty]$, então $-f$ e af são funções mensuráveis.

Demonstração do Lema 3.50. Para qualquer $a\in\mathbb{R}$, podemos escrever

$$\left\{\sup_{n\geqslant 1} f_n\leqslant a\right\}=\bigcap_{n=1}^{\infty}\{f_n\leqslant a\}.$$

Como $\{f_n\leqslant a\}\in\mathcal{F}$ para todo $n\geqslant 1$ e $\mathcal{F}$ é fechada por interseções enumeráveis, $\sup_n f_n$ é mensurável. Observando que

$$\inf_{n\geqslant 1} f_n=-\sup_{n\geqslant 1}(-f_n),$$

concluímos que $\inf_{n\geqslant 1} f_n$ também é uma função mensurável. Da definição de $\limsup$ e $\liminf$, temos

$$\limsup_n f_n=\inf_{n\geqslant 1}\sup_{m\geqslant n} f_m,$$
$$\liminf_n f_n=\sup_{n\geqslant 1}\inf_{m\geqslant n} f_m,$$

logo, ambas são funções mensuráveis. Finalmente, a função f definida no enunciado é dada pelo produto de $\limsup_n f_n$ pela função indicadora do conjunto $\{\limsup_n f_n = \liminf_n f_n\}$ e, como ambas são mensuráveis, o produto também é mensurável, o que conclui a prova. $\square$

D.4 Teoremas de Fubini e de Tonelli

Nesta seção vamos provar o Teorema 5.81 (existência da medida produto), o Lema 5.82, e os Teoremas de Fubini e de Tonelli.

Ao final da seção, vamos estender a teoria para o produto de n espaços, e com isso justificar o Teorema 4.5, a Proposição 4.10 e a Observação 5.89.

Observe que o Lema 5.82 e o Teorema de Tonelli valem trivialmente se g é a função indicadora de algum retângulo $A \times B$, pois, neste caso, $g(x,y) = \mathbb{1}_A(x) \cdot \mathbb{1}_B(y)$ é mensurável em cada variável, $\int_{\Omega_2} g(x,y)\nu(\mathrm{d}y) = \nu(B) \cdot \mathbb{1}_A(x)$ é mensurável em x, $\int_{\Omega_1} g(x,y)\mu(\mathrm{d}x) = \mu(A) \cdot \mathbb{1}_B(y)$ é mensurável em y, e $\int_{\Omega_1}(\int_{\Omega_2} g(x,y)\nu(\mathrm{d}y))\mu(\mathrm{d}x) = \mu(A) \cdot \nu(B) = \int_{\Omega_2}(\int_{\Omega_1} g(x,y)\mu(\mathrm{d}x))\nu(\mathrm{d}y)$.

Lema D.11. *Sejam $(\Omega_1, \mathcal{F}_1, \mu)$ e $(\Omega_2, \mathcal{F}_2, \nu)$ espaços de medida σ-finita. Seja $E \in \mathcal{F}_1 \otimes \mathcal{F}_2$ e defina $g(x,y) = \mathbb{1}_E(x,y)$. Para todo $x \in \Omega_1$ fixo, $g(x,y)$ é uma função mensurável de y. Ademais, $\int_{\Omega_2} g(x,y)\nu(\mathrm{d}y)$ define uma função mensurável de x.*

Demonstração. Para a primeira afirmação, basta observar que a classe de conjuntos E com essa propriedade forma uma σ-álgebra, que, pela observação acima, contém $\mathcal{F}_1 \times \mathcal{F}_2$ e portanto contém $\mathcal{F}_1 \otimes \mathcal{F}_2$.

Passamos à prova da segunda afirmação. Suponhamos inicialmente que ν seja uma medida finita. Seja $\mathcal{D}$ a classe dada pelos conjuntos E para os quais $\int_{\Omega_2} \mathbb{1}_E(x,y)\nu(\mathrm{d}y)$ define uma função mensurável de x. Pela observação feita no início desta seção, $\mathcal{D}$ contém o π-sistema $\mathcal{F}_1 \times \mathcal{F}_2$, que gera $\mathcal{F}_1 \otimes \mathcal{F}_2$. Usando o Teorema π-λ, basta verificar que $\mathcal{D}$ é um λ-sistema. Para $E = \Omega_1 \times \Omega_2$, temos $g(x,y) = 1$ para todo $(x,y) \in \Omega_1 \times \Omega_2$ e a propriedade vale

trivialmente. Se $E \in \mathcal{D}$, então $\int_{\Omega_2} \mathbb{1}_{E^c}(x,y)\nu(\mathrm{d}y) = \nu(\Omega_2) - \int_{\Omega_2} \mathbb{1}_E(x,y)\nu(\mathrm{d}y)$ é mensurável, logo $E^c \in \mathcal{D}$. Sejam $\{E_n\}_n \subseteq \mathcal{D}$ disjuntos. Pelo Teorema da Convergência Monótona, $\int_{\Omega_2} \mathbb{1}_{\cup_n E_n}(x,y)\nu(\mathrm{d}y) = \int_{\Omega_2} \sum_n \mathbb{1}_{E_n}(x,y)\nu(\mathrm{d}y) = \sum_n \int_{\Omega_2} \mathbb{1}_{E_n}(x,y)\nu(\mathrm{d}y)$, que é mensurável pelo Lema 3.50. Logo, $\cup_n E_n \in \mathcal{D}$, e portanto $\mathcal{D}$ é um λ-sistema, o que prova o lema no caso em que ν é finita.

Supondo agora que ν é σ-finita, tome $B_n \uparrow \Omega_2$ tais que $\nu(B_n) < \infty$ para todo n, e defina $\nu_n(B) = \nu(B \cap B_n)$ para todos $B \in \mathcal{F}_2$ e $n \in \mathbb{N}$. Pelo Teorema da Convergência Monótona, $\int_{\Omega_2} \mathbb{1}_E(x,y)\nu(\mathrm{d}y) = \lim_n \int_{\Omega_2} \mathbb{1}_E(x,y)\mathbb{1}_{B_n}(y)\nu(\mathrm{d}y) = \lim_n \int_{\Omega_2} \mathbb{1}_E(x,y)\nu_n(\mathrm{d}y)$, que, pelo caso anterior combinado com o Lema 3.50, é uma função mensurável de x. □

Lema D.12. *Sejam $(\Omega_1, \mathcal{F}_1, \mu)$ e $(\Omega_2, \mathcal{F}_2, \nu)$ dois espaços de medida σ-finita. Então, existe uma única medida $\mu \otimes \nu$ na σ-álgebra $\mathcal{F}_1 \otimes \mathcal{F}_2$ tal que $(\mu \otimes \nu)(A \times B) = \mu(A)\nu(B)$ para todos $A \in \mathcal{F}_1$ e $B \in \mathcal{F}_2$. Essa medida é dada por $(\mu \otimes \nu)(E) = \int_{\Omega_1}(\int_{\Omega_2} \mathbb{1}_E(x,y)\nu(\mathrm{d}y))\mu(\mathrm{d}x)$, para todo $E \in \mathcal{F}_1 \otimes \mathcal{F}_2$.*

Demonstração. A integral iterada está bem definida pelo Lema D.11, é uma função σ-aditiva de E pelo Teorema da Convergência Monótona, e vale zero quando $E = \emptyset$, portanto define uma medida. Esta medida atribui o valor correto aos retângulos devido à observação do início desta seção. Isso prova a existência. Para a unicidade, usaremos novamente que $\mathcal{F}_1 \times \mathcal{F}_2$ é um π-sistema que gera $\mathcal{F}_1 \otimes \mathcal{F}_2$. Tomando $A_n \uparrow \Omega_1$, com $\mu(A_n) < \infty$ para todo n, e $B_n \uparrow \Omega_2$ com $\nu(B_n) < \infty$ para todo n (o que é possível pois μ e ν são σ-finitas), temos que $A_n \times B_n \in \mathcal{F}_1 \times \mathcal{F}_2$ satisfaz $A_n \times B_n \uparrow \Omega_1 \times \Omega_2$ e $(\mu \otimes \nu)(A_n \times B_n) < \infty$ para todo n. Sendo assim, podemos aplicar o Teorema 3.37 (unicidade de medidas), o que conclui esta demonstração. □

Observe que o Teorema 5.81 segue do lema acima.

