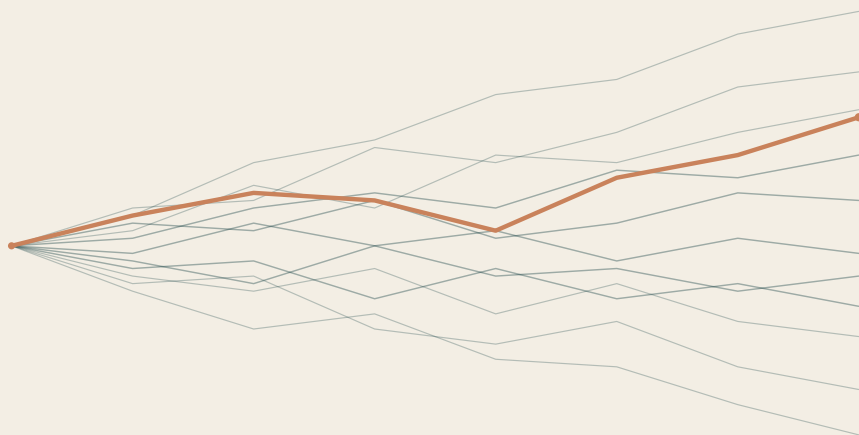


MODELOS, DINÂMICA E APLICAÇÕES

Probabilidade

Notas de aula · versão 0.999



Daniel Miranda · Rafael Grisi

UFABC

22 de setembro de 2026

“It is an old maxim of mine that when you have excluded the impossible, whatever remains, however improbable, must be the truth.”

- Sherlock Holmes

Licença

© 2026 Daniel Miranda.

Exceto quando indicado em contrário, esta obra está licenciada sob a **Creative Commons Atribuição–NãoComercial 4.0 Internacional (CC BY-NC 4.0)**. É permitida a cópia, redistribuição e adaptação do material, em qualquer meio ou formato, desde que seja atribuída a autoria apropriada e que o uso não tenha finalidade comercial.

Uma cópia da licença está disponível em

<https://creativecommons.org/licenses/by-nc/4.0/>.

A obra é fornecida “como está”, sem garantias expressas ou implícitas, nos limites permitidos pela legislação aplicável. Consulte o texto integral da licença para as permissões, condições e limitações correspondentes.

Sumário

Observações	vi
Agradecimentos	vii
Notação	viii
1. Introdução	1
1.1. Nem sempre podemos atribuir probabilidades	3
1.2. A necessidade de álgebras	5
1.3. Álgebras, σ -álgebras e outras classes	6
1.4. Um pouco mais sobre eventos	9
1.5. Classes de Conjuntos	13
1.6. σ -álgebra de Borel	23
2. Medidas de Probabilidade	36
2.1. Continuidade das Medidas de Probabilidade	41
2.2. Teorema de Extensão de Carathéodory	47
2.3. Classes Compactas	56
2.4. Aplicações	60
2.5. Borelianos: como se formam e quantos são?	80
2.6. Medidas em Espaços Métricos	85

3. Independência	94
3.1. Independência	99
3.2. Lei 0–1 de Kolmogorov e lemas de Borel–Cantelli	109
3.3. Aplicações	118
4. Variáveis Aleatórias	130
4.1. Funções Mensuráveis	130
4.2. Variáveis Aleatórias	136
4.3. Funções de Variáveis Aleatórias	139
4.4. Variáveis Aleatórias Geram σ -álgebras	142
4.5. Variáveis Aleatórias Induzem Medidas de Probabilidade em $\mathbb{R}$. . .	147
4.6. Independência de Variáveis Aleatórias	147
5. Integral de Lebesgue	154
5.1. Integral de Lebesgue	154
5.2. Propriedades da integral de Lebesgue	160
5.3. Comparando a Integral de Lebesgue e de Riemann	164
5.4. Integração e Limites	168
5.5. Espaços L_p	172
6. Distribuições	178
6.1. Distribuição de uma Variável Aleatória	178
6.2. A Função Distribuição	182
6.3. Tipos de Distribuições	193
7. Valor Esperado	209
7.1. Calculando a Esperança	215
7.2. Funções de Vetores Aleatórios	220
7.3. Variância, Covariância e Momentos de uma Variável Aleatória . . .	223
7.4. Integração com respeito à Distribuição	230
7.5. Desigualdades de Markov e Chebyshev	234
7.6. Desigualdade Maximal de Kolmogorov	238
7.7. Desigualdade de Jensen, Hölder e Minkowski	239
7.8. Funções Geradoras de Probabilidade	242

7.9. Processos de Ramificação	248
7.10. Funções Geradoras de Momento	252
7.11. Entropia	256
8. Medida Produto	266
8.1. Áreas de Seções Transversais	272
8.2. Integrais Iteradas	273
8.3. Distribuição Conjunta	277
8.4. Extensão de Kolmogorov em sequências	278
8.5. Existência de Variáveis Aleatórias Independentes	285
8.6. Convolução e Soma de Variáveis Aleatórias Independentes	290
8.7. Aplicações	296
8.8. Extensão de Kolmogorov geral	299
9. Acoplamento	306
9.1. Construções conjuntas	307
9.2. Acoplamento de processos	309
9.3. Variação total	310
9.4. A desigualdade de acoplamento	312
9.5. Acoplamentos maximais	314
9.6. Uma família inteira no mesmo espaço: percolação	316
10. Modos de Convergência	320
10.1. Convergências Determinísticas	320
10.2. Convergência em Espaços de Probabilidade	323
10.3. Comparando os Modos de Convergência	336
10.4. Teorema de Representação de Skorokhod	340
10.5. Princípio de Seleção de Helly	342
11. Teoremas-limite	351
11.1. Lei fraca dos grandes números	352
11.2. Lei forte dos grandes números	357
11.3. Integração de Monte Carlo	362
11.4. Polinômios de Bernstein e o teorema de Weierstrass	363

12. Teorema do Limite Central	371
12.1. Funções características	372
12.2. O Teorema do Limite Central	386
12.3. TLC para arranjos triangulares	387
13. Medidas com sinal e o teorema de Radon–Nikodým	398
13.1. Medidas com sinal	399
13.2. Decomposição de Jordan e variação total	404
13.3. Continuidade absoluta e singularidade	407
13.4. Decomposição de Lebesgue–Radon–Nikodým	409
14. Esperança Condicional	417
14.1. Existência e unicidade da esperança condicional	420
14.2. Esperança Condicional em Relação a uma Variável Aleatória	425
14.3. Propriedades da Esperança Condicional	425
14.4. Esperança Condicional como Projecção	430
15. Martingalas	437
15.1. Informação ao longo do tempo	437
15.2. Martingalas, submartingalas e supermartingalas	439
15.3. Exemplos fundamentais	441
15.4. Estratégias previsíveis e processos parados	442
15.5. O teorema da parada de Doob	445
15.6. Desigualdades e convergência de Doob	451
15.7. Decomposição de Doob: o caso discreto de Doob–Meyer	456
16. Movimento browniano	460
16.1. A construção de Paul Lévy	461
16.2. Propriedades de continuidade	469
16.3. Não diferenciabilidade	475
A. História da medida e da probabilidade	487
A.1. O primeiro nascimento: calcular o acaso	488
A.2. Frequências, erros e leis-limite	489

A.3. Medir o que não é um intervalo	490
A.4. O segundo nascimento: probabilidade como medida	491
A.5. A síntese de Kolmogorov	492
A.6. Processos, condicionamento e martingalas	493
A.7. A linguagem que ficou	494

Observações

Estas notas correspondem aos apontamentos dos cursos de Probabilidade ministrados na graduação e na pós-graduação em 2017, 2018, 2019, 2022 e 2023.

Ficaremos muito gratos pelo envio de sugestões e pela indicação de eventuais erros.

Agradecimentos

Agradecemos aos alunos Iuri Correa de Salles Barbosa, Tiago Leite Dias, Kayo Douglas (KY), Carlo Lemos, Bruno Carvalho, Guilherme Afonso Gigeck e Lucas Vasconcellos pelas inúmeras sugestões e correções.

Notação

Notação	Significado
$\wedge$	ínfimo
$\vee$	supremo
$\mathbb{1}_A$	função indicadora do conjunto A
$A_n \downarrow A$	sequência não crescente de conjuntos que converge a A
$A_n \uparrow A$	sequência não decrescente de conjuntos que converge a A
$\sigma(\mathcal{A})$	σ -álgebra gerada por $\mathcal{A}$
$f(\mathcal{A})$	álgebra gerada por $\mathcal{A}$
$\mathcal{M}(\mathcal{A})$	classe monótona gerada por $\mathcal{A}$
$\mathbf{P}$	probabilidade
μ, ν	medidas
$\mathcal{F}$	σ -álgebra
$\mathbf{P}^*, \mu^*$	medida exterior
ξ, η, ζ	variáveis aleatórias
$\mathbf{E}$	valor esperado
Var	variância
$\mu_1 \otimes \mu_2$	medida produto
$\xrightarrow{d}$	convergência em distribuição
$\xrightarrow{\mathbf{P}}$	convergência em probabilidade
$\xrightarrow{q.c.}, \xrightarrow{q.s.}$	convergência quase certa
$\xrightarrow{\mathcal{L}^p}$	convergência em $\mathcal{L}^p$
$*$	convolução
φ_ξ	função característica
$\hat{\mu}$	transformada de Fourier de μ
$\mathbf{E}[\xi \mid \mathcal{G}]$	esperança condicional

Introdução

Há fenômenos para os quais um modelo determinístico descreve com grande precisão o que será observado. Fixadas as condições iniciais, a lei da gravitação, por exemplo, determina a posição e a velocidade de um corpo em queda.

Em outras situações, a previsão individual deixa de ser uma linguagem razoável. Não sabemos quantas partículas um material radioativo emitirá no próximo intervalo de tempo. Mesmo o lançamento de uma moeda, determinístico em princípio, depende de condições iniciais que não controlamos na prática. Ainda assim, ambos os fenômenos exibem regularidades que podem ser descritas por uma lei probabilística. É nesse ponto — quando o resultado isolado escapa, mas o conjunto dos resultados conserva estrutura — que começa a probabilidade.

Para construir um modelo desse tipo, supomos que o resultado do experimento é desconhecido, mas que seus resultados possíveis são conhecidos. O conjunto desses resultados é o **espaço amostral**, geralmente denotado por Ω .

Exemplo 1.1 — .

- 1 Para o lançamento de uma moeda, podemos tomar $\Omega = \{\text{cara}, \text{coroa}\}$ ou, de modo equivalente, $\Omega = \{0, 1\}$, com 0 representando cara e 1, coroa.
- 2 Para modelar o número de pessoas que entram em uma agência bancária ao longo do dia, podemos tomar $\Omega = \{0, 1, 2, 3, \dots\}$.
- 3 Considere o experimento de lançar um dardo em um alvo circular de raio $r > 0$. Neste caso podemos definir o espaço amostral como qualquer círculo de raio r , como por exemplo

$$\Omega = \{(x, y) : x^2 + y^2 \leq r^2\}.$$

- 4 Para o sorteio uniforme de um número em $[0, 1]$, tomamos $\Omega = [0, 1]$.



Em geral, a pergunta não é qual resultado ocorreu, mas se ele possui certa propriedade. Ao retirar uma carta de um baralho, por exemplo, podemos querer saber apenas se ela é vermelha. Uma propriedade desse tipo corresponde a um subconjunto de Ω : o conjunto dos resultados que a satisfaz.

Chamaremos de **eventos** os subconjuntos de Ω aos quais o modelo permite atribuir probabilidade. A ressalva é importante: como veremos em breve, nem todo subconjunto poderá receber uma probabilidade.

1.0.1. Afinal, o que pedimos de uma probabilidade?

Uma probabilidade deve atribuir a cada evento A um número $\mathbf{P}(A) \in [0, 1]$, com $\mathbf{P}(\Omega) = 1$. Além disso, se dois eventos disjuntos A e B não podem ocorrer simultaneamente, é natural exigir que

$$\mathbf{P}[A \cup B] = \mathbf{P}[A] + \mathbf{P}[B].$$

Em espaços finitos, essa propriedade se estende por indução a qualquer união finita. Em espaços gerais, porém, precisamos controlar também uniões enumeráveis.

Se $A_1, A_2, \dots$ são eventos disjuntos dois a dois, exigimos

$$\mathbf{P}\left[\bigcup_{k=1}^{\infty} A_k\right] = \sum_{k=1}^{\infty} \mathbf{P}[A_k].$$

Essa propriedade é a **σ -aditividade**. Sua força ficará clara ao longo do texto: ela é a ponte entre operações finitas e fenômenos definidos por passagem ao limite.

Provisoriamente, portanto, uma probabilidade é uma função $\mathbf{P}$ definida sobre uma família de eventos e tal que

1 $\mathbf{P}[\Omega] = 1;$

2 se $A_1, A_2, \dots \subset \Omega$ são eventos disjuntos, então

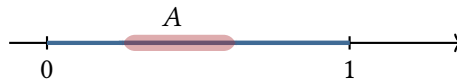
$$\mathbf{P}\left[\bigcup_{k=1}^{\infty} A_k\right] = \sum_{k=1}^{\infty} \mathbf{P}[A_k].$$

1.1. Nem sempre podemos atribuir probabilidades

Tentemos agora construir um modelo para uma situação aparentemente simples: o sorteio **uniforme** de um ponto no intervalo $[0,1)$.



Que condições esse modelo deveria satisfazer? Atribuir probabilidades apenas aos pontos não basta: cada ponto deverá ter probabilidade zero, embora o intervalo inteiro tenha probabilidade um. Precisamos, portanto, atribuir probabilidades a subconjuntos $A \subset [0,1)$.



Para expressar a uniformidade, pedimos invariância por translações módulo 1. Se insistirmos em medir todos os subconjuntos, procuraremos uma função $\mu : \mathcal{P}([0,1)) \rightarrow [0,1]$ que satisfaça:

1 $\mu([0,1)) = 1;$

2 para toda sequência de conjuntos disjuntos $A_1, A_2, \dots,$

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n);$$

3 seja uniforme, isto é, invariante por translações módulo 1: se $A \oplus b := \{\text{dec}(x + b) : x \in A\}$, então $\mu(A \oplus b) = \mu(A)$.

Uma função não negativa e σ -aditiva será chamada de **medida**; os conjuntos aos quais ela pode ser aplicada serão os conjuntos mensuráveis. Essas definições serão precisadas no Capítulo 2.

As três exigências acima parecem modestas. Juntas, porém, pedem demais: tal modelo não existe.

Teorema 1.2 — Conjunto de Vitali

Seja μ uma medida finita e não nula, definida em uma σ -álgebra $\mathcal{F}$ de subconjuntos de $\mathbb{T} = [0,1)$, e suponha que $A \in \mathcal{F}$ implique $A \oplus b \in \mathcal{F}$ e $\mu(A \oplus b) = \mu(A)$ para todo $b \in \mathbb{T}$. Então $\mathcal{F} \neq \mathcal{P}(\mathbb{T})$. Mais precisamente, todo conjunto de representantes das classes de $\mathbb{T}/(\mathbb{Q} \bmod 1)$ é não μ -mensurável.

Demonstração. Em $\mathbb{T}$, considere a relação de equivalência

$$x \sim y \iff x - y \in \mathbb{Q} \pmod{1}.$$

Pelo axioma da escolha, existe um conjunto $V \subset \mathbb{T}$ que contém exatamente um representante de cada classe de equivalência. Esse é um conjunto de Vitali.

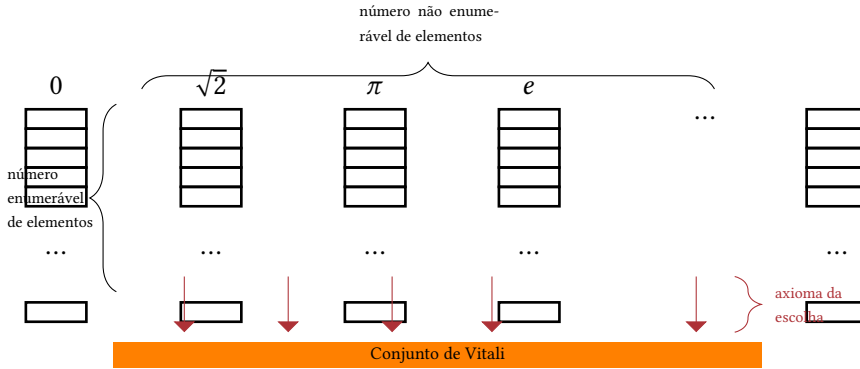
Seja $D = \mathbb{Q} \cap [0,1)$. A família $(V \oplus q)_{q \in D}$ é enumerável, disjunta duas a duas e sua união é $\mathbb{T}$. De fato, todo $x \in \mathbb{T}$ é equivalente a um único $v \in V$, de modo que $x = v \oplus q$ para algum $q \in D$. Se $v_1 \oplus q_1 = v_2 \oplus q_2$, então $v_1 \sim v_2$; pela escolha de um único representante, $v_1 = v_2$, e portanto $q_1 = q_2$ em $\mathbb{T}$.

Suponha, por absurdo, que $V \in \mathcal{F}$. Pela invariância, todos os conjuntos $V \oplus q$

são mensuráveis e têm a mesma medida. A σ -aditividade fornece

$$0 < \mu(\mathbb{T}) = \sum_{q \in D} \mu(V \oplus q) = \sum_{q \in D} \mu(V).$$

Se $\mu(V) = 0$, o lado direito é 0; se $\mu(V) > 0$, ele é infinito. As duas possibilidades contradizem $0 < \mu(\mathbb{T}) < \infty$. Logo $V \notin \mathcal{F}$, como queríamos. □



1.2. A necessidade de álgebras

O exemplo de Vitali mostra que pedimos demais. Para escapar da contradição, alguma das exigências deve ser abandonada:

- a a sigma aditividade $\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n)$ para conjuntos disjuntos;
- b a invariância por translação;
- c o Axioma da Escolha;
- d ou que todos os subconjuntos sejam mensuráveis.

A σ -aditividade é útil demais para ser descartada. A invariância por translações não é necessária em todo modelo, mas é inseparável da ideia de uniformidade neste exemplo. E o axioma da escolha... é o axioma da escolha.

Aceitaremos, então, que alguns subconjuntos não são mensuráveis. Em troca, preservamos a noção de probabilidade numa família ampla de eventos, estável pelas operações de conjuntos de que precisaremos.

1.3. Álgebras, σ -álgebras e outras classes

Fixado o espaço amostral Ω , denotaremos por $\mathcal{F}$ a família dos eventos aos quais podemos atribuir probabilidade. O exemplo de Vitali mostrou que, em geral, $\mathcal{F}$ não pode conter todos os subconjuntos de Ω .

Essa família deve, contudo, ser estável pelas operações naturais entre eventos. Se podemos perguntar se A ou B ocorreu, devemos também poder perguntar se A não ocorreu e se ocorreu ao menos um dos dois. Começemos pela estrutura finita correspondente.

Definição 1.3 —

Uma família não vazia de eventos $\mathcal{F}$ contendo Ω é denominada **álgebra** se é fechada em relação

1 ao complementar

$$A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$$

2 à união finita

$$A, B \in \mathcal{F} \Rightarrow A \cup B \in \mathcal{F}.$$

Essas duas condições também garantem o fechamento por interseções finitas:

$$A, B \in \mathcal{F} \Rightarrow A \cap B \in \mathcal{F}.$$

De fato, pela Lei de De Morgan $(A \cap B)^c = A^c \cup B^c$. Logo, $A \cap B = (A^c \cup B^c)^c$.

Exemplo 1.4 — Álgebra Trivial.

Seja $\mathcal{F} = \{\emptyset, \Omega\}$. Essa é a **álgebra trivial** e é a menor álgebra possível.

Exemplo 1.5 — Álgebra das Partes.

No outro extremo, $\mathcal{F} = \mathcal{P}(\Omega)$. Essa é a maior álgebra possível.

◀

A primeira álgebra distingue pouco; a segunda, no caso dos reais, distingue demais. A estrutura útil costuma estar entre esses extremos.

Exemplo 1.6 — Álgebra de Borel.

Considere $\Omega = \mathbb{R}$. Queremos que todos os intervalos estejam na álgebra. Por causa das propriedades da Definição 1.3, temos que a menor álgebra $\mathcal{F}$ que contém os intervalos será:

$$\mathcal{F} = \{\text{união finita de intervalos em } \mathbb{R}\}.$$

◀

Exemplo 1.7 — Álgebra de partição.

Dado um conjunto qualquer Ω , uma coleção de conjuntos não vazios e dois a dois disjuntos $\mathcal{P} = \{P_1, P_2, \dots, P_n\}$ é dita uma partição se $\Omega = \cup_{j=1}^n P_j$. Nesse caso os conjuntos $P_1, P_2, \dots, P_n$ são ditos átomos da partição.

Nesse caso a família gerada por todas as combinações de átomos é uma álgebra

$$f\langle \mathcal{P} \rangle = \left\{ \bigcup_{j \in J} P_j : J \subseteq \{1, \dots, n\} \right\}$$

Temos também que $|f\langle \mathcal{P} \rangle| = 2^n$.

Se $\mathcal{Q}$ for outra partição, dizemos que $\mathcal{Q}$ refina $\mathcal{P}$ se, para cada $P \in \mathcal{P}$, existem $Q_1, \dots, Q_m \in \mathcal{Q}$ de modo que

$$P = Q_1 \cup Q_2 \cup \dots \cup Q_m$$

◀

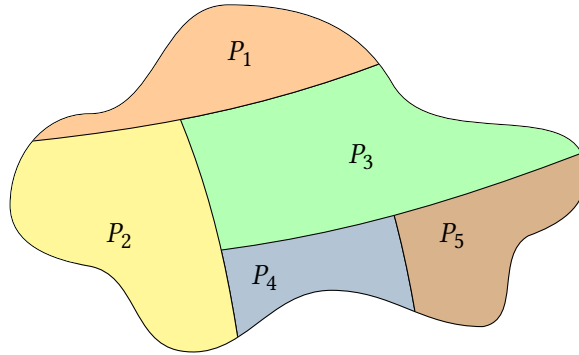


Figura 1.1.: Uma partição de Ω

A álgebra gerada pelos intervalos contém apenas uniões finitas de intervalos. Isso é pouco para a probabilidade: não permite descrever limites de eventos. Admitir uniões arbitrárias, por outro lado, seria demais, pois todo subconjunto dos reais é união de seus pontos:

$$A = \bigcup_{x \in A} \{x\}$$

O compromisso é admitir uniões enumeráveis. Este será o primeiro dogma da teoria da probabilidade.

Definição 1.8 – σ -álgebra

Uma família não vazia $\mathcal{F}$ de eventos é denominada de **σ -álgebra** se é fechada em relação

1 ao complementar

$$A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$$

2 a união enumerável

$$A_1, A_2, \dots \in \mathcal{F} \Rightarrow \bigcup_{n=1}^{\infty} A_n \in \mathcal{F}.$$

Dado Ω e $\mathcal{F}$ uma σ -álgebra de Ω , dizemos que $(\Omega, \mathcal{F})$ é um **espaço mensurável**.

Se voltarmos aos nossos exemplos, quais são σ -álgebras? Os Exemplos 1.4, 1.5 e 1.7 são, mas a álgebra de Borel apresentada no Exemplo 1.6 não é.

- a** A família $\{\emptyset, A, A^c, \Omega\}$ é uma σ -álgebra simples gerada pelo subconjunto A .
- b** A família dos subconjuntos de X que são enumeráveis ou cujos complementos são enumeráveis é uma σ -álgebra.

Exemplo 1.9 — σ -álgebra de Borel.

Considere $\Omega = \mathbb{R}$. A menor σ -álgebra que contém todos os intervalos existe, como mostraremos na Seção 1.5.1.

Essa σ -álgebra é denominada σ -álgebra de Borel. Veremos que ela não contém todos os subconjuntos de $\mathbb{R}$, conforme exigia o exemplo de Vitali.

Existem $2^{2^{\aleph_0}}$ subconjuntos de $\mathbb{R}$. Os conjuntos de Borel são somente $2^{\aleph_0}$. A palavra “somente” pede aqui alguma familiaridade com o infinito. A diferença entre essas cardinalidades garante a existência de subconjuntos não borelianos. A demonstração desse fato é apresentada na Seção 2.5



Exemplo 1.10 — Subespaço.

Seja $(\Omega, \mathcal{F})$ um espaço mensurável e $D \subset \Omega$ um subespaço. A σ -álgebra do subespaço é definida como:

$$\mathcal{F}_D = \{E \cap D : E \in \mathcal{F}\}$$



Em espaços enumeráveis, podemos em geral tomar o conjunto das partes como σ -álgebra; a dificuldade aparece sobretudo nos espaços não enumeráveis.

1.4. Um pouco mais sobre eventos

Uma σ -álgebra organiza as perguntas que o modelo permite fazer. Se $A \in \mathcal{F}$, dizer que A ocorreu significa simplesmente que o resultado observado pertence a

A. Essa tradução entre operações de conjuntos e afirmações sobre o experimento será usada continuamente.

Neste contexto, dados eventos $A, B \in \mathcal{F}$ também dizemos que

- *A não ocorreu ou não foi observado* para nos referirmos a A^c ;
- *ocorreu ou observamos A ou B* para nos referirmos a $A \cup B$;
- *ocorreram ou observamos A e B* para nos referirmos a $A \cap B$.

Analogamente, dada uma sequência de eventos $A_1, A_2, \dots \in \mathcal{F}$

- $\bigcup_{k=1}^{\infty} A_k$ é o mesmo que *ocorreu A_n para algum $n \geq 1$* ;
- $\bigcap_{k=1}^{\infty} A_k$ é o mesmo que *ocorreu A_n para todo $n \geq 1$* ;

Exemplo 1.11 — .

Consideremos o experimento de lançar infinitas moedas. Representaremos coroas por 0 e caras por 1, de modo que o espaço amostral é

$$\Omega = \{0, 1\}^{\infty} = \{(a_1, a_2, a_3, \dots) : a_k \in \{0, 1\}, k \geq 1\}.$$

Assim, o evento $A_k = \{\text{observamos cara no } k\text{-ésimo lançamento}\}$ é dado por

$$A_k = \{(a_1, a_2, \dots) \in \Omega : a_k = 1\}.$$

Deste modo, valem as seguintes descrições.

- $\bigcup_{k=1}^n A_k$ é o mesmo que *ocorreu ao menos uma cara nos n primeiros lançamentos*;
- $\bigcap_{k=1}^n A_k$ é o mesmo que *ocorreu cara em todos os n primeiros lançamentos*;
- $\bigcup_{k=1}^{\infty} A_k$ é o mesmo que *ocorreu ao menos uma cara*;
- $\bigcap_{k=1}^{\infty} A_k$ é o mesmo que *ocorreu cara em todos os lançamentos*;

◁

O fechamento enumerável das σ -álgebras permite ainda descrever eventos que dependem do comportamento de uma sequência inteira.

Definição 1.12 — Limite Superior e Inferior

Suponha que $\mathcal{F}$ é uma σ -álgebra. E considere $A_1, A_2, \dots \subseteq \Omega$.

Então o **limite superior** é o conjunto definido como

$$\begin{aligned}\limsup A_n &:= \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k \\ &= \{\omega \in \Omega : \forall n \exists k \geq n \text{ tal que } \omega \in A_k\} \\ &= \{\omega \in \Omega : \omega \text{ pertence a infinitos } A_n\}\end{aligned}$$

Por outro lado, o **limite inferior** é o conjunto definido como

$$\begin{aligned}\liminf A_n &:= \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k \\ &= \{\omega \in \Omega : \exists n \text{ tal que } \forall k \geq n, \omega \in A_k\} \\ &= \{\omega \in \Omega : \omega \text{ pertencem a todos } A_n \text{ exceto por um número finito deles}\}\end{aligned}$$

Heurística 1.13 — Retornos semanais

Denote por A_n o conjunto das pessoas que visitaram uma biblioteca na n -ésima semana. Algumas aparecem apenas nas primeiras semanas; outras retornam infinitas vezes, mas com intervalos; outras, a partir de certo momento, passam a comparecer toda semana. Então

- $\limsup A_n$ reúne quem retorna em infinitas semanas;
- $\liminf A_n$ reúne quem, depois de alguma semana, não volta a faltar.

A diferença entre os dois limites é a diferença entre voltar sempre e, depois de algum instante, estar sempre presente.

Sempre é verdade que $\liminf A_n \subset \limsup A_n$ e quando estes conjuntos coincidem, dizemos que o limite existe:

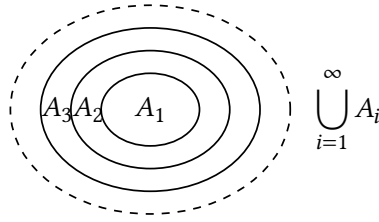


Figura 1.2.: Sequência crescente de eventos e sua união.

Definição 1.14 —

Se $\limsup A_n = \liminf A_n = A$, então dizemos que A_n **converge** para A e escrevemos $A_n \rightarrow A$ ou que $A = \lim A_n$ e nesse caso

$$A = \lim A_n = \limsup A_n = \liminf A_n$$

Exemplo 1.15 — .

Seja $A_1 = A_3 = A_5 = \dots = A$ e $A_2 = A_4 = A_6 = \dots = B$ com B um subconjunto próprio de A . Então $\limsup A_n = A$ e $\liminf A_n = B$.

◁

Definição 1.16 — Eventos não crescentes e não decrescentes

A sequência de eventos $A_1, A_2, \dots$ é dita **não crescente** se $A_1 \supseteq A_2 \supseteq A_3 \supseteq \dots$. Nesse caso escreveremos que $A_n \downarrow$.

De modo análogo, a sequência de eventos $A_1, A_2, \dots$ é dita **não decrescente** se $A_1 \subseteq A_2 \subseteq A_3 \subseteq \dots$. Nesse caso escreveremos que $A_n \uparrow$.

Proposição 1.17 —

1 Se $A_n \uparrow$, então A_n converge, e nesse caso temos que

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

Denotaremos tal fato por $\lim\uparrow A_n = A$ (ou $A_n \uparrow A$).

2 Se $A_n \downarrow$, então A_n converge, e nesse caso temos que

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

Denotaremos tal fato por $\lim\downarrow A_n = A$ (ou $A_n \downarrow A$).

Demonstração. Denotaremos por $A = \limsup A_n$ e por $B = \liminf A_n$. Claramente $B \subset A$.

Suponha que os conjuntos A_n são não decrescentes, ou seja, $A_n \subset A_{n+1}$. Seja $x \in A$ e seja $I_x = \{n \in \mathbb{N} : x \in A_n\}$. O conjunto $I_x \subset \mathbb{N}$ é não vazio e infinito. Seja $n_0 \in I_x$ o menor desses naturais. Se $x \in A_{n_0}$ então $x \in A_n \forall n \geq n_0$. Logo, o conjunto dos A_i em que x não está é finito e, portanto, $x \in B$. Assim, $A \subseteq B$.

Suponha que os conjuntos A_n sejam não crescentes, $A_{n+1} \subset A_n$. Seja $x \in A = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k$. Para cada n , $\bigcup_{k=n}^{\infty} A_k = A_n$ por serem não crescentes, então $A = \bigcap_{n=1}^{\infty} A_n$. Em particular, $x \in A_n$ para todo n ; logo $x \in B$ e, consequentemente, $A \subseteq B$. □

1.5. Classes de Conjuntos

Construir um modelo probabilístico exige escolher uma σ -álgebra e uma medida. A medida virá no próximo capítulo. Nesta seção reunimos ferramentas para reconhecer e construir a σ -álgebra.

Definição 1.18 – λ -sistema

Uma família de eventos $\mathcal{F} \subset \mathcal{P}(\Omega)$ é denominada **λ -sistema** se

$$\boxed{1} \quad \Omega \in \mathcal{F}$$

$$\boxed{2} \quad A \in \mathcal{F} \implies A^c \in \mathcal{F}$$

$$\boxed{3} \quad \underbrace{A_1, A_2, \dots}_{\text{disjuntos}} \in \mathcal{F} \implies \bigcup_{k=1}^{\infty} A_k \in \mathcal{F}.$$

Na definição acima, a condição $\Omega \in \mathcal{F}$ garante que a classe $\mathcal{F}$ não seja vazia.

Proposição 1.19 –

Um λ -sistema $\mathcal{L}$ é fechado em relação a diferenças próprias. Isto é, se $A, B \in \mathcal{L}$ e $B \subseteq A$, então $A \setminus B \in \mathcal{L}$.

Demonstração. Como $A, B \in \mathcal{L}$, também $A^c, B^c \in \mathcal{L}$. A inclusão $B \subseteq A$ implica que A^c e B são disjuntos. Logo $A^c \cup B \in \mathcal{L}$. Tomando novamente o complemento,

$$A \setminus B = (A^c \cup B)^c \in \mathcal{L}.$$

□

Definição 1.20 – π -sistema

Uma família não vazia de eventos $\mathcal{P} \subset \mathcal{P}(\Omega)$ é denominada **π -sistema** se $A, B \in \mathcal{P} \implies A \cap B \in \mathcal{P}$.

Definição 1.21 – Classe Monótona

Uma família não vazia de eventos $\mathcal{M} \subset \mathcal{P}(\Omega)$ é uma **classe monótona** se

$$\boxed{1} \quad A_1, A_2, \dots \in \mathcal{M} \text{ e } A_n \uparrow A \implies A \in \mathcal{M}$$

$$\boxed{2} \quad A_1, A_2, \dots \in \mathcal{M} \text{ e } A_n \downarrow A \implies A \in \mathcal{M}.$$

As classes monótonas servirão para reconhecer σ -álgebras a partir de propriedades de fechamento mais simples. Toda σ -álgebra é uma classe monótona; os resultados a seguir esclarecem quais condições adicionais permitem percorrer o caminho inverso.

Temos a seguinte relação entre as classes de conjuntos

Teorema 1.22 – $\sigma = \lambda + \pi = \mathcal{M} + f$

São equivalentes as seguintes afirmações:

$$\boxed{1} \quad \mathcal{F} \text{ é uma } \sigma\text{-álgebra}$$

$$\boxed{2} \quad \mathcal{F} \text{ é uma álgebra e classe monótona}$$

$$\boxed{3} \quad \mathcal{F} \text{ é um } \lambda\text{-sistema e um } \pi\text{-sistema.}$$

No título, f abrevia *field* (álgebra de conjuntos), e o sinal $+$ significa que as duas propriedades são exigidas simultaneamente. Assim, o teorema diz que ser σ -álgebra equivale tanto a ser λ -sistema e π -sistema quanto a ser classe monótona e álgebra.

Demonstração. Toda σ -álgebra é, ao mesmo tempo, uma álgebra e uma classe monótona; também é um λ -sistema e um π -sistema. Assim, basta provar que cada uma das condições $\boxed{2}$ e $\boxed{3}$ implica $\boxed{1}$.

De $\boxed{3}$ para $\boxed{1}$. Suponha que $\mathcal{F}$ seja um λ -sistema e um π -sistema. A definição de λ -sistema já fornece $\Omega \in \mathcal{F}$ e o fechamento por complementos. Resta obter o

fechamento por uniões enumeráveis que não sejam necessariamente disjuntas.

Dados $A_1, A_2, \dots \in \mathcal{F}$, defina

$$B_1 = A_1, \quad B_k = A_k \setminus \bigcup_{j < k} A_j = A_k \cap \bigcap_{j < k} A_j^c, \quad k \geq 2.$$

Como $\mathcal{F}$ é fechada por complementos e por interseções finitas, cada B_k pertence a $\mathcal{F}$. Além disso, os B_k são disjuntos dois a dois e

$$\bigcup_{k=1}^{\infty} B_k = \bigcup_{k=1}^{\infty} A_k.$$

O fechamento do λ -sistema por uniões enumeráveis disjuntas mostra, portanto, que $\bigcup_k A_k \in \mathcal{F}$. Logo $\mathcal{F}$ é uma σ -álgebra.

De [2](#) para [1](#). Suponha agora que $\mathcal{F}$ seja uma álgebra e uma classe monótona. Se $A_1, A_2, \dots \in \mathcal{F}$, então as uniões finitas

$$S_n = \bigcup_{k=1}^n A_k$$

pertencem à álgebra $\mathcal{F}$ e satisfazem $S_n \uparrow \bigcup_{k=1}^{\infty} A_k$. Pelo fechamento da classe monótona por limites crescentes, $\bigcup_k A_k \in \mathcal{F}$. Assim, $\mathcal{F}$ é uma σ -álgebra, e as três condições são equivalentes. □

1.5.1. Classes Geradas

Proposição 1.23 —

Seja Ω um conjunto qualquer, e $\mathcal{F}_\lambda$ ($\lambda \in \Lambda$) uma família de σ -álgebras em Ω . Então $\bigcap_{\lambda \in \Lambda} \mathcal{F}_\lambda$ é uma σ -álgebra.

Demonstração. Denote $\mathcal{F} = \bigcap_{\lambda \in \Lambda} \mathcal{F}_\lambda$. Como Ω pertence a toda $\mathcal{F}_\lambda$, temos $\Omega \in \mathcal{F}$. Se $A \in \mathcal{F}$, então $A \in \mathcal{F}_\lambda$ para todo λ ; portanto $A^c \in \mathcal{F}_\lambda$ para todo λ , isto é, $A^c \in \mathcal{F}$. Finalmente, se $A_1, A_2, \dots \in \mathcal{F}$, então, para cada λ ,

$$A_n \in \mathcal{F}_\lambda \text{ para todo } n \implies \bigcup_{n=1}^{\infty} A_n \in \mathcal{F}_\lambda.$$

Assim, $\bigcup_{n=1}^{\infty} A_n$ pertence à intersecção de todas as $\mathcal{F}_\lambda$. Os três axiomas de σ -álgebra estão verificados. □

Do mesmo modo podemos provar que:

- intersecção de álgebras é álgebra.
- intersecção de classe monótona é classe monótona
- intersecção de λ -sistema é λ -sistema

Essas observações nos permitem fazer as seguintes definições

Definição 1.24 —

Seja $\mathcal{C}$ uma família arbitrária de subconjuntos de Ω .

- 1** Definimos a **σ -álgebra gerada** por $\mathcal{C}$ como

$$\sigma\langle \mathcal{C} \rangle := \bigcap_{\substack{\mathcal{F} \text{ é } \sigma\text{-álgebra} \\ \mathcal{C} \subset \mathcal{F}}} \mathcal{F}$$

- 2** Definimos a **álgebra gerada** por $\mathcal{C}$ como

$$f\langle \mathcal{C} \rangle := \bigcap_{\substack{\mathcal{F} \text{ é álgebra} \\ \mathcal{C} \subset \mathcal{F}}} \mathcal{F}$$

- 3** Definimos a **classe monótona gerada** por $\mathcal{C}$ como

$$\mathcal{M}\langle \mathcal{C} \rangle := \bigcap_{\substack{\mathcal{M} \text{ é classe monótona} \\ \mathcal{C} \subset \mathcal{M}}} \mathcal{M}$$

4 Definimos o λ -sistema gerado por $\mathcal{C}$ como

$$\lambda\langle\mathcal{C}\rangle := \bigcap_{\substack{\mathcal{L} \text{ é um } \lambda\text{-sistema} \\ \mathcal{C} \subset \mathcal{L}}} \mathcal{L}$$

Assim, $\sigma\langle\mathcal{C}\rangle$ é a menor σ -álgebra que contém $\mathcal{C}$. A definição é implícita: em geral não dispomos de uma lista simples de todos os seus elementos. Ela contém tudo o que as operações enumeráveis de complemento, união e interseção obrigam a incluir — e, em exemplos importantes, essa descrição é mais útil do que uma tentativa de enumerar os conjuntos resultantes.

Exemplo 1.25 — .

Para um exemplo simples, considere o conjunto $\Omega = \{1, 2, 3, 4\}$. Então, a σ -álgebra gerada pelo subconjunto unitário $\{1\}$ é $\{\emptyset, \{1\}, \{2, 3, 4\}, \{1, 2, 3, 4\}\}$.



Exemplo 1.26 — .