Demonstração do Lema 5.82. Se $g = \mathbb{1}_E$ para algum $E \in \mathcal{F}_1 \otimes \mathcal{F}_2$, basta aplicar o Lema D.11 diretamente e também com os papéis de x e y trocados, provando o lema para o caso de funções indicadoras. Estendemos a propriedade para funções simples não-negativas usando a linearidade da

integral e o Lema 3.49. Estendemos para o caso geral usando o Teorema da Convergência Monótona e o Lema 3.50. □

Demonstração do Teorema de Tonelli. Suponha inicialmente que $g = \mathbb{1}_E$ para algum $E \in \mathcal{F}_1 \otimes \mathcal{F}_2$. Aplicando o Lema D.12, temos $\int_{\Omega_1 \times \Omega_2} g \, \mathrm{d}(\mu \otimes \nu) = \int_{\Omega_1} (\int_{\Omega_2} g(x,y)\nu(\mathrm{d}y))\mu(\mathrm{d}x)$. Aplicando o lema com x e y trocados, temos $\int_{\Omega_1 \times \Omega_2} g \, \mathrm{d}(\mu \otimes \nu) = \int_{\Omega_2} (\int_{\Omega_1} g(x,y)\mu(\mathrm{d}x))\nu(\mathrm{d}y)$, o que prova o teorema no caso de funções indicadoras. Estendemos o teorema para uma funções simples não-negativas usando a linearidade das integrais. Finalmente, estendemos o teorema para o caso geral usando o Teorema da Convergência Monótona para cada integral envolvida. □

Demonstração do Teorema de Fubini. Estamos supondo que

$$\int_{\Omega_1} \Big(\int_{\Omega_2} |f(x,y)| \nu(\mathrm{d}y) \Big) \mu(\mathrm{d}x) < \infty.$$

Como $\int_{\Omega_2} |f(x,y)|\nu(\mathrm{d}y)$ é μ-integrável em x, é finito para μ-quase todo $x \in \Omega_1$. Logo, o conjunto N definido por

$$N = \Big\{ x \in \Omega_1 : \int_{\Omega_2} |f(x,y)| \, \nu(\mathrm{d}y) = +\infty \Big\}$$

satisfaz $\mu(N) = 0$. Portanto, $\int_{\Omega_2} f(x,y) \, \nu(\mathrm{d}y)$ está definido para μ-quase todo $x \in \Omega_1$, pois está definido para todo $x \in N^c$.

Observe que, pelo Lema 5.82, $f^+(x,y)$ e $f^-(x,y)$ são funções mensuráveis de y para todo x fixo e, ademais, $\int_{\Omega_2} f^+(x,y)\nu(\mathrm{d}y)$ e $\int_{\Omega_2} f^-(x,y)\nu(\mathrm{d}y)$ são funções mensuráveis de x. Usando que $\mu(N) = 0$, podemos desenvolver

$$\begin{aligned}
\int_{\Omega_1 \times \Omega_2} f \, \mathrm{d}(\mu \otimes \nu) &= \int_{\Omega_1 \times \Omega_2} f^+ \, \mathrm{d}(\mu \otimes \nu) - \int_{\Omega_1 \times \Omega_2} f^- \, \mathrm{d}(\mu \otimes \nu) \\
&= \int_{\Omega_1} \Big(\int_{\Omega_2} f^+(x,y) \, \nu(\mathrm{d}y) \Big) \mu(\mathrm{d}x) - \int_{\Omega_1} \Big(\int_{\Omega_2} f^-(x,y) \, \nu(\mathrm{d}y) \Big) \mu(\mathrm{d}x) \\
&= \int_{N^c} \Big(\int_{\Omega_2} f^+(x,y) \, \nu(\mathrm{d}y) \Big) \mu(\mathrm{d}x) - \int_{N^c} \Big(\int_{\Omega_2} f^-(x,y) \, \nu(\mathrm{d}y) \Big) \mu(\mathrm{d}x)
\end{aligned}$$

$$= \int_{N^c} \Big(\int_{\Omega_2} f(x,y)\, \nu(\mathrm{d}y) \Big) \mu(\mathrm{d}x)$$
$$= \int_{\Omega_1} \Big(\int_{\Omega_2} f(x,y)\, \nu(\mathrm{d}y) \Big) \mu(\mathrm{d}x).$$

A prova de que $\int_{\Omega_1 \times \Omega_2} f \, \mathrm{d}(\mu \otimes \nu) = \int_{\Omega_2} (\int_{\Omega_1} f(x,y) \mu(\mathrm{d}x)) \nu(\mathrm{d}y)$ é análoga, o que conclui esta demonstração. □

Estendemos agora a teoria para o produto de n fatores.

Dados n espaços de medida σ-finita $(\Omega_j, \mathcal{F}_j, \mu_j)$ para $j = 1, \ldots, n$, podemos definir, recursivamente, a σ-álgebra produto como

$$\mathcal{F}_1 \otimes \cdots \otimes \mathcal{F}_n = (\mathcal{F}_1 \otimes \cdots \otimes \mathcal{F}_{n-1}) \otimes \mathcal{F}_n$$

e a medida produto como

$$\mu_1 \otimes \cdots \otimes \mu_n = (\mu_1 \otimes \cdots \otimes \mu_{n-1}) \otimes \mu_n.$$

Mais abaixo vamos mostrar que

$$\mathcal{F}_1 \otimes \cdots \otimes \mathcal{F}_n = \sigma(\mathcal{F}_1 \times \cdots \times \mathcal{F}_n). \tag{D.13}$$

Como $\mathcal{F}_1 \times \cdots \times \mathcal{F}_n$ é um π-sistema, podemos concluir que $\mu_1 \otimes \cdots \otimes \mu_n$ é a única medida ν em $(\Omega_1 \times \cdots \times \Omega_n, \mathcal{F}_1 \otimes \cdots \otimes \mathcal{F}_n)$ tal que

$$\nu(A_1 \times \cdots \times A_n) = \mu_1(A_1) \cdots \mu_n(A_n)$$

para todos $A_1 \in \mathcal{F}_1, \ldots, A_n \in \mathcal{F}_n$.

Proposição D.14. *A σ-álgebra de Borel em $\mathbb{R}^n$, $\mathcal{B}(\mathbb{R}^n)$, aquela gerada pela classe dos conjuntos abertos, também é dada por $\mathcal{B}(\mathbb{R}) \otimes \cdots \otimes \mathcal{B}(\mathbb{R})$. Essa σ-álgebra também é gerada pelos ortantes fechados, i.e., conjuntos da forma $(-\infty, a_1] \times \cdots \times (-\infty, a_n]$ com $a_1, \ldots, a_n \in \mathbb{R}$.*

Dada $\boldsymbol{f} : \Omega \to \mathbb{R}^n$, temos que $\boldsymbol{f}$ é $\mathcal{B}(\mathbb{R}^n)$-mensurável se, e somente se, cada coordenada $f_1, \ldots, f_n$ for uma função real mensurável; portanto a

fórmula (4.2) está bem definida para todo $B \in \mathcal{B}(\mathbb{R}^n)$. Com efeito, $\mathcal{B}(\mathbb{R}^n)$ é gerada por $\mathcal{B}(\mathbb{R}) \times \cdots \times \mathcal{B}(\mathbb{R})$ e, usando o Lema 3.42, verificamos que a mensurabilidade de $\boldsymbol{f}$ é equivalente à mensurabilidade de todas as suas coordenadas.

Como os ortantes fechados formam um π-sistema que gera $\mathcal{B}(\mathbb{R}^n)$ e que contém a sequência $(-\infty, k]^n \uparrow \mathbb{R}^n$, podemos usar o Teorema 3.37 (unicidade de medidas) para concluir que $F_{\boldsymbol{X}} = F_{\boldsymbol{Y}}$ implica $\mathbb{P}_{\boldsymbol{X}} = \mathbb{P}_{\boldsymbol{Y}}$, provando o Teorema 4.5. Da mesma forma, a condição $F_{\boldsymbol{X}}(\boldsymbol{x}) = F_{X_1}(x_1) \dots F_{X_n}(x_n)$ para todo $\boldsymbol{x} \in \mathbb{R}^n$ significa que $\mathbb{P}_{\boldsymbol{X}}(B) = (\mathbb{P}_{X_1} \otimes \cdots \otimes \mathbb{P}_{X_n})(B)$ para todo ortante fechado B, o que implica (pelo Teorema 3.37) que

$$\mathbb{P}_{\boldsymbol{X}} = \mathbb{P}_{X_1} \otimes \cdots \otimes \mathbb{P}_{X_n},$$

o que, por sua vez, implica que $X_1, \dots, X_n$ são independentes. Isso prova a Proposição 4.10.

Como tanto a σ-álgebra produto quanto a medida produto foram definidas recursivamente, o uso recursivo do Teorema de Tonelli feito na Observação 5.89 está bem justificado, assim como a identidade $\mathcal{B}(\mathbb{R}^n) = \mathcal{B}(\mathbb{R}) \otimes \cdots \otimes \mathcal{B}(\mathbb{R})$, diretamente pela a proposição acima.

Para provar (D.13) e a Proposição D.14, usaremos o seguinte fato, cuja prova fica como exercício: dados dois conjuntos não-vazios Ω' e Ω'' e uma classe $\mathcal{E}$ de subconjuntos de Ω', vale $\sigma(\mathcal{E} \times \{\Omega''\}) = \sigma(\mathcal{E}) \times \{\Omega''\}$.