Dado $\Omega = \mathbb{R}$. Então, a σ -álgebra gerada pelos subconjuntos finitos pode ser descrita como:

- a** conjuntos finitos
- b** conjuntos enumeráveis
- c** conjuntos cujo complementar é enumerável ou finito.

Essa é a σ -álgebra dos conjuntos enumeráveis ou co-enumeráveis.



O próximo princípio reúne um argumento recorrente: identificar os conjuntos que possuem a propriedade desejada e estudar a classe que eles formam.

Teorema 1.27 — Princípio dos Bons Conjuntos

Dado $\mathcal{C}$ e $\mathcal{G}$ duas família de subconjuntos de Ω . Se

a $\mathcal{C} \subset \mathcal{G}$;

b $\mathcal{G}$ é uma σ -álgebra

Então $\sigma\langle\mathcal{C}\rangle \subset \mathcal{G}$.

Demonstração. A demonstração é imediata do fato que:

$$\sigma\langle\mathcal{C}\rangle := \bigcap_{\substack{\mathcal{F} \text{ é uma } \sigma\text{-álgebra} \\ \mathcal{C} \subset \mathcal{F}}} \mathcal{F}$$

é a menor σ -álgebra que contém $\mathcal{C}$. □

A geração de σ -álgebras é compatível com a restrição a subespaços de Ω .

Teorema 1.28 — Geradores Restringidos

Dados Ω um espaço amostral e $\mathcal{C}$ uma classe de subconjuntos de Ω . Se $\Omega_0 \subset \Omega$ então

$$\underbrace{\sigma\langle\mathcal{C} \cap \Omega_0\rangle}_{\substack{\sigma\text{-álgebra} \\ \text{em } \Omega_0}} = \underbrace{\sigma\langle\mathcal{C}\rangle}_{\substack{\sigma\text{-álgebra} \\ \text{em } \Omega}} \cap \Omega_0.$$

Demonstração.

$\Rightarrow$ Provaremos inicialmente que $\sigma\langle\mathcal{C} \cap \Omega_0\rangle \subset \sigma\langle\mathcal{C}\rangle \cap \Omega_0$

Essa implicação segue facilmente por *Princípio dos Bons Conjuntos* já que $\mathcal{C} \cap \Omega_0 \subset \sigma\langle\mathcal{C}\rangle \cap \Omega_0$ e $\sigma\langle\mathcal{C}\rangle \cap \Omega_0$ é uma σ -álgebra.

$\Leftarrow$ Agora mostraremos que $\sigma\langle\mathcal{C}\rangle \cap \Omega_0 \subset \sigma\langle\mathcal{C} \cap \Omega_0\rangle$

Observe que essa inclusão é equivalente à afirmação de que, para cada $A \in \sigma\langle\mathcal{C}\rangle$,

$A \cap \Omega_0 \in \sigma(\mathcal{C} \cap \Omega_0)$. Para demonstrar esse fato seja

$$\mathcal{G} := \{A \subset \Omega : A \cap \Omega_0 \in \sigma(\mathcal{C} \cap \Omega_0)\}.$$

Será suficiente demonstrar os quatro itens a seguir e, finalmente, utilizar Princípio dos Bons Conjuntos para concluir que $\sigma(\mathcal{C}) \subset \mathcal{G}$.

$$\square \mathcal{C} \subset \mathcal{G}$$

$$A \in \mathcal{C} \implies A \cap \Omega_0 \in \mathcal{C} \cap \Omega_0 \subset \sigma(\mathcal{C} \cap \Omega_0).$$

$$\square \Omega \in \mathcal{G}$$

$$\Omega_0 \subset \Omega \implies \Omega \cap \Omega_0 = \Omega_0 \in \sigma(\mathcal{C} \cap \Omega_0)$$

pois uma σ -álgebra em Ω_0 deve conter Ω_0

$$\implies \Omega \in \mathcal{G}.$$

$$\square A \in \mathcal{G} \implies A^c \in \mathcal{G}$$

Observe que A^c denota o complementar com respeito a Ω . Assim

$$\begin{aligned} A \in \mathcal{G} &\implies A \cap \Omega_0 \in \sigma(\mathcal{C} \cap \Omega_0) \\ &\implies \underbrace{\Omega_0 - A \cap \Omega_0}_{\text{complemento em } \Omega_0} \in \sigma(\mathcal{C} \cap \Omega_0) \\ &\implies \underbrace{\Omega_0 \cap (A^c \cup \Omega_0^c)}_{=A^c \cap \Omega_0} \in \sigma(\mathcal{C} \cap \Omega_0) \\ &\implies A^c \in \mathcal{G}. \end{aligned}$$

$$\square A_1, A_2, \dots \in \mathcal{G} \implies \bigcup_k A_k \in \mathcal{G}$$

$$\begin{aligned} A_1, A_2, \dots \in \mathcal{G} &\implies A_k \cap \Omega_0 \in \sigma(\mathcal{C} \cap \Omega_0), \forall k \\ &\implies \bigcup_k (A_k \cap \Omega_0) \in \sigma(\mathcal{C} \cap \Omega_0) \\ &\implies \left(\bigcup_k A_k \right) \cap \Omega_0 \in \sigma(\mathcal{C} \cap \Omega_0) \\ &\implies \bigcup_k A_k \in \mathcal{G}. \end{aligned}$$

□

Teorema 1.29 – Teorema da Classe Monótona de Halmos

Se $\mathcal{F}_0$ é uma álgebra então $\mathcal{M}\langle\mathcal{F}_0\rangle = \sigma\langle\mathcal{F}_0\rangle$.

Demonstração. Escreva $\mathcal{M} = \mathcal{M}\langle\mathcal{F}_0\rangle$. Faremos a prova em três passos.

1. *A inclusão imediata.* Como $\sigma\langle\mathcal{F}_0\rangle$ é uma classe monótona que contém $\mathcal{F}_0$, a minimalidade de $\mathcal{M}$ dá

$$\mathcal{M} \subseteq \sigma\langle\mathcal{F}_0\rangle.$$

2. *Fechamento por complementos.* Considere

$$\mathcal{D} = \{A \subseteq \Omega : A^c \in \mathcal{M}\}.$$

Se $A_n \in \mathcal{D}$ e $A_n \uparrow A$, então $A_n^c \in \mathcal{M}$ e $A_n^c \downarrow A^c$; logo $A^c \in \mathcal{M}$, isto é, $A \in \mathcal{D}$. O caso $A_n \downarrow A$ é análogo, com os complementos crescendo. Portanto, $\mathcal{D}$ é uma classe monótona. Como $\mathcal{F}_0$ é uma álgebra, $A \in \mathcal{F}_0$ implica $A^c \in \mathcal{F}_0 \subseteq \mathcal{M}$; assim, $\mathcal{F}_0 \subseteq \mathcal{D}$. Pela minimalidade, $\mathcal{M} \subseteq \mathcal{D}$, e $\mathcal{M}$ é fechada por complementos.

3. *Fechamento por uniões finitas.* Fixe $B \in \mathcal{F}_0$ e defina

$$\mathcal{D}_B = \{A \subseteq \Omega : A \cup B \in \mathcal{M}\}.$$

Se $A_n \uparrow A$, então $A_n \cup B \uparrow A \cup B$; se $A_n \downarrow A$, então $A_n \cup B \downarrow A \cup B$. Isso mostra que $\mathcal{D}_B$ é uma classe monótona. Além disso, $\mathcal{F}_0 \subseteq \mathcal{D}_B$, pois $A, B \in \mathcal{F}_0$ implica $A \cup B \in \mathcal{F}_0 \subseteq \mathcal{M}$. Mais uma vez pela minimalidade, $\mathcal{M} \subseteq \mathcal{D}_B$. Concluimos que

$$A \in \mathcal{M}, B \in \mathcal{F}_0 \implies A \cup B \in \mathcal{M}.$$

Precisamos agora permitir que também o segundo conjunto pertença a $\mathcal{M}$. Para isso, considere

$$\mathcal{E} = \{B \subseteq \Omega : A \cup B \in \mathcal{M} \text{ para todo } A \in \mathcal{M}\}.$$

Para cada $A \in \mathcal{M}$, as identidades do parágrafo anterior mostram que a condição $A \cup B \in \mathcal{M}$ é preservada quando $B_n \uparrow B$ ou $B_n \downarrow B$. Logo $\mathcal{E}$ é uma classe monótona.

A conclusão destacada acima mostra que $\mathcal{F}_0 \subseteq \mathcal{E}$; pela minimalidade, $\mathcal{M} \subseteq \mathcal{E}$. Portanto, $A, B \in \mathcal{M}$ implica $A \cup B \in \mathcal{M}$.

Como $\Omega \in \mathcal{F}_0 \subseteq \mathcal{M}$, os passos 2 e 3 mostram que $\mathcal{M}$ é uma álgebra. Ela já é uma classe monótona; pelo Teorema 1.22, ela é uma σ -álgebra. Como contém $\mathcal{F}_0$, contém também $\sigma\langle \mathcal{F}_0 \rangle$, o que prova a outra inclusão. $\square$

Teorema 1.30 — π — λ de Dynkin

Se $\mathcal{P}$ é um π -sistema então $\lambda\langle \mathcal{P} \rangle = \sigma\langle \mathcal{P} \rangle$.

Demonstração. Escreva $\mathcal{L} = \lambda\langle \mathcal{P} \rangle$. Como toda σ -álgebra é um λ -sistema,

$$\mathcal{L} \subseteq \sigma\langle \mathcal{P} \rangle.$$

Resta provar a inclusão inversa. Para cada $A \subseteq \Omega$, defina

$$\mathcal{G}_A = \{B \subseteq \Omega : A \cap B \in \mathcal{L}\}.$$

Fixe $A \in \mathcal{P}$. Temos $\Omega \in \mathcal{G}_A$, pois $A = A \cap \Omega \in \mathcal{L}$. Se $B \in \mathcal{G}_A$, então

$$A \cap B^c = A \setminus (A \cap B) \in \mathcal{L},$$

pela propriedade de diferenças próprias. Portanto $B^c \in \mathcal{G}_A$. Se $B_1, B_2, \dots$ são disjuntos e pertencem a $\mathcal{G}_A$, então os conjuntos $A \cap B_n$ também são disjuntos e

$$A \cap \bigcup_{n=1}^{\infty} B_n = \bigcup_{n=1}^{\infty} (A \cap B_n) \in \mathcal{L}.$$

Logo $\mathcal{G}_A$ é um λ -sistema. Como $\mathcal{P}$ é um π -sistema, $\mathcal{P} \subseteq \mathcal{G}_A$; pela minimalidade, $\mathcal{L} \subseteq \mathcal{G}_A$.

Considere agora

$$\mathcal{H} = \{A \subseteq \Omega : \mathcal{L} \subseteq \mathcal{G}_A\}.$$

A classe $\mathcal{H}$ é um λ -sistema. Com efeito, $\Omega \in \mathcal{H}$. Se $A \in \mathcal{H}$ e $B \in \mathcal{L}$, então $A \cap B \in \mathcal{L}$ e

$$A^c \cap B = B \setminus (A \cap B) \in \mathcal{L},$$

de modo que $A^c \in \mathcal{H}$. Finalmente, se $A_1, A_2, \dots$ são disjuntos e pertencem a $\mathcal{H}$, então, para cada $B \in \mathcal{L}$, os conjuntos $A_n \cap B$ são disjuntos e

$$\left(\bigcup_{n=1}^{\infty} A_n \right) \cap B = \bigcup_{n=1}^{\infty} (A_n \cap B) \in \mathcal{L}.$$

Assim, $\bigcup_n A_n \in \mathcal{H}$.

O parágrafo anterior mostrou que $\mathcal{P} \subseteq \mathcal{H}$; portanto $\mathcal{L} \subseteq \mathcal{H}$. Se $A, B \in \mathcal{L}$, segue que $B \in \mathcal{G}_A$, isto é, $A \cap B \in \mathcal{L}$. Logo $\mathcal{L}$ é também um π -sistema. Pelo Teorema 1.22, $\mathcal{L}$ é uma σ -álgebra que contém $\mathcal{P}$. Consequentemente, $\sigma\langle \mathcal{P} \rangle \subseteq \mathcal{L}$, concluindo a demonstração. $\square$

O seguinte teorema é uma consequência direta dos Teoremas 1.27 e 1.30.

Teorema 1.31 — Princípio dos Bons Conjuntos 2

Dados $\mathcal{P}$ e $\mathcal{G}$ duas sub-coleções de subconjuntos de Ω . Se

a $\mathcal{P} \subset \mathcal{G}$;

b $\mathcal{P}$ é uma π -sistema;

c $\mathcal{G}$ é uma λ -sistema

então $\sigma\langle \mathcal{P} \rangle \subset \mathcal{G}$.

Demonstração. Como $\mathcal{G}$ é um λ -sistema que contém $\mathcal{P}$, a minimalidade dá $\lambda\langle \mathcal{P} \rangle \subseteq \mathcal{G}$. Pelo Teorema π - λ de Dynkin, $\lambda\langle \mathcal{P} \rangle = \sigma\langle \mathcal{P} \rangle$. Logo $\sigma\langle \mathcal{P} \rangle \subseteq \mathcal{G}$. $\square$

1.6. σ -álgebra de Borel

As σ -álgebras de Borel permeiam a teoria da medida e a probabilidade. Seu traço distintivo é admitir muitos geradores equivalentes: abertos, fechados, bolas, retângulos ou cilindros. A liberdade de escolher o gerador adequado simplifica de-

monstrações e construções; em $(0,1]^d$, por exemplo, retângulos serão a linguagem natural para construir a medida de Lebesgue.

Definição 1.32 – Espaço Métrico

O par (Ω, d) , no qual Ω é um conjunto e d é uma função

$$d : \Omega \times \Omega \rightarrow [0, \infty)$$

é dito **espaço métrico** se

$$1 \quad d(x, y) \geq 0$$

$$2 \quad d(x, y) = 0 \iff x = y$$

$$3 \quad d(x, y) = d(y, x)$$

$$4 \quad d(x, y) \leq d(x, z) + d(z, y)$$

Definição 1.33 – Aberto

Dado Ω um espaço métrico com métrica $d : \Omega \times \Omega \rightarrow [0, \infty]$. Um conjunto $A \subset \Omega$ é dito **aberto** se para cada $x \in A$, existir $\epsilon > 0$ tal que a bola aberta $\{y \in \Omega : d(x, y) < \epsilon\}$ está inteiramente contida em A .

Definição 1.34 – σ -álgebra de Borel $\mathcal{B}(\Omega)$

A **σ -álgebra de Borel** Ω (com respeito à métrica d), denotada por $\mathcal{B}(\Omega)$, é definida como a σ -álgebra gerada pelos conjuntos abertos.

A definição se aplica, em particular, às σ -álgebras de Borel $\mathcal{B}(\mathbb{R}^d)$ e $\mathcal{B}((0,1]^d)$. Para uma classe $\mathcal{C}$ de subconjuntos de Ω , usaremos a notação

$$\mathcal{C}|_{\Omega_0} := \{C \cap \Omega_0 : C \in \mathcal{C}\}.$$

Teorema 1.35 — Restrição de Borel

Se Ω é um espaço métrico e $\Omega_0 \subseteq \Omega$ tem a métrica induzida, então a σ -álgebra de Borel de Ω_0 satisfaz

$$\mathcal{B}(\Omega_0) = \mathcal{B}(\Omega)|_{\Omega_0}.$$

Se, além disso, $\Omega_0 \in \mathcal{B}(\Omega)$, então

$$\mathcal{B}(\Omega_0) = \{B \in \mathcal{B}(\Omega) : B \subseteq \Omega_0\}.$$

Demonstração. Começaremos mostrando que $\mathcal{B}(\Omega_0) = \mathcal{B}(\Omega)|_{\Omega_0}$.

Considere

$\mathcal{G} :=$ subconjuntos abertos de Ω

$\mathcal{G}_0 :=$ subconjuntos abertos de Ω_0

Do curso de topologia você deve se lembrar que:

$$\mathcal{G}_0 = \mathcal{G}|_{\Omega_0}.$$

Logo

$$\begin{aligned} \mathcal{B}(\Omega_0) &= \sigma\langle \mathcal{G}_0 \rangle = \sigma\langle \mathcal{G}|_{\Omega_0} \rangle \\ &= \sigma\langle \mathcal{G} \rangle|_{\Omega_0} \\ &= \mathcal{B}(\Omega)|_{\Omega_0}, \end{aligned}$$

onde a segunda igualdade é o Teorema dos Geradores Restringidos.

Agora suponha $\Omega_0 \in \mathcal{B}(\Omega)$. Se $B \in \mathcal{B}(\Omega)$ e $B \subseteq \Omega_0$, então $B = B \cap \Omega_0$ pertence ao traço. Reciprocamente, se $B = \tilde{B} \cap \Omega_0$, com $\tilde{B} \in \mathcal{B}(\Omega)$, então $B \in \mathcal{B}(\Omega)$, pois tanto $\tilde{B}$ quanto Ω_0 são borelianos, e claramente $B \subseteq \Omega_0$. $\square$

Por *retângulo* em $\mathbb{R}^d$ entenderemos um produto de d intervalos; por *cilindro coordenado*, um conjunto da forma $\pi_j^{-1}(I) = \{x \in \mathbb{R}^d : x_j \in I\}$, onde $I \subseteq \mathbb{R}$ é um intervalo.

Teorema 1.36 – Geradores da σ -álgebra de Borel

$$\begin{aligned}
 \mathcal{B}(\mathbb{R}^d) &= \sigma\langle (-\infty, c_1] \times \cdots \times (-\infty, c_d] : -\infty < c_k < \infty \rangle \\
 &= \sigma\langle \text{bolas abertas de } \mathbb{R}^d \rangle \\
 &= \sigma\langle \text{subconjuntos abertos de } \mathbb{R}^d \rangle \\
 &= \sigma\langle \text{subconjuntos fechados de } \mathbb{R}^d \rangle \\
 &= \sigma\langle \text{subconjuntos compactos de } \mathbb{R}^d \rangle \\
 &= \sigma\langle \text{retângulos } \mathbb{R}^d \rangle \\
 &= \sigma\langle \text{cilindros } \mathbb{R}^d \rangle
 \end{aligned}$$

$$\begin{aligned}
 \mathcal{B}((0,1]) &= \sigma\langle \mathcal{B}_0((0,1]) \rangle \\
 &= \sigma\langle (a,b] : 0 \leq a \leq b \leq 1 \rangle \\
 &= \sigma\langle (a,b) : 0 < a < b < 1 \rangle \\
 &= \sigma\langle [a,b] : 0 < a < b < 1 \rangle \\
 &= \sigma\langle (0,a] : 0 < a < 1 \rangle \\
 &= \sigma\langle \text{subconjuntos abertos de } (0,1] \rangle \\
 &= \sigma\langle \text{subconjuntos fechados de } (0,1] \rangle
 \end{aligned}$$

$$\begin{aligned}
 \mathcal{B}((0,1]^d) &= \mathcal{B}(\mathbb{R}^d) \cap (0,1]^d \\
 &= \{B \in \mathcal{B}(\mathbb{R}^d) : B \subset (0,1]^d\} \\
 &= \sigma\langle (a_1, b_1] \times \cdots \times (a_d, b_d] : 0 \leq a_k < b_k \leq 1 \rangle \\
 &= \sigma\langle \mathcal{B}_0((0,1]^d) \rangle.
 \end{aligned}$$

onde $\mathcal{B}_0((0,1]^d) := \{ (a_1, b_1] \times \cdots \times (a_d, b_d] : 0 \leq a_k < b_k \leq 1 \}$ é a álgebra de Borel em $(0,1]^d$.

A parte mais importante do teorema acima é que $\mathcal{B}((0,1]^d) = \sigma\langle \mathcal{B}_0((0,1]^d) \rangle$.

Esta caracterização permite construir probabilidades em $\mathcal{B}_0((0,1]^d)$, então utilizar o Teorema da Extensão de Carathéodory para estender essas probabilidades para um modelo de probabilidade total em $\mathcal{B}((0,1]^d)$.

Demonstração. Denote por $\mathcal{H}$ a família dos ortantes inferiores

$$(-\infty, c_1] \times \cdots \times (-\infty, c_d], \quad c_1, \dots, c_d \in \mathbb{R}.$$

Cada elemento de $\mathcal{H}$ é fechado, logo $\sigma\langle \mathcal{H} \rangle \subseteq \mathcal{B}(\mathbb{R}^d)$. Para a inclusão inversa, fixe $j \in \{1, \dots, d\}$ e $c \in \mathbb{R}$. O semiespaço coordenado inferior pode ser escrito como

$$\{x \in \mathbb{R}^d : x_j \leq c\} = \bigcup_{m=1}^{\infty} \{x \in \mathbb{R}^d : x_j \leq c \text{ e } x_i \leq m \text{ para } i \neq j\}.$$

Portanto, todos esses semiespaços pertencem a $\sigma\langle \mathcal{H} \rangle$. Diferenças e interseções finitas dão

$$(a_1, b_1] \times \cdots \times (a_d, b_d] = \bigcap_{j=1}^d (\{x : x_j \leq b_j\} \setminus \{x : x_j \leq a_j\}).$$

Os retângulos semiabertos desse tipo geram todos os retângulos abertos e, como os retângulos abertos com extremos racionais formam uma base enumerável da topologia euclidiana, geram $\mathcal{B}(\mathbb{R}^d)$. Assim, $\sigma\langle \mathcal{H} \rangle = \mathcal{B}(\mathbb{R}^d)$.

As bolas abertas também formam uma base: mais precisamente, as bolas com centro em $\mathbb{Q}^d$ e raio racional formam uma base enumerável. Logo todo aberto é união enumerável dessas bolas. Como toda bola aberta é boreliana, segue que as bolas abertas geram $\mathcal{B}(\mathbb{R}^d)$. Por definição, o mesmo vale para a família de todos os abertos. Os fechados geram a mesma σ -álgebra, pois são exatamente os complementos dos abertos.

Todo compacto de $\mathbb{R}^d$ é fechado, de modo que a σ -álgebra gerada pelos compactos está contida em $\mathcal{B}(\mathbb{R}^d)$. Reciprocamente, se F é fechado, então

$$F = \bigcup_{m=1}^{\infty} (F \cap [-m, m]^d),$$

e cada conjunto da união é compacto. Portanto os compactos geram todos os fechados e, conseqüentemente, toda a σ -álgebra de Borel.

Todo retângulo é boreliano, enquanto os retângulos abertos com extremos racionais formam uma base enumerável; portanto os retângulos também geram $\mathcal{B}(\mathbb{R}^d)$. Todo cilindro coordenado é boreliano, pois a projeção π_j é contínua. Reciprocamente, todo retângulo é uma interseção finita de cilindros coordenados. Isso prova também a igualdade relativa aos cilindros.

Passemos a $S = (0,1]$. Pelo Teorema da Restrição de Borel e pelo caso unidimensional já demonstrado, $\mathcal{B}(S)$ é gerada pelos conjuntos $(0,c]$, $0 < c < 1$. Como

$$(a,b] = (0,b] \setminus (0,a],$$

a família dos intervalos semiabertos $(a,b]$, $0 \leq a < b \leq 1$, gera a mesma σ -álgebra. A álgebra $\mathcal{B}_0(S)$ é, por definição, a álgebra gerada por esses intervalos; aplicar primeiro um número finito de operações e depois operações enumeráveis não muda a σ -álgebra gerada. Logo $\sigma\langle \mathcal{B}_0(S) \rangle = \mathcal{B}(S)$.

Os abertos de S geram $\mathcal{B}(S)$ por definição, e os fechados geram a mesma classe por complementação. Para conferir diretamente as duas famílias de intervalos que aparecem no enunciado, fixe $0 < c < 1$. Escolha seqüências $a_n \downarrow 0$ e $b_n \downarrow c$ com $0 < a_n < c < b_n < 1$. Então

$$(0,c] = \bigcap_{n=1}^{\infty} (0,b_n) = \bigcap_{n=1}^{\infty} \bigcup_{m=1}^{\infty} (a_m, b_n).$$

Descartando os poucos pares com $a_m \geq b_n$, todos os intervalos da última expressão têm extremos estritamente entre 0 e 1. Assim, esses intervalos abertos geram $(0,c]$. Por outro lado,

$$(0,c] = \bigcup_{n: 1/n < c} [1/n, c],$$

de modo que os intervalos fechados com extremos interiores também geram $(0,c]$. Como todos os intervalos mencionados são borelianos, ficam provadas todas as igualdades relativas a S .

Finalmente, tome $S_d = (0,1]^d$. Pelo Teorema da Restrição de Borel,

$$\mathcal{B}(S_d) = \mathcal{B}(\mathbb{R}^d)|_{S_d}.$$

Como S_d é boreliano, esse traço é também $\{B \in \mathcal{B}(\mathbb{R}^d) : B \subseteq S_d\}$. Intersectando os ortantes inferiores com S_d , vemos que $\mathcal{B}(S_d)$ é gerada pelos retângulos $(0,b_1] \times \cdots \times$

$(0, b_d]$. Diferenças finitas em cada coordenada mostram que ela é igualmente gerada por todos os retângulos $(a_1, b_1] \times \cdots \times (a_d, b_d]$ do enunciado. A álgebra gerada por esses retângulos é $\mathcal{B}_0(S_d)$, e portanto sua σ -álgebra gerada é novamente $\mathcal{B}(S_d)$. $\square$

A σ -álgebra de Borel é vasta, mas não contém todos os subconjuntos de $\mathbb{R}$. Exibir diretamente um não boreliano não é simples; a medida de Lebesgue fornecerá uma maneira eficiente de provar que eles existem.

Exemplo 1.37 — .

Conjuntos enumeráveis, co-enumeráveis (i.e. complementos de conjuntos enumeráveis), e conjuntos perfeitos de $(0, 1]$ estão em $\mathcal{B}((0, 1])$. Em particular, o conjunto de números irracionais de $(0, 1]$ é um conjunto de Borel.



A mesma construção define a σ -álgebra de Borel em qualquer espaço topológico.

Nesse contexto é interessante o seguinte teorema. Um isomorfismo de espaços mensuráveis é uma bijeção mensurável cuja inversa também é mensurável.

Teorema 1.38 — Classificação dos espaços de Borel poloneses

Seja Ω um espaço polonês, isto é, um espaço cuja topologia é induzida por alguma métrica completa e separável. Então o espaço mensurável $(\Omega, \mathcal{B}(\Omega))$ é isomorfo a exatamente um dos seguintes tipos:

- a** um espaço finito, se Ω é finito;
- b** $\mathbb{Z}$, se Ω é infinito enumerável;
- c** $\mathbb{R}$, se Ω é não enumerável.

Demonstração. Se Ω é finito ou enumerável, todo subconjunto de Ω é uma união enumerável de singletons fechados e, portanto, é boreliano. Nesses dois casos, qualquer bijeção com um conjunto finito da mesma cardinalidade ou com $\mathbb{Z}$, respectivamente, é um isomorfismo mensurável.

Suponha agora que Ω seja não enumerável. Usaremos o seguinte lema de Lusin–Souslin: se X e Y são espaços poloneses e $f : X \rightarrow Y$ é uma função boreliana injetiva, então $f(X)$ é boreliano em Y e $f : X \rightarrow f(X)$ é um isomorfismo mensurável. O mesmo resultado vale quando X é um subconjunto boreliano de um espaço polonês. Esse lema garante, em particular, que imagens de borelianos por aplicações borelianas injetivas são borelianas.

Primeiro construiremos uma injeção boreliana de Ω em $\mathbb{R}$. Seja $(U_n)_{n \geq 1}$ uma base enumerável da topologia de Ω e defina

$$\Phi(x) = (\mathbb{1}_{U_n}(x))_{n \geq 1} \in \{0,1\}^{\mathbb{N}}.$$

Cada coordenada de Φ é boreliana e a base separa pontos, logo Φ é boreliana e injetiva. A aplicação

$$(a_n)_{n \geq 1} \mapsto \sum_{n=1}^{\infty} \frac{2a_n}{3^n}$$

é uma injeção contínua do espaço de Cantor $\{0,1\}^{\mathbb{N}}$ em $[0,1]$. Sua composição com Φ fornece a injeção desejada.

Construiremos agora uma injeção boreliana de $\mathbb{R}$ em Ω . Considere

$$C = \{x \in \Omega : \text{toda vizinhança aberta de } x \text{ é não enumerável}\}.$$

O complementar de C é a união dos elementos enumeráveis da base (U_n) ; portanto é aberto e enumerável. Assim, C é fechado, não vazio e sem pontos isolados. Como é um subespaço fechado de um espaço completamente metrizável, C é completamente metrizável.

Construindo recursivamente, dentro de cada aberto relativo não vazio de C , dois abertos relativos com fechos disjuntos e diâmetros menores que 2^{-n} no nível n , obtemos conjuntos $(V_s)_{s \in \{0,1\}^{<\mathbb{N}}}$ tais que

$$\overline{V}_{s_0} \cup \overline{V}_{s_1} \subseteq V_s, \quad \overline{V}_{s_0} \cap \overline{V}_{s_1} = \emptyset, \quad \text{diam}(V_s) \leq 2^{-|s|}.$$

A completude implica que, para cada sequência $\alpha \in \{0,1\}^{\mathbb{N}}$, a interseção dos fechados $\overline{V}_{\alpha|n}$ contém um único ponto. A aplicação que associa α a esse ponto é contínua e injetiva. Por outro lado, compondo um homeomorfismo de $\mathbb{R}$ em $(0,1)$ com a expansão binária escolhida de modo a não terminar em uma cauda infinita de algarismos 1, obtemos uma injeção boreliana de $\mathbb{R}$ em $\{0,1\}^{\mathbb{N}}$. Logo existe uma injeção boreliana $g : \mathbb{R} \rightarrow \Omega$.

Para concluir, aplicamos a versão boreliana do argumento de Cantor–Schröder–Bernstein. Se $f : \Omega \rightarrow \mathbb{R}$ e $g : \mathbb{R} \rightarrow \Omega$ são as injeções construídas, ponha

$$A_0 = \Omega \setminus g(\mathbb{R}), \quad A_{n+1} = g(f(A_n)), \quad A = \bigcup_{n=0}^{\infty} A_n.$$

Pelo lema de Lusin–Souslin, todos esses conjuntos são borelianos. A função

$$h(x) = \begin{cases} f(x), & x \in A, \\ g^{-1}(x), & x \notin A, \end{cases}$$

é a bijeção usual do argumento de Cantor–Schröder–Bernstein. As duas partes são borelianas, assim como as partes correspondentes de sua inversa; logo h é um isomorfismo entre $(\Omega, \mathcal{B}(\Omega))$ e $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. □

EXERCÍCIOS

Ex. 1.1 — Sejam $\mathcal{A}_1$ e $\mathcal{A}_2$ duas álgebras de subconjuntos do mesmo conjunto Ω . Prove que $\mathcal{A}_1 \cap \mathcal{A}_2$ também é uma álgebra de subconjuntos de Ω .

Ex. 1.2 — Considere $\Omega = (0,1]$ e a família

$$\mathcal{A} = \{\emptyset\} \cup \left\{ \bigcup_{i=1}^n (a_i, b_i] : n \geq 1, 0 \leq a_i < b_i \leq 1, \right. \\ \left. \text{e os intervalos } (a_i, b_i] \text{ são dois a dois disjuntos} \right\}.$$

Prove que $\mathcal{A}$ é uma álgebra de subconjuntos de Ω .

Ex. 1.3 — Prove que a família

$$\mathcal{A} = \{A \subseteq \Omega : A \text{ é finito ou } A^c \text{ é finito}\}$$

é uma álgebra de subconjuntos de Ω .

Ex. 1.4 — Suponha que Ω seja infinito. Mostre que a álgebra

$$\mathcal{A} = \{A \subseteq \Omega : A \text{ é finito ou } A^c \text{ é finito}\}$$

não é uma σ -álgebra. Construa explicitamente uma sequência $(A_n)_{n \geq 1}$ em $\mathcal{A}$ cuja união não pertence a $\mathcal{A}$.

Ex. 1.5 — Suponha que Ω seja infinito e não enumerável. Prove que

$$\mathcal{F} = \{A \subseteq \Omega : A \text{ é enumerável ou } A^c \text{ é enumerável}\}$$

é uma σ -álgebra de subconjuntos de Ω .

Ex. 1.6 — Sejam $(A_n)_{n \geq 1}$ conjuntos de Ω .

1. Suponha que $A_n \uparrow A$, isto é, $A_1 \subseteq A_2 \subseteq \dots$ e $A = \bigcup_{n \geq 1} A_n$. Prove que

$$\liminf_{n \rightarrow \infty} A_n = \limsup_{n \rightarrow \infty} A_n = \bigcup_{n=1}^{\infty} A_n = A.$$

2. Suponha que $A_n \downarrow A$, isto é, $A_1 \supseteq A_2 \supseteq \dots$ e $A = \bigcap_{n \geq 1} A_n$. Prove que

$$\liminf_{n \rightarrow \infty} A_n = \limsup_{n \rightarrow \infty} A_n = \bigcap_{n=1}^{\infty} A_n = A.$$

Ex. 1.7 — Conjunto de Cantor. Seja $C \subseteq [0,1]$ o conjunto de Cantor. Prove que C é fechado e, portanto, boreliano. Em seguida, usando representações na base 3, construa uma injeção de $\{0,1\}^{\mathbb{N}}$ em C e conclua que C não é enumerável.

Ex. 1.8 — Prove o *Princípio dos Bons Conjuntos*: se $\mathcal{C} \subseteq \mathcal{G} \subseteq \mathcal{P}(\Omega)$ e $\mathcal{G}$ é uma σ -álgebra, então

$$\sigma\langle \mathcal{C} \rangle \subseteq \mathcal{G}.$$

Indique precisamente onde é usada a minimalidade de $\sigma\langle \mathcal{C} \rangle$.

Ex. 1.9 — Conjunto de Vitali no plano. Seja $\mathcal{F}$ uma σ -álgebra de subconjuntos de $\mathbb{R}^2$ que contém os borelianos, e seja μ uma medida não nula em $(\mathbb{R}^2, \mathcal{F})$, invariante por translações, tal que

$$0 < \mu([0,1]^2) < \infty.$$

Adapte a construção de Vitali, usando translações por vetores de $\mathbb{Q}^2$, para provar que existe um conjunto $V \subseteq [0,1]^2$ que não pertence a $\mathcal{F}$.

Ex. 1.10 — Prove novamente o Teorema da Classe Monótona de Halmos 1.29. Mais precisamente, dada uma álgebra $\mathcal{F}_0$, escreva $\mathcal{M} = \mathcal{M}\langle\mathcal{F}_0\rangle$ e:

1. prove primeiro que $\mathcal{M} \subseteq \sigma\langle\mathcal{F}_0\rangle$;
2. mostre, por um argumento de classe monótona, que $A \in \mathcal{M}$ implica $A^c \in \mathcal{M}$;
3. use duas vezes o mesmo tipo de argumento para provar que $A, B \in \mathcal{M}$ implica $A \cup B \in \mathcal{M}$;
4. conclua, pelo Teorema 1.22, que $\mathcal{M}\langle\mathcal{F}_0\rangle = \sigma\langle\mathcal{F}_0\rangle$.

Ex. 1.11 — Considere as propriedades seguintes para uma família $\mathcal{L} \subseteq \mathcal{P}(\Omega)$:

$$(\lambda_1) \quad \Omega \in \mathcal{L};$$

$$(\lambda_2) \quad A \in \mathcal{L} \implies A^c \in \mathcal{L};$$

$$(\lambda_3) \quad (A_i)_{i \in I} \subseteq \mathcal{L} \text{ dois a dois disjuntos, } I \text{ não vazio, finito ou enumerável} \implies \bigcup_{i \in I} A_i \in \mathcal{L};$$

$$(\lambda'_2) \quad A, B \in \mathcal{L}, A \subseteq B \implies B \setminus A \in \mathcal{L};$$

$$(\lambda_4) \quad A, B \in \mathcal{L}, A \cap B = \emptyset \implies A \cup B \in \mathcal{L};$$

$$(\lambda_5) \quad A_n \uparrow A, A_n \in \mathcal{L} \implies A \in \mathcal{L};$$

$$(\lambda_6) \quad A_n \downarrow A, A_n \in \mathcal{L} \implies A \in \mathcal{L}.$$

Prove as afirmações abaixo.

1. $\mathcal{L}$ é um λ -sistema se, e somente se, satisfaz (λ_1) , (λ'_2) e (λ_3) .
2. Todo λ -sistema satisfaz também (λ_4) , (λ_5) e (λ_6) .
3. $\mathcal{L}$ é um λ -sistema se, e somente se, satisfaz (λ_1) , (λ'_2) e (λ_6) .
4. Se $\mathcal{L}$ é não vazia e satisfaz (λ_2) e (λ_3) , então $\mathcal{L}$ é um λ -sistema.

Ex. 1.12 — Seja $\emptyset \neq \mathcal{A} \subseteq \mathcal{P}(\Omega)$. Defina

$$\begin{aligned}\mathcal{C} &:= \mathcal{A} \cup \{A^c : A \in \mathcal{A}\}, \\ \mathcal{F} &:= \left\{ \bigcap_{j=1}^n C_j : n \geq 1, C_j \in \mathcal{C} \right\}, \\ \mathcal{U} &:= \left\{ \bigcup_{j=1}^n I_j : n \geq 1, I_j \in \mathcal{F} \right\}.\end{aligned}$$

Prove que a álgebra gerada por $\mathcal{A}$ é $\mathcal{U}$, isto é,

$$f\langle \mathcal{A} \rangle = \mathcal{U}.$$

Ao longo da prova, explique como $\emptyset$ e Ω aparecem em $\mathcal{U}$.

Ex. 1.13 — Seja $S = (0,1]$ e

$$\mathcal{B}_0(S) = f\langle (a,b] : 0 \leq a < b \leq 1 \rangle.$$

Usando o exercício anterior, prove que todo elemento de $\mathcal{B}_0(S)$ pode ser escrito como união finita de intervalos semiabertos $(a_i, b_i]$. Mostre ainda que essa representação pode ser escolhida com os intervalos dois a dois disjuntos.

Ex. 1.14 — Um dado lançado até o primeiro seis. Um dado equilibrado é lançado repetidamente até aparecer a face 6. O experimento deve contemplar também a possibilidade de que isso nunca ocorra. Descreva um espaço amostral que registre exatamente a trajetória observada: uma palavra finita terminada em 6, quando o primeiro 6 aparece, ou uma trajetória infinita, quando ele nunca aparece. Não atribua probabilidades ainda; essa etapa será retomada no capítulo seguinte.

Ex. 1.15 — Três modelos de ocupação. Distribua k partículas por n células distinguíveis. Descreva o espaço de configurações e calcule sua cardinalidade em cada um dos modelos:

1. Maxwell–Boltzmann: partículas distinguíveis e ocupação múltipla permitida;
2. Bose–Einstein: partículas indistinguíveis e ocupação múltipla permitida;

3. Fermi–Dirac: partículas indistinguíveis e no máximo uma partícula por célula. No último caso, trate separadamente a possibilidade $k > n$.

Ex. 1.16 — Três espaços amostrais para uma urna colorida. Uma urna contém $N = n_1 + \cdots + n_r$ bolas fisicamente distintas, das quais n_i têm a cor i . Retiram-se $k \leq N$ bolas sem reposição.

1. Descreva e conte os resultados quando se registra a ordem e a identidade das bolas.
2. Faça o mesmo quando se registram as identidades, mas não a ordem.
3. Se somente as quantidades retiradas de cada cor são registradas, descreva o espaço de resultados e determine a probabilidade de cada vetor de contagens sob amostragem uniforme.

Medidas de Probabilidade

No capítulo anterior vimos por que nem todo subconjunto pode receber uma probabilidade. Agora invertemos a pergunta: dada uma atribuição de massa em uma família simples de conjuntos, quando ela determina uma medida de probabilidade?

O caminho passa por três níveis. Começaremos com probabilidades finitamente aditivas, identificaremos as condições que as transformam em pré-medidas e usaremos o Teorema de Extensão de Carathéodory para obter medidas nas σ -álgebras geradas. A dificuldade não está em distribuir massa finitamente; começa quando os limites entram em cena.

Ao final, aplicaremos esse mecanismo à construção de modelos fundamentais, entre eles a medida de Lebesgue e o espaço de infinitos lançamentos de uma moeda.

Definição 2.1 – Probabilidade Finitamente Aditiva

Se $\mathcal{F}_0$ é uma álgebra em Ω , então $\mathbf{P} : \mathcal{F}_0 \rightarrow [0,1]$ é dita uma **probabilidade finitamente aditiva** ou **atribuição de probabilidade** em $\mathcal{F}_0$ se

$$\boxed{1} \quad \mathbf{P}[\Omega] = 1$$

$$\boxed{2} \quad \mathbf{P}[A \cup B] = \mathbf{P}[A] + \mathbf{P}[B] \text{ para todos os conjuntos } A, B \in \mathcal{F}_0 \text{ disjuntos.}$$

A aditividade finita é o primeiro nível da construção. Em uma álgebra, podemos exigir também aditividade enumerável sempre que a união ainda pertença à álgebra.

Definição 2.2 – Pré-Medida de Probabilidade

Se $\mathcal{F}_0$ é uma álgebra em Ω , então $\mathbf{P} : \mathcal{F}_0 \rightarrow [0,1]$ é dita uma **pré-medida de probabilidade** em $\mathcal{F}_0$ se

$$\boxed{1} \quad \mathbf{P}(\Omega) = 1$$

$$\boxed{2} \quad \mathbf{P}\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} \mathbf{P}(A_k) \text{ para todos conjuntos disjuntos } A_1, A_2, \dots \in \mathcal{F}_0 \text{ tais que } \bigcup_{k=1}^{\infty} A_k \in \mathcal{F}_0.$$

Definição 2.3 –

Uma **medida** em $(\Omega, \mathcal{F})$ é uma aplicação $\mu : \mathcal{F} \rightarrow [0, +\infty]$ que satisfaz:

$$\boxed{1} \quad \mu(\emptyset) = 0$$

$$\boxed{2} \quad \text{para toda sequência de eventos disjuntos } A_1, A_2, \dots \in \mathcal{F}, \text{ temos que}$$

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n).$$

Essa propriedade é denominada de σ – **aditividade**.

Se nas Definições 2.1 e 2.2 removermos a hipótese $\mathbf{P}(\Omega) = 1$, obtemos uma medida finitamente aditiva e uma pré-medida, respectivamente. No caso geral, quando permitimos valores em $[0, \infty]$, incluímos na definição de pré-medida a condição $\mu_0(\emptyset) = 0$. Para pré-medidas de probabilidade essa condição é automática.

Definição 2.4 —

Se $\mathbf{P}$ é uma medida em $(\Omega, \mathcal{F})$ tal que $\mathbf{P}(\Omega) = 1$, então $\mathbf{P}$ é denominada de **medida de probabilidade**.

Nesse caso a tripla $(\Omega, \mathcal{F}, \mathbf{P})$ é denominada de **espaço de probabilidade**.

A definição é abstrata, mas inclui os experimentos aleatórios mais familiares: lançamentos de moedas e dados, sorteios em urnas e muitos outros. Em cada caso, o que muda é o espaço amostral e a família de eventos; a medida de probabilidade é a regra que atribui massa a esses eventos.

A representação seguinte explicita os três ingredientes do modelo: o espaço amostral Ω , a σ -álgebra $\mathcal{F}$ e a medida $\mathbf{P}$.

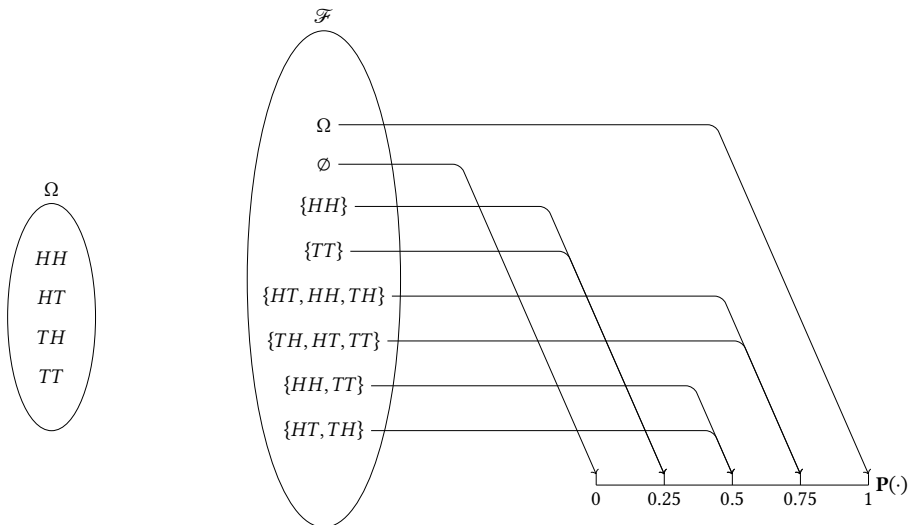


Figura 2.1.: Relação entre o espaço amostral, σ -álgebra e a medida de probabilidade.

Definição 2.5 — Medidas Finitas e σ -finitas

Dado $(\Omega, \mathcal{F}, \mu)$ um espaço de medida.

- a** Se $\mu(\Omega) < \infty$ então μ é dita **medida finita**;
- b** Se $\mu(\Omega) = \infty$ então μ é dita **medida infinita**;
- c** Se existem conjuntos em $\mathcal{F}$ $A_1, A_2, \dots$ tais que $\Omega = \bigcup_{k=1}^{\infty} A_k$ e $\mu(A_k) < \infty$ então μ é dita medida **σ -finita**

Note que toda medida de probabilidade é σ -finita.

Exemplo 2.6 — Espaços de Probabilidade Discretos.

Seja Ω um conjunto enumerável, isto é, finito ou infinito enumerável. Seja $\mathcal{F}$ o conjunto de todos os subconjuntos de Ω e seja uma função $p : \Omega \rightarrow [0,1]$ satisfazendo $\sum_{\omega \in \Omega} p(\omega) = 1$. Definimos então

$$\mathbf{P}(A) = \sum_{\omega \in A} p(\omega) \text{ com } p(\omega) \geq 0 \text{ e } \sum_{\omega \in \Omega} p(\omega) = 1.$$

Esta é a medida de probabilidade mais geral que podemos construir neste espaço.

Quando Ω é um conjunto finito, e $p(\omega) = \frac{1}{|\Omega|}$ onde $|\Omega|$ denota o número de pontos em Ω , temos o que se denomina espaço de probabilidade finito equiprovável.

◀

Muitos modelos dos cursos introdutórios são casos particulares dessa construção.

Exemplo 2.7 — .

Seja $\Omega = \mathbb{N} \cup \{0\}$ e dado $\lambda > 0$ e definida $p : \Omega \rightarrow [0,1]$ por

$$p(k) = e^{-\lambda} \frac{\lambda^k}{k!}.$$

Como

$$\sum_{k=0}^{\infty} p(k) = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} \cdot e^{\lambda} = 1,$$

a medida

$$\mathbf{P}(A) = \sum_{k \in A} p(k),$$

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



Exemplo 2.8 — Lançamento de Moedas.

Vamos considerar o experimento de lançar n moedas. Representando *coroa* por 0 e *cara* por 1, podemos definir o espaço amostral por

$$\Omega = \{0, 1\}^n = \{\omega = (a_1, a_2, \dots, a_n) : a_k \in \{0, 1\}, k = 1, \dots, n\}.$$

Dado $\omega = (a_1, a_2, \dots, a_n) \in \Omega$, defina $N(\omega) = a_1 + \dots + a_n$ ($N(\omega)$ representa o total de 1's em ω)

Fixe agora $q \in (0, 1)$ e defina

$$p(\omega) = q^{N(\omega)}(1 - q)^{n - N(\omega)}.$$

Fica como exercício mostrar que

$$\sum_{\omega \in \Omega} p(\omega) = 1,$$

e portanto pode ser usada para definir um espaço de probabilidade.



As primeiras consequências da definição serão usadas ao longo de todo o texto.

Proposição 2.9 — Propriedades da Probabilidade

1 $\mathbf{P}(A^c) = 1 - \mathbf{P}(A)$.

2 $A \subseteq B$ implica $\mathbf{P}(A) \leq \mathbf{P}(B)$. (**monotonicidade**)

3 Dados eventos A e B então $\mathbf{P}(A \cup B) \leq \mathbf{P}(A) + \mathbf{P}(B)$ (**sub-aditividade**)

$$\boxed{4} \quad \mathbf{P}\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \mathbf{P}(A_n) \text{ } (\sigma\text{-sub-aditividade})$$

$$\boxed{5} \quad \mathbf{P}[A \cup B] = \mathbf{P}[A] + \mathbf{P}[B] - \mathbf{P}[A \cap B]; \text{ (Inclusão-exclusão)}$$

$$\boxed{6} \quad \mathbf{P}(A \triangle B) \geq \max\{\mathbf{P}(A - B), \mathbf{P}(B - A)\}$$

Demonstração.

$\boxed{1}$ $\Omega = A \cup A^c$. Logo, $\mathbf{P}(\Omega) = \mathbf{P}(A) + \mathbf{P}(A^c)$. Agora basta utilizar que $\mathbf{P}(\Omega) = 1$.

$\boxed{2}$ Use a decomposição $B = A \cup (B \setminus A)$.

$\boxed{3}$ Note que $A \cup B$ pode ser escrito como $A \cup (B \setminus A)$. Logo, pela monotonicidade da probabilidade, temos que $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B \setminus A) \leq \mathbf{P}(A) + \mathbf{P}(B)$.

$\boxed{4}$ Defina $B_1 = A_1$ e $B_n = A_n \setminus \bigcup_{k=1}^{n-1} A_k$ para $n \geq 2$. Os eventos B_n são disjuntos, $B_n \subseteq A_n$ e $\bigcup_n B_n = \bigcup_n A_n$. Pela σ -aditividade e pela monotonicidade,

$$\mathbf{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathbf{P}(B_n) \leq \sum_{n=1}^{\infty} \mathbf{P}(A_n).$$

$\boxed{5}$ A decomposição disjunta $A \cup B = (A \setminus B) \dot{\cup} (B \setminus A) \dot{\cup} (A \cap B)$ e as decomposições correspondentes de A e B fornecem

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

$\boxed{6}$ Como $A \triangle B = (A \setminus B) \dot{\cup} (B \setminus A)$,

$$\mathbf{P}(A \triangle B) = \mathbf{P}(A \setminus B) + \mathbf{P}(B \setminus A),$$

e essa soma é maior ou igual a cada uma das duas parcelas.

□

2.1. Continuidade das Medidas de Probabilidade

A σ -aditividade também pode ser lida como compatibilidade com limites de seqüências monótonas de eventos. Essa forma de continuidade será uma de nossas ferramentas principais.

Teorema 2.10 — Continuidade da Probabilidade

Dado uma sequência de eventos $A_1, A_2, \dots \in \mathcal{F}$. Então

- 1 Se $A_n \uparrow A$, então $\mathbf{P}(A_n) \uparrow \mathbf{P}(A)$ quando $n \rightarrow \infty$.
- 2 Se $A_n \downarrow A$, então $\mathbf{P}(A_n) \downarrow \mathbf{P}(A)$ quando $n \rightarrow \infty$.

Demonstração. Começemos pela continuidade por baixo. Dada uma sequência crescente $A_1 \subseteq A_2 \subseteq \dots$, defina $B_1 = A_1$ e, para $k \geq 2$, $B_k = A_k \setminus A_{k-1}$.

Os conjuntos B_n são disjuntos e $\bigcup_{k=1}^j B_k = A_j$ para todo $j \geq 1$. Além disso,

$$A = \bigcup_{k=1}^{\infty} B_k.$$

Portanto, pela aditividade finita e pela σ -aditividade,

$$\mathbf{P}(A_n) = \sum_{k=1}^n \mathbf{P}(B_k) \uparrow \sum_{k=1}^{\infty} \mathbf{P}(B_k) = \mathbf{P}(A).$$

Para a continuidade por cima, se $A_n \downarrow A$, então $A_n^c \uparrow A^c$. Pela primeira parte,

$$\mathbf{P}(A_n) = 1 - \mathbf{P}(A_n^c) \downarrow 1 - \mathbf{P}(A^c) = \mathbf{P}(A).$$

□

Teorema 2.11 —

Dado $A_1, A_2, \dots$ uma sequência de eventos. Então

- 1 $\mathbf{P}(\liminf A_n) \leq \liminf \mathbf{P}(A_n) \leq \limsup \mathbf{P}(A_n) \leq \mathbf{P}(\limsup A_n)$.
- 2 Se $A_n \rightarrow A$ então $\mathbf{P}(A_n) \rightarrow \mathbf{P}(A)$ quando $n \rightarrow \infty$.

Demonstração. Defina

$$B_n = \bigcap_{m=n}^{\infty} A_m \quad \text{e} \quad C_n = \bigcup_{m=n}^{\infty} A_m.$$

Então $B_n \uparrow \liminf_n A_n$, $C_n \downarrow \limsup_n A_n$ e $B_n \subseteq A_n \subseteq C_n$. Pela monotonicidade e pela continuidade da probabilidade,

$$\mathbf{P}(B_n) \leq \mathbf{P}(A_n) \leq \mathbf{P}(C_n),$$

e

$$\mathbf{P}(B_n) \longrightarrow \mathbf{P}(\liminf_n A_n), \quad \mathbf{P}(C_n) \longrightarrow \mathbf{P}(\limsup_n A_n).$$

Tomando limite inferior e limite superior na primeira desigualdade, obtemos

$$\mathbf{P}(\liminf_n A_n) \leq \liminf_n \mathbf{P}(A_n) \leq \limsup_n \mathbf{P}(A_n) \leq \mathbf{P}(\limsup_n A_n).$$

Se $A_n \rightarrow A$, então $\liminf_n A_n = \limsup_n A_n = A$, e as desigualdades anteriores implicam $\mathbf{P}(A_n) \rightarrow \mathbf{P}(A)$. □

Teorema 2.12 — Caracterizações da σ -aditividade

Dado $\mathbf{P} : \mathcal{F}_0 \rightarrow [0,1]$ uma probabilidade finitamente aditiva numa álgebra $\mathcal{F}_0$. Então as seguintes afirmações são equivalentes

- 1** $\mathbf{P}$ é uma pré-medida de probabilidade;
- 2** $\mathbf{P}$ é contínua por baixo. Isto é dados $A_1, A_2, \dots \in \mathcal{F}_0$ e $A_n \uparrow A \in \mathcal{F}_0$ então $\mathbf{P}(A_n) \uparrow \mathbf{P}(A)$;
- 3** $\mathbf{P}$ é contínua por cima. Isto é dados $A_1, A_2, \dots \in \mathcal{F}_0$ e $A_n \downarrow A \in \mathcal{F}_0$ então $\mathbf{P}(A_n) \downarrow \mathbf{P}(A)$;
- 4** Contínua por cima em $\emptyset$. Isto é dados $A_1, A_2, \dots \in \mathcal{F}_0$ e $A_n \downarrow \emptyset$ então $\mathbf{P}(A_n) \downarrow 0$.

Demonstração. Já provamos que $1. \implies 2.$ e é imediato que $3. \implies 4.$

($4. \implies 3.$) Suponha que $\mathbf{P}$ é contínua por cima em $\emptyset$ e que $A_1, A_2, \dots \in \mathcal{F}_0$ e $A_n \downarrow A \in \mathcal{F}_0$.

$$\begin{aligned} A_n \downarrow A &\implies A_n - A \downarrow \emptyset \\ &\implies \mathbf{P}[A_n] - \mathbf{P}[A] = \mathbf{P}[A_n - A] \downarrow 0, \text{ pelo item 4.} \\ &\implies \mathbf{P}[A_n] \downarrow \mathbf{P}[A] \end{aligned}$$

($2. \iff 3.$) Para provar esse fato basta observar que $A_n \uparrow A \iff A_n^c \downarrow A^c$ e que $\mathbf{P}[A] = 1 - \mathbf{P}[A^c]$.

($2. \implies 1.$) Suponha que $A_1, A_2, \dots$ sejam disjuntos em $\mathcal{F}_0$ e que $A = \bigcup_{k=1}^{\infty} A_k \in \mathcal{F}_0$. Os conjuntos $S_n = \bigcup_{k=1}^n A_k$ pertencem à álgebra e satisfazem $S_n \uparrow A$. Pela continuidade por baixo e pela aditividade finita,

$$\mathbf{P}(A) = \lim_{n \rightarrow \infty} \mathbf{P}(S_n) = \lim_{n \rightarrow \infty} \sum_{k=1}^n \mathbf{P}(A_k) = \sum_{k=1}^{\infty} \mathbf{P}(A_k).$$

Logo $\mathbf{P}$ é uma pré-medida de probabilidade. □

Lema 2.13 – Aproximação por uma álgebra

Seja $(\Omega, \mathcal{F}, \mathbf{P})$ um espaço de probabilidade e seja $\mathcal{A} \subseteq \mathcal{F}$ uma álgebra tal que $\sigma(\mathcal{A}) = \mathcal{F}$. Para todo $E \in \mathcal{F}$ e todo $\varepsilon > 0$, existe $A \in \mathcal{A}$ tal que

$$\mathbf{P}(E \triangle A) < \varepsilon.$$

Demonstração. Considere a família

$$\mathcal{M} = \{E \in \mathcal{F} : \text{para todo } \varepsilon > 0 \text{ existe } A \in \mathcal{A} \text{ com } \mathbf{P}(E \triangle A) < \varepsilon\}.$$

Claramente $\mathcal{A} \subseteq \mathcal{M}$. Se $E \in \mathcal{M}$ e $A \in \mathcal{A}$ aproxima E , então

$$E^c \triangle A^c = E \triangle A,$$

logo $\mathcal{M}$ é fechada por complementos.

Se $E_1, E_2, \dots \in \mathcal{M}$ e $E = \bigcup_{j \geq 1} E_j$, ponha $E^{(N)} = \bigcup_{j=1}^N E_j$. Pela continuidade por baixo, $\mathbf{P}(E \setminus E^{(N)}) \rightarrow 0$. Dado $\varepsilon > 0$, escolha N tal que $\mathbf{P}(E \setminus E^{(N)}) < \varepsilon/2$ e, para $1 \leq j \leq N$, escolha $A_j \in \mathcal{A}$ com

$$\mathbf{P}(E_j \triangle A_j) < \frac{\varepsilon}{2N}.$$

Como $A = \bigcup_{j=1}^N A_j \in \mathcal{A}$,

$$E \triangle A \subseteq (E \setminus E^{(N)}) \cup \bigcup_{j=1}^N (E_j \triangle A_j),$$

e, portanto, $\mathbf{P}(E \triangle A) < \varepsilon$. Assim $\mathcal{M}$ é uma σ -álgebra que contém $\mathcal{A}$. Consequentemente, $\mathcal{F} = \sigma\langle \mathcal{A} \rangle \subseteq \mathcal{M}$. □

Teorema 2.14 — Unicidade de Medidas

Dado $\mathcal{P}$ uma família de subconjuntos de Ω . Se μ e ν são duas medidas em $(\Omega, \sigma\langle \mathcal{P} \rangle)$ que concordam em $\mathcal{P}$, e $\mathcal{P}$ é um π -sistema, então elas concordam em toda a σ -álgebra gerada desde que valha uma das condições seguintes:

- 1 μ e ν são medidas de probabilidade;
- 2 existem $D_n \in \mathcal{P}$ tais que $D_n \uparrow \Omega$ e $\mu(D_n) = \nu(D_n) < \infty$ para todo n .

Demonstração. No caso de probabilidades, defina

$$\mathcal{G} := \{A \in \sigma\langle \mathcal{P} \rangle : \mu(A) = \nu(A)\}.$$

Como $\mu(\Omega) = \nu(\Omega) = 1$, a classe $\mathcal{G}$ contém Ω . Se $A \in \mathcal{G}$, então $\mu(A^c) = 1 - \mu(A) = 1 - \nu(A) = \nu(A^c)$. Finalmente, se $A_1, A_2, \dots \in \mathcal{G}$ são disjuntos, a σ -aditividade de ambas as medidas mostra que sua união pertence a $\mathcal{G}$. Portanto $\mathcal{G}$ é um λ -sistema que contém $\mathcal{P}$. Pelo Teorema π - λ de Dynkin, $\sigma\langle \mathcal{P} \rangle \subseteq \mathcal{G}$.

No segundo caso, fixe n e considere as medidas finitas $\mu_n(A) = \mu(A \cap D_n)$ e $\nu_n(A) = \nu(A \cap D_n)$. Elas têm a mesma massa total e concordam em $\mathcal{P}$, pois $A \cap D_n \in \mathcal{P}$. A demonstração do parágrafo anterior vale para medidas finitas com a mesma

massa total (se essa massa for positiva, basta normalizá-las; se for nula, ambas são identicamente nulas). Assim $\mu(A \cap D_n) = \nu(A \cap D_n)$ para todo $A \in \sigma\langle \mathcal{P} \rangle$. Como $D_n \uparrow \Omega$, a continuidade por baixo fornece $\mu(A) = \nu(A)$. $\square$

Definição 2.15 — Medida nula e μ -negligenciável

Dado $(\Omega, \mathcal{F}, \mu)$ um espaço de medida. Então

- a** Um conjunto $A \in \mathcal{F}$ é dito de medida nula se $\mu(A) = 0$.
- b** Um conjunto $A \in \mathcal{P}(\Omega)$ é dito μ -**negligenciável** se existe um conjunto μ -nulo $B \in \mathcal{F}$ tal que $A \subset B$.

Definição 2.16 — Espaços Completos

Um espaço de medida $(\Omega, \mathcal{F}, \mu)$ é dito **completo** se todos os conjuntos μ -negligenciáveis pertencem à $\mathcal{F}$.

Todo espaço de medida pode ser completado.

Teorema 2.17 — O completamento $(\Omega, \overline{\mathcal{F}}, \overline{\mu})$

Sejam $(\Omega, \mathcal{F}, \mu)$ um espaço de medida e $\mathcal{N}_\mu$ a família dos conjuntos μ -negligenciáveis.

Então

- a** $\overline{\mathcal{F}} := \sigma\langle \mathcal{F}, \mathcal{N}_\mu \rangle = \{F \cup N : F \in \mathcal{F}, N \in \mathcal{N}_\mu\}$;
- b** A função $\overline{\mu}$ em $\overline{\mathcal{F}}$ definida por $\overline{\mu}(F \cup N) = \mu(F)$ para $F \in \mathcal{F}$ e $N \in \mathcal{N}_\mu$ é a única extensão de μ para uma medida em $(\Omega, \overline{\mathcal{F}})$;
- c** O espaço de medida $(\Omega, \overline{\mathcal{F}}, \overline{\mu})$ é completo.

A tripla $(\Omega, \overline{\mathcal{F}}, \overline{\mu})$ é denominada **completamento** de $(\Omega, \mathcal{F}, \mu)$.

Demonstração. A família $\mathcal{N}_\mu$ é um σ -ideal: todo subconjunto de um conjunto negligenciável é negligenciável e, se $N_j \subseteq Z_j \in \mathcal{F}$ com $\mu(Z_j) = 0$, então $\bigcup_j N_j \subseteq \bigcup_j Z_j$ e $\mu(\bigcup_j Z_j) = 0$.

Seja $\mathcal{C} = \{F \cup N : F \in \mathcal{F}, N \in \mathcal{N}_\mu\}$. Essa classe contém $\mathcal{F}$ e $\mathcal{N}_\mu$ e é fechada por uniões enumeráveis. Para verificar complementos, escreva $N \subseteq Z \in \mathcal{F}$ com $\mu(Z) = 0$; então

$$(F \cup N)^c = (F^c \setminus Z) \cup (F^c \cap (Z \setminus N)),$$

onde o primeiro conjunto pertence a $\mathcal{F}$ e o segundo é negligenciável. Logo $\mathcal{C}$ é uma σ -álgebra e, pela minimalidade da σ -álgebra gerada, $\overline{\mathcal{F}} = \sigma\langle \mathcal{F}, \mathcal{N}_\mu \rangle = \mathcal{C}$.

A definição de $\bar{\mu}$ independe da representação. De fato, se $F \cup N = G \cup M$, com N e M negligenciáveis, então $F \setminus G \subseteq M$ e $G \setminus F \subseteq N$. Esses dois conjuntos têm medida zero, e portanto $\mu(F) = \mu(G)$.

Para provar a σ -aditividade, sejam $E_j = F_j \cup N_j$ disjuntos. As interseções $F_i \cap F_j$, $i \neq j$, são negligenciáveis. Defina $G_1 = F_1$ e $G_j = F_j \setminus \bigcup_{i < j} F_i$ para $j \geq 2$. Então os G_j são disjuntos, $\mu(G_j) = \mu(F_j)$, e $\bigcup_j E_j$ difere de $\bigcup_j G_j$ apenas por um conjunto negligenciável. Consequentemente,

$$\bar{\mu}\left(\bigcup_j E_j\right) = \mu\left(\bigcup_j G_j\right) = \sum_j \mu(G_j) = \sum_j \bar{\mu}(E_j).$$

Além disso, $\bar{\mu}(F) = \mu(F)$ para $F \in \mathcal{F}$.

Se $E = F \cup N \in \overline{\mathcal{F}}$ e $\bar{\mu}(E) = 0$, então $\mu(F) = 0$; logo todo subconjunto de E é negligenciável e pertence a $\overline{\mathcal{F}}$. Assim o espaço completado é completo. Por fim, qualquer extensão de μ a $\overline{\mathcal{F}}$ deve atribuir medida zero a cada $N \in \mathcal{N}_\mu$ e, portanto, deve atribuir a $F \cup N$ o valor $\mu(F)$. Isso prova a unicidade. $\square$

2.2. Teorema de Extensão de Carathéodory

O Teorema de Extensão de Carathéodory transforma uma pré-medida definida numa álgebra $\mathcal{F}_0$ em uma medida na σ -álgebra gerada por ela. É a ponte que permite começar com comprimentos de intervalos e terminar com a medida de Lebesgue nos borelianos.

Seguiremos a construção em três etapas: associar uma medida exterior à função inicial, identificar os conjuntos mensuráveis para essa medida exterior e restringi-la a eles. A exposição será feita para probabilidades e medidas finitas; o caso σ -finito exige apenas os cuidados indicados nos enunciados.

2.2.1. Medidas Exteriores

Partimos de uma função de conjuntos $\mathbf{P}$ definida apenas em uma família $\mathcal{A}$. Para atribuir um custo a um conjunto arbitrário, cobrimo-lo por elementos de $\mathcal{A}$ e tomamos o menor custo total possível. A fórmula funciona para quase toda função não negativa; as hipóteses de aditividade entram depois, quando queremos recuperar a função original e obter uma medida.

Definição 2.18 — Função Exterior Induzida

Considere $\mathcal{A}$ uma família qualquer de subconjuntos de um espaço Ω , e $\mathbf{P} : \mathcal{A} \rightarrow [0, \infty]$ ser uma função de conjunto não-negativa. Suponha $\emptyset \in \mathcal{A}$ e $\mathbf{P}(\emptyset) = 0$.

A **função exterior** induzida por $\mathbf{P}$ é a função:

$$\mathbf{P}^*(E) = \inf \left\{ \sum_{n=1}^{\infty} \mathbf{P}(A_n) \mid A_1, A_2, \dots \in \mathcal{A} \text{ cobrem } E \right\},$$

definida para todos os conjuntos $E \subseteq \Omega$.

As propriedades básicas de $\mathbf{P}^*$ motivam a definição seguinte.

Definição 2.19 — Medida Exterior

Uma função $\mathbf{Q} : \mathcal{P}(\Omega) \rightarrow [0, \infty]$ é uma **medida exterior** se:

- 1 $\mathbf{Q}(\emptyset) = 0$.
- 2 Monotonicidade $\mathbf{Q}(E) \leq \mathbf{Q}(F)$ quando $E \subseteq F \subseteq \Omega$.
- 3 Subaditividade enumerável Se $E_1, E_2, \dots \subseteq \Omega$, então $\mathbf{Q}(\bigcup_n E_n) \leq$

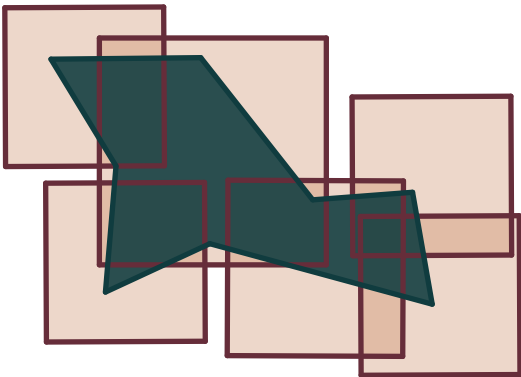


Figura 2.2.: Medida externa

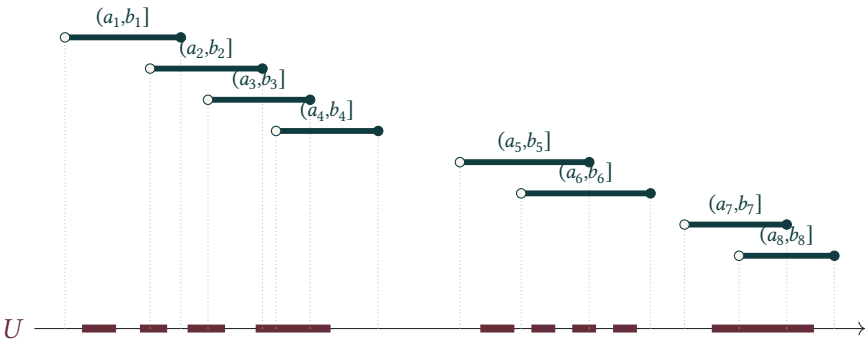


Figura 2.3.: Os intervalos $(a_1,b_1], \dots, (a_8,b_8]$ cobrem U . A extremidade esquerda é aberta e a direita é fechada.

$$\sum_n Q(E_n).$$

Teorema 2.20 — Propriedades da Função Exterior Induzida

A função exterior induzida $\mathbf{P}^*$ é uma medida exterior satisfazendo $\mathbf{P}^*(A) \leq \mathbf{P}(A)$ para todo $A \in \mathcal{A}$.

Demonstração. A cobertura vazia dá $\mathbf{P}^*(\emptyset) = 0$, e toda cobertura de um conjunto também cobre seus subconjuntos, o que prova a monotonicidade. Para $A \in \mathcal{A}$, o próprio conjunto A é uma cobertura admissível; logo $\mathbf{P}^*(A) \leq \mathbf{P}(A)$.

A propriedade 3 é demonstrada por argumentos de aproximação. Se $\sum_n \mathbf{P}^*(E_n) = +\infty$, a desigualdade é automática. Suponha, portanto, que essa soma seja finita. Seja $\varepsilon > 0$. Para cada E_n , pela definição de $\mathbf{P}^*$, existem conjuntos $\{A_{n,m}\}_m \subseteq \mathcal{A}$ cobrindo E_n , com

$$\sum_m \mathbf{P}(A_{n,m}) \leq \mathbf{P}^*(E_n) + \frac{\varepsilon}{2^n}.$$

Todos os conjuntos $A_{n,m}$ juntos cobrem $\bigcup_n E_n$, então temos

$$\mathbf{P}^*\left(\bigcup_n E_n\right) \leq \sum_{n,m} \mathbf{P}(A_{n,m}) = \sum_n \sum_m \mathbf{P}(A_{n,m}) \leq \sum_n \mathbf{P}^*(E_n) + \varepsilon.$$

Como $\varepsilon > 0$ é arbitrário, temos $\mathbf{P}^*\left(\bigcup_n E_n\right) \leq \sum_n \mathbf{P}^*(E_n)$. □

Precisamos agora identificar um domínio no qual $\mathbf{Q} = \mathbf{P}^*$ seja σ -aditiva. Em geral, precisaremos restringir o domínio, mas isso não basta: a família resultante deve ser uma σ -álgebra suficientemente ampla para o modelo que queremos construir.

O critério de Carathéodory fornece essa escolha de domínio para $\mathbf{Q}$:

Definição 2.21 — Conjuntos Mensuráveis para a Medida Exterior

Seja $\mathbf{Q} : \mathcal{P}(\Omega) \rightarrow [0, \infty]$ uma medida exterior. Então

$$\mathcal{M} = \{B \in \mathcal{P}(\Omega) \mid \mathbf{Q}(B \cap E) + \mathbf{Q}(B^c \cap E) = \mathbf{Q}(E) \quad \text{para todo } E \subseteq \Omega\}$$

é a família de **conjuntos mensuráveis para a medida exterior $\mathbf{Q}$** .

Aplicando a subaditividade de $\mathbf{Q}$, a seguinte definição é equivalente:

$$\mathcal{M} = \{B \in \mathcal{P}(\Omega) \mid \mathbf{Q}(B \cap E) + \mathbf{Q}(B^c \cap E) \leq \mathbf{Q}(E) \quad \text{para todo } E \subseteq \Omega\}.$$

Heurística 2.22 – [1]

Suponha que Q^* seja uma medida externa finita em Ω

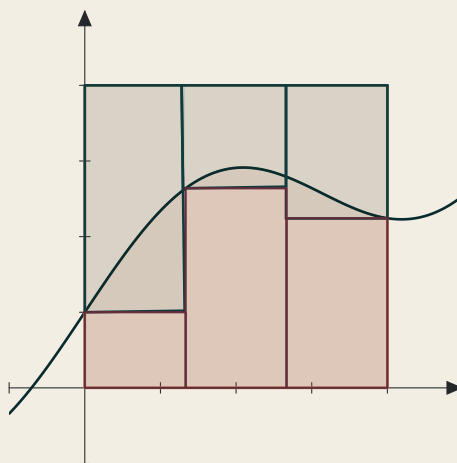
Podemos definir uma "medida interna" Q_* em Ω por

$$Q_*(A) = Q^*(\Omega) - Q^*(A^c)$$

Se a medida externa Q^* , for induzida a partir de uma medida σ -aditiva definida em alguma álgebra de conjuntos de Ω , então um subconjunto de Ω será mensurável no sentido de Carathéodory se e somente se sua medida externa e medida interna concordarem, veja o Exercício 2.2.

A partir deste ponto de vista, a construção da medida (bem como da σ -álgebra de conjuntos mensuráveis) é apenas uma generalização da construção natural da integral de Riemann em $\mathbb{R}$ - você tenta aproximar a área de um conjunto limitado E usando retângulos finitos e o conjunto é "mensurável no sentido de Riemann" se a melhor aproximação externa de sua área concorda com a melhor aproximação interna de sua área.

O ponto fundamental aqui é que o conceito de área interna é redundante e poderia ser definido em termos da área externa, como feito acima. Se a função é limitada, considere um retângulo contendo o conjunto abaixo dessa função e defina a medida interna como a medida externa do complemento do conjunto em relação a este retângulo.



A construção de Carathéodory não exige que $\mathbf{Q}^*$ seja finita. O argumento acima serve como motivação; a definição, porém, alcança também o caso geral.

A denominação “conjunto mensurável” é justificada pelo seguinte resultado sobre $\mathcal{M}$:

Teorema 2.23 —

A família $\mathcal{M}$ é uma σ -álgebra.

Demonstração. É imediato a partir da definição que $\mathcal{M}$ é fechado em relação ao complementar, e que $\emptyset \in \mathcal{M}$. Primeiro demostramos que $\mathcal{M}$ é fechado sob interseções finitas e, portanto, sob união finita. Considere $A, B \in \mathcal{M}$ então:

$$\mathbf{Q}((A \cap B) \cap E) + \mathbf{Q}((A \cap B)^c \cap E) \quad (2.1)$$

$$= \mathbf{Q}(A \cap B \cap E) + \mathbf{Q}((A^c \cap B \cap E) \cup (A \cap B^c \cap E) \cup (A^c \cap B^c \cap E)) \quad (2.2)$$

$$\leq \mathbf{Q}(A \cap B \cap E) + \mathbf{Q}(A^c \cap B \cap E) + \mathbf{Q}(A \cap B^c \cap E) + \mathbf{Q}(A^c \cap B^c \cap E) \quad (2.3)$$

$$= \mathbf{Q}(B \cap E) + \mathbf{Q}(B^c \cap E), \quad \text{pela definição de } A \in \mathcal{M} \quad (2.4)$$

$$= \mathbf{Q}(E), \quad \text{pela definição de } B \in \mathcal{M}. \quad (2.5)$$

Portanto, $A \cap B \in \mathcal{M}$.

Agora temos que demonstrar que se $B_1, B_2, \dots \in \mathcal{M}$, então $\bigcup_n B_n \in \mathcal{M}$. Podemos assumir que os conjuntos B_n são disjuntos, caso contrário, considere em vez $B'_n = B_n \setminus (B_1 \cup \dots \cup B_{n-1})$.

Para conveniência de notação, seja:

$$D_N = \bigcup_{n=1}^N B_n \in \mathcal{M}, \quad D_N \uparrow D_\infty = \bigcup_{n=1}^{\infty} B_n.$$

Precisamos também do seguinte fato

$$\mathbf{Q}\left(\bigcup_{n=1}^N B_n \cap E\right) = \sum_{n=1}^N \mathbf{Q}(B_n \cap E), \quad \text{para todo } E \subseteq \Omega \text{ e } N = 1, 2, \dots$$

que pode ser demonstrado por indução direta. O caso inicial $N = 1$ é trivial. Para $N > 1$,

$$\mathbf{Q}(D_N \cap E) = \mathbf{Q}(D_{N-1} \cap (D_N \cap E)) + \mathbf{Q}(D_{N-1}^c \cap (D_N \cap E)) \quad (2.6)$$

$$= \mathbf{Q}(D_{N-1} \cap E) + \mathbf{Q}(B_N \cap E) \quad (2.7)$$

$$= \sum_{n=1}^N \mathbf{Q}(B_n \cap E), \quad \text{por hipótese de indução.} \quad (2.8)$$

Assim

$$\mathbf{Q}(E) = \mathbf{Q}(D_N \cap E) + \mathbf{Q}(D_N^c \cap E) \quad (2.9)$$

$$= \sum_{n=1}^N \mathbf{Q}(B_n \cap E) + \mathbf{Q}(D_N^c \cap E) \quad (2.10)$$

$$\geq \sum_{n=1}^N \mathbf{Q}(B_n \cap E) + \mathbf{Q}(D_\infty^c \cap E), \quad \text{por monotonicidade.} \quad (2.11)$$

Tomando $N \rightarrow \infty$, obtemos

$$\mathbf{Q}(E) \geq \sum_{n=1}^{\infty} \mathbf{Q}(B_n \cap E) + \mathbf{Q}(D_\infty^c \cap E) \geq \mathbf{Q}(D_\infty \cap E) + \mathbf{Q}(D_\infty^c \cap E).$$

Mas isso implica $D_\infty = \bigcup_{n=1}^{\infty} B_n \in \mathcal{M}$. □

Teorema 2.24 – Aditividade Enumerável da Medida Exterior

Dado uma medida exterior $\mathbf{Q}$, então $\mathbf{Q}$ é σ -aditiva em $\mathcal{M}$. Isso é, se $B_1, B_2, \dots \in \mathcal{M}$ forem disjuntos, então $\mathbf{Q}(\bigcup_n B_n) = \sum_n \mathbf{Q}(B_n)$.

Demonstração. O caso finito $\mathbf{Q}(\bigcup_{n=1}^N B_n) = \sum_{n=1}^N \mathbf{Q}(B_n)$ já foi provado, bastando substituir $E = \Omega$ na equação.

Por monotonicidade, $\sum_{n=1}^N \mathbf{Q}(B_n) \leq \mathbf{Q}(\bigcup_{n=1}^{\infty} B_n)$. Fazendo $N \rightarrow \infty$ temos $\sum_{n=1}^{\infty} \mathbf{Q}(B_n) \leq \mathbf{Q}(\bigcup_{n=1}^{\infty} B_n)$.

A desigualdade na outra direção está implícita na subaditividade. □

Resumimos o que obtivemos até agora

Corolário 2.25 — Teorema de Carathéodory

Se $\mathbf{Q}$ for uma medida exterior (2.19), então a restrição de $\mathbf{Q}$ para os conjuntos mensuráveis $\mathcal{M}$ (2.21) produz uma medida em $\mathcal{M}$.