Demonstração de (D.13). Vamos supor que a identidade vale com $n-1$. Como $\mathcal{F}_1 \times \cdots \times \mathcal{F}_n \subseteq (\mathcal{F}_1 \otimes \cdots \otimes \mathcal{F}_{n-1}) \times \mathcal{F}_n$, segue que $\sigma(\mathcal{F}_1 \times \cdots \times \mathcal{F}_n) \subseteq \mathcal{F}_1 \otimes \cdots \otimes \mathcal{F}_n$. Resta mostrar a inclusão oposta. Seja $B \in (\mathcal{F}_1 \otimes \cdots \otimes \mathcal{F}_{n-1}) \times \mathcal{F}_n$ e escreva $B = A' \times A''$, com $A' \in \mathcal{F}_1 \otimes \cdots \otimes \mathcal{F}_{n-1}$ e $A'' \in \mathcal{F}_n$. Observe que

$$B = (A' \times \Omega_n) \cap (\Omega_1 \times \cdots \times \Omega_{n-1} \times A'').$$

Como $A' \times \Omega_n \in \sigma(\mathcal{F}_1 \times \cdots \times \mathcal{F}_{n-1}) \times \{\Omega_n\} = \sigma(\mathcal{F}_1 \times \cdots \times \mathcal{F}_{n-1} \times \{\Omega_n\})$ e $\Omega_1 \times \cdots \times \Omega_{n-1} \times A'' \in \mathcal{F}_1 \times \cdots \times \mathcal{F}_n$, segue que $B \in \sigma(\mathcal{F}_1 \times \cdots \times \mathcal{F}_n)$.

Logo, $(\mathcal{F}_1 \otimes \cdots \otimes \mathcal{F}_{n-1}) \times \mathcal{F}_n \subseteq \sigma(\mathcal{F}_1 \times \cdots \times \mathcal{F}_n)$ e, portanto, $\mathcal{F}_1 \otimes \cdots \otimes \mathcal{F}_n \subseteq \sigma(\mathcal{F}_1 \times \cdots \times \mathcal{F}_n)$, o que conclui a prova. □

Demonstração da Proposição D.14. Definimos $\mathcal{C}_n$ como a classe dos ortantes fechados em $\mathbb{R}^n$, $\mathcal{D}_n$ como a classe dos conjuntos abertos de $\mathbb{R}^n$, e $\mathcal{E}_n = \mathcal{B}(\mathbb{R}^{n-1}) \times \mathcal{B}(\mathbb{R})$. Vamos provar que $\sigma(\mathcal{C}_n) \subseteq \sigma(\mathcal{D}_n) \subseteq \sigma(\mathcal{E}_n) \subseteq \sigma(\mathcal{C}_n)$.

Seja $C \in \mathcal{C}_n$. Como C é fechado, $C^c \in \mathcal{D}_n$, logo $C \in \sigma(\mathcal{D}_n)$. Ou seja, $\mathcal{C}_n \subseteq \sigma(\mathcal{D}_n)$, portanto $\sigma(\mathcal{C}_n) \subseteq \sigma(\mathcal{D}_n)$. Seja $D \in \mathcal{D}_n$. Pelo Teorema A.11, podemos escrever $D = \cup_{k\in\mathbb{N}} E_k$, onde $E_k \in \mathcal{E}_n$ para cada k, logo $D \in \sigma(\mathcal{E}_n)$. Ou seja, $\mathcal{D}_n \subseteq \sigma(\mathcal{E}_n)$, portanto $\sigma(\mathcal{D}_n) \subseteq \sigma(\mathcal{E}_n)$.

Resta mostrar que $\sigma(\mathcal{E}_n) \subseteq \sigma(\mathcal{C}_n)$, o que faremos por indução. O caso $n = 1$ é dado pela Proposição 1.43. Supondo que a inclusão vale para um $n \in \mathbb{N}$ fixo, vamos mostrar que $\mathcal{B}(\mathbb{R}^n) \times \mathcal{B}(\mathbb{R}) \subseteq \sigma(\mathcal{C}_{n+1})$. Afirmamos que $\mathcal{B}(\mathbb{R}^n) \times \{\mathbb{R}\} \subseteq \sigma(\mathcal{C}_{n+1})$. Com efeito, dado $A \in \mathcal{C}_n$, podemos tomar $B_k = A \times (-\infty, k] \in \mathcal{C}_{n+1}$, que satisfaz $\cup_k B_k = A \times \mathbb{R}$, donde $A \times \mathbb{R} \in \sigma(\mathcal{C}_{n+1})$, logo $\mathcal{C}_n \times \{\mathbb{R}\} \subseteq \sigma(\mathcal{C}_{n+1})$, portanto $\sigma(\mathcal{C}_n) \times \{\mathbb{R}\} = \sigma(\mathcal{C}_n \times \{\mathbb{R}\}) \subseteq \sigma(\mathcal{C}_{n+1})$. Afirmamos também que $\{\mathbb{R}^n\} \times \mathcal{B}(\mathbb{R}) \subseteq \sigma(\mathcal{C}_{n+1})$, e a prova é análoga. Finalmente, dados $A \in \mathcal{B}(\mathbb{R}^n)$ e $B \in \mathcal{B}(\mathbb{R})$, podemos escrever $A \times B = (A \times \mathbb{R}) \cap (\mathbb{R}^n \times B) \in \sigma(\mathcal{C}_{n+1})$. Ou seja, $\mathcal{E}_{n+1} \subseteq \sigma(\mathcal{C}_{n+1})$, portanto $\sigma(\mathcal{E}_{n+1}) \subseteq \sigma(\mathcal{C}_{n+1})$, concluindo a prova. □

D.5 Teorema de Radon-Nikodým

Nesta seção provaremos o Teorema de Radon-Nikodým.

Começamos motivando a ideia da prova. Suponha que μ e ν são medidas finitas e $\nu \ll \mu$. Estamos à busca de uma função mensurável f tal que

$$\nu(A) = \int_A f \, \mathrm{d}\mu \quad \forall A \in \mathcal{F}. \tag{D.15}$$

Descrever f é o mesmo que descrever os subconjuntos de Ω onde f assume

determinados valores. Seja $D_{a,b} = \{\omega : a \leqslant f(\omega) \leqslant b\}$. Por (D.15),

$$a \cdot \mu(A \cap D_{a,b}) \leqslant \nu(A \cap D_{a,b}) \leqslant b \cdot \mu(A \cap D_{a,b}) \quad \forall A \in \mathcal{F}.$$

Observe que a relação acima não faz referência explícita a f. Para percorrer o caminho oposto, gostaríamos de ir fatiando Ω em conjuntos da forma $D_{a,b}$ com essa propriedade, e usar esses conjuntos para ir construindo f. O primeiro passo seria decompor $\Omega = N_a \cup N_a^c$ de forma que

$$\nu(A \cap N_a) \leqslant a \cdot \mu(A \cap N_a) \text{ e } \nu(A \cap N_a^c) \geqslant a \cdot \mu(A \cap N_a^c) \quad \forall A \in \mathcal{F}. \tag{D.16}$$

Para isso, vamos estudar a função $\lambda : \mathcal{F} \to \mathbb{R}$ dada por

$$\lambda(A) = \nu(A) - a\mu(A).$$

Uma *carga* (ou medida com sinal) em $(\Omega, \mathcal{F})$ é uma função $\lambda : \mathcal{F} \to [-\infty, +\infty]$ tal que $\lambda(\emptyset) = 0$ e, para toda sequência de conjuntos $(A_n)_n$ em $\mathcal{F}$ disjuntos, pelo menos uma das somas $\sum_n \lambda(A_n)^+$ e $\sum_n \lambda(A_n)^-$ é finita, e $\lambda(\cup_n A_n) = \sum_n \lambda(A_n)$. Dizemos que $N \in \mathcal{F}$ é um *conjunto negativo para* λ se $\lambda(A \cap N) \leqslant 0$ para todo $A \in \mathcal{F}$. O conjunto vazio é um exemplo trivial de conjunto negativo.

Lema D.17. *Seja λ uma carga em $(\Omega, \mathcal{F})$. Para todo $D \in \mathcal{F}$, existe $A \in \mathcal{F}$ tal que A é negativo, $A \subseteq D$ e $\lambda(A) \leqslant \lambda(D)$.*

Demonstração. Podemos supor que $\lambda(D) < 0$, pois caso contrário podemos tomar $A = \emptyset$. Vamos definir uma sequência decrescente $(A_n)_n$ e verificar que $A = \cap_n A_n$ é o conjunto procurado. Definimos inicialmente $A_0 = D$. Se A_n já foi definido, tome $t_n = \sup\{\lambda(B) : B \in \mathcal{F}, B \subseteq A_n\}$ e note que $t_n \geqslant \lambda(\emptyset) = 0$. Seja $\varepsilon_n = \min\{\frac{t_n}{2}, 1\}$ (precisamos do mínimo para assegurar-nos de que $\varepsilon_n < t_n$ no caso $t_n = +\infty$) e tome $B_n \in \mathcal{F}$ tal que $B_n \subseteq A_n$ e $\lambda(B_n) \geqslant \varepsilon_n$. Defina $A_{n+1} = A_n \setminus B_n$ e $A = \cap_n A_n \subseteq D$.