Demonstração. Pelo Teorema 2.23, $\mathcal{M}$ é uma σ -álgebra. Pelo Teorema 2.24, a restrição de $\mathbf{Q}$ a $\mathcal{M}$ é σ -aditiva; como $\mathbf{Q}(\emptyset) = 0$, essa restrição é uma medida. $\square$

2.2.2. Definindo uma Medida por Extensão

A medida exterior já está definida em todos os subconjuntos de Ω , mas não é aditiva em todos eles. O critério de Carathéodory seleciona exatamente a σ -álgebra na qual essa aditividade reaparece.

Teorema 2.26 — Teorema de Extensão de Carathéodory

Seja $\mu_0 : \mathcal{F}_0 \rightarrow [0, +\infty]$ uma pré-medida numa álgebra $\mathcal{F}_0$ de subconjuntos de Ω . Então existe uma medida μ em $\sigma\langle\mathcal{F}_0\rangle$ cuja restrição a $\mathcal{F}_0$ é μ_0 . Se μ_0 é σ -finita, isto é, se existem $D_n \in \mathcal{F}_0$ tais que $\Omega = \bigcup_{n \geq 1} D_n$ e $\mu_0(D_n) < \infty$, essa extensão é única. Em particular, toda pré-medida de probabilidade possui uma única extensão para uma medida de probabilidade em $\sigma\langle\mathcal{F}_0\rangle$.

Demonstração. Considere a medida exterior μ^* induzida por μ_0 . A prova tem três passos.

1. Os conjuntos da álgebra inicial são mensuráveis para μ^* . Fixe $A \in \mathcal{F}_0$ e $E \subseteq \Omega$. Pela subaditividade de μ^* , sempre temos

$$\mu^*(E) \leq \mu^*(E \cap A) + \mu^*(E \cap A^c).$$

Para obter a desigualdade inversa, se $\mu^*(E) = +\infty$ não há nada a provar. Suponha, então, que $\mu^*(E) < \infty$. Dado $\varepsilon > 0$, escolha $B_1, B_2, \dots \in \mathcal{F}_0$ que cubram E e satisfaçam

$$\sum_n \mu_0(B_n) \leq \mu^*(E) + \varepsilon.$$

As famílias $(A \cap B_n)_n$ e $(A^c \cap B_n)_n$ cobrem, respectivamente, $A \cap E$ e $A^c \cap E$. Além disso, todos esses conjuntos pertencem a $\mathcal{F}_0$. Para todo $C \in \mathcal{F}_0$, a cobertura formada pelo próprio C mostra, diretamente da definição de μ^* , que $\mu^*(C) \leq \mu_0(C)$. Portanto, sem usar ainda a igualdade que será provada no passo seguinte,

$$\mu^*(A \cap E) + \mu^*(A^c \cap E) \leq \mu^*\left(A \cap \bigcup_n B_n\right) + \mu^*\left(A^c \cap \bigcup_n B_n\right) \quad (2.12)$$

$$\leq \sum_n \mu^*(A \cap B_n) + \sum_n \mu^*(A^c \cap B_n) \quad (2.13)$$

$$\leq \sum_n \mu_0(A \cap B_n) + \sum_n \mu_0(A^c \cap B_n) \quad (2.14)$$

$$= \sum_n \mu_0(B_n), \quad \text{pela aditividade finita de } \mu_0, \quad (2.15)$$

$$\leq \mu^*(E) + \varepsilon. \quad (2.16)$$

Como ε é arbitrário, obtemos a desigualdade inversa e, assim, A satisfaz o critério de Carathéodory. Logo $\mathcal{F}_0$ está contida na σ -álgebra $\mathcal{M}$ dos conjuntos mensuráveis para μ^* ; como $\mathcal{M}$ é uma σ -álgebra,

$$\sigma\langle \mathcal{F}_0 \rangle \subseteq \mathcal{M}.$$

Consequentemente, μ^* restrita a $\sigma\langle \mathcal{F}_0 \rangle$ já é uma medida.

2. *A medida construída prolonga μ_0 .* Fixe $A \in \mathcal{F}_0$ e considere uma cobertura arbitrária de A por $B_1, B_2, \dots \in \mathcal{F}_0$. Disjuntifique essa cobertura dentro de A , definindo

$$G_1 = A \cap B_1 \quad \text{e} \quad G_n = (A \cap B_n) \setminus \bigcup_{j=1}^{n-1} B_j, \quad n \geq 2.$$

Os conjuntos G_n pertencem a $\mathcal{F}_0$, são disjuntos, têm união A e satisfazem $G_n \subseteq B_n$. Como μ_0 é uma pré-medida e a união A pertence a $\mathcal{F}_0$,

$$\mu_0(A) = \sum_n \mu_0(G_n) \leq \sum_n \mu_0(B_n).$$

Tomando o ínfimo sobre todas as coberturas, obtemos $\mu_0(A) \leq \mu^*(A)$. A cobertura $(A, \emptyset, \emptyset, \dots)$ fornece a desigualdade contrária. Portanto $\mu^*(A) = \mu_0(A)$ em $\mathcal{F}_0$, e a medida do passo 1 é a extensão procurada.

3. *Unicidade no caso σ -finito.* Substituindo D_n por $H_n = \bigcup_{j=1}^n D_j$, podemos supor que $H_n \uparrow \Omega$ e $\mu_0(H_n) < \infty$ para todo n . Se μ e ν são duas extensões, fixe n e defina, para $E \in \sigma(\mathcal{F}_0)$,

$$\mu_n(E) = \mu(E \cap H_n), \quad \nu_n(E) = \nu(E \cap H_n).$$

Essas são medidas finitas com a mesma massa total $\mu_0(H_n)$. Elas concordam em $\mathcal{F}_0$, pois, para $A \in \mathcal{F}_0$, também $A \cap H_n \in \mathcal{F}_0$ e

$$\mu_n(A) = \mu_0(A \cap H_n) = \nu_n(A).$$

Pelo Teorema 2.14, μ_n e ν_n concordam em toda $\sigma(\mathcal{F}_0)$. Portanto, para cada $E \in \sigma(\mathcal{F}_0)$,

$$\mu(E \cap H_n) = \nu(E \cap H_n), \quad n \geq 1.$$

Como $E \cap H_n \uparrow E$, a continuidade por baixo fornece $\mu(E) = \nu(E)$. No caso de uma pré-medida de probabilidade, basta tomar $H_1 = \Omega$; além disso, a extensão construída tem massa total $\mu(\Omega) = \mu_0(\Omega) = 1$. □

2.3. Classes Compactas

Para mostrar que uma função finitamente aditiva é uma pré-medida, precisamos controlar uma condição enumerável. A noção de classe compacta foi concebida precisamente para realizar essa passagem: uma interseção enumerável vazia de conjuntos da classe já é detectada por uma subfamília finita. Depois de aproximarmos os conjuntos da álgebra por elementos da classe compacta, a aditividade finita basta para controlar os erros dessa subfamília finita. Assim, obtemos continuidade no vazio, portanto uma pré-medida, e só então aplicamos o Teorema de Extensão de Carathéodory. A classe compacta é, desse modo, o mecanismo que torna verificável a hipótese necessária para a extensão, e não uma simples reformulação daquele teorema.

Definição 2.27 — Classe Compacta

Dizemos que uma família $\mathcal{K} \subseteq \mathcal{P}(\Omega)$ é uma **classe compacta** se, para toda sequência $(C_n)_{n \geq 1}$ de elementos de $\mathcal{K}$ tal que

$$\bigcap_{n=1}^{\infty} C_n = \emptyset,$$

existe $m \in \mathbb{N}$ para o qual $C_1 \cap \dots \cap C_m = \emptyset$.

Equivalentemente, toda subfamília enumerável de $\mathcal{K}$ com interseção vazia possui uma subfamília finita com interseção vazia.

Adicionar $\emptyset$ a uma classe compacta preserva essa propriedade. Por isso, quando for conveniente, suporemos sem perda de generalidade que $\emptyset \in \mathcal{K}$.

Exemplo 2.28 — .

A família dos conjuntos compactos de um espaço métrico é uma classe compacta. De fato, suponha, por contradição, que $(C_n)_{n \geq 1}$ seja uma sequência de compactos para a qual toda interseção finita seja não vazia. Defina

$$B_n = \bigcap_{j=1}^n C_j$$

e escolha $x_n \in B_n$. Como $x_n \in C_1$ para todo n e C_1 é compacto, a sequência possui uma subsequência $(x_{n_\ell})_{\ell \geq 1}$ que converge para algum $x \in C_1$. Fixado $r \geq 1$, temos $x_{n_\ell} \in C_r$ sempre que $n_\ell \geq r$. Como C_r é fechado, $x \in C_r$. Portanto $x \in \bigcap_{r \geq 1} C_r$, uma contradição.

O mesmo resultado vale, mais geralmente, em qualquer espaço topológico de Hausdorff, sem que o espaço ambiente precise ser compacto. Com efeito, seja $(C_n)_{n \geq 1}$ uma sequência de compactos tal que

$$\bigcap_{n \geq 1} C_n = \emptyset.$$

Se $C_1 = \emptyset$, a conclusão é imediata. Caso contrário, fixe C_1 . Num espaço de Hausdorff, cada compacto C_n é fechado; logo

$$\{C_1 \setminus C_n : n \geq 2\}$$

é uma cobertura por abertos relativos do compacto C_1 . Extraíndo uma subcobertura finita, existem $n_1, \dots, n_k \geq 2$ tais que

$$C_1 = \bigcup_{j=1}^k (C_1 \setminus C_{n_j}).$$

Consequentemente,

$$C_1 \cap C_{n_1} \cap \dots \cap C_{n_k} = \emptyset.$$

O conjunto C_1 deve ser incluído nessa subfamília finita. É a compacidade de C_1 , e não a do espaço ambiente, que produz a subcobertura finita.

◁

A estabilidade das classes compactas sob essas duas operações é tratada no Exercício 2.19, ao final do capítulo.

Teorema 2.29 — Teorema de Extensão Compacta

Seja $\mu : \mathcal{F}_0 \rightarrow [0, +\infty]$ uma função finitamente aditiva numa álgebra $\mathcal{F}_0$ de subconjuntos de Ω , com $\mu(\emptyset) = 0$, e seja $\mathcal{K} \subseteq \mathcal{P}(\Omega)$ uma classe compacta que contém $\emptyset$. Suponha que, para cada $A \in \mathcal{F}_0$ e cada $\varepsilon > 0$, existam $C \in \mathcal{K}$ e $B \in \mathcal{F}_0$ tais que

$$B \subseteq C \subseteq A \quad \text{e} \quad \mu(A \setminus B) < \varepsilon. \quad (2.17)$$

Então μ é uma pré-medida em $\mathcal{F}_0$.

Em particular, se $\mathcal{K} \subseteq \mathcal{F}_0$ e $\mu(\Omega) < \infty$, basta supor que, para cada $A \in \mathcal{F}_0$,

$$\mu(A) = \sup\{\mu(C) : C \in \mathcal{K} \text{ e } C \subseteq A\}. \quad (2.18)$$

Demonstração. Começemos pela hipótese (2.17), que não exige que μ seja finita. Seja $(A_n)_{n \geq 1}$ uma sequência em $\mathcal{F}_0$ tal que $A_n \downarrow \emptyset$, e fixe $\varepsilon > 0$. Para cada n , escolha $C_n \in \mathcal{K}$ e $B_n \in \mathcal{F}_0$ de modo que

$$B_n \subseteq C_n \subseteq A_n \quad \text{e} \quad \mu(A_n \setminus B_n) < \frac{\varepsilon}{2^n}.$$

Como

$$\bigcap_{n \geq 1} C_n \subseteq \bigcap_{n \geq 1} A_n = \emptyset,$$

a compacidade da classe fornece $m \geq 1$ tal que $C_1 \cap \cdots \cap C_m = \emptyset$. Como $B_i \subseteq C_i$, também $B_1 \cap \cdots \cap B_m = \emptyset$. Para todo $n \geq m$, o encaixe dos conjuntos A_n implica

$$A_n \subseteq \bigcup_{i=1}^m (A_n \setminus B_i) \subseteq \bigcup_{i=1}^m (A_i \setminus B_i).$$

Pela monotonicidade e pela subaditividade finita,

$$\mu(A_n) \leq \sum_{i=1}^m \mu(A_i \setminus B_i) < \sum_{i=1}^m \frac{\varepsilon}{2^i} < \varepsilon.$$

Logo $\mu(A_n) \downarrow 0$. Este é exatamente o ponto em que a classe compacta converte a interseção enumerável vazia num controle finito dos erros.

Para concluir diretamente, sem pressupor que μ seja uma probabilidade, sejam $E_1, E_2, \dots \in \mathcal{F}_0$ disjuntos, com $E = \bigcup_{j \geq 1} E_j \in \mathcal{F}_0$, e defina

$$R_n = E \setminus \bigcup_{j=1}^n E_j.$$

Então $R_n \downarrow \emptyset$, de modo que $\mu(R_n) \rightarrow 0$, e a aditividade finita dá

$$\mu(E) = \sum_{j=1}^n \mu(E_j) + \mu(R_n).$$

Se algum E_j tem medida infinita, a igualdade enumerável é imediata. Caso contrário, $\mu(R_n) < \infty$ para n suficientemente grande; a igualdade anterior mostra então que $\mu(E) < \infty$, e, fazendo $n \rightarrow \infty$, obtemos

$$\mu(E) = \sum_{j=1}^{\infty} \mu(E_j).$$

Portanto μ é uma pré-medida.

Finalmente, suponha $\mu(\Omega) < \infty$ e (2.18). Então $\mu(A) < \infty$ para todo $A \in \mathcal{F}_0$. Dado $\varepsilon > 0$, podemos escolher $C \in \mathcal{X}$, $C \subseteq A$, tal que $\mu(C) > \mu(A) - \varepsilon$. Tomando $B = C$ e usando a aditividade finita,

$$\mu(A \setminus C) = \mu(A) - \mu(C) < \varepsilon,$$

o que reduz o caso finito à primeira parte da demonstração. □

Observação 2.30 — Pré-medidas gerais e pré-medidas de probabilidade

A finitude não é usada no argumento de classe compacta quando se assume diretamente a aproximação por sanduíche e o controle residual de (2.17). Ela é usada somente para deduzir esse controle, no caso $\mathcal{K} \subseteq \mathcal{F}_0$, a partir da identidade pelo supremo (2.18): é preciso que $\mu(A) < \infty$ para escrever $\mu(A \setminus C) = \mu(A) - \mu(C)$. Se $\mu(A) = \infty$, a igualdade pelo supremo, sozinha, não controla a medida do resíduo, pois envolveria a expressão indeterminada $\infty - \infty$.

Para uma probabilidade, $\mu(\Omega) = 1$, e portanto a finitude é automática: a formulação pelo supremo é suficiente quando $\mathcal{K} \subseteq \mathcal{F}_0$. Para uma função finitamente aditiva geral, há duas alternativas precisas: supor que μ seja finita e usar essa regularidade interna, ou manter a generalidade e exigir diretamente (2.17). Em ambos os casos, a classe compacta prova primeiro que μ é uma pré-medida. O Teorema de Extensão de Carathéodory fornece então uma medida em $\sigma\langle \mathcal{F}_0 \rangle$; no caso geral, a unicidade requer σ -finitude, enquanto no caso de probabilidade ela é automática.

2.4. Aplicações

2.4.1. Teorema dos Números Normais de Borel

Construiremos agora a medida de Lebesgue em $(0,1]$. O mesmo modelo descreverá tanto a escolha uniforme de um número quanto uma sequência infinita de lançamentos de moeda. O argumento segue um roteiro que reaparecerá no livro:

Construção de Probabilidades

- a** Começaremos construindo uma probabilidade finitamente aditiva. Essa construção em geral é simples e intuitiva.
- b** Provaremos a σ -aditividade. Em alguns casos pode ser mais fácil provar a

condição equivalente de continuidade no vazio. Em outros provar a condição do Teorema da Extensão Compacta.

c Usaremos o Teorema de Extensão de Carathéodory para obter um espaço de probabilidade.

Começaremos construindo uma probabilidade finitamente aditiva nos intervalos. Essa função atribui a cada intervalo seu comprimento.

Definição 2.31 — Álgebra de Borel

Seja $\mathcal{B}_0((0,1])$ a álgebra de Borel, isto é a família formada por uniões finitas (possivelmente vazia) de intervalos da forma $(a,b] \subset (0,1]$.

Definição 2.32 —

Se $A \in \mathcal{B}_0((0,1])$ possui uma representação disjunta

$$A = \bigcup_{j=1}^m (a_j, b_j],$$

defina

$$\mathbf{P}(A) = \sum_{j=1}^m (b_j - a_j),$$

e defina $\mathbf{P}(\emptyset) = 0$.

Teorema 2.33 —

A função $\mathbf{P}$ está bem definida e é uma probabilidade finitamente aditiva em $\mathcal{B}_0((0,1])$.

Demonstração. Considere duas decomposições finitas de um mesmo conjunto A em intervalos semiabertos disjuntos. Ordene todos os extremos que aparecem nas duas decomposições, acrescentando 0 e 1 se necessário: $0 = t_0 < t_1 < \dots < t_r = 1$. Em cada átomo $(t_{k-1}, t_k]$, a função indicadora de A é constante. Portanto qualquer uma das duas decomposições atribui a A a soma de $t_k - t_{k-1}$ exatamente sobre os mesmos átomos. As duas somas coincidem, e $\mathbf{P}$ está bem definida.

O mesmo refinamento mostra que, se A e B são disjuntos, os átomos que compõem $A \cup B$ são a união disjunta dos átomos que compõem A e dos que compõem B . Logo $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B)$. Finalmente, $\mathbf{P}((0,1]) = 1$, de modo que $\mathbf{P}$ é uma probabilidade finitamente aditiva. $\square$

Podemos ver um número real $\omega \in (0,1]$ como a realização do lançamento de uma moeda infinitas vezes. Para isso considere a expansão binária desse número. Nos pontos diádicos, que admitem duas expansões, adotaremos a expansão que não é eventualmente nula; em particular, representamos 1 por 0,111 Dessa forma $\omega \in (0,1]$ determina uma sequência de dígitos em $\{0,1\}^{\mathbb{N}}$, omitindo-se o dígito inicial antes da vírgula.

Nosso primeiro resultado será sobre a distribuição de dígitos 0 e 1 num número qualquer $\omega \in (0,1]$. Para realizar essa contagem faremos inicialmente uma conversão dos dígitos dessa sequência. Converteremos os $0 \rightarrow -1$ obtendo a sequência de funções de Rademacher z_k .

Definição 2.34 —

Para todo número $\omega \in (0,1]$, a função $d_k(\omega)$ denotará o k -ésimo dígito da representação de ω . Seja $z_k(\omega) := 2d_k(\omega) - 1$ e

$$s_n(\omega) := \sum_{k=1}^n z_k(\omega) \equiv \text{excesso de 1's nos } n \text{ primeiros dígitos.}$$

Observação 2.35 —

A sequência $s_n(\omega)$ definida acima pode ser vista como um passeio aleatório simétrico nos pontos de coordenadas inteiras da reta, isto é, em $\mathbb{Z}$, que começa em 0 e a cada passo move para a direita ou esquerda, +1 ou -1, com probabilidade igual. Este passeio pode ser ilustrado da seguinte maneira. Um ponto é colocado no zero e uma moeda justa é jogada. Se sair cara, o marcador é movido uma unidade para a direita e se sair coroa o marcador é movido uma unidade para a esquerda.

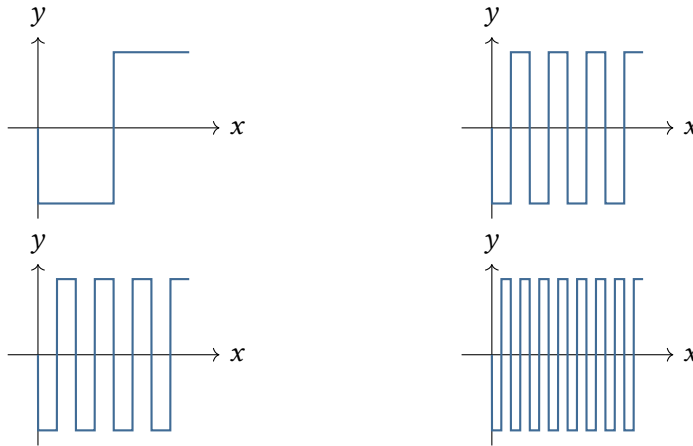


Figura 2.4.: Funções de Rademacher z_k .

Provaremos que

$$\lim_{n \rightarrow \infty} \mathbf{P}\left[\{\omega \in (0,1] : |\text{excesso de 1's nos } n \text{ primeiros dígitos}| \geq n\varepsilon\}\right] = 0.$$

Na expressão acima, o limite está fora da probabilidade. Como veremos, o evento dentro da probabilidade pertence à álgebra de Borel; portanto, a expressão faz sentido com o que desenvolvemos até este ponto.

Teorema 2.36 — Lei Fraca de Bernoulli

Para todo $\varepsilon > 0$,

$$\lim_{n \rightarrow \infty} \mathbf{P}\left[\{\omega \in (0,1] : |s_n(\omega)/n| \geq \varepsilon\}\right] = 0.$$

Demonstração. Qualquer evento baseado nos valores $z_1(\omega), \dots, z_n(\omega)$ pode ser descrito como uma união disjunta de intervalos diádicos da forma $\left(\frac{k-1}{2^n}, \frac{k}{2^n}\right]$. Logo $\{\omega \in (0,1] : |s_n(\omega)/n| \geq \varepsilon\} \in \mathcal{B}_0((0,1])$.

Nesta prova, a integral de uma função constante em intervalos de $\mathcal{B}_0((0,1])$ denota simplesmente a soma dos valores constantes multiplicados pelos comprimentos correspondentes. Cada intervalo diádico de nível $i-1$ se divide em dois no nível i . Logo z_i toma valor $+1$ em metade do intervalo e valor -1 na outra metade. Se $i < j$, então, em cada intervalo diádico de nível $j-1$, z_i é constante, enquanto z_j vale -1 numa metade e $+1$ na outra. O produto $z_i z_j$, portanto, tem integral zero em cada um desses intervalos, e

$$\int_0^1 z_k(\omega) z_j(\omega) d\omega = \begin{cases} 1 & \text{quando } k = j \\ 0 & \text{quando } k \neq j. \end{cases}$$

Isso implica que $\int_0^1 s_n^2(\omega) d\omega = \int_0^1 \sum_{k,j=1}^n z_k(\omega) z_j(\omega) d\omega = n$ e logo

$$\begin{aligned} n &= \int_0^1 s_n^2(\omega) d\omega \geq \int_{|s_n/n| \geq \varepsilon} s_n^2(\omega) d\omega \\ &\geq \int_{|s_n/n| \geq \varepsilon} n^2 \varepsilon^2 d\omega = n^2 \varepsilon^2 \mathbf{P}[|s_n/n| \geq \varepsilon]. \end{aligned}$$

Logo $\mathbf{P}[|s_n/n| \geq \varepsilon] \leq 1/(n\varepsilon^2) \rightarrow 0$ quando $n \rightarrow \infty$. □

Dado uma base, b um **número normal** é um número real cujos algarismos aparecem todos com a mesma frequência, ou seja, cuja sequência infinita de dígitos

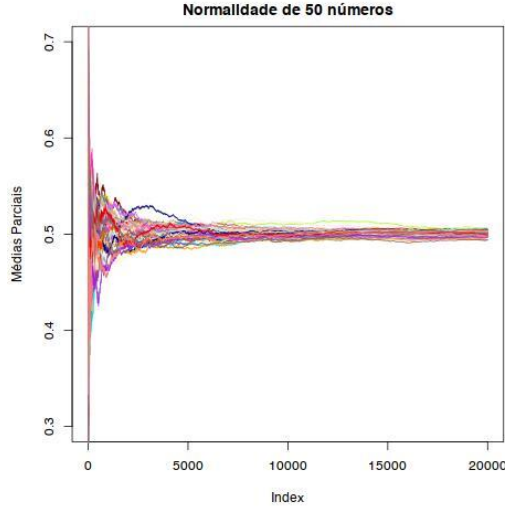


Figura 2.5.: Dado uma base, um número normal é um número real cujos algarismos aparecem todos com a mesma frequência.

na base b é distribuída uniformemente no sentido de que cada um dos valores algarismos tem a mesma frequência $1/b$. Também significa que nenhum algarismo, ou combinação (finita) de algarismos, ocorre com mais frequência do que qualquer outra. No que se segue trataremos apenas de números normais na base 2, mas os argumentos podem ser adaptados facilmente para outras bases.

Definição 2.37 —

O conjunto dos números **simplesmente normais na base 2** em $(0,1]$ é definido como

$$\begin{aligned} N_1 &:= \{\omega \in (0,1] : \lim_{n \rightarrow \infty} s_n(\omega)/n = 0\} \\ &= \{\omega \in (0,1] : \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n d_k(\omega) = \frac{1}{2}\}. \end{aligned}$$

Seu complemento será denotado por $A_1 := (0,1] \setminus N_1$.

Provaremos que os números anormais formam um conjunto negligenciável.

Definição 2.38 – Conjunto Negligenciável

Um subconjunto $B \subset (0,1]$ é dito **negligenciável** se para todo $\varepsilon > 0$, existem subconjuntos $B_1, B_2, \dots$ de $\mathcal{B}_0((0,1])$ tal que

$$B \subset \bigcup_{k=1}^{\infty} B_k \quad \text{e} \quad \sum_{k=1}^{\infty} \mathbf{P}[B_k] \leq \varepsilon.$$

O próximo teorema também é conhecido como a Lei Forte dos Grandes Números para o Lançamento de Moedas

Teorema 2.39 – Lei forte para a frequência dos algarismos

O conjunto A_1 dos números que não são simplesmente normais na base 2 é negligenciável.

Demonstração.

1 Começaremos notando as seguintes inclusões:

Dado $\varepsilon_k \downarrow 0$ quando $k \rightarrow \infty$. Então

$$\begin{aligned} \{\omega : \left| \frac{s_{k^2}(\omega)}{k^2} \right| < \varepsilon_k \text{ para } k \text{ suficientemente grande} \} \\ \subset \{\omega : \lim_k \frac{s_{k^2}(\omega)}{k^2} = 0\} \\ \subset \underbrace{\{\omega : \lim_n \frac{s_n(\omega)}{n} = 0\}}_{=N_1} \end{aligned}$$

Para justificar a última inclusão, se $k^2 \leq n < (k+1)^2$, então

$$\frac{|s_n|}{n} \leq \frac{|s_{k^2}|}{k^2} + \frac{n - k^2}{k^2} \leq \frac{|s_{k^2}|}{k^2} + \frac{2k+1}{k^2}.$$

Logo a convergência ao longo dos quadrados implica a convergência da sequência inteira.

2 Agora por (1.) temos que

$$A_1 = N_1^c \subset \{\omega : \left| \frac{s_{k^2}(\omega)}{k^2} \right| \geq \varepsilon_k \text{ para infinitos } k\}$$

$$\subset \bigcup_{k=j}^{\infty} \underbrace{\{\omega : \left| \frac{s_{k^2}(\omega)}{k^2} \right| \geq \varepsilon_k\}}_{=: B_k}, \quad \text{para todo } j$$

onde $B_k \in \mathcal{B}_0((0,1])$. Da demonstração da Lei Fraca,

$$\mathbf{P}[B_k] \leq \frac{1}{k^2 \varepsilon_k^2} = \frac{1}{k^{3/2}}$$

se tomarmos $\varepsilon_k := k^{-1/4}$. Logo $\sum_{k=1}^{\infty} \mathbf{P}[B_k] < \infty$ e assim $\sum_{k=j}^{\infty} \mathbf{P}[B_k] \rightarrow 0$ quando $j \rightarrow \infty$. Portanto A_1 é negligenciável. $\square$

Definição 2.40 – Número normal na base 2

Fixe $r \geq 1$ e considere uma palavra

$$u = (u_1, \dots, u_r) \in \{0,1\}^r.$$

Defina

$$N_n(u, \omega) = \#\{1 \leq j \leq n : (d_j(\omega), \dots, d_{j+r-1}(\omega)) = u\}.$$

Dizemos que ω é **normal na base 2** se, para todo $r \geq 1$ e toda palavra $u \in \{0,1\}^r$,

$$\frac{N_n(u, \omega)}{n} \longrightarrow 2^{-r}.$$

Denotaremos por N o conjunto dos números normais e por $A = (0,1] \setminus N$ o conjunto dos números anormais.

Teorema 2.41 — Teorema dos Números Normais de Borel

O conjunto A dos números anormais na base 2 é negligenciável.

Demonstração. Fixe $r \geq 1$, uma palavra $u \in \{0,1\}^r$ e uma classe de resíduos $\ell \in \{1, \dots, r\}$. Para $m \geq 0$, seja $Y_m(\omega)$ a indicadora do evento em que a palavra u começa na posição $\ell + mr$. Esses eventos dependem de blocos disjuntos de r dígitos. Pela definição de $\mathbf{P}$ nos intervalos diádicos,

$$\mathbf{P}(Y_m = 1) = 2^{-r} \quad \text{e} \quad \mathbf{P}(Y_m = 1, Y_q = 1) = 2^{-2r} \quad (m \neq q).$$

Escrevendo $p = 2^{-r}$ e $S_n = \sum_{m=0}^{n-1} Y_m$, a mesma soma finita usada na prova da Lei Fraca dá

$$\int_0^1 (S_n - np)^2 d\omega = np(1-p).$$

Consequentemente,

$$\mathbf{P}\left(\left|\frac{S_n}{n} - p\right| \geq \varepsilon\right) \leq \frac{p(1-p)}{n\varepsilon^2}.$$

Aplicando o argumento dos índices quadrados do Teorema 2.39, concluímos que o conjunto em que S_n/n não converge para p é negligenciável.

A união enumerável de conjuntos negligenciáveis é negligenciável: dada uma lista (E_j) e $\eta > 0$, cubra E_j por elementos da álgebra com soma de medidas menor que $\eta/2^j$ e reúna todas essas coberturas. Podemos, portanto, excluir simultaneamente os conjuntos excepcionais correspondentes a todos os trios (r, u, ℓ) .

Fora dessa união excepcional, para cada palavra u as frequências ao longo das r classes de posições módulo r convergem todas para 2^{-r} . Somando as classes e desprezando no máximo $r-1$ posições na extremidade, obtemos $N_n(u, \omega)/n \rightarrow 2^{-r}$. Isso vale para toda palavra finita, e portanto $\omega \in N$. Logo A é negligenciável. $\square$

É até aqui que a aditividade finita nos leva. Para “passar o limite para dentro” e concluir que os números anormais formam um conjunto de medida nula, precisamos da aditividade enumerável.

A extensão a uma medida de probabilidade começa, portanto, pela verificação da continuidade.

Teorema 2.42 —

A aplicação $\mathbf{P} : \mathcal{B}_0((0,1]) \rightarrow [0,1]$ definida pela Definição 2.32 é uma pré-medida de probabilidade.

Demonstração. Seja $\mathcal{K}$ a família formada por $\emptyset$ e pelas uniões finitas de intervalos fechados contidos em $(0,1]$. Cada elemento de $\mathcal{K}$ é compacto em $\mathbb{R}$; portanto, pelo exemplo da Seção 2.3, $\mathcal{K}$ é uma classe compacta.

Fixe $A \in \mathcal{B}_0((0,1])$ e $\varepsilon > 0$, e escreva $A = \bigcup_{j=1}^m (a_j, b_j]$. Escolha números $0 < \delta_j < b_j - a_j$ tais que $\sum_j \delta_j < \varepsilon$ e defina

$$C = \bigcup_{j=1}^m [a_j + \delta_j, b_j] \quad \text{e} \quad B = \bigcup_{j=1}^m (a_j + \delta_j, b_j].$$

Então $B \in \mathcal{B}_0((0,1])$, $C \in \mathcal{K}$, $B \subseteq C \subseteq A$ e

$$\mathbf{P}(A \setminus B) = \sum_{j=1}^m \delta_j < \varepsilon.$$

O Teorema de Extensão Compacta, aplicado a esse sanduíche, mostra que $\mathbf{P}$ é uma pré-medida. Em termos da estratégia da prova, se $A_n \downarrow \emptyset$, os compactos C_n têm interseção vazia e uma subfamília finita — incluindo o primeiro compacto quando ele é o compacto fixado — já tem interseção vazia. Os conjuntos B_n permitem então medir, por aditividade finita, o erro dessa subfamília finita. $\square$

Teorema 2.43 — Medida de Lebesgue

Dado $\mathbf{P} : \mathcal{B}_0((0,1]) \rightarrow [0,1]$ como na Definição 2.32. Então

- 1** $\mathbf{P}$ admite uma única extensão para uma medida de probabilidade em $\mathcal{B}((0,1])$ (também denotada $\mathbf{P}$);
- 2** $\mathbf{P}$ é a única medida em $\mathcal{B}((0,1])$ que satisfaz $\mathbf{P}[(0,x]] = x$ para todo $x \in (0,1]$;

Demonstração.

□ Demonstração do item 2.43.1

O Teorema de Extensão de Carathéodory com o Teorema 2.42 mostra que existe uma única extensão $\mathbf{P} : \mathcal{B}_0((0,1]) \rightarrow [0,1]$ para $\mathbf{P} : \mathcal{B}((0,1]) \rightarrow [0,1]$ como $\mathcal{B}((0,1]) = \sigma\langle \mathcal{B}_0((0,1]) \rangle$.

□ Demonstração do item 2.43.2

Esse item segue do Teorema de Unicidade de medida pois $\mathcal{B}((0,1]) = \sigma\langle (0,x] : x \in (0,1] \rangle$ e $\{(0,x] : x \in (0,1]\}$ é uma π -sistema.

□

Como corolário da construção da medida de Lebesgue temos que o conjunto dos números normais tem probabilidade 1 e surpreendentemente temos que a σ -álgebra de Borel não é o conjunto das partes! Isso ocorre pois existe uma medida de probabilidade não trivial e invariante na σ -álgebra de Borel o que sabemos pela construção do conjunto de Vitali não ocorre no conjunto das partes.

Corolário 2.44 —

Dado $\mathbf{P} : \mathcal{B}_0((0,1]) \rightarrow [0,1]$ como na Definição 2.32. Então

- 1 $N \in \mathcal{B}((0,1])$ e $\mathbf{P}[N] = 1$ onde N é o conjunto dos números normais em $(0,1]$;
- 2 $\mathcal{B}_0((0,1]) \subsetneq \mathcal{B}((0,1]) \subsetneq \overline{\mathcal{B}((0,1])} \subsetneq \mathcal{P}(\Omega)$. Conjuntos em $\mathcal{B}((0,1])$ são denominados **Borel mensuráveis**. Conjunto em $\overline{\mathcal{B}((0,1])}$ são denominados **Lebesgue mensuráveis**.

Demonstração.

□ Demonstração do item 2.44.1

Para cada palavra finita u , o evento $\{|N_n(u, \omega)/n - 2^{-|u|}| < 1/m\}$ depende de um número finito de dígitos e pertence a $\mathcal{B}_0((0,1])$. Assim,

$$\left\{ \omega : \frac{N_n(u, \omega)}{n} \rightarrow 2^{-|u|} \right\} = \bigcap_{m=1}^{\infty} \bigcup_{j=1}^{\infty} \bigcap_{n=j}^{\infty} \left\{ \omega : \left| \frac{N_n(u, \omega)}{n} - 2^{-|u|} \right| < \frac{1}{m} \right\}$$

é boreliano. Intersectando sobre a família enumerável de palavras finitas, concluímos que N e N^c pertencem a $\mathcal{B}((0,1])$.

Pelo Teorema 2.41, N^c é negligenciável. Isto é, dado $\varepsilon > 0$, existem $B_n \in \mathcal{B}_0((0,1])$ tais que $N^c \subseteq \bigcup_{n=1}^{\infty} B_n$ e $\sum_{n=1}^{\infty} \mathbf{P}(B_n) \leq \varepsilon$. Pela subaditividade,

$$\mathbf{P}(N^c) \leq \mathbf{P}\left(\bigcup_{n=1}^{\infty} B_n\right) \leq \sum_{n=1}^{\infty} \mathbf{P}(B_n) \leq \varepsilon.$$

Como ε é arbitrário, $\mathbf{P}(N^c) = 0$ e $\mathbf{P}(N) = 1$.

□ Demonstração do item 2.44.2

A inclusão $\mathcal{B}_0((0,1]) \subsetneq \mathcal{B}((0,1])$ é própria porque todo singleton é boreliano, enquanto nenhum singleton é união finita de intervalos semiabertos não degenerados. A Proposição 2.55, junto com o Teorema da cardinalidade dos borelianos provado adiante, mostra que $\mathcal{B}((0,1]) \subsetneq \overline{\mathcal{B}((0,1])}$: todo subconjunto de um conjunto de Cantor de medida zero é Lebesgue mensurável, mas nem todos esses subconjuntos podem ser borelianos.

Para a última inclusão, usaremos a invariância por translações da medida de Lebesgue, demonstrada adiante no Teorema 2.48. Essa invariância passa à completção: se $E = B \cup N$, com B boreliano e N contido em um boreliano nulo, então $E + x = (B + x) \cup (N + x)$ continua Lebesgue mensurável e tem a mesma medida. A translação módulo 1 em $(0,1]$ é obtida cortando o conjunto no ponto de quebra e transladando as duas partes da maneira usual; portanto a medida completada em $(0,1]$ também é invariante por essa operação. O Teorema de Vitali do Capítulo 1 mostra então que essa medida não pode estar definida em todo $\mathcal{P}((0,1])$. Logo $\overline{\mathcal{B}((0,1])} \subsetneq \mathcal{P}((0,1])$.

□

Qual a diferença entre a lei fraca e a lei forte de grandes números?

Lei Fraca: $\lim_{n \rightarrow \infty} \mathbf{P}\left(\left|\frac{s_n}{n}\right| < \varepsilon\right) = 1$, para todo $\varepsilon > 0$;

Lei Forte : $\mathbf{P}\left(\lim_{n \rightarrow \infty} \frac{s_n}{n} = 0\right) = 1$.

A lei fraca fixa o n e analisa o conjunto dos $s_n(\omega)/n$ sobre $\omega \in (0,1]$. Em particular, a lei fraca diz que para grandes valores de n torna-se cada vez mais raro encontrar ω 's que satisfazem $|s_n(\omega)/n| \geq \varepsilon$. Por outro lado, a lei forte fixa cada $\omega \in (0,1]$ e analisa os conjuntos dos $s_n(\omega)/n$ sobre n . Em particular para quase todo ω , $s_n(\omega)/n \rightarrow 0$ quando $n \rightarrow \infty$.

2.4.2. Moedas II - Medida de Probabilidade em $\{0, 1\}^\infty$

Construiremos o mesmo experimento diretamente no espaço das sequências. Para infinitos lançamentos de moedas honestas, queremos que cada configuração possível dos primeiros n resultados tenha probabilidade 2^{-n} .

O espaço amostral para esta experiência é o conjunto $\{0, 1\}^\infty$ de todas as sequências infinitas de zeros e uns.

$$\omega = (\omega_1, \omega_2, \dots) \quad \omega_i \in \{0, 1\}$$

Como já discutimos, de modo geral pode não ser possível atribuir uma probabilidade a todo subconjunto do espaço amostral. Assim nossa abordagem será um pouco mais cuidadosa. Começaremos criando uma álgebra de subconjuntos, atribuiremos probabilidades finitas aos conjuntos que pertencem a esta álgebra e, em seguida, estenderemos para uma medida de probabilidade na σ álgebra gerada por essa álgebra.

Álgebras determinadas pelos primeiros lançamentos

Considere $\mathcal{F}_n$ a coleção de eventos cuja ocorrência pode ser decidida olhando apenas para os resultados dos primeiros lançamentos. Por exemplo, o evento $\{\omega | \omega_1 = 1 \text{ e } \omega_2 = \omega_6\}$ pertence a $\mathcal{F}_6$ (e deste modo pertence a $\mathcal{F}_k$ para todo $k \geq 6$).

Seja B um subconjunto arbitrário de $\{0, 1\}^n$. Considere o conjunto

$$A = \{\omega \in \{0, 1\}^\infty | (\omega_1, \omega_2, \dots, \omega_n) \in B\}.$$

Podemos expressar $A \subset \{0, 1\}^\infty$ na forma

$$A = B \times \{0, 1\}^\infty.$$

(Isto é toda sequência em A pode ser vista como um par que consiste numa sequência de tamanho n que pertence a B , seguida de uma sequência infinita arbitrária. Nesse caso B é dito base do cilindro A .)

Os elementos de $\mathcal{F}_n$ são precisamente os cilindros dessa forma. A representação também permite verificar que $\mathcal{F}_n$ é uma σ -álgebra.

O Exercício 2.20, ao final do capítulo, formaliza a verificação de que $\mathcal{F}_n$ é uma σ -álgebra.

Cada σ -álgebra $\mathcal{F}_n$ descreve apenas os primeiros n lançamentos. Para reunir os eventos que podem ser decididos após algum número finito de lançamentos, sem fixar esse número de antemão para toda a família, definimos

$$\mathcal{F}_0 = \bigcup_{n=1}^{\infty} \mathcal{F}_n,$$

a coleção dos conjuntos que pertencem a $\mathcal{F}_n$ para algum n . Assim, $A \in \mathcal{F}_0$ quando sua ocorrência pode ser decidida após um número fixo de lançamentos, que depende do evento.

Exemplo 2.45 — .

Seja $A_n = \{\omega | \omega_n = 1\}$, o evento que o n -ésimo lance resulta em 1. Observamos que $A_n \in \mathcal{F}_n$. Seja $A = \bigcup_{n=1}^{\infty} A_n$, que é o evento que existe pelo menos um 1 na sequência de lançamentos infinitos. O evento A não pertence a $\mathcal{F}_n$, para qualquer n . (Intuitivamente, tendo observado uma sequência de n zeros isso não nos permite decidir se haverá um 1 subsequente ou não.)

◁

O exemplo anterior mostra que $\mathcal{F}_0$ não é uma σ -álgebra. Por outro lado, podemos mostrar que $\mathcal{F}_0$ é uma álgebra.

A estrutura algébrica de $\mathcal{F}_0$ é verificada no Exercício 2.21, ao final do capítulo.

Queremos atribuir probabilidades a todos os eventos em $\mathcal{F}_n$, para cada n . Precisamos, portanto, de uma σ -álgebra que contenha $\mathcal{F}_0$. A escolha natural é a menor possível: ela inclui os eventos que dependem de um número finito de coordenadas e apenas os subconjuntos de $\{0,1\}^\infty$ exigidos pelas operações da teoria. Tomamos, assim, F como a σ -álgebra gerada por $\mathcal{F}_0$.

2.4.2.1. Definindo a medida de Probabilidade

Começamos definindo uma função finitamente aditiva $\mathbf{P}_0$ na álgebra $\mathcal{F}_0$ que satisfaz $\mathbf{P}_0(\{0,1\}^\infty) = 1$. Isso será realizado da seguinte forma. Todo conjunto A em $\mathcal{F}_0$ é da forma $B \times \{0,1\}^\infty$, para algum $B \subset \{0,1\}^n$. Então, definimos $\mathbf{P}_0(A) = \frac{|B|}{2^n}$.

Observe que desta forma ao evento que nos primeiros n lançamentos ocorre uma sequência particular $\{\omega_1, \omega_2, \dots, \omega_n\}$, é atribuída a probabilidade $1/2^n$. Em particular, todas as sequências possíveis de comprimento fixo são atribuídas a mesma probabilidade, conforme desejado.

A definição deve independer da representação: o mesmo conjunto A pode pertencer a $\mathcal{F}_n$ para vários valores de n . Precisamos verificar que essas escolhas dão o mesmo valor a $\mathbf{P}_0(A)$. Na verdade, suponha que $A \in \mathcal{F}_m$, que implica que $A \in \mathcal{F}_n$, para $n > m$. Neste caso,

$$A = B \times \{0,1\}^\infty = C \times \{0,1\}^\infty,$$

onde $B \subset \{0,1\}^n$ e $C \subset \{0,1\}^m$. Assim, $B = C \times \{0,1\}^{n-m}$ e $|B| = |C| \cdot 2^{n-m}$. Uma aplicação da definição produz $\mathbf{P}_0(A) = |B|/2^n$, e outra produz $\mathbf{P}_0(A) = |C|/2^m$. Como $|B| = |C| \cdot 2^{n-m}$, ambos produzem o mesmo valor.

A contagem fornece $\mathbf{P}_0(\Omega) = 1$ e mostra que $\mathbf{P}_0$ é finitamente aditiva. Na topologia produto discreta, $\Omega = \{0,1\}^\infty$ é um espaço compacto de Hausdorff, e cada elemento de $\mathcal{F}_0$ é um cilindro fechado, portanto compacto. Pelo exemplo da Seção 2.3, $\mathcal{F}_0$ é uma classe compacta. Além disso, como o próprio A participa do supremo,

$$\mathbf{P}_0(A) = \sup\{\mathbf{P}_0(C) : C \in \mathcal{F}_0 \text{ e } C \subseteq A\}.$$

Logo o Teorema de Extensão Compacta mostra que $\mathbf{P}_0$ é uma pré-medida. Aplicando em seguida o Teorema de Extensão de Carathéodory, concluímos que existe uma única medida de probabilidade em $\mathcal{F}$, a σ -álgebra gerada por $\mathcal{F}_0$, que concorda com $\mathbf{P}_0$ em $\mathcal{F}_0$. Esta medida de probabilidade atribui a mesma probabilidade, $1/2^n$, a cada sequência possível de comprimento n , conforme desejado.

Os dois modelos de lançamentos de moedas que construímos são equivalentes.

2.4.3. Medida de Lebesgue em $\mathbb{R}^d$

Para todo $\mathbf{i} = (i_1, \dots, i_d) \in \mathbb{Z}^d$ seja $(\mathbf{i}, \mathbf{i} + 1]$ o cubo unitário em $\mathbb{R}^d$ transladado por $\mathbf{i}$ ou seja

$$(\mathbf{i}, \mathbf{i} + 1] \equiv (i_1, i_1 + 1] \times \dots \times (i_d, i_d + 1].$$

Esses conjuntos particionam o espaço,

$$\mathbb{R}^d = \bigcup_{\mathbf{i} \in \mathbb{Z}^d} (\mathbf{i}, \mathbf{i} + 1],$$

e assim $\mathbb{R}^d$ pode ser decomposto como união enumerável de cubos unitários disjuntos. Seja $\mathcal{B}_0^{(\mathbf{i}, \mathbf{i} + 1]}$ a álgebra das uniões finitas de retângulos semiabertos disjuntos contidos em $(\mathbf{i}, \mathbf{i} + 1]$, e seja $\mathcal{B}((\mathbf{i}, \mathbf{i} + 1]) = \sigma\langle \mathcal{B}_0^{(\mathbf{i}, \mathbf{i} + 1]} \rangle$ a σ -álgebra de Borel relativa desse cubo. Denote por $\mathbf{P}_i$ a medida de probabilidade uniforme em $\mathcal{B}((\mathbf{i}, \mathbf{i} + 1])$ que atribui o volume euclidiano aos retângulos, isto é,

$$\mathbf{P}_i((a_1, b_1] \times \dots \times (a_d, b_d]) = \prod_{k=1}^d (b_k - a_k)$$

sempre que $(a_1, b_1] \times \dots \times (a_d, b_d] \subset (\mathbf{i}, \mathbf{i} + 1]$.

A construção de $\mathbf{P}_i$ repete, em dimensão d , o argumento usado em $(0, 1]$. Numa união disjunta de retângulos, define-se primeiro a soma dos volumes. Um refinamento por todos os hiperplanos coordenados que contêm suas faces mostra que a soma não depende da decomposição e é finitamente aditiva. Uniões finitas de caixas fechadas formam uma classe compacta e, recuando ligeiramente cada face esquerda, obtemos o sanduíche da condição (2.17). O Teorema de Extensão Compacta fornece a pré-medida; o Teorema de Extensão de Carathéodory fornece então a única medida de probabilidade $\mathbf{P}_i$ na σ -álgebra relativa de Borel.

Para $A \in \mathcal{B}(\mathbb{R}^d)$, defina

$$\mu_{\mathcal{L}}^d(A) := \sum_{\mathbf{i} \in \mathbb{Z}^d} \mathbf{P}_{\mathbf{i}}((\mathbf{i}, \mathbf{i} + 1] \cap A). \quad (2.19)$$

Agora provaremos que $\mu_{\mathcal{L}}^d$ é uma medida em $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$.

Teorema 2.46 — Medida de Lebesgue

$\mu_{\mathcal{L}}^d$ é uma medida em $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$.

Demonstração. Demonstraremos que $\mu_{\mathcal{L}}^d$ satisfaz os três axiomas (i), (ii) e (iii):

- **(i)** $\mu_{\mathcal{L}}^d(A) \in [0, \infty]$: Trivial.
- **(ii)** $\mu_{\mathcal{L}}^d(\emptyset) = 0$: Imediato pois $\mathbf{P}_{\mathbf{i}}((\mathbf{i}, \mathbf{i} + 1] \cap \emptyset) = 0$.
- **(iii)** σ -aditividade: Suponha $A_1, A_2, \dots \in \mathcal{B}(\mathbb{R}^d)$ são disjuntos. Então

$$\begin{aligned} \mu_{\mathcal{L}}^d\left(\bigcup_{k=1}^{\infty} A_k\right) &= \sum_{\mathbf{i} \in \mathbb{Z}^d} \mathbf{P}_{\mathbf{i}}\left((\mathbf{i}, \mathbf{i} + 1] \cap \bigcup_{k=1}^{\infty} A_k\right) \\ &= \sum_{\mathbf{i} \in \mathbb{Z}^d} \mathbf{P}_{\mathbf{i}}\left(\bigcup_{k=1}^{\infty} (\mathbf{i}, \mathbf{i} + 1] \cap A_k\right) \\ &= \sum_{\mathbf{i} \in \mathbb{Z}^d} \sum_{k=1}^{\infty} \mathbf{P}_{\mathbf{i}}((\mathbf{i}, \mathbf{i} + 1] \cap A_k) \\ &= \sum_{k=1}^{\infty} \sum_{\mathbf{i} \in \mathbb{Z}^d} \mathbf{P}_{\mathbf{i}}((\mathbf{i}, \mathbf{i} + 1] \cap A_k) \\ &= \sum_{k=1}^{\infty} \mu_{\mathcal{L}}^d(A_k) \end{aligned}$$

□

Teorema 2.47 —

$\mu_{\mathcal{L}}^d$ é a única medida em $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ que atribui o volume euclidiano padrão aos retângulos semiabertos limitados, isto é,

$$\mu_{\mathcal{L}}^d((a_1, b_1] \times \cdots \times (a_d, b_d]) = \prod_{k=1}^d (b_k - a_k)$$

para $-\infty < a_k < b_k < \infty$.

Demonstração. Seja ν outra medida com a propriedade do enunciado e seja $\mathcal{P}$ o π -sistema formado por $\emptyset$ e pelos retângulos semiabertos limitados. Então $\mathcal{B}(\mathbb{R}^d) = \sigma\langle \mathcal{P} \rangle$, e as duas medidas concordam em $\mathcal{P}$. Além disso, os cubos $D_n = (-n, n]^d$ pertencem a $\mathcal{P}$, crescem para $\mathbb{R}^d$ e satisfazem

$$\nu(D_n) = \mu_{\mathcal{L}}^d(D_n) = (2n)^d < \infty.$$

A versão σ -finita do Teorema 2.14 implica $\nu = \mu_{\mathcal{L}}^d$ em $\mathcal{B}(\mathbb{R}^d)$. □

Teorema 2.48 —

Para todo $A \in \mathcal{B}(\mathbb{R}^d)$ e $x \in \mathbb{R}^d$, o conjunto $A + x := \{a + x : a \in A\}$ está em $\mathcal{B}(\mathbb{R}^d)$ e

$$\mu_{\mathcal{L}}^d(A + x) = \mu_{\mathcal{L}}^d(A)$$

Demonstração. A translação $\tau_x(y) = y + x$ é um homeomorfismo de $\mathbb{R}^d$, com inversa τ_{-x} . Portanto leva borelianos em borelianos. Defina $\mu_x(A) = \mu_{\mathcal{L}}^d(A + x)$. Como a translação preserva uniões disjuntas, μ_x é uma medida em $\mathcal{B}(\mathbb{R}^d)$. Para cada retângulo semiaberto limitado R , o translado $R + x$ é outro retângulo com os mesmos comprimentos laterais; logo $\mu_x(R) = \mu_{\mathcal{L}}^d(R)$. Pela unicidade provada no Teorema 2.47, as duas medidas coincidem em toda a σ -álgebra de Borel. □

Teorema 2.49 —

Se $T : \mathbb{R}^d \rightarrow \mathbb{R}^d$ é linear e não singular, então $A \in \mathcal{B}(\mathbb{R}^d)$ implica que $TA := \{T(a) : a \in A\} \in \mathcal{B}(\mathbb{R}^d)$ e

$$\mu_{\mathcal{L}}^d(TA) = |\det T| \mu_{\mathcal{L}}^d(A).$$

Demonstração. Como T é um homeomorfismo, leva borelianos em borelianos. Usaremos primeiro o seguinte fato geométrico: se R é um retângulo semiaberto limitado, então

$$\mu_{\mathcal{L}}^d(TR) = |\det T| \mu_{\mathcal{L}}^d(R). \quad (2.20)$$

Eis uma justificativa compatível com a construção anterior. A fronteira do paralelepípedo TR está contida numa união finita de hiperplanos afins, que têm medida nula pela Proposição 2.51. Considere as uniões finitas I_m e O_m dos cubos da malha $2^{-m}\mathbb{Z}^d$ que, respectivamente, estão contidos em TR e encontram TR . Então $I_m \subseteq TR \subseteq O_m$, e $O_m \setminus I_m$ fica contido numa vizinhança de raio $\sqrt{d} 2^{-m}$ da fronteira, dentro de um conjunto limitado fixo. Pela continuidade por cima e pela nulidade da fronteira, $\mu(O_m \setminus I_m) \rightarrow 0$. As medidas de I_m e O_m são as somas dos volumes dos cubos que os compõem; essas são precisamente as somas interna e externa de Jordan. Neste ponto usamos um resultado clássico da Álgebra Linear: o volume euclidiano de um paralelepípedo obtido pela transformação linear T é $|\det T|$ vezes o volume do paralelepípedo original. Assim, o limite comum das somas de Jordan é $|\det T| \mu_{\mathcal{L}}^d(R)$, e (2.20) segue. Essa relação determinante-volume é tomada aqui como resultado prévio de Álgebra Linear.

Defina agora

$$\nu(A) = \frac{\mu_{\mathcal{L}}^d(TA)}{|\det T|}, \quad A \in \mathcal{B}(\mathbb{R}^d).$$

A bijetividade de T mostra que ν é uma medida. Pela equação (2.20), ela concorda com $\mu_{\mathcal{L}}^d$ em todos os retângulos semiabertos limitados. A unicidade do Teorema 2.47 dá $\nu = \mu_{\mathcal{L}}^d$, que é a identidade desejada. $\square$

Teorema 2.50 —

Dado $(\Omega, \mathcal{F}, \mu)$ uma medida σ -finita. Então $\mathcal{F}$ não pode conter uma família de conjuntos disjuntos não enumeráveis de μ -medida positiva

Demonstração. Seja $\{B_i : i \in \mathcal{I}\}$ uma família disjunta com $\mu(B_i) > 0$. Escolha $A_k \in \mathcal{F}$ tais que $\Omega = \bigcup_k A_k$ e $\mu(A_k) < \infty$. Para $k, m \geq 1$, defina

$$I_{k,m} = \{i \in \mathcal{I} : \mu(A_k \cap B_i) > 1/m\}.$$

Esse conjunto é finito: se contivesse N índices distintos, a aditividade finita daria $\mu(A_k) > N/m$, o que é impossível para N arbitrariamente grande. Portanto $I_k = \{i : \mu(A_k \cap B_i) > 0\} = \bigcup_{m=1}^{\infty} I_{k,m}$ é enumerável.

Para cada $i \in \mathcal{I}$, a σ -subaditividade e $B_i = \bigcup_k (A_k \cap B_i)$ mostram que algum k satisfaz $\mu(A_k \cap B_i) > 0$; caso contrário, $\mu(B_i) = 0$. Logo $\mathcal{I} = \bigcup_k I_k$, uma união enumerável de conjuntos enumeráveis. $\square$

Proposição 2.51 —

Se $0 \leq k < d$, então todo subespaço afim k -dimensional de $\mathbb{R}^d$ tem medida de Lebesgue nula.

Demonstração. Escreva $A = a + V$, onde V é um subespaço linear de dimensão $k < d$, e escolha $v \notin V$. Os conjuntos $A + tv$, $t \in \mathbb{R}$, são fechados, dois a dois disjuntos e têm todos a mesma medida, pela invariância por translação. Como a medida de Lebesgue é σ -finita, o Teorema 2.50 impede que essa família não enumerável seja formada por conjuntos de medida positiva. Logo $\mu_{\mathcal{L}}^d(A) = 0$. $\square$

Definição 2.52 — Conjuntos mensuráveis de Borel versus Lebesgue

Dado $(\mathbb{R}^d, \overline{\mathcal{B}(\mathbb{R}^d)}, \overline{\mu_{\mathcal{L}}^d})$ o completamento de $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \mu_{\mathcal{L}}^d)$. Se $A \in \mathcal{B}(\mathbb{R}^d)$ então A é dito Borel mensurável. Se $A \in \overline{\mathcal{B}(\mathbb{R}^d)}$ então A é dito Lebesgue mensurável.

2.5. Borelianos: como se formam e quantos são?

Nesta seção opcional estudaremos a estrutura e a cardinalidade dos conjuntos borelianos e dos conjuntos mensuráveis de Lebesgue. Os resultados não serão usados no restante do texto.

Começaremos com uma descrição elementar dos abertos e fechados.

Proposição 2.53 —

- Todo subconjunto aberto de $\mathbb{R}$ é uma união enumerável de intervalos abertos e, portanto, é boreliano.
- Todo subconjunto fechado de $\mathbb{R}$ é boreliano.

Demonstração. Se $U \subseteq \mathbb{R}$ é aberto, para cada $x \in U$ existem racionais $p_x < x < q_x$ tais que $(p_x, q_x) \subseteq U$. Logo U é a união dos intervalos (p, q) , com extremos racionais, que estão contidos em U . Essa família é enumerável. Como intervalos abertos geram $\mathcal{B}(\mathbb{R})$, segue que U é boreliano. Se F é fechado, então $F = \mathbb{R} \setminus F^c$, e a segunda afirmação decorre da primeira. □

Conjunto de Cantor Para construir o conjunto Cantor, começamos com o intervalo unitário: $C_0 = [0, 1]$. Em seguida, removemos o terço médio desse intervalo, deixando uma união de dois intervalos fechados:

$$C_1 = [0, 1/3] \cup [2/3, 1]$$

Em seguida, removemos o terço médio de cada um desses intervalos, deixando uma união de quatro intervalos fechados:

$$C_2 = [0, 1/9] \cup [2/9, 1/3] \cup [2/3, 7/9] \cup [8/9, 1]$$

Procedendo desta maneira, obtemos uma sequência encaixante $C_0 \supset C_1 \supset C_2 \supset \dots$ de conjuntos fechados, onde C_n é a união de 2^n intervalos fechados, como ilustrado na Figura 2.6. Então a interseção

$$C = \bigcap_n C_n$$

é o conjunto de Cantor.

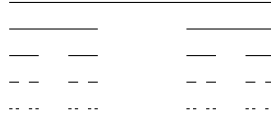


Figura 2.6.: As primeiras etapas da construção do conjunto de Cantor.

Proposição 2.54 —

O conjunto de Cantor C tem as seguintes propriedades:

- 1 C é fechado, e logo é um Boreliano.
- 2 C é não enumerável. De fato, $|C| = |\mathbb{R}|$.
- 3 $\mu(C) = 0$.

Demonstração. Como $C = \bigcap_n C_n$ e cada C_n é fechado, C é fechado. A aplicação

$$(a_n)_{n \geq 1} \mapsto \sum_{n=1}^{\infty} \frac{2a_n}{3^n}$$

é uma bijeção de $\{0, 1\}^{\mathbb{N}}$ sobre C ; logo $|C| = |\{0, 1\}^{\mathbb{N}}| = |\mathbb{R}|$. Finalmente, C_n é a união de 2^n intervalos de comprimento 3^{-n} , e portanto

$$0 \leq \mu_{\mathcal{L}}(C) \leq \mu_{\mathcal{L}}(C_n) = \left(\frac{2}{3}\right)^n.$$

Fazendo $n \rightarrow \infty$, obtemos $\mu_{\mathcal{L}}(C) = 0$. □

Proposição 2.55 — Cardinalidade dos conjuntos mensuráveis

Seja $\mathcal{L}$ a coleção de todos os subconjuntos Lebesgue mensuráveis de $\mathbb{R}$. Então

$$|\mathcal{L}| = |\mathcal{P}(\mathbb{R})|.$$

Demonstração. Claramente $|\mathcal{L}| \leq |\mathcal{P}(\mathbb{R})|$. Como o conjunto de Cantor C tem medida de Lebesgue nula, todo subconjunto de C é Lebesgue mensurável, isto é, $\mathcal{P}(C) \subseteq \mathcal{L}$. Como $|C| = |\mathbb{R}|$, temos $|\mathcal{P}(C)| = |\mathcal{P}(\mathbb{R})|$ e, portanto, $|\mathcal{P}(\mathbb{R})| \leq |\mathcal{L}|$. O teorema de Cantor–Bernstein conclui a igualdade. $\square$

Definição 2.56 — Hierarquia Finita de Borel

A hierarquia finita do Borel consiste em duas sequências Σ_n e Π_n de subconjuntos de $\mathcal{B}$ definidos da seguinte forma:

- $\square$ Σ_1 é a coleção de todos os conjuntos abertos em $\mathbb{R}$ e Π_1 é a coleção de todos os conjuntos fechados em $\mathbb{R}$.
- $\square$ Para cada $n \geq 1$, a coleção Σ_{n+1} consiste de todas as uniões enumeráveis de conjuntos de Π_n , e a coleção Π_{n+1} consiste de todas as interseções enumeráveis de conjuntos de Σ_n .

Exemplo 2.57 — .

Assim, cada conjunto em Σ_3 pode ser escrito como

$$\bigcup_{m \in \mathbb{N}} \bigcap_{n \in \mathbb{N}} U_{m,n}$$

para conjuntos abertos $U_{m,n}$ e cada conjunto em Π_3 pode ser escrito como

$$\bigcap_{m \in \mathbb{N}} \bigcup_{n \in \mathbb{N}} F_{m,n}$$

para conjuntos fechados $F_{m,n}$.

Obtemos, portanto, $\Sigma_1 \subseteq \Sigma_2, \Sigma_1 \subseteq \Pi_2, \Pi_1 \subseteq \Sigma_2$ e $\Pi_1 \subseteq \Pi_2$. E além disso, todas as quatro inclusões são próprias.

Teorema 2.58 — Propriedades de Σ_n e Π_n

Para cada $n \in \mathbb{N}$, as seguintes afirmações são válidas.

1 Para todo $S \subseteq \mathbb{R}$, temos $S \in \Sigma_n$ se e somente se $S^c \in \Pi_n$.

2 $\Sigma_n \subseteq \Sigma_{n+1}, \Sigma_n \subseteq \Pi_{n+1}, \Pi_n \subseteq \Sigma_{n+1}$, e $\Pi_n \subseteq \Pi_{n+1}$.

Além disso, todas as inclusões são próprias.

Demonstração. A primeira afirmação segue por indução das leis de De Morgan. Para as inclusões, observe inicialmente que, em $\mathbb{R}$, todo aberto é uma união enumerável de fechados: se U é aberto, tome, por exemplo,

$$F_m = \{x \in [-m, m] : d(x, U^c) \geq 1/m\}, \quad U = \bigcup_{m=1}^{\infty} F_m.$$

Logo $\Sigma_1 \subseteq \Sigma_2$; tomando complementos, $\Pi_1 \subseteq \Pi_2$. As inclusões cruzadas seguem repetindo o mesmo conjunto numa união ou interseção enumerável. A indução nas definições de Σ_{n+1} e Π_{n+1} fornece todas as inclusões do item 2.

Registramos também por que nenhuma delas é uma igualdade. No espaço de Baire $\mathcal{N} = \mathbb{N}^{\mathbb{N}}$ existe, para cada n , um conjunto universal $U_n \in \Sigma_n(\mathcal{N} \times \mathcal{N})$: cada conjunto de $\Sigma_n(\mathcal{N})$ é uma seção $(U_n)_x$. Para $n = 1$, basta enumerar os cilindros básicos e usar as coordenadas de x para indicar quais entram na união. Passar de n para $n + 1$ consiste em tomar complementos e uniões enumeráveis; uma única sequência de naturais codifica uma sequência enumerável de parâmetros, pois $\mathcal{N} \cong \mathcal{N}^{\mathbb{N}}$. Essa é uma indução explícita da propriedade universal.

O conjunto diagonal

$$D_n = \{x \in \mathcal{N} : (x, x) \notin U_n\}$$

pertence a Π_n , mas não a Σ_n : se $D_n = (U_n)_y$ para algum y , teríamos $y \in D_n$ se e somente se $y \notin D_n$. Portanto $\Sigma_n \neq \Pi_n$. Como $\mathcal{N}$ é homeomorfo ao conjunto dos

irracionais, uma subclasse G_δ de $\mathbb{R}$, a igualdade de qualquer uma das inclusões em $\mathbb{R}$ forçaria a igualdade correspondente nas classes relativas de $\mathcal{N}$. A contradição diagonal, juntamente com as inclusões já provadas e a passagem aos complementos, mostra que as quatro inclusões são próprias. $\square$

Se $B \subseteq \mathbb{R}$ for boreliano e pertencer a algum nível finito, sua **classificação de Borel finita** é o menor n tal que $B \in \Sigma_n \cup \Pi_n$. Assim, conjuntos abertos ou fechados têm classificação no máximo 1.

Nem todo boreliano tem classificação finita. A hierarquia prossegue pelos ordinais contáveis. Para um sucessor $\alpha + 1$, define-se $\Sigma_{\alpha+1}$ por uniões enumeráveis de conjuntos de Π_α , e $\Pi_{\alpha+1}$ por complementação. Se λ é um ordinal limite contável, Σ_λ é formada pelas uniões enumeráveis de conjuntos que pertencem a classes Π_α com $\alpha < \lambda$, e novamente Π_λ é a classe dos complementos. A mesma construção universal e diagonal usada acima mostra que a hierarquia é própria em cada nível contável; em particular, existem borelianos de classificação não finita. Além disso,

$$\mathcal{B}(\mathbb{R}) = \bigcup_{1 \leq \alpha < \omega_1} \Sigma_\alpha = \bigcup_{1 \leq \alpha < \omega_1} \Pi_\alpha,$$

onde ω_1 é o primeiro ordinal não contável.

Teorema 2.59 — Cardinalidade da σ -álgebra de Borel

Seja $\mathcal{B}$ a σ -álgebra de Borel em $\mathbb{R}$. Então $|\mathcal{B}| = |\mathbb{R}|$.

Demonstração. Como todo singleton é fechado, a aplicação $x \mapsto \{x\}$ injeta $\mathbb{R}$ em $\mathcal{B}(\mathbb{R})$; portanto $|\mathbb{R}| \leq |\mathcal{B}(\mathbb{R})|$. Para a desigualdade contrária, fixe uma base enumerável de intervalos com extremos racionais. Todo boreliano pode ser descrito por um código: uma árvore bem fundada e enumerável cujas folhas são elementos da base e cujos vértices registram complementos ou uniões enumeráveis. Essas árvores podem ser vistas como subconjuntos de $\mathbb{N}^{<\mathbb{N}}$ com rótulos naturais. Há no máximo $2^{\aleph_0} = |\mathbb{R}|$ códigos desse tipo. Logo $|\mathcal{B}(\mathbb{R})| \leq |\mathbb{R}|$, e a igualdade segue. $\square$

Proposição 2.60 —

A cardinalidade da coleção de conjuntos Lebesgue mensuráveis é igual a $|\mathcal{P}(\mathbb{R})|$.

Demonstração. Esta é precisamente a Proposição 2.55. $\square$

Como essa cardinalidade é maior que a cardinalidade da σ -álgebra de Borel, obtemos o seguinte corolário.

Corolário 2.61 —

Existe um conjunto mensurável de Lebesgue que não é um conjunto de Borel.

Demonstração. Pelo Teorema 2.59, existem $|\mathbb{R}|$ borelianos. Pela Proposição 2.55, existem $|\mathcal{P}(\mathbb{R})|$ conjuntos Lebesgue mensuráveis. O teorema de Cantor dá $|\mathcal{P}(\mathbb{R})| > |\mathbb{R}|$, e portanto nem todo conjunto Lebesgue mensurável pode ser boreliano. $\square$

2.6. Medidas em Espaços Métricos

Em espaços métricos separáveis e completos, a topologia permite aproximar conjuntos mensuráveis por abertos e compactos. Essa regularidade será útil sempre que precisarmos controlar medidas por objetos geometricamente mais simples.

Definição 2.62 — Espaço polonês

Um espaço métrico é **polonês** se é separável e admite uma métrica completa que gera sua topologia.

Definição 2.63 — Regularidade e tightness

Seja S um espaço métrico e seja μ uma medida finita em $\mathcal{B}(S)$. Dizemos que μ é **regular** se, para todo $A \in \mathcal{B}(S)$,

$$\mu(A) = \inf\{\mu(U) : A \subseteq U, U \text{ aberto}\} = \sup\{\mu(F) : F \subseteq A, F \text{ fechado}\}.$$

Uma medida de probabilidade μ é **tight** se, para todo $\varepsilon > 0$, existe um compacto $K \subseteq S$ tal que $\mu(K) > 1 - \varepsilon$. Quando a aproximação interna na igualdade anterior pode ser feita por compactos, a medida é também chamada de **medida de Radon**.

Teorema 2.64 — Regularidade de medidas em espaços poloneses

Toda medida de probabilidade boreliana num espaço polonês é regular, tight e de Radon.

Demonstração. Seja (S, d) polonês e escolha um conjunto enumerável denso. Para cada $n \geq 1$, as bolas de raio 2^{-n} centradas nesse conjunto cobrem S . Como $\mu(S) = 1$, uma união finita dessas bolas, denotada por U_n , pode ser escolhida de modo que $\mu(U_n) > 1 - \varepsilon/2^{n+1}$. Seja F_n a união das bolas fechadas com os mesmos centros e raios e defina $K = \bigcap_n F_n$. Então

$$\mu(K) \geq 1 - \sum_{n=1}^{\infty} \mu(F_n^c) > 1 - \varepsilon.$$

O conjunto K é fechado em um espaço completo e, para cada n , pode ser coberto por um número finito de bolas de raio 2^{-n} . Logo é completo e totalmente limitado, portanto compacto. Isso prova a tightness.

Provemos agora a regularidade. Se F é fechado, os abertos $U_n = \{x : d(x, F) < 1/n\}$ decrescem para F . Pela continuidade por cima da medida finita, $\mu(U_n) \downarrow \mu(F)$; logo todo fechado é regular por fora. A finitude mostra, por complementação, que todo aberto é regular por dentro por conjuntos fechados.

Seja $\mathcal{R}$ a classe dos borelianos que são regulares por fora por abertos e por dentro

por fechados. Ela contém abertos e fechados e é fechada por complementos. Se $A_j \in \mathcal{R}$, escolha abertos $U_j \supseteq A_j$ com $\mu(U_j \setminus A_j) < \varepsilon/2^j$; então $\bigcup_j U_j$ dá regularidade externa para $\bigcup_j A_j$. Para a regularidade interna, escolha primeiro m de modo que $\mu(\bigcup_j A_j) - \mu(\bigcup_{j=1}^m A_j) < \varepsilon/2$ e aproxime cada A_j , $j \leq m$, por um fechado; a união finita desses fechados dá o aproximante desejado. Assim $\mathcal{R}$ é uma σ -álgebra que contém os abertos, e portanto $\mathcal{R} = \mathcal{B}(S)$.

Finalmente, dado $A \in \mathcal{B}(S)$, escolha um fechado $F \subseteq A$ com $\mu(A \setminus F) < \varepsilon/2$ e, pela tightness, um compacto K com $\mu(K^c) < \varepsilon/2$. O conjunto $F \cap K$ é compacto, está contido em A e satisfaz $\mu(A \setminus (F \cap K)) < \varepsilon$. Logo a regularidade interna pode ser feita por compactos. $\square$

EXERCÍCIOS

Ex. 2.1 — O primeiro seis: construção da probabilidade. Retome o espaço amostral Ω do Exercício 1.14. Denote por W o conjunto das palavras finitas que terminam no primeiro 6 e, para todo $A \subseteq \Omega$, defina

$$\mathbf{P}(A) = \sum_{w \in A \cap W} 6^{-|w|},$$

onde $|w|$ é o comprimento da palavra w .

1. Prove que $\mathbf{P}$ é uma medida de probabilidade em $(\Omega, \mathcal{P}(\Omega))$.
2. Se T é o instante do primeiro 6, com $T = \infty$ nas trajetórias infinitas, determine a lei de T e prove que $T < \infty$ quase certamente.

Ex. 2.2 — Seja μ_0 uma pré-medida numa álgebra $\mathcal{A}$ de subconjuntos de X , com $X \in \mathcal{A}$ e $\mu_0(X) < \infty$, e seja μ^* a medida exterior induzida por μ_0 . Prove que um conjunto $E \subseteq X$ é mensurável no sentido de Carathéodory se, e somente se,

$$\mu^*(E) + \mu^*(E^c) = \mu_0(X),$$

onde o complemento é tomado em X .

Ex. 2.3 — Medida de Lebesgue–Stieltjes Considere a álgebra

$$\mathcal{A}_0 = f \langle (-\infty, a] : a \in \mathbb{R} \rangle$$

de subconjuntos de $\mathbb{R}$, e seja P uma probabilidade finitamente aditiva em $\mathcal{A}_0$. Defina

$$F(x) = P((-\infty, x]), \quad x \in \mathbb{R}.$$

Prove que P é uma pré-medida em $\mathcal{A}_0$ se, e somente se, F é não decrescente, contínua à direita e satisfaz

$$\lim_{x \rightarrow -\infty} F(x) = 0, \quad \lim_{x \rightarrow +\infty} F(x) = 1.$$

Conclua que, nessas condições, P possui uma única extensão a uma probabilidade em $\mathcal{B}(\mathbb{R})$.

Ex. 2.4 — Subaditividade enumerável. Se $(A_n)_{n \geq 1}$ é uma sequência de eventos, prove que

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

Ex. 2.5 — Continuidade da probabilidade. Sejam $A_1, A_2, \dots$ eventos. Prove as afirmações abaixo.

1. Se $A_n \uparrow A$, então $\mathbf{P}(A_n) \uparrow \mathbf{P}(A)$.
2. Se $A_n \downarrow A$, então $\mathbf{P}(A_n) \downarrow \mathbf{P}(A)$.
- 3.

$$\mathbf{P}(\liminf A_n) \leq \liminf_{n \rightarrow \infty} \mathbf{P}(A_n) \leq \limsup_{n \rightarrow \infty} \mathbf{P}(A_n) \leq \mathbf{P}(\limsup A_n).$$

4. Se $A_n \rightarrow A$ no sentido de que $\liminf A_n = \limsup A_n = A$, então $\mathbf{P}(A_n) \rightarrow \mathbf{P}(A)$.

Ex. 2.6 — Um precursor do primeiro lema de Borel–Cantelli. Se $(A_n)_{n \geq 1}$ é uma sequência de eventos e

$$\mathbf{P}(A_n \text{ i.v.}) = 1,$$

prove que

$$\sum_{n=1}^{\infty} \mathbf{P}(A_n) = \infty.$$

Faça a prova usando apenas subaditividade enumerável e continuidade da probabilidade.

Ex. 2.7 — Considere um espaço de probabilidade cujo espaço amostral é $\mathbb{R}$. Use a continuidade da probabilidade para provar que

1.

$$\mathbf{P}([0, \infty)) = \lim_{n \rightarrow \infty} \mathbf{P}([0, n]);$$

2.

$$\lim_{n \rightarrow \infty} \mathbf{P}([n, \infty)) = 0.$$

Ex. 2.8 — Uma condição insuficiente. Sejam

$$\Omega = \{1, 2, 3, 4\}, \quad \mathcal{A} = \{\{1, 2\}, \{1, 3\}\}.$$

1. Mostre que $\sigma(\mathcal{A}) = \mathcal{P}(\Omega)$.

2. Construa duas probabilidades distintas P_1 e P_2 em $\mathcal{P}(\Omega)$ tais que

$$P_1(E) = P_2(E) \quad \text{para todo } E \in \mathcal{A}.$$

Este exercício mostra que a coincidência em uma família geradora não basta para garantir a igualdade de duas medidas. Compare-o com o Teorema de Unicidade de Medidas, no qual a família geradora é um π -sistema.

Ex. 2.9 — Desigualdades de Bonferroni. Dados eventos $A_1, \dots, A_n$, defina, para $1 \leq j \leq n$,

$$S_j = \sum_{1 \leq i_1 < \dots < i_j \leq n} \mathbf{P}(A_{i_1} \cap \dots \cap A_{i_j}).$$

Prove que, para $1 \leq k \leq n$,

$$\mathbf{P}\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{j=1}^k (-1)^{j-1} S_j \quad \text{se } k \text{ é ímpar,}$$

e

$$\mathbf{P}\left(\bigcup_{i=1}^n A_i\right) \geq \sum_{j=1}^k (-1)^{j-1} S_j \quad \text{se } k \text{ é par.}$$

Se a preguiça estiver particularmente forte, os casos $k = 2$ e $k = 3$ já contêm a ideia central.

Ex. 2.10 — Problema do Pareamento Ao final do dia, n pais chegam a uma escola para buscar seus n filhos. Suponha que as crianças sejam entregues segundo uma permutação uniforme das n crianças entre os n pais. Use a fórmula de inclusão-exclusão para provar que a probabilidade de pelo menos um pai receber o próprio filho é

$$\sum_{k=1}^n \frac{(-1)^{k-1}}{k!} = 1 - \frac{1}{2!} + \frac{1}{3!} - \dots + (-1)^{n-1} \frac{1}{n!}.$$

Mostre que essa probabilidade converge para $1 - e^{-1}$ quando $n \rightarrow \infty$.

Ex. 2.11 — Escolha um ponto uniformemente no interior de um triângulo equilátero. Qual é a probabilidade de que sua distância ao centro do triângulo seja menor do que sua distância à borda? Justifique geometricamente o valor obtido.

Ex. 2.12 — Agulha de Buffon. Um plano é dividido por retas paralelas consecutivas, separadas por uma distância $d > 0$. Uma agulha de comprimento d é lançada ao acaso sobre o plano.

1. Modele precisamente o experimento por meio da distância do centro da agulha à reta mais próxima e do ângulo que a agulha forma com as retas. Especifique o espaço amostral e a medida de probabilidade.
2. Prove que a probabilidade de a agulha interceptar alguma das retas é $2/\pi$. Conclua que a probabilidade de não haver interseção é $1 - 2/\pi$.

Ex. 2.13 — Fixe uma base inteira $b \geq 2$. Expresse, por meio de uniões e interseções enumeráveis de eventos determinados por um número finito de dígitos, o conjunto dos números de $(0,1]$ normais na base b . Conclua que tanto o conjunto dos números normais quanto o dos números não normais na base b pertencem a $\mathcal{B}((0,1])$. Mostre também que o conjunto dos números absolutamente normais é boreliano.