Observando que $\cup_n B_n \subseteq D$ e que esta última união é disjunta,

$$\lambda(D) = \lambda(A) + \sum_n \lambda(B_n) \geqslant \lambda(A) + \sum_n \varepsilon_n \geqslant \lambda(A).$$

Falta mostrar que A é negativo. Observe que $\sum_n \lambda(B_n) < +\infty$, pois caso contrário teríamos $\lambda(D) = +\infty$. Portanto, $\lambda(B_n) \to 0$ quando $n \to \infty$ e, por conseguinte, $t_n \to 0$. Agora seja $C \in \mathcal{F}$ tal que $C \subseteq A = \cap_n A_n$. Como $C \subseteq A_n$, pela definição de t_n, temos $\lambda(C) \leqslant t_n$ para todo $n \in \mathbb{N}$. Portanto, $\lambda(C) \leqslant \inf_n t_n = 0$, o que mostra que A é um conjunto negativo. $\square$

Teorema D.18 (Teorema de Decomposição de Hahn)**.** *Seja λ uma carga em $(\Omega, \mathcal{F})$. Então existe um conjunto $N \in \mathcal{F}$ tal que $\lambda(B \cap N) \leqslant 0$ e $\lambda(B \cap N^c) \geqslant 0$ para todo $B \in \mathcal{F}$.*

Demonstração. Como preliminar, observamos que λ não pode assumir ambos os valores $-\infty$ e $+\infty$. Com efeito, se $\lambda(A) = -\infty$ para algum $A \in \mathcal{F}$, segue que $\lambda(\Omega) = \lambda(A) + \lambda(A^c) = -\infty$ e, analogamente, se $\lambda(B) = +\infty$ para algum $B \in \mathcal{F}$, então $\lambda(\Omega) = +\infty$. Portanto, podemos supor, sem perda de generalidade, que $\lambda(A) \neq -\infty$ para todo $A \in \mathcal{F}$ (caso contrário bastaria considerar $-\lambda$ ao invés de λ e tomar N^c no lugar de N).

Seja $\alpha = \inf\{\lambda(A) : A \in \mathcal{F}$ é um conjunto negativo para $\lambda\}$ e note que $\alpha \leqslant \lambda(\emptyset) = 0$. Tome uma sequência $(A_n)_n$ de conjuntos negativos tal que $\lambda(A_n) \to \alpha$, e defina $N = \cup_n A_n$. Como a união enumerável de conjuntos negativos é negativa (exercício!), segue que N é um conjunto negativo.

Afirmamos que $\alpha = \lambda(N)$ e, em particular, $\alpha \neq -\infty$. Com efeito,

$$\alpha \leqslant \lambda(N) = \lambda(A_n) + \lambda(N \setminus A_n) \leqslant \lambda(A_n) \to \alpha,$$

onde as desigualdades seguem da definição de α e da negatividade de N.

Finalmente, seja $B \in \mathcal{F}$. Como N é negativo, segue que $\lambda(B \cap N) \leqslant 0$, e resta mostrar que $\lambda(B \cap N^c) \geqslant 0$. Pelo Lema D.17, existe $A \in \mathcal{F}$ tal que $A \subseteq B \cap N^c$, A é negativo e $\lambda(A) \leqslant \lambda(B \cap N^c)$. Para mostrar que $\lambda(A) \geqslant 0$, observamos que, como $A \cup N$ é um conjunto negativo,

$$\alpha \leqslant \lambda(A \cup N) = \lambda(A) + \lambda(N) = \lambda(A) + \alpha,$$

e podemos subtrair α pois $-\infty < \alpha \leqslant 0$, completando a prova. $\square$

Prova do Teorema de Radon-Nikodým. Vamos supor inicialmente que ν e μ sejam medidas finitas, e posteriormente extrapolar para o caso geral.

Para cada $a > 0$, pelo Teorema de Decomposição de Hahn existe $N_a \in \mathcal{F}$ satisfazendo (D.16). Definimos $D_{0,1} = N_1$, $D_{1,2} = N_2 \cap N_1^c$, $D_{2,3} = N_3 \cap N_1^c \cap N_2^c$, e assim por diante. Observe que esses conjuntos são disjuntos e $\cup_k D_{k-1,k} = \cup_k N_k$. Defina $S = \Omega \setminus \cup_k N_k$.

Observamos que $S, D_{0,1}, D_{1,2}, D_{2,3}, \ldots$ formam uma partição de Ω. Além disso, como $S \subseteq N_k^c$ temos por (D.16) que $\nu(S) \geqslant k\mu(S)$ e, como isso vale para todo k e $\nu(S) < \infty$, temos $\mu(S) = 0$. Como $\nu \ll \mu$, segue que $\nu(S) = 0$. Como observação à parte, sem a hipótese de $\nu \ll \mu$, poderíamos decompor $\nu = \nu_{|S} + \nu_{|S^c}$ e encontrar a derivada de $\nu_{|S^c}$ com respeito a μ.

Nossa primeira aproximação para f será dada por

$$f_0(\omega) = \sum_j j \cdot \mathbb{1}_{D_{j,j+1}}(\omega).$$

De (D.16), obtemos, para todo $A \in \mathcal{F}$ e todo $j \in \mathbb{N}$,

$$\nu(A \cap D_{j,j+1}) \geqslant j\mu(A \cap D_{j,j+1}) = \int_{A \cap D_{j,j+1}} f_0 \, \mathrm{d}\mu$$

e

$$\nu(A \cap D_{j,j+1}) \leqslant (j+1)\mu(A \cap D_{j,j+1}) = \int_{A \cap D_{j,j+1}} (f_0 + 1) \, \mathrm{d}\mu.$$

Somando sobre j, temos

$$\int_A f_0 \, \mathrm{d}\mu \leqslant \nu(A) \leqslant \int_A f_0 \, \mathrm{d}\mu + \mu(A) \quad \forall A \in \mathcal{F}.$$

Vamos agora subdividir cada intervalo ao meio. Para cada $j \in \mathbb{N}_0$, definimos $D_{j,j+1/2} = D_{j,j+1} \cap N_{j+1/2}$ e $D_{j+1/2,j+1} = D_{j,j+1} \cap N_{j+1/2}^c$. Novamente, a

família $(D_{j/2,(j+1)/2})_{j\in\mathbb{N}_0}$ forma uma partição de $\Omega \setminus S$. Definimos

$$f_1(\omega) = \sum_j 2^{-1} j \cdot \mathbb{1}_{D_{2^{-1}j,2^{-1}(j+1)}}(\omega).$$

Com essa definição, temos $f_1 \geqslant f_0 \geqslant 0$ e, de forma análoga à relação anterior,

$$\int_A f_1 \,\mathrm{d}\mu \leqslant \nu(A) \leqslant \int_A f_1 \,\mathrm{d}\mu + 2^{-1}\mu(A) \quad \forall A \in \mathcal{F}.$$

Partindo novamente os intervalos ao meio, construímos $f_2 \geqslant f_1$ tal que

$$\int_A f_2 \,\mathrm{d}\mu \leqslant \nu(A) \leqslant \int_A f_2 \,\mathrm{d}\mu + 2^{-2}\mu(A) \quad \forall A \in \mathcal{F},$$

e assim por diante obtemos $f_k \geqslant f_{k-1}$ tal que

$$\int_A f_k \,\mathrm{d}\mu \leqslant \nu(A) \leqslant \int_A f_k \,\mathrm{d}\mu + 2^{-k}\mu(A) \quad \forall A \in \mathcal{F}.$$

Tomando $f = \lim_k f_k$, pelo Teorema da Convergência Monótona temos (D.15) pois $\mu(A) < \infty$, o que prova o teorema no caso de μ e ν finitas.

Mostraremos agora o caso geral, ou seja, quando μ e ν são σ-finitas. Neste caso, existe uma partição $(A_n)_{n\in\mathbb{N}}$ de Ω tal que $\mu_n = \mu_{|A_n}$ e $\nu_n = \nu_{|A_n}$ são medidas finitas e $\nu_n \ll \mu_n$. Tome $f_n = \frac{\mathrm{d}\nu_n}{\mathrm{d}\mu_n}$, e defina $f = \sum_n f_n \mathbb{1}_{A_n}$. Verificar que $f = \frac{\mathrm{d}\nu}{\mathrm{d}\mu}$ é imediato:

$$\nu(A) = \sum_n \nu_n(A) = \sum_n \int_A f_n \,\mathrm{d}\mu_n = \sum_n \int_A f_n \mathbb{1}_{A_n} \,\mathrm{d}\mu = \int_A \sum_n f_n \mathbb{1}_{A_n} \,\mathrm{d}\mu.$$

Isso conclui a prova do teorema. □

D.6 Distribuição condicional regular

Nesta seção vamos provar os Teoremas 11.58 e 11.59.