Ex. 2.14 — Prove que todo subconjunto enumerável, todo subconjunto coenumerável e todo subconjunto perfeito de $(0,1]$ é boreliano. Conclua, em particular, que o conjunto dos irracionais de $(0,1]$ é boreliano.

Ex. 2.15 — Seja (Ω, d) um espaço métrico. Recorde que Ω é *separável* quando possui um subconjunto enumerável e denso.

1. Mostre que

$$\sigma\langle \text{bolas abertas de } \Omega \rangle \subseteq \mathcal{B}(\Omega).$$

2. Se Ω é separável, prove a inclusão inversa e conclua que as bolas abertas geram $\mathcal{B}(\Omega)$.
3. Mostre que $\mathbb{Q}$ é denso em $\mathbb{R}$ com a métrica usual.
4. Conclua que $\sigma\langle \text{bolas abertas de } \mathbb{R} \rangle = \mathcal{B}(\mathbb{R})$.

Ex. 2.16 — Sejam μ_1 e μ_2 medidas em $\sigma\langle \mathcal{F}_0 \rangle$.

1. Suponha que $\mathcal{F}_0$ seja uma álgebra, que ambas as medidas sejam σ -finitas em $\mathcal{F}_0$ e que

$$\mu_1(A) \leq \mu_2(A) \quad \text{para todo } A \in \mathcal{F}_0. \quad (2.21)$$

Prove que a mesma desigualdade vale para todo $A \in \sigma\langle \mathcal{F}_0 \rangle$.

Dica. Trate primeiro o caso em que $\mu_1 + \mu_2$ é finita e use aproximação em diferença simétrica por elementos da álgebra. Depois localize o argumento numa exaustão de medida finita.

2. Construa contraexemplos mostrando que a conclusão pode falhar nas situações abaixo:
 - a)) $\mathcal{F}_0$ é uma álgebra, mas as medidas não são σ -finitas em $\mathcal{F}_0$;
 - b)) as medidas são finitas, mas $\mathcal{F}_0$ é apenas um π -sistema gerador, e não uma álgebra.

Ex. 2.17 — Princípio de Littlewood. Seja $\mathcal{F}_0$ uma álgebra, e seja μ uma medida σ -finita em $\mathcal{F}_0$, definida em $\sigma\langle \mathcal{F}_0 \rangle$.

1. Dados $B \in \sigma\langle \mathcal{F}_0 \rangle$ e $\varepsilon > 0$, prove que existem conjuntos dois a dois disjuntos $A_1, A_2, \dots \in \mathcal{F}_0$ tais que

$$B \subseteq \bigcup_{n=1}^{\infty} A_n, \quad \mu \left(\left(\bigcup_{n=1}^{\infty} A_n \right) \setminus B \right) \leq \varepsilon.$$

2. Se, além disso, $\mu(B) < \infty$, prove que existe $A \in \mathcal{F}_0$ tal que $\mu(A \triangle B) \leq \varepsilon$.

3. Dê um exemplo mostrando que a conclusão do item anterior pode falhar quando $\mu(B) = \infty$.

Ex. 2.18 — Seja I um conjunto arbitrário e $f : I \rightarrow [0, \infty]$. Defina

$$\sum_{i \in I} f(i) := \sup \left\{ \sum_{i \in H} f(i) : H \subseteq I \text{ é finito} \right\}.$$

Prove as afirmações abaixo.

1. Se $I = \bigcup_{k \in K} I_k$ é uma partição de I em conjuntos não vazios, então

$$\sum_{i \in I} f(i) = \sum_{k \in K} \left(\sum_{i \in I_k} f(i) \right),$$

onde a soma sobre K é entendida pela mesma definição.

2. Se I é enumerável e $I = \{i_1, i_2, \dots\}$, então

$$\sum_{i \in I} f(i) = \lim_{n \rightarrow \infty} \sum_{m=1}^n f(i_m).$$

3. Mostre que

$$\mu(A) := \sum_{a \in A} f(a), \quad A \subseteq I,$$

define uma medida em $(I, \mathcal{P}(I))$.

Ex. 2.19 — Se $\mathcal{K}$ é uma classe compacta, denote por $\widehat{\mathcal{K}}$ a menor família de subconjuntos de Ω que contém $\mathcal{K}$ e é fechada sob uniões finitas e interseções enumeráveis. Prove que $\widehat{\mathcal{K}}$ também é uma classe compacta.

Dica. Mostre primeiro que todo elemento de $\widehat{\mathcal{K}}$ é uma interseção enumerável de uniões finitas de elementos de $\mathcal{K}$; depois use a árvore finitamente ramificada das escolhas possíveis.

Ex. 2.20 — No espaço $\Omega = \{0,1\}^{\mathbb{N}}$, seja $\mathcal{F}_n$ a família dos eventos cuja ocorrência é determinada pelas primeiras n coordenadas. Prove que $\mathcal{F}_n$ é uma σ -álgebra, identificando-a com a família das imagens inversas de subconjuntos de $\{0,1\}^n$ pela projeção nas primeiras n coordenadas.

Ex. 2.21 — Com a notação do exercício anterior, prove que

$$\mathcal{F}_0 = \bigcup_{n \geq 1} \mathcal{F}_n$$

é uma álgebra. Explique também por que esse argumento não prova fechamento sob uniões enumeráveis.

Ex. 2.22 — Probabilidade enumerável-coenumerável. Seja Ω um conjunto não enumerável e defina

$$\mathcal{F} = \{A \subseteq \Omega : A \text{ é enumerável ou } A^c \text{ é enumerável}\}, \quad \mathbf{P}(A) = \begin{cases} 0, & A \text{ enumerável,} \\ 1, & A^c \text{ enumerável.} \end{cases}$$

Prove que $\mathcal{F}$ é uma σ -álgebra e que $\mathbf{P}$ é uma medida de probabilidade bem definida em $\mathcal{F}$.

Independência

Dois eventos são independentes quando conhecer a ocorrência de um deles não altera a probabilidade do outro. Antes de transformar essa ideia em uma definição, precisamos formalizar o efeito de uma informação parcial sobre o modelo probabilístico.

Suponha que observamos um evento B , com $\mathbf{P}(B) > 0$. Os resultados fora de B deixam de ser possíveis. Dentro de B , porém, queremos preservar as proporções anteriores: se $A \cap B$ era duas vezes mais provável que $C \cap B$, essa razão não deve mudar apenas porque descobrimos que B ocorreu. Isso nos leva a restringir o espaço a B e renormalizar a massa restante.

Definição 3.1 — Probabilidade condicional

Dado um espaço de probabilidade $(\Omega, \mathcal{F}, \mathbf{P})$ e eventos $A, B \in \mathcal{F}$, com $\mathbf{P}(B) > 0$, a **probabilidade condicional** de A dado B é

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

A definição satisfaz exatamente as propriedades sugeridas pela interpretação anterior:

$$\mathbf{P}(B \mid B) = 1,$$

e, se $A, C \in \mathcal{F}$ e $\mathbf{P}(C \cap B) > 0$,

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

Assim, condicionar em B preserva as proporções internas e transforma B no novo evento certo.

A probabilidade condicional é, de fato, uma nova probabilidade no espaço amostral B . O próximo resultado formaliza essa interpretação.

Teorema 3.2 —

Dado um espaço de probabilidade $(\Omega, \mathcal{F}, \mathbf{P})$ e $B \in \mathcal{F}$ com $\mathbf{P}(B) > 0$, defina a σ -álgebra traço

$$\mathcal{F}_B = \{A \cap B : A \in \mathcal{F}\}.$$

Então

$$(B, \mathcal{F}_B, \mathbf{P}(\cdot \mid B))$$

é um espaço de probabilidade.

Demonstração. A família $\mathcal{F}_B$ contém $B = \Omega \cap B$. Se $D = A \cap B \in \mathcal{F}_B$, então seu complemento relativo a B é

$$B \setminus D = A^c \cap B \in \mathcal{F}_B.$$

Além disso, se $D_n = A_n \cap B \in \mathcal{F}_B$, então

$$\bigcup_{n \geq 1} D_n = \left(\bigcup_{n \geq 1} A_n \right) \cap B \in \mathcal{F}_B.$$

Logo, $\mathcal{F}_B$ é uma σ -álgebra em B .

Para $D \in \mathcal{F}_B$, ponha

$$\mathbf{P}_B(D) = \frac{\mathbf{P}(D)}{\mathbf{P}(B)}.$$

Como $D \subseteq B$, essa expressão coincide com $\mathbf{P}(D \mid B)$. Temos $\mathbf{P}_B(B) = 1$. Se $D_1, D_2, \dots \in \mathcal{F}_B$ são dois a dois disjuntos, então

$$\mathbf{P}_B\left(\bigcup_{n \geq 1} D_n\right) = \frac{1}{\mathbf{P}(B)} \sum_{n \geq 1} \mathbf{P}(D_n) = \sum_{n \geq 1} \mathbf{P}_B(D_n).$$

Portanto, $\mathbf{P}_B$ é uma probabilidade em $(B, \mathcal{F}_B)$. □

Exemplo 3.3 — Inspeção automática com tempo máximo.

Uma inspeção automática pode terminar antes do prazo ou, se a análise não for concluída antecipadamente, permanecer ativa até o tempo máximo de 40 minutos. Seja T sua duração. Suponha que

$$\mathbf{P}(T < t) = \frac{t}{60}, \quad 0 \leq t \leq 40.$$

Sabendo que a inspeção ainda está em andamento após 30 minutos, qual é a probabilidade de ela permanecer ativa até os 40 minutos?

Seja $F = \{T = 40\}$ o evento de a inspeção atingir o tempo máximo. Como

$$\mathbf{P}(T < 40) = \frac{40}{60} = \frac{2}{3},$$

temos

$$\mathbf{P}(F) = 1 - \frac{2}{3} = \frac{1}{3}.$$

Por outro lado,

$$\mathbf{P}(T \geq 30) = 1 - \mathbf{P}(T < 30) = 1 - \frac{30}{60} = \frac{1}{2}.$$

Como $F \subseteq \{T \geq 30\}$,

$$\mathbf{P}(F \mid T \geq 30) = \frac{\mathbf{P}(F)}{\mathbf{P}(T \geq 30)} = \frac{1/3}{1/2} = \frac{2}{3}.$$

O modelo possui, portanto, uma parte contínua antes de 40 minutos e uma massa de probabilidade no próprio tempo máximo: toda inspeção que não termina antes do limite fica acumulada em $T = 40$.

◀

A regra seguinte é uma das maneiras mais úteis de desfazer um condicionamento. Se o espaço é dividido em casos mutuamente exclusivos, calculamos a probabilidade de A em cada caso e depois fazemos a média ponderada desses valores.

Teorema 3.4 – Lei da probabilidade total

Se $\{B_1, B_2, \dots\}$ é uma partição enumerável de Ω por eventos de probabilidade positiva, então, para todo $A \in \mathcal{F}$,

$$\mathbf{P}(A) = \sum_{i \geq 1} \mathbf{P}(A \mid B_i) \mathbf{P}(B_i).$$

Demonstração. Como $\{B_i\}_{i \geq 1}$ é uma partição de Ω ,

$$A = A \cap \Omega = A \cap \left(\bigcup_{i \geq 1} B_i \right) = \bigcup_{i \geq 1} (A \cap B_i).$$

Essa união é disjunta. Pela σ -aditividade,

$$\mathbf{P}(A) = \sum_{i \geq 1} \mathbf{P}(A \cap B_i).$$

Por definição de probabilidade condicional,

$$\mathbf{P}(A \cap B_i) = \mathbf{P}(A \mid B_i) \mathbf{P}(B_i),$$

e portanto

$$\mathbf{P}(A) = \sum_{i \geq 1} \mathbf{P}(A \mid B_i) \mathbf{P}(B_i).$$

□

A mesma identidade de interseção pode ser lida na direção oposta:

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

Essa observação fornece a fórmula de Bayes.

Teorema 3.5 — Fórmula de Bayes

Se $A, B \in \mathcal{F}$, com $P(A) > 0$ e $P(B) > 0$, então

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

Demonstração. Pela definição de probabilidade condicional,

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

Basta dividir por $P(A)$. □

Quando as hipóteses possíveis formam uma partição, a lei da probabilidade total substitui o denominador por uma soma explícita.

Teorema 3.6 — Fórmula de Bayes para uma partição

Se $A_1, \dots, A_n$ formam uma partição de Ω , $P(A_i) > 0$ para todo i , e $B \in \mathcal{F}$ satisfaz $P(B) > 0$, então

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

Demonstração. Pela definição de probabilidade condicional,

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

Por outro lado, a lei da probabilidade total aplicada a B dá

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

Substituindo essa expressão no denominador, obtemos a fórmula. □

3.1. Independência

Agora podemos voltar à pergunta que motivou o capítulo. Quando a informação de que B ocorreu não muda a probabilidade de A , temos

$$\mathbf{P}(A \mid B) = \mathbf{P}(A).$$

Essa é a leitura mais intuitiva de independência, embora ainda não seja a melhor definição formal.

Definição 3.7 – Eventos independentes – definição provisória

Em um espaço de probabilidade $(\Omega, \mathcal{F}, \mathbf{P})$, sejam $A, B \in \mathcal{F}$ com $\mathbf{P}(A) > 0$ e $\mathbf{P}(B) > 0$. Dizemos provisoriamente que A e B são independentes quando

$$\mathbf{P}(A \mid B) = \mathbf{P}(A).$$

A aparente assimetria é apenas notacional. Se $\mathbf{P}(A \mid B) = \mathbf{P}(A) > 0$, então

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

Além disso,

$$\mathbf{P}(A \mid B) = \mathbf{P}(A) \iff \mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B).$$

A igualdade à direita continua fazendo sentido mesmo quando um dos eventos tem probabilidade zero e torna a simetria explícita. Por isso, ela é adotada como definição.

Definição 3.8 – Eventos independentes

Dois eventos $A, B \in \mathcal{F}$ são **independentes** quando

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

Proposição 3.9 —

Um evento A é independente dele mesmo se, e somente se,

$$\mathbf{P}(A) \in \{0,1\}.$$

Demonstração. A independência de A consigo mesmo equivale a

$$\mathbf{P}(A) = \mathbf{P}(A \cap A) = \mathbf{P}(A)^2.$$

Como $0 \leq \mathbf{P}(A) \leq 1$, isso ocorre exatamente para $\mathbf{P}(A) = 0$ ou $\mathbf{P}(A) = 1$. □

Exemplo 3.10 — Três eventos independentes dois a dois.

Considere dois dados justos e

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

com os 36 resultados equiprováveis. Defina

$$A = \{\text{o primeiro dado mostra } 2\} = \{(2,j) : j = 1, \dots, 6\},$$

$$B = \{\text{o segundo dado mostra } 4\} = \{(i,4) : i = 1, \dots, 6\},$$

$$C = \{\text{a soma é } 7\} = \{(i,j) \in \Omega : i + j = 7\}.$$

Temos

$$\mathbf{P}(A) = \mathbf{P}(B) = \mathbf{P}(C) = \frac{1}{6}$$

e

$$\mathbf{P}(A \cap B) = \mathbf{P}(A \cap C) = \mathbf{P}(B \cap C) = \frac{1}{36}.$$

Logo, cada par é independente.

Entretanto, os três eventos não são independentes como família. De fato, $A \cap B$ força a soma a ser 6, de modo que

$$\mathbf{P}(C \mid A \cap B) = 0 \neq \mathbf{P}(C).$$

A independência de cada par não impede, portanto, uma dependência que só aparece quando os três eventos são considerados juntos.



O exemplo mostra que, para uma coleção de eventos, devemos exigir a fatoração em toda subcoleção finita, não apenas em cada par. A igualdade para a interseção de todos os eventos também não é suficiente, como mostra o Exercício 3.2.

Definição 3.11 — Família de eventos independentes

Uma família $\{A_k\}_{k \in K} \subseteq \mathcal{F}$ é **independente** se, para todo subconjunto finito não vazio $H \subseteq K$,

$$\mathbf{P}\left(\bigcap_{h \in H} A_h\right) = \prod_{h \in H} \mathbf{P}(A_h).$$

Exemplo 3.12 — Dois lançamentos de uma moeda.

Lance uma moeda justa duas vezes. Seja A o evento “o primeiro lançamento é cara” e B o evento “o segundo lançamento é coroa”. Então

$$\mathbf{P}(A) = \mathbf{P}(B) = \frac{1}{2}, \quad \mathbf{P}(A \cap B) = \frac{1}{4},$$

e, portanto, A e B são independentes.



Exemplo 3.13 — Amostragem com e sem reposição.

Uma urna contém cinco bolas brancas e cinco pretas. Retiramos duas bolas.

Se a retirada é feita sem reposição, sejam

$$A = \{\text{a primeira bola é branca}\}, \quad B = \{\text{a segunda bola é branca}\}.$$

Temos $\mathbf{P}(B) = 1/2$, por simetria, mas

$$\mathbf{P}(B \mid A) = \frac{4}{9}.$$

Logo, A e B não são independentes.

Com reposição, a composição da urna volta a ser a mesma antes da segunda retirada. Nesse caso,

$$\mathbf{P}(B \mid A) = \mathbf{P}(B) = \frac{1}{2},$$

e os dois eventos são independentes.



A independência também pode fazer parte da própria construção de um modelo. O exemplo seguinte mostra isso no caso clássico de infinitas moedas.

Exemplo 3.14 — Infinitas moedas.

Queremos construir um espaço de probabilidade para uma sequência infinita de lançamentos independentes de uma moeda justa. Tome

$$\Omega = \{0,1\}^{\mathbb{N}},$$

onde 1 representa cara e 0, coroa. Para $n \geq 1$, seja $\mathcal{F}_n$ a σ -álgebra dos eventos determinados pelas primeiras n coordenadas, e ponha

$$\mathcal{F}_0 = \bigcup_{n \geq 1} \mathcal{F}_n, \quad \mathcal{F} = \sigma\langle \mathcal{F}_0 \rangle.$$

A família $\mathcal{F}_0$ é uma álgebra.

Para $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n) \in \{0,1\}^n$, defina o cilindro

$$C(\varepsilon) = \{\omega \in \Omega : \omega_k = \varepsilon_k, k = 1, \dots, n\}.$$

Se os lançamentos devem ser justos e independentes, a probabilidade desse cilindro é necessariamente

$$\mathbf{P}_0(C(\varepsilon)) = 2^{-n}.$$

Essas massas são consistentes quando um cilindro é descrito em níveis diferentes, pois

$$2^{-n} = 2^{-(n+1)} + 2^{-(n+1)}.$$

Por aditividade finita, elas determinam uma probabilidade finitamente aditiva $\mathbf{P}_0$ em $\mathcal{F}_0$.

A construção por classes compactas do capítulo anterior mostra que $\mathbf{P}_0$ é uma pré-medida. O teorema de extensão de Carathéodory fornece, portanto, uma única probabilidade $\mathbf{P}$ em $\mathcal{F}$ que estende essas massas cilíndricas.

Agora, se

$$H_k = \{\omega : \omega_k = 1\},$$

então $\mathbf{P}(H_k) = 1/2$. Para índices distintos $i_1, \dots, i_r$,

$$\mathbf{P}(H_{i_1} \cap \dots \cap H_{i_r}) = 2^{-r} = \prod_{j=1}^r \mathbf{P}(H_{i_j}),$$

de modo que os lançamentos são independentes. A independência, portanto, não foi usada circularmente para definir a medida: ela é uma consequência das probabilidades cilíndricas com que a medida foi construída.

◀

A independência de uma família finita estende-se de maneira natural a uma interseção enumerável.

Proposição 3.15 —

Se $E_1, E_2, \dots$ são eventos independentes, então

$$\mathbf{P}\left(\bigcap_{i \geq 1} E_i\right) = \prod_{i \geq 1} \mathbf{P}(E_i),$$

onde o produto infinito é o limite dos produtos parciais.

Demonstração. Os eventos

$$D_n = \bigcap_{i=1}^n E_i$$

formam uma sequência decrescente. Pela continuidade da probabilidade por cima,

$$\mathbf{P}\left(\bigcap_{i \geq 1} E_i\right) = \lim_{n \rightarrow \infty} \mathbf{P}(D_n) = \lim_{n \rightarrow \infty} \prod_{i=1}^n \mathbf{P}(E_i).$$

◻

Também podemos complementar qualquer número finito de eventos sem destruir a independência.

Proposição 3.16 —

Se $\{E_i : i \in I\}$ é uma família de eventos independentes e, para cada $i \in I$,

$$F_i \in \{E_i, E_i^c\},$$

então $\{F_i : i \in I\}$ é independente.

Demonstração. Basta verificar a identidade de independência em uma subfamília finita. Fixe índices distintos $i_1, \dots, i_n$. Partimos de $E_{i_1}, \dots, E_{i_n}$ e substituímos, um a um, os eventos que devem ser complementados.

Suponha que todos os fatores, exceto talvez o primeiro, já tenham sido escolhidos como evento ou complemento, e denote por D a interseção desses $n-1$ fatores. Pela hipótese de indução,

$$\mathbf{P}(E_{i_1} \cap D) = \mathbf{P}(E_{i_1})\mathbf{P}(D).$$

Consequentemente,

$$\mathbf{P}(E_{i_1}^c \cap D) = \mathbf{P}(D) - \mathbf{P}(E_{i_1} \cap D) = (1 - \mathbf{P}(E_{i_1}))\mathbf{P}(D).$$

Isso substitui E_{i_1} por seu complemento preservando a fatoração. Repetindo o argumento, obtemos a identidade para qualquer escolha de complementos. $\square$

A independência pode até transformar um problema de teoria dos números em um cálculo probabilístico.

Teorema 3.17 — Fórmula dos números primos de Euler

Para $s > 1$,

$$\zeta(s) := \sum_{n \geq 1} \frac{1}{n^s} = \prod_{p \in \mathbb{P}} \frac{1}{1 - p^{-s}},$$

onde $\mathbb{P}$ denota o conjunto dos números primos.

Demonstração. Como $s > 1$, a série que define $\zeta(s)$ converge. Em $\Omega = \mathbb{N}$, defina

$$\mathbf{P}(\{n\}) = \frac{1}{\zeta(s)n^s}.$$

Para $k \geq 1$, seja

$$E_k = \{n \in \mathbb{N} : k \text{ divide } n\}.$$

Então

$$\mathbf{P}(E_k) = \sum_{j \geq 1} \frac{1}{\zeta(s)(jk)^s} = k^{-s}.$$

Se $p_1, \dots, p_r$ são primos distintos,

$$E_{p_1} \cap \dots \cap E_{p_r} = E_{p_1 \dots p_r},$$

e, portanto,

$$\mathbf{P}(E_{p_1} \cap \dots \cap E_{p_r}) = \prod_{j=1}^r p_j^{-s} = \prod_{j=1}^r \mathbf{P}(E_{p_j}).$$

Logo, os eventos E_p , $p \in \mathbb{P}$, são independentes.

O único inteiro positivo não divisível por nenhum primo é 1. Pela proposição anterior e pela fórmula para interseções enumeráveis,

$$\frac{1}{\zeta(s)} = \mathbf{P}(\{1\}) = \mathbf{P}\left(\bigcap_{p \in \mathbb{P}} E_p^c\right) = \prod_{p \in \mathbb{P}} (1 - p^{-s}).$$

Tomando recíprocos, obtemos a fórmula. □

Para trabalhar com informação gerada por coleções inteiras de eventos, precisamos de uma versão da independência que não dependa de escolher um único evento por índice.

Definição 3.18 — Famílias independentes

Dada uma família $\mathcal{A}_k \subseteq \mathcal{F}$, $k \in K$, dizemos que $\{\mathcal{A}_k\}_{k \in K}$ é uma família de **classes independentes** se, para todo subconjunto finito não vazio $H \subseteq K$

e toda escolha $A_k \in \mathcal{A}_k, k \in H$,

$$\mathbf{P}\left(\bigcap_{k \in H} A_k\right) = \prod_{k \in H} \mathbf{P}(A_k).$$

Teorema 3.19 —

Sejam $\{\mathcal{A}_k\}_{k \in K}$ famílias de subconjuntos de $\mathcal{F}$.

1 Se $\mathcal{A}_k \subseteq \mathcal{B}_k \subseteq \mathcal{F}$ para todo k e $\{\mathcal{B}_k\}_{k \in K}$ é independente, então $\{\mathcal{A}_k\}_{k \in K}$ é independente.

2 A família $\{\mathcal{A}_k\}_{k \in K}$ é independente se, e somente se,

$$\{\mathcal{A}_k \cup \{\Omega\}\}_{k \in K}$$

é independente.

3 Se $\mathcal{A}_1, \dots, \mathcal{A}_n$ contêm Ω , então são independentes se, e somente se,

$$\mathbf{P}\left(\bigcap_{k=1}^n A_k\right) = \prod_{k=1}^n \mathbf{P}(A_k)$$

para toda escolha $A_k \in \mathcal{A}_k$.

Demonstração. A primeira afirmação é imediata: qualquer escolha feita nas $\mathcal{A}_k$ também é admissível nas $\mathcal{B}_k$.

Acrescentar Ω não altera nenhuma identidade, pois $A \cap \Omega = A$ e $\mathbf{P}(\Omega) = 1$. A recíproca segue da primeira afirmação.

Na terceira afirmação, a identidade exibida para todos os n índices implica as identidades de qualquer subcoleção: basta escolher $A_j = \Omega$ nos índices que desejamos omitir. A implicação contrária faz parte da definição. $\square$

A mesma noção aplica-se diretamente a σ -álgebras.

Definição 3.20 — σ -álgebras independentes

Uma família de sub- σ -álgebras $\{\mathcal{G}_k\}_{k \in K}$ de $\mathcal{F}$ é independente se, para todo subconjunto finito não vazio $H \subseteq K$ e toda escolha $G_k \in \mathcal{G}_k, k \in H$,

$$\mathbf{P}\left(\bigcap_{k \in H} G_k\right) = \prod_{k \in H} \mathbf{P}(G_k).$$

O próximo resultado é a ferramenta que permite passar da independência de geradores à independência da informação que eles geram.

Teorema 3.21 — Independência das σ -álgebras geradas

Suponha que $\{\mathcal{A}_k\}_{k \in K}$ seja uma família independente e que cada $\mathcal{A}_k$ seja um π -sistema. Então

$$\{\sigma\langle \mathcal{A}_k \rangle\}_{k \in K}$$

é uma família de σ -álgebras independentes.

Demonstração. Pelo resultado anterior, podemos substituir cada $\mathcal{A}_k$ por $\mathcal{A}_k \cup \{\Omega\}$. As famílias continuam independentes, continuam sendo π -sistemas e agora contêm Ω .

Considere primeiro apenas m índices, renomeados como $1, \dots, m$. Fixe $A_j \in \mathcal{A}_j, j < m$, e defina

$$\mathcal{D} = \left\{ B \in \mathcal{F} : \mathbf{P}\left(B \cap \bigcap_{j=1}^{m-1} A_j\right) = \mathbf{P}(B) \prod_{j=1}^{m-1} \mathbf{P}(A_j) \right\}.$$

A independência original mostra que $\mathcal{A}_m \subseteq \mathcal{D}$. Vamos verificar que $\mathcal{D}$ é um λ -sistema.

Primeiro, como as classes são independentes,

$$\mathbf{P}\left(\bigcap_{j=1}^{m-1} A_j\right) = \prod_{j=1}^{m-1} \mathbf{P}(A_j),$$

de modo que $\Omega \in \mathcal{D}$.

Agora suponha $B \in \mathcal{D}$. Escrevendo

$$D_0 = \bigcap_{j=1}^{m-1} A_j,$$

temos

$$\mathbf{P}(B^c \cap D_0) = \mathbf{P}(D_0) - \mathbf{P}(B \cap D_0).$$

Como $B \in \mathcal{D}$,

$$\mathbf{P}(B \cap D_0) = \mathbf{P}(B) \prod_{j=1}^{m-1} \mathbf{P}(A_j),$$

enquanto

$$\mathbf{P}(D_0) = \prod_{j=1}^{m-1} \mathbf{P}(A_j).$$

Portanto,

$$\mathbf{P}(B^c \cap D_0) = \mathbf{P}(B^c) \prod_{j=1}^{m-1} \mathbf{P}(A_j),$$

e $B^c \in \mathcal{D}$.

Finalmente, se $B_1, B_2, \dots \in \mathcal{D}$ são dois a dois disjuntos, então

$$\mathbf{P}\left(\left(\bigcup_{n \geq 1} B_n\right) \cap D_0\right) = \sum_{n \geq 1} \mathbf{P}(B_n \cap D_0)$$

e, como cada $B_n \in \mathcal{D}$,

$$\sum_{n \geq 1} \mathbf{P}(B_n \cap D_0) = \left(\sum_{n \geq 1} \mathbf{P}(B_n)\right) \prod_{j=1}^{m-1} \mathbf{P}(A_j).$$

Pela disjunção,

$$\sum_{n \geq 1} \mathbf{P}(B_n) = \mathbf{P}\left(\bigcup_{n \geq 1} B_n\right),$$

logo $\bigcup_n B_n \in \mathcal{D}$. Assim, $\mathcal{D}$ é um λ -sistema.

Pelo teorema π - λ ,

$$\sigma\langle \mathcal{A}_m \rangle \subseteq \mathcal{D}.$$

Portanto, podemos substituir a última classe por sua σ -álgebra gerada sem perder a independência.

Repetindo o mesmo argumento uma classe de cada vez, substituímos todas as $\mathcal{A}_j$ por $\sigma\langle\mathcal{A}_j\rangle$. Como a definição de independência envolve apenas subcoleções finitas, o argumento se aplica a um conjunto de índices K arbitrário. $\square$

Proposição 3.22 —

Se $\{A_k : k \in K\}$ é uma família de eventos independentes e $I, J \subseteq K$ são disjuntos, então

$$\sigma\langle A_k : k \in I \rangle \quad \text{e} \quad \sigma\langle A_k : k \in J \rangle$$

são independentes.

Demonstração. Considere os π -sistemas

$$\mathcal{C}_I = \left\{ \bigcap_{i \in H} A_i : H \subseteq I \text{ finito} \right\}, \quad \mathcal{C}_J = \left\{ \bigcap_{j \in L} A_j : L \subseteq J \text{ finito} \right\},$$

com a convenção de que a interseção vazia é Ω . Se

$$C = \bigcap_{i \in H} A_i \in \mathcal{C}_I, \quad D = \bigcap_{j \in L} A_j \in \mathcal{C}_J,$$

então, como $I \cap J = \emptyset$,

$$\mathbf{P}(C \cap D) = \prod_{i \in H} \mathbf{P}(A_i) \prod_{j \in L} \mathbf{P}(A_j) = \mathbf{P}(C)\mathbf{P}(D).$$

Logo, $\mathcal{C}_I$ e $\mathcal{C}_J$ são independentes. O teorema anterior fornece a independência das σ -álgebras que elas geram. $\square$

3.2. Lei 0–1 de Kolmogorov e lemas de Borel–Cantelli

Dada uma sequência de eventos $A_1, A_2, \dots$, há perguntas que não dependem de nenhum trecho inicial da sequência. Elas são determinadas apenas pelo que acontece arbitrariamente longe.

Definição 3.23 — σ -álgebra caudal

A **σ -álgebra caudal** da sequência $A_1, A_2, \dots$ é

$$\mathcal{T} = \bigcap_{n \geq 1} \sigma\langle A_n, A_{n+1}, A_{n+2}, \dots \rangle.$$

Seus elementos são chamados de **eventos caudais**.

Se o índice representa o tempo, $\sigma\langle A_n, A_{n+1}, \dots \rangle$ contém a informação disponível depois de descartarmos os primeiros $n - 1$ eventos. A interseção sobre todos os n retém apenas o que não depende de nenhum início finito.

Exemplo 3.24 — .

Em uma sequência infinita de lançamentos de moeda, seja H_n o evento “o n -ésimo lançamento é cara”. São eventos caudais, por exemplo:

- ☐ ocorrerem infinitas caras;
- ☐ ocorrerem apenas finitas coroas;
- ☐ ocorrerem infinitas caras nos lançamentos pares;
- ☐ aparecerem infinitas vezes blocos de 100 caras consecutivas.

Alterar um número finito de lançamentos não muda nenhum desses eventos.

**Teorema 3.25 — Lei 0–1 de Kolmogorov**

Se $A_1, A_2, \dots$ são eventos independentes e $\mathcal{T}$ é sua σ -álgebra caudal, então todo $A \in \mathcal{T}$ satisfaz

$$\mathbf{P}(A) \in \{0, 1\}.$$

Demonstração. Fixe $A \in \mathcal{T}$. Para cada n , a proposição anterior mostra que

$$\sigma\langle A_1, \dots, A_{n-1} \rangle \quad \text{e} \quad \sigma\langle A_n, A_{n+1}, \dots \rangle$$

são independentes. Como A pertence à segunda σ -álgebra, toda subfamília finita de

$$A_1, A_2, \dots, A$$

é independente: basta escolher n maior que todos os índices finitos envolvidos.

Logo, a família $\{A_1, A_2, \dots, A\}$ é independente. Aplicando novamente a proposição anterior, concluímos que

$$\sigma\langle A_1, A_2, \dots \rangle \quad \text{e} \quad \sigma\langle A \rangle$$

são independentes. Mas

$$A \in \mathcal{T} \subseteq \sigma\langle A_1, A_2, \dots \rangle \quad \text{e} \quad A \in \sigma\langle A \rangle.$$

Portanto, A é independente de si mesmo. Pela proposição anterior sobre autoindependência, $\mathbf{P}(A) \in \{0, 1\}$. □

Corolário 3.26 — Lei 0–1 para σ -álgebras independentes

Se $(\mathcal{G}_n)_{n \geq 1}$ é uma sequência de σ -álgebras independentes e

$$\mathcal{T} = \bigcap_{n \geq 1} \sigma\langle \mathcal{G}_n, \mathcal{G}_{n+1}, \dots \rangle,$$

então todo $A \in \mathcal{T}$ satisfaz $\mathbf{P}(A) \in \{0, 1\}$.

Demonstração. Se $I, J \subseteq \mathbb{N}$ são disjuntos, considere as classes formadas por interseções finitas de eventos pertencentes às $\mathcal{G}_i, i \in I$, e, respectivamente, às $\mathcal{G}_j, j \in J$. Pela independência da família $(\mathcal{G}_n)$, essas duas classes são π -sistemas independentes; pelo Teorema 3.21, as σ -álgebras que elas geram também são independentes. Em particular,

$$\sigma\langle \mathcal{G}_1, \dots, \mathcal{G}_{n-1} \rangle \quad \text{e} \quad \sigma\langle \mathcal{G}_n, \mathcal{G}_{n+1}, \dots \rangle$$

são independentes para todo n .

Fixe $A \in \mathcal{T}$. Então A é independente de cada $\sigma\langle \mathcal{G}_1, \dots, \mathcal{G}_{n-1} \rangle$. Essas σ -álgebras crescem com n , e sua união é um π -sistema que gera $\sigma\langle \mathcal{G}_1, \mathcal{G}_2, \dots \rangle$. Para A fixo, a classe

$$\mathcal{D}_A = \{B \in \mathcal{F} : \mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B)\}$$

é um λ -sistema. Como ela contém aquela união, o teorema π - λ mostra que A é independente de toda $\sigma\langle \mathcal{G}_1, \mathcal{G}_2, \dots \rangle$. Como $A \in \mathcal{T} \subseteq \sigma\langle \mathcal{G}_1, \mathcal{G}_2, \dots \rangle$, segue que A é independente de si mesmo; logo $\mathbf{P}(A) \in \{0, 1\}$. $\square$

Para formular os lemas de Borel–Cantelli, precisamos distinguir dois tipos de comportamento assintótico.

Definição 3.27 – Limite superior e inferior de eventos

Para eventos $A_1, A_2, \dots$, definimos

$$\limsup_{n \rightarrow \infty} A_n = \bigcap_{m \geq 1} \bigcup_{n \geq m} A_n,$$

o evento em que A_n ocorre infinitas vezes, e

$$\liminf_{n \rightarrow \infty} A_n = \bigcup_{m \geq 1} \bigcap_{n \geq m} A_n,$$

o evento em que, a partir de algum índice, todos os A_n ocorrem.

Escreveremos A_n i.v. para $\limsup_{n \rightarrow \infty} A_n$. Para o $\liminf$, preferimos a descrição “eventualmente sempre”, evitando confundi-la com a abreviação “q.s.” de “quase seguramente” usada em outros capítulos.

Teorema 3.28 – Primeiro lema de Borel–Cantelli

Se $E_1, E_2, \dots \in \mathcal{F}$ e

$$\sum_{n \geq 1} \mathbf{P}(E_n) < \infty,$$

então

$$\mathbf{P}(E_n \text{ i.v.}) = 0.$$

Demonstração. Pela definição de limite superior,

$$\{E_n \text{ i.v.}\} = \limsup_{n \rightarrow \infty} E_n = \bigcap_{N \geq 1} \bigcup_{n \geq N} E_n.$$

Assim, para todo N ,

$$\{E_n \text{ i.v.}\} \subseteq \bigcup_{n \geq N} E_n.$$

Pela subaditividade,

$$\mathbf{P}(E_n \text{ i.v.}) \leq \mathbf{P}\left(\bigcup_{n \geq N} E_n\right) \leq \sum_{n \geq N} \mathbf{P}(E_n).$$

Como a série $\sum_n \mathbf{P}(E_n)$ converge, sua cauda tende a zero. Fazendo $N \rightarrow \infty$, obtemos

$$\mathbf{P}(E_n \text{ i.v.}) = 0.$$

□

O lema seguinte compara as probabilidades dos limites superior e inferior com os limites das probabilidades individuais.

Teorema 3.29 — Lema de Fatou para conjuntos

Para $E_1, E_2, \dots \in \mathcal{F}$,

$$\mathbf{P}\left(\liminf_{n \rightarrow \infty} E_n\right) \leq \liminf_{n \rightarrow \infty} \mathbf{P}(E_n) \leq \limsup_{n \rightarrow \infty} \mathbf{P}(E_n) \leq \mathbf{P}\left(\limsup_{n \rightarrow \infty} E_n\right).$$

Demonstração. Defina

$$B_n = \bigcap_{k \geq n} E_k, \quad C_n = \bigcup_{k \geq n} E_k.$$

Então

$$B_n \uparrow \liminf_{n \rightarrow \infty} E_n, \quad C_n \downarrow \limsup_{n \rightarrow \infty} E_n.$$

Como $B_n \subseteq E_k \subseteq C_n$ para todo $k \geq n$,

$$\mathbf{P}(B_n) \leq \inf_{k \geq n} \mathbf{P}(E_k), \quad \sup_{k \geq n} \mathbf{P}(E_k) \leq \mathbf{P}(C_n).$$

A continuidade da probabilidade por baixo e por cima fornece as duas desigualdades extremas. A desigualdade central vale para toda sequência real. $\square$

O primeiro lema de Borel–Cantelli diz mais do que “probabilidades pequenas”: a hipótese relevante é a somabilidade. Quando $\sum_n \mathbf{P}(E_n) < \infty$, com probabilidade um apenas um número finito dos eventos E_n ocorre.

A relação com os complementos fica transparente pela lei de De Morgan:

$$\left(\limsup_{n \rightarrow \infty} E_n \right)^c = \left(\bigcap_{N \geq 1} \bigcup_{n \geq N} E_n \right)^c = \bigcup_{N \geq 1} \bigcap_{n \geq N} E_n^c = \liminf_{n \rightarrow \infty} E_n^c.$$

Portanto, dizer que E_n ocorre apenas finitas vezes é o mesmo que dizer que, a partir de algum índice, todos os complementos E_n^c ocorrem.

Exemplo 3.30 — Passeio aleatório assimétrico.

Considere um passeio em $\mathbb{Z}$ que começa em 0 e, a cada passo, vai uma unidade para a direita com probabilidade $p \in (0,1)$ e uma unidade para a esquerda com probabilidade $1 - p$.