Para isso, vamos construir uma função $F_{X|Y}$ mensurável em ambas as variáveis e tal que

$$F_{X,Y}(x,y) = \int_{-\infty}^{y} F_{X|Y}(x|y)\,\mathbb{P}_Y(\mathrm{d}y)$$

para todos $x, y \in \mathbb{R}$. Começamos estudando uma função $G : \mathbb{R}^2 \to \mathbb{R}$ que terá determinadas propriedades em "quase todos" os pontos x e y, depois usar G para construir $F_{X|Y}$, e finalmente usar $F_{X|Y}$ para construir $\mathbb{P}_{X|Y}$.

Sejam X e Y variáveis aleatórias definidas em um mesmo espaço de probabilidade.

Lema D.19. *Para todo $x \in \mathbb{R}$, existe uma função mensurável $G(x\,|\,\cdot\,) : \mathbb{R} \to [0,1]$, tal que*

$$\mathbb{P}(X \leqslant x, Y \in C) = \int_C G(x|y)\,\mathbb{P}_Y(\mathrm{d}y)$$

para todo $C \in \mathcal{B}$.

Demonstração. Seja $x \in \mathbb{R}$ fixo. Defina

$$\nu(C) = \mathbb{P}(X \leqslant x, Y \in C), \qquad \mu(C) = \mathbb{P}(Y \in C)$$

para todo $C \in \mathcal{B}$. Como $\nu \ll \mu$, pelo Teorema de Radon-Nikodým, existe uma função mensurável $G(x|\,\cdot\,) : \mathbb{R} \to \mathbb{R}$, tal que

$$\mathbb{P}(X \leqslant x, Y \in C) = \int_C G(x,y)\,\mathbb{P}_Y(\mathrm{d}y) \quad \text{para todo } C \in \mathcal{B}.$$

Como essa integral nunca é negativa, $G(x,y) \geqslant 0$ para $\mathbb{P}_Y$-quase todo $y \in \mathbb{R}$. Como $\mathbb{P}(X > x, Y \in C) = \int_C [1 - G(x,y)]\mathbb{P}_Y(\mathrm{d}y)$ nunca é negativo, segue que $G(x,y) \leqslant 1$ para $\mathbb{P}_Y$-quase todo $y \in \mathbb{R}$. Modificando $G(x,y)$ em um conjunto de medida $\mathbb{P}_Y$ nula, podemos supor que $G(x,y) \in [0,1]$ para todo $y \in \mathbb{R}$. $\square$

Para construir $F_{X|Y}$ a partir de G, usaremos o fato de que $\mathbb{Q}$ é enumerável e denso em $\mathbb{R}$, e que uniões enumeráveis de conjuntos de medida nula têm medida nula.

Lema D.20. *Existe uma função $F_{X|Y}(\cdot|\cdot) : \mathbb{R}^2 \to \mathbb{R}$ com as seguintes propriedades:*

(1) $F_{X|Y}(\,\cdot\,|y)$ *é uma função de distribuição para todo $y \in \mathbb{R}$ fixo,*
(2) $F_{X|Y}(q|\,\cdot\,)$ *é uma função mensurável para todo $q \in \mathbb{Q}$ fixo,*
(3) $F_{X,Y}(q,y) = \int_{-\infty}^{y} F_{X|Y}(q|s)\,\mathbb{P}_Y(\mathrm{d}s)$ *para todos $q \in \mathbb{Q}$ e $y \in \mathbb{R}$.*

Demonstração. Para cada $q \in \mathbb{Q}$, seja $G(q|\cdot)$ a função dada pelo Lema D.19. Agora, para cada par de racionais $r < q$, seja

$$A_{1,r,q} = \{y \in \mathbb{R} : G(r\,|\,y) \leqslant G(q\,|\,y)\} \in \mathcal{B}.$$

Definindo $A_1 = \cap_{r,q} A_{1,r,q}$, temos que $\mathbb{P}_Y(A_1) = 1$, pois $\mathbb{P}_Y(A_{1,r,q}) = 1$ para todos r < q. Com efeito, $\int_C [G(q|y) - G(r|y)]\mathbb{P}_Y(\mathrm{d}y) = \mathbb{P}(r < X \leqslant q, Y \in C) \geqslant 0$ para todo $C \in \mathcal{B}$, donde concluímos que o integrando é não-negativo para $\mathbb{P}_Y$-quase todo $y \in \mathbb{R}$.

Em seguida, para cada $q \in \mathbb{Q}$, definimos

$$A_{2,q} = \{y \in A_1 : G(q + \tfrac{1}{n}\,|\,y) \to G(q\,|\,y)\} \in \mathcal{B}.$$

Observe que

$$\int_{\mathbb{R}} G(q + \tfrac{1}{n}|y)\,\mathbb{P}_Y(\mathrm{d}y) = \mathbb{P}(X \leqslant q + \tfrac{1}{n}) \to \mathbb{P}(X \leqslant q) = \int_{\mathbb{R}} G(q|y)\,\mathbb{P}_Y(\mathrm{d}y),$$

logo, pelo Teorema da Convergência Dominada,

$$\int_{\mathbb{R}} [\lim_n G(q + \tfrac{1}{n}|y)]\,\mathbb{P}_Y(\mathrm{d}y) = \int_{\mathbb{R}} G(q|y)\,\mathbb{P}_Y(\mathrm{d}y),$$

e, como o integrando do lado direito é cotado superiormente pelo integrando do lado esquerdo, eles têm que ser iguais para $\mathbb{P}_Y$-quase todo y, ou seja, $\mathbb{P}(A_{2,q}) = 1$. Novamente, tomamos $A_2 = \cap_{q\in\mathbb{Q}} A_{2,q}$ e definimos agora

$$A = \{y \in A_2 : \lim_{k\to+\infty} G(k\,|\,y) = 1, \lim_{k\to-\infty} G(k\,|\,y) = 0\} \in \mathcal{B}$$

e, pelo mesmo argumento utilizado para $A_{2,q}$, pode-se mostrar que $\mathbb{P}_Y(A) = 1$.

Finalmente definimos, para $y \in A$ e $x \in \mathbb{R}$

$$F_{X|Y}(x|y) = \inf\{G(q|y) : q \in \mathbb{Q}, q > x\}.$$

Para $y \in A^c$, definimos $F_{X|Y}(x|y) = F_X(x)$.

Observe que, pela definição de G e $F_{X|Y}$, a função $F_{X|Y}(\,\cdot\,|y)$ é uma função de distribuição para todo $y \in \mathbb{R}$, como afirmado. Agora seja $x \in \mathbb{Q}$ fixo. Observe que $F_{X|Y}(x|y) = G(x|y)\mathbb{1}_A(y) + F_X(x)\mathbb{1}_{A^c}(y)$ e, pelo Lema D.19, isso define uma função mensurável de y. Ademais, como $\mathbb{P}_Y(A) = 1$, o item (3) também segue diretamente do Lema D.19. □

Na demonstração acima, a função $F_{X|Y}$ foi obtida através do Lema D.19, que não diz como calculá-la. Ressaltamos que a função definida por (11.57) também satisfaz a essas três propriedades, porém a demonstração desse fato exige ferramentas de Teoria da Medida que vão muito além do escopo deste livro (ver Teorema 6.66 em Giaquinta e Modica (2009)).

Demonstração do Teorema 11.58. Seja $F_{X|Y}$ como dada pelo Lema D.20.

Preliminarmente, afirmamos que, para todo $q \in \mathbb{Q}$, vale

$$\mathbb{P}(X \leqslant q, Y \in C) = \int_C F_{X|Y}(q|s)\,\mathbb{P}_Y(\mathrm{d}s) \quad \text{para todo } C \in \mathcal{B}. \tag{D.21}$$

Com efeito, ambos os lados determinam medidas de probabilidade em C e, pelo item (3) do Lema D.20, essas medidas coincidem na classe $\{(-\infty, y]\}_{y\in\mathbb{R}}$, que forma um π-sistema e contém $\{(-\infty, n]\}_n$, cuja união é $\mathbb{R}$, logo, pelo Teorema 3.37 (unicidade de medidas), elas coincidem para todo $C \in \mathcal{B}$.

Para cada $y \in \mathbb{R}$, definimos $\mathbb{P}_{X|Y}(\cdot\,|\,y)$ como a medida de probabilidade em $\mathbb{R}$ correspondente à função de distribuição $F_{X|Y}(\cdot\,|\,y)$. Seja $\mathcal{D}$ a classe de conjuntos $B \in \mathcal{B}$ tais que $\mathbb{P}_{X|Y}(B\,|\,y)$ é uma função mensurável de y e tais

que

$$\mathbb{P}(X \in B, Y \in C) = \int_C \mathbb{P}_{X|Y}(B \,|\, y)\, \mathbb{P}_Y(\mathrm{d}y) \quad \text{para todo } C \in \mathcal{B}. \tag{D.22}$$

Observe que $\mathcal{D}$ é um λ-sistema (exercício!). Seja $\mathcal{C}$ a classe dos conjuntos da forma $B = (-\infty, q]$ com $q \in \mathbb{Q}$. Para $B \in \mathcal{C}$, $\mathbb{P}_{X|Y}(B \,|\, y)$ é uma função mensurável de y pelo item (2) do Lema D.20; ademais, (D.22) se reduz a (D.21). Ou seja, $\mathcal{C} \subseteq \mathcal{D}$. Como $\mathcal{C}$ é um π-sistema e $\sigma(\mathcal{C}) = \mathcal{B}$, segue do Teorema π-λ que $\mathcal{D} = \mathcal{B}$, o que prova o Teorema 11.58. □

Demonstração do Teorema 11.59. A ideia da prova é estudar a classe das funções g para as quais vale o enunciado do teorema. Consideramos inicialmente $g(x, y) = \mathbb{1}_A(x, y)$ para $A \in \mathcal{B}(\mathbb{R}^2)$. Seja $\mathcal{D}$ a classe dos conjuntos $A \in \mathcal{B}(\mathbb{R}^2)$ para os quais a integral interna em (11.60), com $\mathbb{1}_A$ no lugar de g, fornece uma função mensurável de y e vale a igualdade (11.60). Observe que $\mathcal{D}$ é um λ-sistema (exercício!). Seja $\mathcal{C} = \{B \times C : B, C \in \mathcal{B}(\mathbb{R})\}$. Pela Definição 11.56, temos que, para todos $B, C \in \mathcal{B}(\mathbb{R})$

$$\int_{\mathbb{R}} \mathbb{1}_{B\times C}(x, y)\, \mathbb{P}_{X|Y}(\mathrm{d}x|y) = \mathbb{1}_C(y)\mathbb{P}_{X|Y}(B|y),$$

que é uma função mensurável de y e cuja integral com respeito a $\mathbb{P}_Y$ é igual a $\mathbb{E}[\mathbb{1}_{B\times C}(X, Y)]$. Ou seja, $\mathcal{C} \subseteq \mathcal{D}$. Como $\mathcal{C}$ é um π-sistema e $\sigma(\mathcal{C}) = \mathcal{B}(\mathbb{R}^2)$, segue do Teorema π-λ que $\mathcal{D} = \mathcal{B}(\mathbb{R}^2)$. Isso prova o teorema para funções mensuráveis g que apenas assumem valores 0 e 1. Por linearidade, vale o teorema para funções simples não-negativas. Finalmente, seja $g : \mathbb{R}^2 \to [0, +\infty]$ uma função mensurável. Tome $0 \leqslant g_n \uparrow g$, onde as funções g_n são simples. Pelo Teorema da Convergência Monótona, $\int_{\mathbb{R}} g(x, y)\, \mathbb{P}_{X|Y}(\mathrm{d}x|y) = \lim_n \int_{\mathbb{R}} g_n(x, y)\, \mathbb{P}_{X|Y}(\mathrm{d}x|y)$ e, como limite de funções mensuráveis é mensurável, segue que a integral interna em (11.60) fornece uma função mensurável de y. Como

$$\mathbb{E}[g_n(X, Y)] = \int_{\mathbb{R}} \Big(\int_{\mathbb{R}} g_n(x, y)\, \mathbb{P}_{X|Y}(\mathrm{d}x|y) \Big) \mathbb{P}_Y(\mathrm{d}y)$$

para todo n, pelo Teorema da Convergência Monótona, vale a igualdade em (11.60). Isso conclui prova do Teorema 11.59. $\square$

D.7 Desigualdades de Hölder e de Minkowski

Nesta seção vamos enunciar e provar as Desigualdades de Hölder e Minkowski. Começamos pela Desigualdade de Young, que será usada na prova. Salientamos que nesta seção estamos no contexto de espaços de medida que não são necessariamente espaços de probabilidade.

Dado um espaço de medida $(\Omega, \mathcal{F}, \mu)$, uma função mensurável $f : \Omega \to [-\infty, +\infty]$ e $p \geqslant 1$, definimos

$$\|f\|_p = \Big(\int_\Omega |f|^p \, \mathrm{d}\mu \Big)^{1/p}$$

caso a integral seja finita, e $\|f\|_p = +\infty$ caso contrário. Definimos $\mathcal{L}^p$ como o conjunto das funções mensuráveis $f : \Omega \to [-\infty, +\infty]$ tais que $\|f\|_p < \infty$. No caso $p = 1$, vale a desigualdade triangular pois

$$\|f + g\|_1 = \int_\Omega |f + g| \, \mathrm{d}\mu \leqslant \int_\Omega (|f| + |g|) \, \mathrm{d}\mu = \|f\|_1 + \|g\|_1.$$

Daqui em diante vamos supor que $p > 1$.

Seja $p > 1$ fixo. Tome $q > 1$ tal que

$$\frac{1}{p} + \frac{1}{q} = 1.$$

Dessa identidade, seguem: $p - 1 = \frac{1}{q-1}$ e $(p-1)q = p$.

Teorema D.23 (Desigualdade de Young)**.** *Para $a, b \geqslant 0$,*

$$ab \leqslant \frac{a^p}{p} + \frac{b^q}{q}.$$

Demonstração. Considere a curva $s = r^{p-1}$, ou seja, $r = s^{q-1}$, no quadrante $\{(r,s) \in [0,\infty)^2\}$. A Desigualdade de Young segue de

$$\frac{a^p}{p} + \frac{b^q}{q} = \int_0^a r^{p-1}\mathrm{d}r + \int_0^b s^{q-1}\mathrm{d}s \geqslant ab,$$

que vale porque as integrais acima correspondem a áreas de regiões cuja união contém o retângulo $[0,a] \times [0,b]$. □

Teorema D.24 (Desigualdade de Hölder)**.** *Dados um espaço de medida* $(\Omega, \mathcal{F}, \mu)$ *e funções mensuráveis* f *e* $g : \Omega \to [-\infty, +\infty]$*, vale*

$$\int_\Omega |fg| \,\mathrm{d}\mu \leqslant \|f\|_p \cdot \|g\|_q.$$

Demonstração. Podemos supor que $\|f\|_p > 0$ e $\|g\|_q > 0$, caso contrário $f \cdot g = 0$ q.t.p. e a desigualdade vale trivialmente. Também podemos supor que $\|f\|_p < \infty$ e $\|g\|_q < \infty$, caso contrário $\|f\|_p \cdot \|g\|_q = +\infty$, e a desigualdade vale trivialmente. Dividindo f por $\|f\|_p$ e g por $\|g\|_q$, podemos supor que $\|f\|_p = 1$ e $\|g\|_q = 1$. Aplicando a desigualdade de Young,

$$\int_\Omega |fg| \,\mathrm{d}\mu \leqslant \int_\Omega \Big(\frac{|f|^p}{p} + \frac{|g|^q}{q}\Big) \,\mathrm{d}\mu = \frac{1}{p} + \frac{1}{q} = 1 = \|f\|_p \|g\|_q,$$

o que prova a Desigualdade de Hölder. □

Teorema D.25 (Desigualdade de Minkowski)**.** *Dados um espaço de medida* $(\Omega, \mathcal{F}, \mu)$ *e funções* $f, g \in \mathcal{L}^p$*,* $f + g$ *está definida q.t.p. e*

$$\|f + g\|_p \leqslant \|f\|_p + \|g\|_p.$$

Demonstração. Observe que f e g são finitas q.t.p., donde $f + g$ está definida q.t.p. Ademais, $|f + g|^p \leqslant |2f|^p + |2g|^p$, logo $\|f + g\|_p < \infty$. Podemos supor também que $\|f + g\|_p > 0$, caso contrário a desigualdade vale trivialmente.