A construção de cilindros usada para as infinitas moedas, agora com pesos p e $1 - p$, fornece um espaço de probabilidade no qual os incrementos $\omega_1, \omega_2, \dots \in \{-1, 1\}$ são independentes e

$$\mathbf{P}(\omega_k = 1) = p, \quad \mathbf{P}(\omega_k = -1) = 1 - p.$$

Defina

$$S_0 = 0, \quad S_n = \omega_1 + \dots + \omega_n.$$

Para retornar à origem no instante $2n$, o passeio deve dar exatamente n passos para a direita e n para a esquerda. Há $\binom{2n}{n}$ ordenações possíveis desses passos e, pela independência, cada uma tem probabilidade $p^n(1-p)^n$. Portanto,

$$\mathbf{P}(S_{2n} = 0) = \binom{2n}{n} p^n (1-p)^n.$$

A fórmula de Stirling,

$$n! \sim \sqrt{2\pi n} \left(\frac{n}{e}\right)^n,$$

dá

$$\binom{2n}{n} = \frac{(2n)!}{(n!)^2} \sim \frac{4^n}{\sqrt{\pi n}},$$

e, consequentemente,

$$\mathbf{P}(S_{2n} = 0) \sim \frac{(4p(1-p))^n}{\sqrt{\pi n}}.$$

Para aplicar Borel–Cantelli, basta uma cota mais simples. Como

$$\binom{2n}{n} \leq 4^n,$$

temos

$$\mathbf{P}(S_{2n} = 0) \leq (4p(1-p))^n.$$

Se $p \neq 1/2$, então $r := 4p(1-p) < 1$, e

$$\sum_{n \geq 1} \mathbf{P}(S_{2n} = 0) \leq \sum_{n \geq 1} r^n < \infty.$$

Pelo primeiro lema de Borel–Cantelli, o passeio retorna à origem apenas um número finito de vezes, quase certamente.

No caso simétrico $p = 1/2$,

$$\mathbf{P}(S_{2n} = 0) = \binom{2n}{n} 4^{-n} \sim \frac{1}{\sqrt{\pi n}},$$

e portanto

$$\sum_{n \geq 1} \mathbf{P}(S_{2n} = 0) = \infty.$$

Isso, sozinho, não prova recorrência: os eventos $\{S_{2n} = 0\}$ não são independentes, de modo que o segundo lema de Borel–Cantelli não pode ser aplicado diretamente.

◁

A independência é exatamente a hipótese que fornece a recíproca do primeiro lema.

Teorema 3.31 – Segundo lema de Borel–Cantelli

Se $E_1, E_2, \dots \in \mathcal{F}$ são independentes e

$$\sum_{n \geq 1} \mathbf{P}(E_n) = \infty,$$

então

$$\mathbf{P}(E_n \text{ i.v.}) = 1.$$

Demonstração. Queremos mostrar que o complemento do evento “ E_n ocorre infinitas vezes” tem probabilidade zero. Pela lei de De Morgan,

$$\{E_n \text{ i.v.}\}^c = \left(\bigcap_{N \geq 1} \bigcup_{n \geq N} E_n \right)^c = \bigcup_{N \geq 1} \bigcap_{n \geq N} E_n^c.$$

Portanto, basta provar que, para cada N ,

$$\mathbf{P}\left(\bigcap_{n \geq N} E_n^c\right) = 0.$$

Como os E_n são independentes, seus complementos também são. Pela fórmula para interseções enumeráveis,

$$\mathbf{P}\left(\bigcap_{n \geq N} E_n^c\right) = \prod_{n \geq N} (1 - \mathbf{P}(E_n)).$$

Usando $1 - x \leq e^{-x}$ para $0 \leq x \leq 1$,

$$\prod_{n \geq N} (1 - \mathbf{P}(E_n)) \leq \exp\left(-\sum_{n \geq N} \mathbf{P}(E_n)\right).$$

Como $\sum_n \mathbf{P}(E_n) = \infty$, toda cauda dessa série também diverge. Consequentemente,

$$\mathbf{P}\left(\bigcap_{n \geq N} E_n^c\right) = 0 \quad \text{para todo } N.$$

Por continuidade por baixo,

$$\mathbf{P}\left(\bigcup_{N \geq 1} \bigcap_{n \geq N} E_n^c\right) = 0.$$

Esse evento é $\{E_n \text{ i.v.}\}^c$, e portanto

$$\mathbf{P}(E_n \text{ i.v.}) = 1.$$

□

Assim, para uma sequência de eventos independentes,

$$\mathbf{P}(E_n \text{ i.v.}) = \begin{cases} 0, & \sum_n \mathbf{P}(E_n) < \infty, \\ 1, & \sum_n \mathbf{P}(E_n) = \infty. \end{cases}$$

A primeira linha não precisa da hipótese de independência; a segunda, sim.

Exemplo 3.32 — Infinitas caras e infinitas coroas.

Em lançamentos independentes de uma moeda justa, seja

$$A_n = \{\text{o } n\text{-ésimo lançamento é cara}\}.$$

Os eventos A_n são independentes e $\mathbf{P}(A_n) = 1/2$. Logo,

$$\sum_{n \geq 1} \mathbf{P}(A_n) = \infty,$$

e o segundo lema de Borel–Cantelli dá

$$\mathbf{P}(A_n \text{ i.v.}) = 1.$$

Assim, quase certamente aparecem infinitas caras. Aplicando o mesmo argumento aos complementos A_n^c , obtemos também infinitas coroas.

◁

O Exercício 3.13 propõe uma aplicação direta desse argumento a blocos disjuntos de lançamentos.

3.3. Aplicações

Sequências de caras

Em uma sequência infinita de moedas, qualquer comprimento fixo de caras consecutivas reaparece infinitas vezes. Uma pergunta mais interessante é saber como o maior comprimento observado pode crescer com o tempo.

Seja $\ell(n)$ o número de caras consecutivas que termina no n -ésimo lançamento, com $\ell(n) = 0$ quando o n -ésimo resultado é coroa. Para todo inteiro fixo k ,

$$\mathbf{P}(\ell(n) \geq k \text{ i.v.}) = 1.$$

Se permitimos que o comprimento exigido dependa de n , surge uma escala crítica: $\log_2 n$.

Teorema 3.33 —

1 Para todo $\varepsilon > 0$,

$$\mathbf{P}(\ell(n) \geq (1 + \varepsilon) \log_2 n \text{ i.v.}) = 0.$$

2

$$\mathbf{P}(\ell(n) \geq \log_2 n \text{ i.v.}) = 1.$$

Demonstração. Para a primeira afirmação, fixe $\varepsilon > 0$ e ponha

$$r_n = \left\lceil (1 + \varepsilon) \log_2 n \right\rceil.$$

Se $r_n \leq n$, o evento $\{\ell(n) \geq r_n\}$ exige cara nos últimos r_n lançamentos e, portanto, tem probabilidade 2^{-r_n} . Se $r_n > n$, o evento é vazio. Em ambos os casos,

$$\mathbf{P}(\ell(n) \geq (1 + \varepsilon) \log_2 n) \leq n^{-(1+\varepsilon)}.$$

A série do lado direito converge; o primeiro lema de Borel–Cantelli conclui a primeira parte.

Para a segunda afirmação, os eventos naturais se sobrepõem e não são independentes. Vamos extrair deles uma família de blocos disjuntos.

Para $m \geq 2$, considere

$$I_m = \{2^m + 1, \dots, 2^{m+1}\}.$$

Divida os primeiros termos de I_m em

$$N_m = \left\lfloor \frac{2^m}{m+1} \right\rfloor$$

blocos disjuntos de comprimento $m+1$. Seja E_m o evento de que ao menos um desses blocos seja formado apenas por caras. Faixas diferentes usam lançamentos disjuntos, logo os eventos E_m são independentes. Dentro de uma faixa, os blocos também são independentes, e

$$\mathbf{P}(E_m) = 1 - \left(1 - 2^{-(m+1)}\right)^{N_m}.$$

Para m suficientemente grande,

$$N_m \geq \frac{2^{m-1}}{m+1}.$$

Usando $1 - u \leq e^{-u}$ e $1 - e^{-x} \geq x/2$ para $0 \leq x \leq 1$, obtemos

$$\mathbf{P}(E_m) \geq \frac{1}{8(m+1)}.$$

Logo,

$$\sum_m \mathbf{P}(E_m) = \infty.$$

Pelo segundo lema de Borel–Cantelli, E_m ocorre para infinitos valores de m , quase certamente.

Quando E_m ocorre, existe um bloco de $m+1$ caras que termina em algum $n \leq 2^{m+1}$. Portanto,

$$\ell(n) \geq m+1 \geq \log_2 n.$$

Isso prova a segunda afirmação. □

Teorema do macaco infinito

Substitua a moeda por escolhas independentes e uniformes em um alfabeto finito. O infinito não melhora a datilografia; ele apenas torna inevitável cada bloco finito.

Teorema 3.34 – Teorema do macaco infinito

Considere uma sequência de escolhas independentes e uniformes em um alfabeto finito. Quase certamente, toda palavra finita aparece na sequência infinitas vezes.

Demonstração. Considere um alfabeto com q símbolos e fixe uma palavra

$$w = (w_1, \dots, w_r)$$

de comprimento r . Para $j \geq 1$, seja E_j o evento de que o bloco de posições

$$(j-1)r + 1, \dots, jr$$

coincide com w . Os eventos E_j dependem de blocos disjuntos e são independentes, com

$$\mathbf{P}(E_j) = q^{-r}.$$

Logo,

$$\sum_j \mathbf{P}(E_j) = \infty,$$

e o segundo lema de Borel–Cantelli mostra que w ocorre infinitas vezes com probabilidade 1.

Há apenas um conjunto enumerável de palavras finitas sobre um alfabeto finito. Enumerando-as como $w^{(1)}, w^{(2)}, \dots$ e escrevendo G_k para o evento de que $w^{(k)}$ aparece infinitas vezes, temos $\mathbf{P}(G_k) = 1$ para todo k . Assim,

$$\mathbf{P}\left(\bigcap_{k \geq 1} G_k\right) = 1 - \mathbf{P}\left(\bigcup_{k \geq 1} G_k^c\right) = 1.$$

□

Teoria de percolação

Considere a rede quadrada

$$G = (V, E), \quad V = \mathbb{Z}^2,$$

em que E é o conjunto das arestas entre vizinhos mais próximos. Na **percolação de elos**, cada aresta é declarada aberta, independentemente, com probabilidade p . A pergunta física é simples: se um fluido puder passar apenas pelas ligações abertas, será possível atravessar a rede indefinidamente?

Usaremos a seguinte terminologia:

- os vértices são chamados de **sítios** e as arestas, de **elos**;
- um elo com estado 1 é **aberto**; com estado 0, é **fechado**;
- o subgrafo formado pelos elos abertos é o **subgrafo aberto**;
- as componentes conexas do subgrafo aberto são os **clusters abertos**.

Há também um modelo de percolação de sítios, no qual são os vértices que recebem estados aleatórios. Nesta seção trabalharemos apenas com percolação de elos.

O teorema de Harris–Kesten, que citamos sem prova, afirma que para a percolação de elos em $\mathbb{Z}^2$ a probabilidade crítica é

$$p_c = \frac{1}{2}.$$

Para $p > 1/2$, existe um cluster aberto infinito quase certamente; para $p \leq 1/2$, não existe. A demonstração completa exige ferramentas de percolação planar que estão fora do escopo destas notas. Nosso objetivo aqui é construir rigorosamente o modelo e provar a dicotomia 0-1 em torno do parâmetro crítico.

Construção da medida de percolação

Defina

$$\Omega = \{0,1\}^E.$$

Uma configuração $\omega \in \Omega$ associa a cada aresta $e \in E$ um estado $\omega_e \in \{0,1\}$.

Para $F \subseteq E$ finito e $\sigma \in \{0,1\}^F$, defina o cilindro

$$C(F, \sigma) = \{\omega \in \Omega : \omega_f = \sigma_f \text{ para todo } f \in F\}.$$

Se

$$N_1(\sigma) = |\{f \in F : \sigma_f = 1\}|,$$

a independência desejada das arestas força

$$\mathbf{P}_0(C(F, \sigma)) = p^{N_1(\sigma)}(1-p)^{|F|-N_1(\sigma)}.$$

Se $\mathcal{F}_0$ é a álgebra gerada pelos cilindros, essa prescrição é consistente: ao adicionar uma coordenada livre, as duas extensões possíveis têm massas cuja soma é multiplicada por

$$p + (1-p) = 1.$$

Logo, por somas sobre uniões finitas disjuntas de cilindros, obtemos uma probabilidade finitamente aditiva $\mathbf{P}_0$ em $\mathcal{F}_0$.

Equipe Ω com a topologia produto das cópias discretas de $\{0,1\}$. Pelo teorema de Tychonoff, Ω é compacto. Cada cilindro é aberto e fechado e, portanto, compacto. Como os elementos de $\mathcal{F}_0$ são uniões finitas de cilindros, também são compactos.

Se $D_n \in \mathcal{F}_0$ e $D_n \downarrow \emptyset$, a propriedade da interseção finita para compactos implica que algum D_N já é vazio. Logo,

$$\mathbf{P}_0(D_n) \rightarrow 0.$$

A continuidade no vazio transforma a aditividade finita de $\mathbf{P}_0$ em σ -aditividade na álgebra: $\mathbf{P}_0$ é uma pré-medida. O teorema de extensão de Carathéodory fornece então uma única probabilidade $\mathbf{P}_p$ em

$$\mathcal{F} = \sigma\langle \mathcal{F}_0 \rangle$$

com essas probabilidades cilíndricas.

A tripla $(\Omega, \mathcal{F}, \mathbf{P}_p)$ é o modelo de percolação de elos de parâmetro p em $\mathbb{Z}^2$. A mesma construção funciona em qualquer grafo com conjunto enumerável de arestas.

Probabilidade crítica

Para um sítio x , denote por C_x o cluster aberto que contém x e defina

$$\theta_x(p) = \mathbf{P}_p(|C_x| = \infty).$$

Proposição 3.35 —

Existe $p_c \in [0,1]$ tal que, para todo sítio x ,

$$\theta_x(p) = 0 \quad \text{se } p < p_c, \quad \theta_x(p) > 0 \quad \text{se } p > p_c.$$

O enunciado não afirma o que ocorre em $p = p_c$.

Demonstração. Primeiro mostramos que, para $p > 0$, a positividade de $\theta_x(p)$ não depende do sítio. Fixe um caminho γ de y a x , com m arestas, e condicione as arestas fora de γ . O evento $\{|C_\gamma| = \infty\}$ é crescente em relação aos estados das arestas de γ . Se ele ocorre para alguma configuração dessas arestas, continua ocorrendo quando todas são abertas. Como a configuração totalmente aberta em γ tem probabilidade condicional p^m ,

$$\mathbf{P}_p(|C_\gamma| = \infty, \gamma \text{ aberta}) \geq p^m \theta_y(p).$$

Nesse evento, x está ligado a y , logo $|C_x| = \infty$. Portanto,

$$\theta_x(p) \geq p^m \theta_y(p).$$

Trocando x e y , concluímos que, para cada $p > 0$, ou todas as $\theta_x(p)$ são nulas, ou todas são positivas.

Agora construímos simultaneamente os modelos de todos os parâmetros. Associe a cada aresta e uma marca independente $U_e \sim \text{Uni}[0,1]$ e declare e aberta no nível p quando $U_e \leq p$. Se $p_1 \leq p_2$, toda aresta aberta no nível p_1 também está aberta no nível p_2 . Consequentemente,

$$p \mapsto \theta_x(p)$$

é não decrescente.

Fixe x_0 e ponha

$$p_c = \inf\{p \in [0,1] : \theta_{x_0}(p) > 0\}.$$

Temos $\theta_{x_0}(0) = 0$ e $\theta_{x_0}(1) = 1$. Pela monotonicidade, $p < p_c$ implica $\theta_{x_0}(p) = 0$, enquanto $p > p_c$ implica $\theta_{x_0}(p) > 0$. A primeira parte da prova estende essas conclusões a todo sítio x .

Mostrar diretamente que $0 < p_c < 1$ exige argumentos adicionais. Para $\mathbb{Z}^2$, o teorema de Harris–Kesten citado acima fornece o valor exato $p_c = 1/2$.

A construção simultânea por marcas uniformes é um exemplo de acoplamento; voltaremos a essa técnica no Capítulo 9. $\square$

Resta relacionar a positividade de $\theta_x(p)$ com a existência, em toda a rede, de algum cluster infinito.

Teorema 3.36 —

Seja

$$I = \{\text{existe um cluster aberto infinito}\}.$$

Então

$$\mathbf{P}_p(I) = 0 \quad \text{se } p < p_c, \quad \mathbf{P}_p(I) = 1 \quad \text{se } p > p_c.$$

Demonstração. Começamos pela mensurabilidade. Para um sítio x , o evento $\{|C_x| = \infty\}$ é a interseção, sobre $r \geq 1$, dos eventos “existe um caminho aberto de comprimento r partindo de x ”. Cada um desses eventos é uma união enumerável de eventos cilíndricos. Como $\mathbb{Z}^2$ é localmente finita, a existência de caminhos arbitrariamente longos equivale à infinitude de C_x . Assim,

$$I = \bigcup_{x \in \mathbb{Z}^2} \{|C_x| = \infty\} \in \mathcal{F}.$$

Mostremos agora que I é caudal. Enumere as arestas como

$$E = \{e_0, e_1, \dots\}$$

e defina

$$A_k = \{\omega : \omega_{e_k} = 1\}.$$

Os eventos A_k são independentes e

$$\mathcal{F} = \sigma\langle A_0, A_1, \dots \rangle.$$

Para $n \geq 1$, retire as arestas $e_0, \dots, e_{n-1}$ e denote por I_n o evento de que o subgrafo restante contém uma componente infinita. Então

$$I_n \in \sigma\langle A_n, A_{n+1}, \dots \rangle.$$

Retirar um número finito de arestas não altera a existência de uma componente infinita. De fato, cada aresta removida pode aumentar em no máximo uma unidade o número de componentes de uma componente dada; após retirar n arestas, uma componente infinita é dividida em no máximo $n + 1$ componentes, e pelo menos uma delas continua infinita. A inclusão inversa é imediata. Logo $I = I_n$ para todo n , e portanto I é um evento caudal.

Pela lei 0-1 de Kolmogorov,

$$\mathbf{P}_p(I) \in \{0, 1\}.$$

Se $p < p_c$, então $\theta_x(p) = 0$ para todo x , e

$$\mathbf{P}_p(I) \leq \sum_{x \in \mathbb{Z}^2} \theta_x(p) = 0.$$

Se $p > p_c$, então, para qualquer sítio x ,

$$\mathbf{P}_p(I) \geq \theta_x(p) > 0.$$

A dicotomia 0-1 obriga $\mathbf{P}_p(I) = 1$. □

A independência, introduzida como uma relação entre informação e probabilidade, passou de eventos a famílias e σ -álgebras. Dela extraímos consequências para sequências inteiras: leis 0-1 para eventos caudais e critérios de Borel–Cantelli para ocorrências infinitas. Na percolação, essas ferramentas se encontram para descrever um fenômeno global a partir de escolhas locais.

EXERCÍCIOS

Ex. 3.1 — Seja $\Omega = \{1, 2, \dots, 9\}$, com todos os pontos de probabilidade $1/9$, e considere

$$A = \{1, 2, 3\}, \quad B = \{1, 4, 5\}, \quad C = \{2, 4, 6\}.$$

Mostre que A , B e C são independentes dois a dois, mas que a família $\{A, B, C\}$ não é independente.

Ex. 3.2 — Construa três eventos A , B e C que não formem uma família independente, embora satisfaçam

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

Sua construção deve exibir explicitamente ao menos um par que não seja independente.

Ex. 3.3 — Seja $A_1, \dots, A_n$ uma família independente de eventos. Prove que, para qualquer escolha

$$B_i \in \{A_i, A_i^c\}, \quad i = 1, \dots, n,$$

a família $B_1, \dots, B_n$ é independente. Em particular, $A_1^c, \dots, A_n^c$ são independentes.

Ex. 3.4 — Fixe um inteiro $k \geq 2$. Construa um espaço de probabilidade e k eventos, cada um com probabilidade $1/2$, tais que toda subfamília com no máximo $k - 1$ eventos seja independente, mas a família completa não seja. Escolha um espaço amostral com o menor número possível de pontos e justifique essa minimalidade.

Ex. 3.5 — Uma urna contém inicialmente uma bola preta e $m \geq 1$ bolas brancas. A cada etapa, retira-se uma bola uniformemente. Se ela for preta, ela é simplesmente recolocada; se for branca, ela é recolocada juntamente com uma nova bola branca. Seja A_n o evento de que a bola preta não seja retirada nas primeiras n etapas.

1. Calcule $\mathbf{P}(A_n)$ e mostre que $\sum_{n \geq 1} \mathbf{P}(A_n) = \infty$.
2. Determine $\mathbf{P}(A_n \text{ i.v.})$.
3. Explique por que o exemplo mostra que a independência não pode ser omitida da recíproca fornecida pelo segundo lema de Borel–Cantelli.

Ex. 3.6 — Sejam $A_1, A_2, \dots$ subconjuntos borelianos de $[0, 1]$ e λ a medida de Lebesgue. Suponha que exista $\eta > 0$ tal que $\lambda(A_n) \geq \eta$ para todo n . Prove que

$$\lambda(\limsup_{n \rightarrow \infty} A_n) \geq \eta.$$

Conclua que algum ponto de $[0, 1]$ pertence a infinitos conjuntos A_n .

Ex. 3.7 — Para $k = 1, \dots, n$, seja $\mathcal{P}_k$ uma partição enumerável de Ω formada por eventos. Prove que as σ -álgebras

$$\sigma\langle\mathcal{P}_1\rangle, \dots, \sigma\langle\mathcal{P}_n\rangle$$

são independentes se, e somente se,

$$\mathbf{P}\left(\bigcap_{k=1}^n A_k\right) = \prod_{k=1}^n \mathbf{P}(A_k)$$

para toda escolha de uma célula $A_k \in \mathcal{P}_k$, $k = 1, \dots, n$.

Ex. 3.8 — Sejam $\mathcal{A}_1, \dots, \mathcal{A}_n$ π -sistemas contidos em $\mathcal{F}$ e suponha que

$$\mathbf{P}\left(\bigcap_{k=1}^n A_k\right) = \prod_{k=1}^n \mathbf{P}(A_k) \quad \text{para toda escolha } A_k \in \mathcal{A}_k. \quad (3.1)$$

1. Dê um exemplo mostrando que, sem hipóteses adicionais, (3.1) não garante a independência das σ -álgebras $\sigma\langle\mathcal{A}_k\rangle$.
2. Suponha agora que, para cada k , o espaço Ω seja união enumerável de elementos de $\mathcal{A}_k$. Prove que $\sigma\langle\mathcal{A}_1\rangle, \dots, \sigma\langle\mathcal{A}_n\rangle$ são independentes.

Dica. Fixe todas as coordenadas menos uma, passe primeiro a uniões finitas usando inclusão-exclusão e depois use o Teorema π - λ .

Ex. 3.9 — Seja $(A_n)_{n \geq 1}$ uma sequência de eventos e defina

$$N(\omega) = \sum_{n=1}^{\infty} \mathbb{1}_{A_n}(\omega).$$

Prove diretamente que

$$\mathbf{P}(A_n \text{ i.v.}) < 1$$

se, e somente se, existe um evento B com $\mathbf{P}(B) > 0$ e um inteiro $r \geq 0$ tais que $N \leq r$ em B . Deduza a equivalência

$$\mathbf{P}(A_n \text{ i.v.}) = 1 \iff \sum_{n=1}^{\infty} \mathbf{P}(A_n \mid B) = \infty \quad \text{para todo } B \text{ com } \mathbf{P}(B) > 0.$$

Ex. 3.10 — Medida de Lebesgue. Construa a medida de Lebesgue em $\mathbb{R}$ a partir da pré-medida comprimento na álgebra das uniões finitas de intervalos semiabertos. Na etapa em que é preciso provar a continuidade no vazio, use o Teorema de Extensão Compacta e explicita a classe compacta escolhida.

Ex. 3.11 — Infinitas moedas I. Seja $\Omega = \{0,1\}^{\mathbb{N}}$ e, para $\varepsilon_1, \dots, \varepsilon_n \in \{0,1\}$, defina o cilindro

$$C_n(\varepsilon_1, \dots, \varepsilon_n) = \{\omega \in \Omega : \omega_1 = \varepsilon_1, \dots, \omega_n = \varepsilon_n\}.$$

Seja $\mathcal{F}_0$ a álgebra das uniões finitas desses cilindros. Fixe $p \in [0,1]$ e prescreva

$$\mathbf{P}_p(C_n(\varepsilon_1, \dots, \varepsilon_n)) = p^{\sum_{i=1}^n \varepsilon_i} (1-p)^{n-\sum_{i=1}^n \varepsilon_i}.$$

1. Mostre que essa prescrição é consistente quando um mesmo cilindro é representado em níveis diferentes e que sua extensão por aditividade finita está bem definida em $\mathcal{F}_0$.
2. Equipe Ω com a topologia produto discreta. Mostre que Ω é compacto e que os elementos de $\mathcal{F}_0$ são abertos e fechados. Use o Teorema de Extensão Compacta para concluir que $\mathbf{P}_p$ é uma pré-medida em $\mathcal{F}_0$.
3. Aplique o Teorema de Extensão de Carathéodory para provar que $\mathbf{P}_p$ possui uma única extensão a $\sigma(\mathcal{F}_0)$.
4. Mostre que as projeções coordenadas

$$X_n(\omega) = \omega_n, \quad n \geq 1,$$

formam uma sequência i.i.d. de variáveis de Bernoulli de parâmetro p .

Ex. 3.12 — Percolação de elos. Seja $G = (V, E)$ um grafo infinito e $\Omega = \{0,1\}^E$, onde $\omega_e = 1$ significa que a aresta e está aberta. Para $F \subseteq E$ finito e $\tau \in \{0,1\}^F$, defina

$$C(F, \tau) = \{\omega \in \Omega : \omega_e = \tau_e \text{ para todo } e \in F\}.$$

Seja $\mathcal{F}_0$ a álgebra dos cilindros e prescreva

$$\mathbf{P}_p(C(F, \tau)) = \prod_{e \in F: \tau_e=1} p \prod_{e \in F: \tau_e=0} (1-p), \quad p \in [0,1].$$

1. Mostre que a prescrição é consistente e se estende de maneira única a uma probabilidade finitamente aditiva em $\mathcal{F}_0$.
2. Use cilindros com base compacta e o Teorema de Extensão Compacta para mostrar que essa função é uma pré-medida.

3. Conclua, pelo Teorema de Extensão de Carathéodory, que existe uma única probabilidade em $\sigma\langle\mathcal{F}_0\rangle$ com as probabilidades cilíndricas acima.

Ex. 3.13 — Em uma sequência infinita de lançamentos independentes de uma moeda justa, seja

$$B_n = \{\text{os lançamentos } n, n+1, \dots, n+99 \text{ são caras}\}.$$

Calcule $\mathbf{P}(B_n \text{ i.v.})$. Como os blocos consecutivos se sobrepõem, comece por uma subsequência de blocos disjuntos.

Ex. 3.14 — Raio de convergência de uma série aleatória. Sejam $Z_0, Z_1, \dots$ variáveis aleatórias reais e considere a série de potências

$$\sum_{n=0}^{\infty} Z_n z^n.$$

Com as convenções $1/0 = \infty$ e $1/\infty = 0$, defina

$$R = \left(\limsup_{n \rightarrow \infty} |Z_n|^{1/n} \right)^{-1}.$$

1. Prove que R , com valores em $[0, \infty]$, é uma variável aleatória e coincide com o raio de convergência da série.
2. Suponha agora que os Z_n sejam independentes. Prove que existe $r \in [0, \infty]$ tal que $R = r$ quase certamente.

Variáveis Aleatórias

Até aqui, os resultados de um experimento permaneceram no espaço abstrato Ω . Para calcular com eles, precisamos transformá-los em números, vetores ou outros objetos conhecidos. Uma variável aleatória realiza essa tradução. A exigência de mensurabilidade garante que, ao formularmos uma pergunta sobre seu valor, a resposta corresponda a um evento do modelo.

4.1. Funções Mensuráveis

Uma função entre espaços mensuráveis deve respeitar as famílias de conjuntos que escolhemos observar. Essa compatibilidade é expressa pelas pré-imagens.

Definição 4.1 —

Dados $(\Omega_1, \mathcal{F}_1)$ e $(\Omega_2, \mathcal{F}_2)$ espaços mensuráveis, uma função $f : \Omega_1 \rightarrow \Omega_2$ é dita $\mathcal{F}_1 - \mathcal{F}_2$ -**mensurável** ou simplesmente **mensurável** se a pré-imagem de todo conjunto mensurável é mensurável. Ou seja, se

$$A \in \mathcal{F}_2 \implies f^{-1}(A) \in \mathcal{F}_1.$$

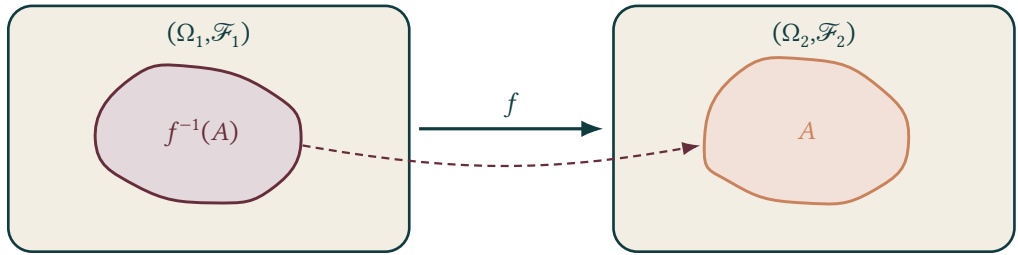


Figura 4.1.: A mensurabilidade exige que a pré-imagem $f^{-1}(A)$ de cada conjunto mensurável $A \in \mathcal{F}_2$ pertença a $\mathcal{F}_1$.

Observação 4.2 —

- Dado uma função $f : \Omega_1 \rightarrow \Omega_2$ que é $\mathcal{F}_1 - \mathcal{F}_2$ -**mensurável** então se $\mathcal{F}_1 \subset \mathcal{F}_1'$ claramente f é $\mathcal{F}_1' - \mathcal{F}_2$ -**mensurável**.
- O mesmo não ocorre se aumentarmos a σ -álgebra do contradomínio. Por exemplo, a identidade de $\mathbb{R}$ em $\mathbb{R}$, embora contínua, não é $\mathcal{B}(\mathbb{R}) - \overline{\mathcal{B}(\mathbb{R})}$ -mensurável: a pré-imagem de um conjunto Lebesgue mensurável não boreliano é o próprio conjunto. Em contraste, toda função contínua $\mathbb{R} \rightarrow \mathbb{R}$ é $\overline{\mathcal{B}(\mathbb{R})} - \mathcal{B}(\mathbb{R})$ -mensurável.

A família das funções mensuráveis é ampla: para funções reais, por exemplo, a

continuidade é suficiente. Verificar diretamente a pré-imagem de todo conjunto mensurável, contudo, seria impraticável. O resultado seguinte permite trabalhar apenas com uma família geradora.

Teorema 4.3 —

Considere uma função $f : (\Omega_1, \mathcal{F}_1) \rightarrow (\Omega_2, \mathcal{F}_2)$, e $\mathcal{A} \subset \mathcal{F}_2$ tal que $\mathcal{F}_2 = \sigma\langle \mathcal{A} \rangle$. Nestas condições, se $f^{-1}(A) \in \mathcal{F}_1$ para todo $A \in \mathcal{A}$, então f é mensurável. Ou seja, se

$$A \in \mathcal{A} \Rightarrow f^{-1}(A) \in \mathcal{F}_1.$$

então

$$A \in \sigma\langle \mathcal{A} \rangle \Rightarrow f^{-1}(A) \in \mathcal{F}_1,$$

Demonstração. Considere $\mathcal{G} := \{A \in \sigma\langle \mathcal{A} \rangle : f^{-1}(A) \in \mathcal{F}_1\}$. Iremos demonstrar que $\mathcal{G} = \sigma\langle \mathcal{A} \rangle$.

Sabemos que $\mathcal{A} \subseteq \mathcal{G} \subseteq \sigma\langle \mathcal{A} \rangle$. Além disso temos que $\sigma\langle \mathcal{A} \rangle$ é a σ -álgebra minimal contendo $\mathcal{A}$. Assim, basta demonstrar que $\mathcal{G}$ é uma σ -álgebra.

1 Temos $\Omega_2 \in \mathcal{G}$, pois $f^{-1}(\Omega_2) = \Omega_1 \in \mathcal{F}_1$.

2 Demonstraremos que se $A \in \mathcal{G} \Rightarrow A^c \in \mathcal{G}$.

Para isso observe que $f^{-1}(A^c) = (f^{-1}(A))^c$.

3 Demonstraremos que se $A_1, A_2, \dots \in \mathcal{G}$, então $\bigcup_{n=1}^{\infty} A_n \in \mathcal{G}$. Mas

$$f^{-1}\left(\bigcup_n A_n\right) = \bigcup_{n=1}^{\infty} f^{-1}(A_n)$$

Logo, $\mathcal{G}$ é uma σ -álgebra e $\mathcal{A} \subseteq \mathcal{G} \subseteq \sigma\langle \mathcal{A} \rangle$, o que implica $\mathcal{G} = \sigma\langle \mathcal{A} \rangle$. □

Corolário 4.4 —

Sejam $(\Omega, \mathcal{F})$ um espaço mensurável e $f : \Omega \rightarrow \mathbb{R}$ uma função real. Nestas condições são equivalentes as seguintes afirmativas:

- a** f é $\mathcal{F} - \mathcal{B}(\mathbb{R})$ mensurável (mensurável a Borel);
- b** $\{\omega \in \Omega : f(\omega) \leq x\} = f^{-1}((-\infty, x]) \in \mathcal{F}$ para todo $x \in \mathbb{R}$;
- c** $\{\omega \in \Omega : f(\omega) < x\} = f^{-1}((-\infty, x)) \in \mathcal{F}$ para todo $x \in \mathbb{R}$;
- d** $\{\omega \in \Omega : f(\omega) \geq x\} = f^{-1}([x, +\infty)) \in \mathcal{F}$ para todo $x \in \mathbb{R}$;
- e** $\{\omega \in \Omega : f(\omega) > x\} = f^{-1}((x, +\infty)) \in \mathcal{F}$ para todo $x \in \mathbb{R}$;

Demonstração. Basta lembrar que

$$\begin{aligned}
 \mathcal{B}(\mathbb{R}) &= \sigma\langle\{(-\infty, x] : x \in \mathbb{R}\}\rangle \\
 &= \sigma\langle\{(-\infty, x) : x \in \mathbb{R}\}\rangle \\
 &= \sigma\langle\{[x, +\infty) : x \in \mathbb{R}\}\rangle \\
 &= \sigma\langle\{(x, +\infty) : x \in \mathbb{R}\}\rangle
 \end{aligned}$$

□

Os exemplos seguintes mostram como a mensurabilidade aparece nas construções mais simples.

Exemplo 4.5 — Função Indicadora de um Evento.

Dado um espaço mensurável $(\Omega, \mathcal{F})$ e $A \in \mathcal{F}$, a função indicadora de A , geralmente denotada por 1_A ou por $I\{A\}$ é definida por:

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

Agora, dado $B \in \mathcal{B}(\mathbb{R})$ note que

$$\mathbb{1}_A^{-1}(B) = \begin{cases} \emptyset & , \quad 0, 1 \notin B; \\ \Omega & , \quad 0, 1 \in B; \\ A & , \quad 0 \notin B \text{ e } 1 \in B; \\ A^c & , \quad 1 \notin B \text{ e } 0 \in B. \end{cases}$$

E portanto, $\mathbb{1}_A$ é uma função mensurável.

◁

Exemplo 4.6 — Funções Simples.

Podemos modificar o exemplo anterior de modo a criar uma função ξ cuja imagem seja um conjunto finito $\{a_1, \dots, a_n\}$.

Para isso tomemos uma partição $A_1, A_2, \dots, A_n \in \mathcal{F}$ de Ω e defina

$$\xi(\omega) = \sum_{k=1}^n a_k \mathbb{1}_{A_k}(\omega).$$

Neste caso, tomando os a_k 's distintos, teríamos que $A_k = \xi^{-1}(\{a_k\})$ e a mensurabilidade de ξ pode ser verificada observando que

$$\xi^{-1}(B) = \bigcup_{k: a_k \in B} A_k \in \mathcal{F},$$

onde a união acima é vazia se $\{k : a_k \in B\} = \emptyset$.

◁

A próxima proposição amplia a coleção de exemplos ao relacionar continuidade e mensurabilidade.

Proposição 4.7 — Mensurabilidade de aplicações contínuas

Seja (Ω_1, d_1) e (Ω_2, d_2) espaços métricos e seja $f : \Omega_1 \rightarrow \Omega_2$ uma aplicação contínua. Então f é $\mathcal{B}(\Omega_1) - \mathcal{B}(\Omega_2)$ -mensurável.

Demonstração. Da continuidade de f , sabemos que $f^{-1}(U)$ é aberto em Ω_1 sempre que U é aberto em Ω_2 . Como os abertos geram $\mathcal{B}(\Omega_2)$, o Teorema 4.3 mostra que f é $\mathcal{B}(\Omega_1) - \mathcal{B}(\Omega_2)$ -mensurável. $\square$

Além de reconhecer funções mensuráveis, precisamos saber como combiná-las. O próximo resultado trata da composição.

Proposição 4.8 —

Dadas funções mensuráveis $g : \Omega_1 \rightarrow \Omega_2$ e $f : \Omega_2 \rightarrow \Omega_3$. Então a composição $f \circ g$ é mensurável.

Demonstração. Para todo conjunto mensurável A do espaço de chegada de f ,

$$(f \circ g)^{-1}(A) = g^{-1}(f^{-1}(A)).$$

Como f é mensurável, $f^{-1}(A)$ é mensurável no espaço intermediário; como g é mensurável, a pré-imagem desse conjunto por g é mensurável no espaço inicial. $\square$

Exemplo 4.9 — .

Se $f : (\Omega, \mathcal{F}) \rightarrow (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ é uma função mensurável, tal que

$$f(\omega) = (f_1(\omega), \dots, f_n(\omega)),$$

com $f_k : \Omega \rightarrow \mathbb{R}, k = 1, \dots, n$, então f_k é uma função mensurável, para cada $k = 1, \dots, n$.

Para ver isso tome a projeção canônica $\pi_k : \mathbb{R}^n \rightarrow \mathbb{R}$ dada por $\pi_k(x_1, \dots, x_n) = x_k$ e note que

$$\pi_k^{-1}((-\infty, x]) = \mathbb{R}^{k-1} \times (-\infty, x] \times \mathbb{R}^{n-k} \in \mathcal{B}(\mathbb{R}^n),$$

e, pelo corolário 4.4, π_k é mensurável.

Para ver que f_k é mensurável basta notar que $f_k = \pi_k \circ f$, e o resultado segue da proposição 4.8.



4.2. Variáveis Aleatórias

Uma variável aleatória representa uma quantidade associada a um experimento. Ao lançar um dado duas vezes, podemos observar o primeiro resultado, o segundo ou a soma dos dois. Cada escolha define uma variável aleatória sobre o mesmo espaço amostral.

Formalmente, definimos uma variável aleatória como uma função real mensurável quando colocamos em $\mathbb{R}$ a σ -álgebra de Borel $\mathcal{B}(\mathbb{R})$.

Definição 4.10 —

Uma **variável aleatória** é uma função mensurável que toma valores em $\mathbb{R}$

$$\xi : (\Omega, \mathcal{F}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R})).$$

Observação 4.11 —

Em muitas aplicações, interessa a distribuição dos valores de uma variável aleatória, enquanto a descrição particular do espaço amostral fica em segundo plano. A função continua sendo mensurável; muda o aspecto que colocamos em evidência.

No caso em que o espaço de chegada é $\mathbb{R}^n$, a função mensurável ξ é denominada **veter aleatório**. Neste caso é comum denotarmos $X = (X_1, \dots, X_n)$, onde cada $X_k, k = 1, \dots, n$ é uma variável aleatória, como vimos no exemplo 4.9.