Desenvolvendo as integrais e usando a Desigualdade de Hölder,

$$\begin{aligned}
\int_\Omega |f+g|^p \,\mathrm{d}\mu &= \int_\Omega |f+g| \cdot |f+g|^{p-1} \,\mathrm{d}\mu \\
&\leqslant \int_\Omega |f| \cdot |f+g|^{p-1} \,\mathrm{d}\mu + \int_\Omega |g| \cdot |f+g|^{p-1} \,\mathrm{d}\mu \\
&\leqslant (\|f\|_p + \|g\|_p)\Big(\int_\Omega (|f+g|^{p-1})^q \,\mathrm{d}\mu\Big)^{\frac{1}{q}} \\
&= (\|f\|_p + \|g\|_p)\Big(\int_\Omega |f+g|^p \,\mathrm{d}\mu\Big)^{1-\frac{1}{p}} \\
&= \frac{\|f\|_p + \|g\|_p}{\|f+g\|_p} \int_\Omega |f+g|^p \,\mathrm{d}\mu
\end{aligned}$$

e, simplificando, obtemos a Desigualdade de Minkowski. $\square$

Lista de Figuras

1.1 Tabuleiro de Galton. 18
1.2 Modelo do azulejo usado no Exemplo 1.3. 20
1.3 Agulha de Buffon. 27
1.4 Correspondência entre árvores e excursões. 39
1.5 Princípio da reflexão. 40

3.1 Gráfico de uma função de distribuição discreta. 75
3.2 Gráfico de uma função de distribuição contínua. 76
3.3 Ilustração de como é obtida a função quantil. 79

4.1 Exemplo de função de distribuição conjunta. 111

5.1 A esperança de X como o centro de massa de p_X. 136
5.2 Esperança a partir da função de distribuição. 142
5.3 Gráfico de ψ_2 e a aproximação $\psi_n(z) \uparrow z$ para um z fixo. . . . 155
5.4 Integral de Riemann e integral de Lebesgue. 161

6.1 Reta que minimiza o erro quadrático médio. 197
6.2 Prova da desigualdade de Jensen. 200

7.1 Primeiros elementos do contra-exemplo da “onda dançante”. . 209
7.2 Diagrama de implicações entre os tipos de convergência. . . . 221

9.1 Aproximação de binomial a normal. 254

11.1 Ilustração da definição de esperança condicional. 304
11.2 Diagrama ilustrando a esperança condicional iterada. 308

12.1 Argumento de travessias ascendentes completas. 356

15.1 Obtenção da função taxa a partir da função $\log M_X$. 422

Notação

$\#A$ Cardinalidade de A 19

$A_n \downarrow A$ Interseção decrescente 43

$A_n \uparrow A$ União crescente 43

$x_n \downarrow x$ Limite decrescente 76

$x_n \uparrow x$ Limite crescente 76

A^c Complementar de A 23

$a \wedge b$ Mínimo entre a e b 347

$A \triangle B$ Diferença simétrica 377

Bernoulli(p) Distribuição de Bernoulli 82

$\binom{n}{k}$ Combinações de n, k a k 35

Binom(n, p) Distribuição binomial 83

$\frac{\mathrm{d}\mu}{\mathrm{d}\nu}$ Derivada de Radon-Nikodým 176

Exp(λ) Distribuição exponencial 88

$F(a-)$ Limite lateral: $F_X(a-) = \lim_{z \to a^-} F_X(z)$. 77

$\text{Geom}(p)$ Distribuição geométrica 82

$\mathbb{1}_A$ Função indicadora 72

i.i.d. Independentes e identicamente distribuídas 113

$\mathbb{N}$ Inteiros positivos, $\mathbb{N} = \{1, 2, 3, \dots\}$ 22

$\mathbb{N}_0$ Inteiros não-negativos, $\mathbb{N}_0 = \{0, 1, 2, \dots\}$ 22

$\mathcal{N}(\mu, \sigma^2)$ Distribuição normal 89

$o(\cdot)$ ordem pequena; qualquer função satisfazendo $\frac{o(w)}{|w|} \to 0$ 418

$\mathcal{P}(\Omega)$ Conjunto das partes 24

$\text{Poisson}(\lambda)$ Distribuição de Poisson 84

$\mathcal{U}[a, b]$ Distribuição uniforme 88

$\boldsymbol{X}$ Vetor aleatório 110

$\boldsymbol{x}$ Um vetor com n coordenadas 110

$x_n \downarrow a$ Sequência não-crescente que converge para a 430

$x_n \uparrow a$ Sequência não-decrescente que converge para a 430

$X \sim Y$ X e Y têm a mesma distribuição 74

X, Y Variáveis aleatórias 72

Bibliografia

Bartle, R. G. (1995). *The elements of integration and Lebesgue measure.* Wiley Classics Library. John Wiley & Sons, Inc., New York.

Billingsley, P. (1968). *Convergence of Probability Measures.* New York: Wiley.

Billingsley, P. (1995). *Probability and measure.* Third. John Wiley & Sons, Inc., New York.

Chung, K. L. (2001). *A Course in Probability Theory.* 3ª ed. Academic Press.

Chung, K. L. e F. AitSahlia (2003). *Elementary probability theory.* 4ª ed. Undergraduate Texts in Mathematics. New York: Springer-Verlag.

Cohn, D. L. (2013). *Measure theory.* Second. Birkhäuser Advanced Texts: Basel Textbooks. Birkhäuser/Springer, New York.

Durrett, R. (2019). *Probability: theory and examples.* 5ª ed. Vol. 49. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge.

Feller, W. (1971). *An Introduction to Probability Theory and Its Applications.* 2ª ed. Vol. 2. New York: Wiley.

Giaquinta, M. e G. Modica (2009). *Mathematical analysis.* Birkhäuser Boston, Ltd., Boston, MA.

Grimmett, G. e D. Welsh (2014). *Probability – an introduction.* 2ª ed. Oxford University Press, Oxford.

Grimmett, G. R. e D. R. Stirzaker (2020). *Probability and random processes.* 4ª ed. Oxford University Press, Oxford.

Gut, A. (2013). *Probability: a graduate course.* 2ª ed. Springer Texts in Statistics. Springer, New York.

Isnard, C. (2013). *Introdução à medida e integração.* 3ª ed. Projeto Euclides. Instituto de Matemática Pura e Aplicada (IMPA), Rio de Janeiro.

James, B. R. (2004). *Probabilidade: Um Curso em Nível Intermediário.* 3ª ed. Rio de Janeiro: IMPA.

Klenke, A. (2014). *Probability theory.* Second. Universitext. Springer, London.

Kubrusly, C. S. (2015). *Essentials of measure theory.* Springer, Cham.

Pitman, J. (1993). *Probability.* Springer New York.

Ross, S. (2009). *Probabilidade: Um Curso Moderno com Aplicações.* Bookman.

Shiryaev, A. N. (1996). *Probability.* 2ª ed. Vol. 95. Graduate Texts in Mathematics. New York: Springer-Verlag.

Tao, T. (2011). *An introduction to measure theory.* Vol. 126. Graduate Studies in Mathematics. American Mathematical Society.

Walsh, J. B. (2012). *Knowing the odds: An introduction to probability.* Vol. 139. Graduate Studies in Mathematics. American Mathematical Society.

Williams, D. (1991). *Probability with martingales.* Cambridge Mathematical Textbooks. Cambridge University Press, Cambridge.

Tabela Normal

No verso apresentamos os valores da distribuição normal $\Phi(z) = \mathbb{P}(X \leqslant z)$ se X tem distribuição normal padrão.

Para formar a tabela escrevemos $z = x + y$, onde x corresponde à parte inteira de z e à primeira casa decimal, e y corresponde à segunda casa decimal de z. Nas linhas temos x variando de 0,0 a 3,4 e nas colunas y varia de 0,00 a 0,09.

Portanto, temos 350 valores de z, de 0,00 a 3,49.

É importante observar que a tabela tem sérias limitações. Primeiro, se z é conhecido somente até a segunda casa decimal (tendo, pois, dois ou três algarismos significativos), é necessário comparar com os valores de $z \pm 0{,}01$ para saber quantas casas decimais de precisão terá $\Phi(z)$. Há que se notar também que $1 - \Phi(z)$, que dá a probabilidade de $X \geqslant z$, pode ter muito menos dígitos significativos que $\Phi(z)$, principalmente para $z \geqslant 2{,}3$ e ainda pior para $z \geqslant 3{,}1$.

Tabela Normal: $\Phi(x+y)$, onde x são os valores das linhas e y os das colunas.

	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,0	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990
3,1	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,2	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995
3,3	0,9995	0,9995	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997
3,4	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998

Índice Remissivo

agulha de Buffon, 28
álgebra, 376
axiomas de Kolmogorov, 41

Bayes, *veja* fórmula
Bernoulli, *veja* distribuição
Bernstein, *veja* polinômio
Beta, *veja* distribuição
Birkhoff, *veja* teorema ergódico
Borel, *veja* σ-álgebra de Borel, *veja* lema
borelianos, *veja* σ-álgebra de Borel
Buffon, *veja* agulha

Cantelli, *veja* lema, *veja* lei dos grandes números, *veja* teorema
Carathéodory, *veja* teorema
Catalan, *veja* número
Cauchy, *veja* distribuição, *veja* desigualdade
Chebyshev, *veja* Tchebyshev
Chernoff, *veja* desigualdade
coeficiente binomial, **37**
coeficiente de correlação, **194**
combinações, 35
côncava, *veja* função côncava
condicional, *veja* probabilidade condicional, *veja* esperança condicional
conjunto
  das partes, 24
  denso, 435
  enumerável, 24
  pequeno, 94
continuidade
  por baixo, 48
convergência
  de variáveis aleatórias, 207
  em $\mathcal{L}^p$, 210
  em distribuição, 289
  em probabilidade, 207
  quase certa, 209
  relações de implicação, 221
  unicidade do limite, 223
convexa, *veja* função convexa
correlação, 194
covariância, **192**
Cramér, *veja* princípio dos grandes desvios

De Moivre, *veja* teorema

densidade, *veja* função de densidade
desigualdade
de Cauchy-Schwarz, **201**
de Chernoff, 418
de Doob-Kolmogorov, 350
de Hölder, 474
de Jensen, **199**, 324
de Lyapunov, **201**
de Markov, **198**
de Minkowski, 474
de Paley-Zygmund, 203
de Tchebyshev, **198**
de Young, 473
maximal de Kolmogorov, **387**
desvio-padrão, **191**
Dirac, *veja* medida
Dirichlet, *veja* integral
distribuição
Beta, 91
binomial, **83**
condicional
dado um evento, **92**
regular, 328
de Bernoulli, **82**
de Cauchy, 91
de Gumbel, 91
de Laplace, 91
de Poisson, **84**
de uma variável aleatória, 73
exponencial, **88**
função de, *veja* função de distribuição
Gama, 91
geométrica, **82**
normal, **89**
soma de, 125
tabela, 484
singular, *veja* variável aleatória singular
uniforme, **88**
Doob, *veja* desigualdade, *veja* teorema de decomposição
Dynkin, *veja* teorema π-λ

enumerável, *veja* conjunto
equiprovável, 25
espaço
amostral, **21**
discreto, 26
de probabilidade, 17, **43**
mensurável, **47**
espaço de probabilidade
induzido, 73, 110
esperança
caso contínuo, 145
caso discreto, 145, 149
condicional
dada uma partição, **304**
dada uma σ-álgebra, **319**
dado um evento, 158
propriedades, 320
de variáveis independentes, 141, **147**
de variável aleatória simples, **136**
definição, **143**

iterada, 305, 318, 320
linearidade, 146
monotonicidade, 146
propriedades, 146, 147
unitariedade, 146
Euler, *veja* fórmula
evento
aleatório, **22**
certo, 23
impossível, 23
incompatível, 23
expansão de Taylor, **433**
com resto de Lagrange, **434**
da função característica, **288**

Fatou, *veja* lema
fórmula
de Bayes, **60**
de Euler, **284**
de Stirling, **437**
Fourier, *veja* transformada
Fubini, *veja* teorema
função
característica, **286**
côncava, 199
contínua, 431
contínua por partes, 431
convexa, 199
de densidade, **86**
condicional, 93, **314**
conjunta, **116**
marginal, **116**
de distribuição, **74**
condicional, 92
conjunta, **110**
marginal, **111**
propriedades, 76
de probabilidade, **80**
condicional, 92, **308**
marginal, **115**
degrau, 431
geradora de momentos, **279**
indicadora, 72
mensurável, **97**
Riemann-integráel, 172, **432**
taxa, 420

Galton, *veja* tabuleiro
Glivenko, *veja* teorema
grandes desvios, *veja* princípio
grandes números, *veja* lei
Gumbel, *veja* distribuição

Hahn, *veja* teorema
Hewitt, *veja* lei 0-1
Hölder, *veja* desigualdade

identidade de Wald, 354
independência
de eventos, **62**
coletiva, 64
dois a dois, 63
de σ-álgebras, 378
de variáveis aleatórias, **112**
caso contínuo, 117
caso discreto, 115

critério, 113
indicadora, *veja* função indicadora
infinitas vezes, 212
integral
de Dirichlet, 174, 296, 298
de Lebesgue, **152**
de Riemann, **432**
iterada, 178

Jacobi, *veja* método do jacobiano
jacobiano, *veja* método do jacobiano
Jensen, *veja* desigualdade

Kolmogorov, *veja* axiomas, *veja* lei dos grandes números, *veja* lei 0-1, *veja* desigualdade

Lagrange, *veja* multiplicador, *veja* expansão
λ-sistema, 448
Laplace, *veja* distribuição, *veja* teorema, *veja* transformada
Lebesgue, *veja* integral, *veja* teorema
lei
da probabilidade total, **59**
de um vetor aleatório, 110
de uma variável aleatória, 73
lei 0-1
de Hewitt-Savage, **382**
de Kolmogorov, **380**
lei dos grandes números, **232**
de Cantelli, 235
de Kolmogorov, 236, 239
de Tchebyshev, 233
forte, **234**
fraca, **233**
lema
de Borel-Cantelli, 214
de Fatou, 170
Lévy, *veja* teorema
Lindeberg, *veja* teorema
Lyapunov, *veja* desigualdade, *veja* teorema

Markov, *veja* desigualdade
martingale, **343**
média, *veja* esperança
medida, **47**
de contagem, 47
de Dirac, 47
de Lebesgue, 48
de probabilidade, 24, **41**
finita, 47
produto, 178
σ-finita, 47
método
de Monte Carlo, 243
do jacobiano, **123**
do primeiro momento, 203
do segundo momento, 203
dos mínimos quadrados, 196
Minkowski, *veja* desigualdade
momentos, **189**
Monte Carlo, *veja* método
mudança de variável, *veja* método do jacobiano

multiplicador de Lagrange, 103

normal, *veja* distribuição, *veja* número
número normal, 245
número de Catalan, 38

Paley, *veja* desigualdade
partes, *veja* conjunto
partição, **59**
  mais fina, 307
  mensurabilidade, 306
permutações, 34
π-sistema, 96
Poincaré, *veja* teorema
Poisson, *veja* distribuição
polinômio
  de Bernstein, 243
  de Taylor, *veja* expansão
  trigonométrico, 295
princípio
  da reflexão, 38
  dos grandes desvios, 420
probabilidade, **24**
  condicional, **56**
  total, *veja* Lei da Probabilidade Total
produto de Wallis, 440

quase certamente, 147

regra
  da cadeia, **176**
  do produto, **58**
relação de Stifel, 35
Riemann, *veja* soma

Savage, *veja* lei 0-1
Schwarz, *veja* ver desigualdade
σ-álgebra, **41**
  de Borel, **45**, 73, 110
    em $\mathbb{R}^{\mathbb{N}}$, **376**
  gerada, 44, 97
  produto, 177
σ-aditividade, 42, 47
singular, *veja* variável aleatória singular, *veja* vetor aleatório singular
soma de Riemann, 432
Stifel, *veja* relação
Stirling, *veja* fórmula
subaditividade, 48

tabela normal, 484
tabuleiro de Galton, 17
Taylor, *veja* expansão de
Tchebyshev, *veja* desigualdade, *veja* lei dos grandes números
teorema
  binomial, 37
  central do limite, *veja* teorema do limite central
  da amostragem opcional, **352**
  da continuidade de Lévy, **289**
  da convergência
    dominada, 156, 171
    monótona, 152, 163

das linhas, 36
de aproximação por álgebras, **377**
de convergência de martingales, 355
de convergência de Vitali, 361
de decomposição de Doob, 367
de decomposição de Hahn, 466
de extensão de Carathéodory, 451
de Fubini, 180
de Glivenko-Cantelli, 241
de Helly-Bray, 224
de recorrência de Poincaré, 396
de Tonelli, 179
de unicidade de medidas, 96
de Weierstrass, 242
do limite central, **251**
de De Moivre-Laplace, **255**
de Lindeberg, **261**
de Lyapunov, **258**
para variáveis i.i.d., **252**
do martingale parado, **347**
ergódico
de Birkhoff, 398
de von Neumann, 412
π-λ, 449
Tonelli, *veja* teorema
transformada, 279
de Fourier, *veja* função característica
de Laplace, *veja* função geradora de momentos

valor esperado, *veja* esperança
variável aleatória, **72**
absolutamente contínua, **86**
complexa, **284**
densidade, *veja* função de densidade
discreta, **80**
estendida, **100**
esperança, 168
independente, *veja* independência
mista, 94, 150
não-correlacionada, 192
simples, 136
singular, 94
variância, **190**
vetor aleatório, **109**
absolutamente contínuo, **116**
discreto, **115**
Vitali, *veja* teorema de convergência
von Neumann, *veja* teorema ergódico

Wald, *veja* identidade
Wallis, *veja* produto
Weierstrass, *veja* teorema

Young, *veja* desigualdade

Zygmund, *veja* desigualdade