Exemplo 4.12 — .

O primeiro exemplo que consideraremos é o lançamento de moeda. Nesse caso, temos que o espaço amostral é $\Omega = \{H, T\}$ e $\mathcal{F} = \mathcal{P}(\Omega)$.

Considere a variável aleatória

$$\xi : \Omega \rightarrow \mathbb{R}$$

definida como:

$$\xi(H) = 1,$$

$$\xi(T) = 0.$$

◁

As funções simples da seção anterior recebem, no contexto probabilístico, a denominação seguinte.

Definição 4.13 — Variável Aleatória Simples

Seja $(\Omega, \mathcal{F})$ um espaço mensurável. Uma aplicação $X : \Omega \rightarrow \mathbb{R}$ é dita variável aleatória simples se houver um $n \in \mathbb{N}$ e conjuntos mensuráveis mutuamente disjuntos $A_1, \dots, A_n \in \mathcal{F}$, assim como um conjunto finito $\{a_k, k = 1, \dots, n\} \subset \mathbb{R}$, tal que

$$X = \sum_{k=1}^n a_k \mathbb{1}_{A_k}.$$

É importante observar que uma variável aleatória simples pode ser expressa de várias maneiras diferentes, bastando para isso que escolhamos diferentes conjuntos mensuráveis mutuamente disjuntos. Além disso, para transformar $A_1, \dots, A_n \in \mathcal{F}$ em uma partição, como usamos no exemplo de função simples, basta fazer $A_0 = (\cup_{k=1}^n A_k)^c$ e $a_0 = 0$. Com isso teremos

$$X = \sum_{k=1}^n a_k \mathbb{1}_{A_k} = \sum_{k=0}^n a_k \mathbb{1}_{A_k}.$$

Retomamos agora o resultado da seção anterior na notação de variáveis aleatórias.

Teorema 4.14 —

Considere a função $\xi : \Omega \rightarrow \mathbb{R}$ no espaço de probabilidade $(\Omega, \mathcal{F}, \mathbf{P})$. Então ξ é uma variável aleatória se, e somente se, $\{\omega : \xi(\omega) \leq a\} \in \mathcal{F}$ para todo $a \in \mathbb{R}$.

Demonstração. A implicação direta segue da definição. Para a recíproca, use o Teorema 4.3 e a identidade

$$\mathcal{B}(\mathbb{R}) = \sigma\{\{(-\infty, a] : a \in \mathbb{R}\}\}.$$

Com efeito, $\xi^{-1}((-\infty, a]) = \{\omega : \xi(\omega) \leq a\} \in \mathcal{F}$ por hipótese. □

Adotaremos uma notação que permite formular os eventos diretamente em termos dos valores das variáveis.

Para uma variável aleatória ξ e um boreliano $A \in \mathcal{B}(\mathbb{R})$, escreveremos:

$$\{\xi \in A\} := \{\omega : \xi(\omega) \in A\} = \xi^{-1}(A).$$

Do mesmo modo, para $a, b \in \mathbb{R}$,

$$\{\xi \leq a\} := \{\omega : \xi(\omega) \leq a\} = \xi^{-1}((-\infty, a]),$$

$$\{\xi > b\} := \{\omega : \xi(\omega) > b\} = \xi^{-1}((b, +\infty)),$$

$$\{b \leq \xi \leq a\} := \{\omega : b \leq \xi(\omega) \leq a\} = \xi^{-1}([b, a]),$$

e assim por diante.

Para variáveis aleatórias X, Y , faremos

$$\{X \in A, Y \in B\} = \{X \in A\} \cap \{Y \in B\}.$$

A notação destaca os valores de ξ , mas continua designando eventos no espaço amostral. O domínio fica implícito, não desaparece.

4.3. Funções de Variáveis Aleatórias

Depois de construir variáveis aleatórias, queremos combiná-las sem abandonar a mensurabilidade. Vetores, somas, produtos, máximos e composições com funções borelianas preservam a estrutura necessária.

Proposição 4.15 —

Se $\xi_1, \dots, \xi_n$ são variáveis aleatórias reais, então $(\xi_1, \dots, \xi_n)$ é um vetor aleatório.

Demonstração. Utilizando o Teorema 4.3, basta verificar a mensurabilidade do mapa

$$\omega \rightarrow (\xi_1(\omega), \dots, \xi_n(\omega))$$

em uma classe de geradores da σ -álgebra de Borel em $\mathbb{R}^n$ como por exemplo os geradores

$$\{(-\infty, a_1] \times (-\infty, a_2] \times \dots \times (-\infty, a_n] : a_1, a_2, \dots, a_n \in \mathbb{R}\}.$$

Para esse fim consideraremos as pré-imagens:

$$\begin{aligned} & \{(\xi_1, \dots, \xi_n) \in (-\infty, a_1] \times (-\infty, a_2] \times \dots \times (-\infty, a_n]\} \\ &= \{\xi_1 \in (-\infty, a_1], \dots, \xi_n \in (-\infty, a_n]\} \\ &= \bigcap_{k=1}^n \{\xi_k \in (-\infty, a_k]\} = \bigcap_{k=1}^n \{\xi_k \leq a_k\}. \end{aligned}$$

Logo, cada conjunto $\{\xi_k \leq a\}$ é mensurável (pois ξ_k é uma variável aleatória), e suas intersecções são mensuráveis pelas propriedades de σ -álgebra. □

Teorema 4.16 — Funções de Variáveis Aleatórias

Dadas variáveis aleatórias reais $\xi_1, \dots, \xi_n$ e $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função mensurável. Então, $f(\xi_1, \dots, \xi_n)$ é a variável aleatória real.

Demonstração. A função $f(\xi_1, \dots, \xi_n)$ é composição de duas funções mensuráveis:

□ a função

$$\omega \mapsto (\xi_1(\omega), \dots, \xi_n(\omega))$$

é mensurável por 4.15.

□ a função

$$(x_1, \dots, x_n) \mapsto f(x_1, \dots, x_n)$$

é mensurável por hipótese .

Como a composição de funções mensuráveis é mensurável temos o resultado desejado. □

Exemplo 4.17 — .

Como corolário do Teorema anterior temos que se ξ é a variável aleatória, então:

□ ξ^2

□ $\sin(\xi)$

□ e^ξ

são variáveis aleatórias.

Se $\xi_1, \dots, \xi_n$ são variáveis aleatórias, então:

□ $\xi_1 + \dots + \xi_n$

□ $\xi_1 \cdot \xi_2 \cdot \dots \cdot \xi_n$

são variáveis aleatórias.

Teorema 4.18 — Limites de Variáveis Aleatórias

Se $\xi_1, \xi_2, \dots$ são variáveis aleatórias, então $\sup_n \xi_n$, $\inf_n \xi_n$, $\limsup_n \xi_n$ e $\liminf_n \xi_n$ são mensuráveis sempre que tomem valores reais. Em particular, se $\xi_n(\omega)$ converge para um limite real $\xi(\omega)$ para todo ω , então ξ é uma variável aleatória.

Demonstração. Para todo $a \in \mathbb{R}$,

$$\left\{ \sup_n \xi_n \leq a \right\} = \bigcap_{n=1}^{\infty} \{ \xi_n \leq a \},$$

logo o supremo é mensurável. De modo análogo,

$$\left\{ \inf_n \xi_n < a \right\} = \bigcup_{n=1}^{\infty} \{ \xi_n < a \},$$

e o Corolário 4.4 mostra que o ínfimo é mensurável. Finalmente,

$$\limsup_{n \rightarrow \infty} \xi_n = \inf_n \sup_{k \geq n} \xi_k, \quad \liminf_{n \rightarrow \infty} \xi_n = \sup_n \inf_{k \geq n} \xi_k,$$

de modo que ambos são mensuráveis pelas duas primeiras partes. Quando o limite pontual real existe, ele coincide com o limite superior e com o limite inferior, portanto também é mensurável. $\square$

Exemplo 4.19 — .

Se $\xi_1, \xi_2, \dots$ são variáveis aleatórias e a série converge pontualmente para um valor real, então

$$\sum_{n=1}^{\infty} \xi_n.$$

também o é.

Esse fato é verdadeiro porque as somas parciais são variáveis aleatórias e a soma é seu limite pontual. Se, além disso, $\xi_n \geq 0$ para todo n , a soma também é o supremo das somas parciais.

4.4. Variáveis Aleatórias Geram σ -álgebras

Uma variável aleatória não registra necessariamente toda a informação do experimento. Nesta seção identificaremos precisamente a informação que pode ser recuperada a partir de seus valores.

Começemos com um exemplo simples.

Exemplo 4.20 — .

Considere o exemplo do lançamento de uma moeda n vezes, de maneira independente. Como de costume, tomamos o espaço amostral $\Omega = \{0, 1\}^n$ (“0” representando coroa e “1” representando cara), e munimos com a σ -álgebra 2^Ω , uma vez que o espaço é finito.

Seja então a variável aleatória N = número de caras observadas. Explicitamente, $N : \Omega \rightarrow \mathbb{R}$ é dada por

$$N(\omega) = \sum_{k=1}^n \omega_k.$$

Esta variável só enxerga o número de caras, e portanto não carrega consigo nenhuma informação sobre quais lançamentos tais caras foram observadas ou se existem ou não lançamentos consecutivos de cara.

Assim, se soubermos que $N = 3$, por exemplo, sabemos apenas que o resultado está no evento

$$\{N = 3\} = \{\omega \in \Omega : N(\omega) = 3\} = \{\omega \in \Omega : \omega_1 + \dots + \omega_n = 3\},$$

e nada mais.

Qualquer evento cuja ocorrência possa ser decidida conhecendo apenas N tem a forma $\{N \in A\}$, com $A \in \mathcal{B}(\mathbb{R})$. Neste caso,

$$\{N \in A\} = \bigcup_{k \in A \cap \{1, \dots, n\}} \{N = k\},$$

e portanto toda informação dada por N pode ser vista como união de eventos $\{N = k\}$.

É interessante observar que a família $\{N \in A\} : A \in \mathcal{B}(\mathbb{R})\}$, que coleciona toda a *informação* contida na variável N , é de fato uma σ -álgebra. Com efeito, ela contém $\Omega = N^{-1}(\mathbb{R})$ e, como pré-imagens preservam complementos e uniões enumeráveis,

$$N^{-1}(A^c) = N^{-1}(A)^c, \quad N^{-1}\left(\bigcup_j A_j\right) = \bigcup_j N^{-1}(A_j).$$

Com isso estamos prontos para a próxima definição.

Definição 4.21 —

Dada uma variável aleatória ξ definida em um espaço mensurável $(\Omega, \mathcal{F})$, definimos a σ -álgebra gerada por ξ , e denotamos por $\sigma\langle\xi\rangle$, como a família

$$\begin{aligned} \sigma\langle\xi\rangle &:= \{\xi^{-1}(A) : A \in \mathcal{B}(\mathbb{R})\} \\ &= \{\{\xi \in A\} : A \in \mathcal{B}(\mathbb{R})\} \end{aligned}$$

Analogamente, dado um vetor aleatório $X = (X_1, X_2, \dots, X_n)$, a σ -álgebra gerada por $X_1, X_2, \dots, X_n$ é a família

$$\begin{aligned} \sigma\langle X_1, \dots, X_n \rangle &:= \sigma\langle X \rangle \\ &= \{X^{-1}(A) : A \in \mathcal{B}(\mathbb{R}^n)\} \\ &= \{\{(X_1, \dots, X_n) \in A\} : A \in \mathcal{B}(\mathbb{R}^n)\} \end{aligned}$$

Como comentamos anteriormente, $\sigma\langle\xi\rangle$ pode ser vista como a σ -álgebra que carrega toda a informação encapsulada pela variável aleatória ξ . Além disso, $\sigma\langle\xi\rangle$ é a menor σ -álgebra de subconjuntos de Ω para a qual ξ é mensurável.

O próximo resultado permite obter a σ -álgebra gerada por uma variável a partir de uma família geradora de borelianos.

Proposição 4.22 —

Seja ξ uma variável aleatória e $\mathcal{A} \subset \mathcal{B}(\mathbb{R})$ tal que $\mathcal{B}(\mathbb{R}) = \sigma\langle \mathcal{A} \rangle$. Nestas condições

$$\sigma\langle \xi \rangle = \sigma\langle \{\{\xi \in A\} : A \in \mathcal{A}\} \rangle.$$

O mesmo vale se ξ é um vetor aleatório e $\mathcal{A} \subset \mathcal{B}(\mathbb{R}^n)$ tal que $\mathcal{B}(\mathbb{R}^n) = \sigma\langle \mathcal{A} \rangle$.

Demonstração. Defina

$$\mathcal{F} := \sigma\langle \{\{\xi \in A\} : A \in \mathcal{A}\} \rangle, \quad \text{e}$$

$$\mathcal{G} := \{A \in \mathcal{B}(\mathbb{R}) : \{\xi \in A\} \in \mathcal{F}\}.$$

Observe que $\mathcal{A} \subseteq \mathcal{G} \subseteq \mathcal{B}(\mathbb{R})$. A classe $\mathcal{G}$ contém $\mathbb{R}$, pois $\{\xi \in \mathbb{R}\} = \Omega \in \mathcal{F}$. Se $A \in \mathcal{G}$, então

$$\{\xi \in A^c\} = \{\xi \in A\}^c \in \mathcal{F};$$

e, se $A_j \in \mathcal{G}$, então

$$\left\{ \xi \in \bigcup_j A_j \right\} = \bigcup_j \{\xi \in A_j\} \in \mathcal{F}.$$

Logo, $\mathcal{G}$ é uma σ -álgebra que contém $\mathcal{A}$, donde $\mathcal{G} = \mathcal{B}(\mathbb{R})$. Isso prova a primeira afirmação. O argumento para vetores é idêntico, substituindo $\mathbb{R}$ por $\mathbb{R}^n$. $\square$

Corolário 4.23 —

Se $X = (X_1, X_2, \dots, X_n)$, então

$$\sigma\langle X_1, \dots, X_n \rangle = \sigma\langle \{X_1 \in A_1, \dots, X_n \in A_n\} : A_1, \dots, A_n \in \mathcal{B}(\mathbb{R}) \rangle.$$

Demonstração. Basta notar que $\mathcal{B}(\mathbb{R}^n)$ é gerada por $\{A_1 \times \cdots \times A_n : A_k \in \mathcal{B}(\mathbb{R}), k = 1, \dots, n\}$ e que

$$\{X_1 \in A_1, \dots, X_n \in A_n\} = \{X \in A_1 \times \cdots \times A_n\}.$$

□

Exemplo 4.24 – .

Dados $\Omega = \{0, 1\}^{\mathbb{N}}$ e $\omega = (\omega_1, \omega_2, \dots) \in \Omega$, considere a variável aleatória $\xi_n : \Omega \rightarrow \mathbb{R}$ definida por $\xi_n(\omega) = \omega_n$ e seja $\mathcal{F}_n = \sigma\langle \xi_1, \dots, \xi_n \rangle$.

Calcule explicitamente $\mathcal{F}_1$ e $\mathcal{F}_2$

Solução: Como ξ_1 só toma dois valores 0 ou 1, $\sigma\langle \xi_1 \rangle$ é gerado por

$$\xi_1^{-1}(\{0\}) = A_1 = \{\omega : \omega_1 = 0\} \quad \text{e}$$

$$\xi_1^{-1}(\{1\}) = B_1 = \{\omega : \omega_1 = 1\} = A_1^c$$

Logo $\mathcal{F}_1 = \{\Omega, \emptyset, A_1, A_1^c\}$

De modo análogo, $\sigma\langle \xi_2 \rangle$ é gerado por

$$\xi_2^{-1}(\{0\}) = A_2 = \{\omega : \omega_2 = 0\} \quad \text{e}$$

$$\xi_2^{-1}(\{1\}) = B_2 = \{\omega : \omega_2 = 1\} = A_2^c$$

Assim $\sigma\langle \xi_2 \rangle = \{\Omega, \emptyset, A_2, A_2^c\}$

Os quatro átomos de $\mathcal{F}_2$ são

$$A_1 \cap A_2, \quad A_1 \cap A_2^c, \quad A_1^c \cap A_2, \quad A_1^c \cap A_2^c.$$

Logo, $\mathcal{F}_2$ é a família de todas as uniões desses quatro átomos e possui $2^4 = 16$ elementos. Essa descrição inclui, em particular, os dois eventos de paridade que faltariam numa lista formada apenas por linhas, colunas, átomos e complementos de átomos.

◁

Proposição 4.25 — Lema de fatoração

Dados ξ uma variável aleatória em Ω , $\sigma\langle\xi\rangle$ a σ -álgebra gerada por ξ e η uma variável aleatória em Ω . Então η é mensurável com respeito a $\sigma\langle\xi\rangle$ se e somente se existe uma função mensurável $f : \mathbb{R} \rightarrow \mathbb{R}$ tal que $\eta = f(\xi)$.

Demonstração. Se $\eta = f(\xi)$ para alguma função boreliana f , então ξ é $\sigma\langle\xi\rangle$ - $\mathcal{B}(\mathbb{R})$ -mensurável por definição. A composição $f \circ \xi$ é, portanto, $\sigma\langle\xi\rangle$ -mensurável.

Reciprocamente, suponha que η seja $\sigma\langle\xi\rangle$ -mensurável. Considere a bijeção contínua

$$h : \mathbb{R} \longrightarrow (0,1), \quad h(t) = \frac{1}{2} + \frac{1}{\pi} \arctan t,$$

e ponha $Z = h(\eta)$. Para $n \geq 1$, defina a aproximação diádica

$$Z_n = 2^{-n} \lfloor 2^n Z \rfloor.$$

Ela assume apenas os valores $k2^{-n}$, $0 \leq k < 2^n$. Cada conjunto $A_{n,k} := \{Z_n = k2^{-n}\}$ pertence a $\sigma\langle\xi\rangle$; logo, pela definição dessa σ -álgebra, existe $B'_{n,k} \in \mathcal{B}(\mathbb{R})$ tal que $A_{n,k} = \xi^{-1}(B'_{n,k})$.

Podemos tornar esses borelianos disjuntos sem alterar suas pré-imagens: ponha $B_{n,0} = B'_{n,0}$ e, para $k \geq 1$,

$$B_{n,k} = B'_{n,k} \setminus \bigcup_{j < k} B'_{n,j}.$$

De fato, as pré-imagens $A_{n,k}$ já são disjuntas. A função boreliana

$$g_n(x) = \sum_{k=0}^{2^n-1} k2^{-n} \mathbb{1}_{B_{n,k}}(x)$$

satisfaz $g_n(\xi) = Z_n$. Como $Z_n \rightarrow Z$, a função boreliana $g := \limsup_n g_n$ satisfaz $g(\xi) = Z$ e toma valores em $(0,1)$ sobre $\xi(\Omega)$. Finalmente, defina

$$f(x) = \begin{cases} h^{-1}(g(x)), & 0 < g(x) < 1, \\ 0, & \text{caso contrário.} \end{cases}$$

A função f é boreliana e, para todo $\omega \in \Omega$, $f(\xi(\omega)) = \eta(\omega)$. □

A σ -álgebra gerada reúne a informação registrada pela variável. Para trabalhar no espaço de seus valores, falta transportar também a probabilidade. O espaço original pode sair de cena; suas probabilidades, não.

4.5. Variáveis Aleatórias Induzem Medidas de Probabilidade em $\mathbb{R}$

Seja ξ uma variável aleatória real. Para todo $A \in \mathcal{B}(\mathbb{R})$, defina

$$\mathbf{P}_\xi(A) := \mathbf{P}(\xi \in A) = \mathbf{P}(\xi^{-1}(A)).$$

Essa função é uma probabilidade em $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. De fato, $\mathbf{P}_\xi(\mathbb{R}) = \mathbf{P}(\Omega) = 1$ e, para borelianos disjuntos $A_1, A_2, \dots$,

$$\mathbf{P}_\xi\left(\bigcup_n A_n\right) = \mathbf{P}\left(\bigcup_n \xi^{-1}(A_n)\right) = \sum_n \mathbf{P}(\xi^{-1}(A_n)) = \sum_n \mathbf{P}_\xi(A_n).$$

A probabilidade $\mathbf{P}_\xi$ é denominada **distribuição** de ξ .

Na teoria da medida, o objeto central é o estudo dos espaços de medida $(\Omega, \mathcal{F}, \mu)$. Na teoria da probabilidade, em muitas situações tentamos esquecer rapidamente os espaços da medida e discutir as distribuições.

Falaremos mais sobre distribuições no Capítulo 6.

4.6. Independência de Variáveis Aleatórias

A independência de variáveis aleatórias significa que a informação produzida por uma delas não altera as probabilidades associadas às demais. Formalmente, isso se traduz na independência das σ -álgebras que elas geram.

Definição 4.26 — Variáveis aleatórias independentes

As variáveis aleatórias $\xi_1, \dots, \xi_n$ são ditas independentes se

$$\mathbf{P}(\xi_1 \in B_1, \dots, \xi_n \in B_n) = \mathbf{P}(\xi_1 \in B_1) \cdot \dots \cdot \mathbf{P}(\xi_n \in B_n)$$

para todos os conjuntos de Borel $B_1, \dots, B_n$ em $\mathbb{R}$.

Exemplo 4.27 — Construindo duas moedas independentes.

Dado o espaço amostral de uma moeda $\Omega = \{1, 0\}$, considere $\Omega_1 = \Omega \times \Omega$, munido da σ -álgebra das partes e da probabilidade uniforme. Ou seja, para $\omega = (\omega_1, \omega_2) \in \Omega_1$, vale $\mathbf{P}(\{\omega\}) = 1/4$.

Considere as variáveis aleatórias:

$$\xi(\omega_1, \omega_2) = \omega_1$$

$$\eta(\omega_1, \omega_2) = \omega_2$$

As variáveis ξ e η são independentes.

**Teorema 4.28 —**

As variáveis aleatórias $\xi_1, \dots, \xi_n$ são independentes se e somente se $\sigma\langle \xi_1 \rangle, \dots, \sigma\langle \xi_n \rangle$ são independentes.

Demonstração. Para cada k , a família

$$\sigma\langle \xi_k \rangle = \{\{\xi_k \in B\} : B \in \mathcal{B}(\mathbb{R})\}$$

é precisamente a σ -álgebra gerada por ξ_k . Assim, a identidade que define a independência dessas σ -álgebras, aplicada a uma escolha $\{\xi_k \in B_k\}$ de cada uma, é exatamente a identidade da Definição 4.26. Isso prova as duas implicações. $\square$

Como consequência do Teorema anterior temos que:

Proposição 4.29 —

As variáveis aleatórias $\xi_1, \dots, \xi_n$ são independentes se e somente se

$$\mathbf{P}(\xi_1 \leq x_1, \dots, \xi_n \leq x_n) = \mathbf{P}(\xi_1 \leq x_1) \cdot \dots \cdot \mathbf{P}(\xi_n \leq x_n)$$

para todo $x_1, \dots, x_n \in \mathbb{R}$.

Demonstração. A implicação direta é imediata, tomando $B_k = (-\infty, x_k]$ na definição de independência.

Para a recíproca, suponha a identidade do enunciado. Para cada k , defina

$$\mathcal{A}_k := \{\{\xi_k \leq x\} : x \in \mathbb{R}\}.$$

Cada $\mathcal{A}_k$ é um π -sistema. A hipótese também implica as identidades de independência para qualquer subcoleção dos índices. De fato, para omitir os índices fora de $H \subseteq \{1, \dots, n\}$, substitua seus limiares por t e faça $t \rightarrow \infty$; a continuidade por baixo fornece

$$\mathbf{P}\left(\bigcap_{k \in H} \{\xi_k \leq x_k\}\right) = \prod_{k \in H} \mathbf{P}(\xi_k \leq x_k).$$

Portanto, as famílias $\mathcal{A}_k$ são independentes. Como

$$\sigma\langle \mathcal{A}_k \rangle = \sigma\langle \xi_k \rangle,$$

o teorema sobre π -sistemas independentes do capítulo anterior mostra que $\sigma\langle \xi_1 \rangle, \dots, \sigma\langle \xi_n \rangle$ são independentes. Pelo teorema precedente, as variáveis aleatórias são independentes. □

Proposição 4.30 —

Dadas funções $\{f_k\}_{k=1}^n$ mensuráveis, Se $\xi_1, \dots, \xi_n$ são variáveis aleatórias independentes, então $f_1(\xi_1), \dots, f_n(\xi_n)$ são independentes.

Demonstração. Para borelianos $B_1, \dots, B_n$, a mensurabilidade de cada f_k dá $f_k^{-1}(B_k) \in \mathcal{B}(\mathbb{R})$. Além disso,

$$\{f_k(\xi_k) \in B_k\} = \{\xi_k \in f_k^{-1}(B_k)\}.$$

Pela independência das ξ_k ,

$$\begin{aligned} \mathbf{P}(f_1(\xi_1) \in B_1, \dots, f_n(\xi_n) \in B_n) \\ = \prod_{k=1}^n \mathbf{P}(\xi_k \in f_k^{-1}(B_k)) = \prod_{k=1}^n \mathbf{P}(f_k(\xi_k) \in B_k), \end{aligned}$$

que é a identidade desejada. □

Como consequência da última proposição temos que se ξ, η são independentes, então ξ^2, η^4 são independentes.

Teorema 4.31 — Lei 0-1 de Kolmogorov para Variáveis Aleatórias

Suponha que $(\xi_n : n \in \mathbb{N})$ é uma sequência de variáveis aleatórias independentes. Então a σ -álgebra caudal $\mathcal{T}$ de $(\sigma\langle \xi_n \rangle : n \in \mathbb{N})$ contém somente eventos de probabilidade 0 ou 1.

Mais ainda, toda variável aleatória mensurável com respeito a $\mathcal{T}$ é constante quase certamente.

Demonstração. Aplicando o Teorema 4.28 a cada subcoleção finita, vemos que as σ -álgebras $\sigma\langle \xi_n \rangle$, $n \geq 1$, são independentes. A primeira afirmação é, portanto, exatamente a Lei 0-1 de Kolmogorov aplicada a essa sequência de σ -álgebras.

Seja agora η uma variável aleatória $\mathcal{T}$ -mensurável. Para todo $y \in \mathbb{R}$, o evento $\{\eta \leq y\}$ pertence a $\mathcal{T}$; logo,

$$F(y) := \mathbf{P}(\eta \leq y) \in \{0, 1\}.$$

Como $F(y) \rightarrow 0$ quando $y \rightarrow -\infty$ e $F(y) \rightarrow 1$ quando $y \rightarrow \infty$, o número

$$c := \inf\{y \in \mathbb{R} : F(y) = 1\}$$

é real. Pela monotonicidade de F , temos $F(y) = 0$ para $y < c$ e $F(y) = 1$ para $y > c$. Escolhendo sequências racionais que se aproximam de c pelos dois lados e usando

continuidade da probabilidade,

$$\mathbf{P}(\eta < c) = 0, \quad \mathbf{P}(\eta > c) = 0.$$

Consequentemente, $\mathbf{P}(\eta = c) = 1$.

□

EXERCÍCIOS

Ex. 4.1 — Sejam $(\Omega, \mathcal{F})$, $(S, \mathcal{S})$ e $(T, \mathcal{T})$ espaços mensuráveis. Prove que a composição de duas funções mensuráveis

$$f : (\Omega, \mathcal{F}) \rightarrow (S, \mathcal{S}), \quad g : (S, \mathcal{S}) \rightarrow (T, \mathcal{T})$$

é mensurável.

Ex. 4.2 — Operações com variáveis aleatórias. Se X e Y são variáveis aleatórias reais no mesmo espaço de probabilidade $(\Omega, \mathcal{F}, \mathbf{P})$, prove que também são variáveis aleatórias:

1. $|X|$;
2. $X + Y$;
3. XY ;
4. $\max\{X, Y\}$;
5. $\min\{X, Y\}$.

Ex. 4.3 — Seja X uma variável aleatória tal que $\mathbf{P}(X > 0) > 0$. Prove que existe $\delta > 0$ para o qual $\mathbf{P}(X > \delta) > 0$.

Ex. 4.4 — Amostragem de uma variável exponencial. Seja $U \sim U(0,1)$ e fixe $\lambda > 0$. Encontre uma função mensurável $g : (0,1) \rightarrow [0, \infty)$ tal que $X = g(U)$ tenha distribuição $\text{Exp}(\lambda)$. Prove sua afirmação calculando a função de distribuição de X .

Ex. 4.5 — A propriedade de falta de memória. Diremos que uma variável aleatória real não negativa X possui a propriedade de falta de memória quando

$$\mathbf{P}(X > s + t \mid X > s) = \mathbf{P}(X > t)$$

para todos $s, t \geq 0$ para os quais $\mathbf{P}(X > s) > 0$.

1. Mostre que essa identidade é equivalente a

$$\mathbf{P}(X > u \mid X > s) = \mathbf{P}(X > u - s), \quad 0 \leq s < u.$$

em todos os casos em que $\mathbf{P}(X > s) > 0$.

2. Prove que toda variável com distribuição exponencial possui a propriedade de falta de memória.
3. Reciprocamente, suponha que X seja finita quase certamente, não degenerada e tenha a propriedade de falta de memória. Prove que $X \sim \text{Exp}(\lambda)$ para algum $\lambda > 0$.

Dica. Para $G(t) = \mathbf{P}(X > t)$, obtenha $G(s + t) = G(s)G(t)$. Resolva primeiro a equação nos racionais não negativos e use a monotonicidade de G para passar aos reais.

Ex. 4.6 — Seja X uma variável aleatória independente de si mesma, isto é, suponha que $\sigma(X)$ seja independente de $\sigma(X)$. Prove que X é constante quase certamente.

Ex. 4.7 — Fixe $\varepsilon, \delta > 0$ e seja $(X_i)_{i \geq 1}$ uma sequência de variáveis aleatórias não negativas e independentes tal que

$$\mathbf{P}(X_i > \delta) \geq \varepsilon \quad \text{para todo } i \geq 1.$$

Prove que

$$\sum_{i=1}^{\infty} X_i = \infty \quad \text{quase certamente.}$$

Ex. 4.8 — Se X e Y são independentes, prove que cada uma das famílias

$$\{X^+, Y^+\}, \quad \{X^+, Y^-\}, \quad \{X^-, Y^+\}, \quad \{X^-, Y^-\}$$

é independente.

Ex. 4.9 — Variáveis mensuráveis numa partição enumerável. Seja $(E_n)_{n \geq 1}$ uma partição enumerável de Ω em conjuntos não vazios e ponha $\mathcal{F} = \sigma(E_n : n \geq 1)$.

1. Prove que

$$\mathcal{F} = \left\{ \bigcup_{n \in I} E_n : I \subseteq \mathbb{N} \right\}.$$

2. Caracterize todas as funções $X : (\Omega, \mathcal{F}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mensuráveis.

Integral de Lebesgue

5.1. Integral de Lebesgue

A integral de Lebesgue é definida em um espaço de medida arbitrário $(\Omega, \mathcal{F}, \mu)$. Não suporemos que $\mu(\Omega)$ seja finita. Quando a finitude da medida for necessária, ela aparecerá explicitamente no enunciado; no caso de medidas de probabilidade, essa hipótese é automaticamente satisfeita.

Quando permitirmos funções com valores em $[0, \infty]$, esse conjunto será munido da σ -álgebra gerada pelos intervalos $(a, \infty]$, $a \geq 0$. Equivalentemente, trata-se da σ -álgebra de Borel da reta estendida, restrita a $[0, \infty]$. Assim, $f : \Omega \rightarrow [0, \infty]$ é mensurável se, e somente se, $\{f > a\} \in \mathcal{F}$ para todo $a \geq 0$. Na aritmética de $[0, \infty]$ adotaremos ainda as convenções

$$0 \cdot \infty = \infty \cdot 0 = 0, \quad a \cdot \infty = \infty \quad \text{para } a > 0.$$

A integral de Riemann organiza a soma particionando o domínio. Num espaço abstrato, essa geometria pode não existir. Lebesgue desloca a aproximação para os

valores da função. A construção avança por camadas, começando pelas funções simples e chegando às funções mensuráveis gerais.

Se f é simples e não negativa, escrevemos

$$f = \sum_{k=1}^n a_k \mathbb{1}_{A_k}$$

para alguma partição mensurável $\Omega = \bigcup_{k=1}^n A_k$, com $a_k \geq 0$, e definimos

$$\int f \, d\mu = \sum_{k=1}^n a_k \mu(A_k).$$

Essa soma assume valores em $[0, \infty]$. Para uma função simples com valores reais, a mesma fórmula é usada sempre que a parte positiva ou a parte negativa tiver integral finita; em particular, ela é sempre bem definida para funções simples integráveis. Essa ressalva evita a forma indeterminada $\infty - \infty$ quando a medida não é finita. Dois infinitos não se cancelam por cortesia da notação.

Para uma função mensurável não negativa f , tomamos o supremo das integrais das funções simples que aproximam f por baixo:

$$\int f \, d\mu = \sup_{0 \leq \phi \leq f} \int \phi \, d\mu$$

onde ϕ percorre as funções simples não negativas. Se essa integral for finita, diremos que f é **integrável**.

Finalmente, para uma função mensurável geral, separamos suas partes positiva e negativa,

$$f = f^+ - f^-,$$

onde $f^+ = \max(f, 0)$ e $f^- = \max(-f, 0)$, e integramos cada parte. Assim, a construção inteira se apoia no primeiro caso: funções simples não negativas.

A construção exige primeiro uma verificação: a integral de uma função simples deve depender da função, não da representação escolhida.

Proposição 5.1 —

Dadas duas representações de uma função simples não negativa f por partições mensuráveis,

$$f = \sum_{i=1}^n a_i \mathbb{1}_{A_i} \quad \text{e} \quad f = \sum_{j=1}^m b_j \mathbb{1}_{B_j}$$

então

$$\sum_{i=1}^n a_i \mu(A_i) = \sum_{j=1}^m b_j \mu(B_j)$$

Demonstração. Os conjuntos $A_i \cap B_j$ formam um refinamento comum das duas partições. Se $A_i \cap B_j \neq \emptyset$, então $f(x) = a_i = b_j$ nesse conjunto. Se a interseção é vazia, sua medida é zero, de modo que o termo correspondente também não contribui. Portanto,

$$\sum_{i=1}^n a_i \mu(A_i) = \sum_{i=1}^n \sum_{j=1}^m a_i \mu(A_i \cap B_j) \quad (5.1)$$

$$= \sum_{j=1}^m \sum_{i=1}^n b_j \mu(A_i \cap B_j) \quad (5.2)$$

$$= \sum_{j=1}^m b_j \mu(B_j) \quad (5.3)$$

As igualdades valem também quando o valor comum é $+\infty$, pois todos os termos são não negativos. Para funções simples integráveis, aplica-se o mesmo argumento separadamente às partes positiva e negativa. $\square$

Proposição 5.2 — Aproximação por Funções Simples

Toda função mensurável não negativa $f : \Omega \rightarrow [0, \infty]$ é o limite pontual de uma sequência monótona crescente de funções simples não negativas. Reciprocamente, o limite pontual de uma sequência monótona crescente de funções simples é mensurável.

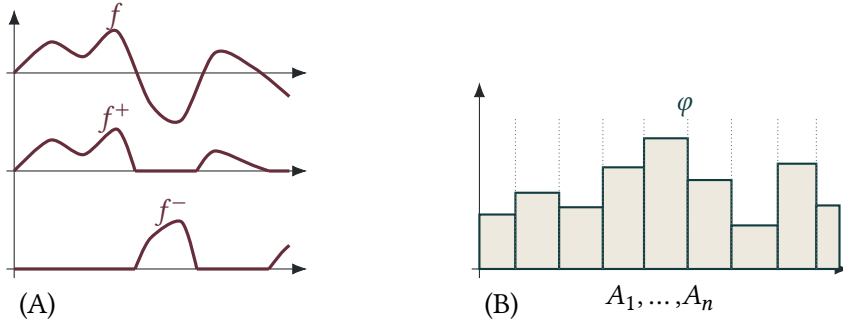


Figura 5.1.: Partes positiva e negativa de uma função e representação de uma função simples numa partição finita. Adaptado de [26].

Demonstração.

$\Rightarrow$ Considere f uma função mensurável não negativa definida sobre o espaço de medida $(\Omega, \mathcal{F}, \mu)$. Para cada $n \in \mathbb{N}$, subdivida $[0, \infty]$ em $2^{2n} + 1$ intervalos, 2^{2n} dos quais têm comprimento 2^{-n} . Para cada n , considere

$$I_{n,k} = \left[\frac{k-1}{2^n}, \frac{k}{2^n} \right)$$

para $k = 1, 2, \dots, 2^{2n}$ e $I_{n, 2^{2n}+1} = [2^n, \infty]$.

Observe que, para n fixo, os conjuntos $I_{n,k}$ são disjuntos e cobrem a reta real não negativa.

Agora, defina os conjuntos mensuráveis

$$A_{n,k} = f^{-1}(I_{n,k}) \text{ para } k = 1, 2, \dots, 2^{2n} + 1.$$

Então as funções simples

$$f_n = \sum_{k=1}^{2^{2n}+1} \frac{k-1}{2^n} \mathbf{1}_{A_{n,k}}$$

satisfazem $0 \leq f_n \leq f_{n+1} \leq f$. De fato, elas são o arredondamento diádico de f por baixo, truncado na altura 2^n ; nos pontos em que f é finita,

$$f_n = (2^{-n} \lfloor 2^n f \rfloor) \wedge 2^n.$$

Para $f(x) < \infty$, vale $0 \leq f(x) - f_n(x) < 2^{-n}$ tão logo $f(x) < 2^n$; se $f(x) = \infty$, então $f_n(x) = 2^n$. Logo $f_n \uparrow f$ pontualmente. Se f for limitada, a convergência é uniforme.

◀ Já provamos que o limite de funções mensuráveis é mensurável. ◻

Ou seja, a classe das funções mensuráveis é o fecho por limites pontuais das funções simples.

Adaptando a aproximação anterior, obtemos:

Proposição 5.3 — Aproximação Uniforme por Funções Simples

Sejam $A \in \mathcal{F}$ e $f : A \rightarrow [0, \infty)$ uma função mensurável e limitada, considerando em A a σ -álgebra traço. Para todo $\varepsilon > 0$ existem funções simples mensuráveis ϕ_ε e ψ_ε em A tais que:

1 $\phi_\varepsilon(x) \leq f(x) \leq \psi_\varepsilon(x)$ para todo $x \in A$.

2 $0 \leq \psi_\varepsilon(x) - \phi_\varepsilon(x) < \varepsilon$ para todo $x \in A$.

Demonstração. Escolha $N \in \mathbb{N}$ tal que $f < N\varepsilon$ em A e ponha

$$A_k = \{x \in A : k\varepsilon/2 \leq f(x) < (k+1)\varepsilon/2\}, \quad k = 0, \dots, 2N-1.$$

Esses conjuntos são mensuráveis e formam uma partição de A . Defina

$$\phi_\varepsilon = \sum_{k=0}^{2N-1} \frac{k\varepsilon}{2} \mathbb{1}_{A_k}, \quad \psi_\varepsilon = \sum_{k=0}^{2N-1} \frac{(k+1)\varepsilon}{2} \mathbb{1}_{A_k}.$$

Em cada A_k , a primeira função é o extremo inferior e a segunda o extremo superior do intervalo correspondente. Assim, $\phi_\varepsilon \leq f \leq \psi_\varepsilon$ e $\psi_\varepsilon - \phi_\varepsilon = \varepsilon/2 < \varepsilon$. ◻

Lema 5.4 – Passagem ao limite para aproximações simples

Se $0 \leq s_n \uparrow f$ e cada s_n é simples, então

$$\int s_n \, d\mu \uparrow \int f \, d\mu.$$

Demonstração. Pela monotonicidade que decorre diretamente da definição por supremo, $\alpha := \lim_n \int s_n \, d\mu$ existe e satisfaz $\alpha \leq \int f \, d\mu$. Para a desigualdade inversa, fixe uma função simples $0 \leq s \leq f$ e $0 < c < 1$. Os conjuntos

$$E_n = \{s_n \geq cs\}$$

crescem para Ω . Com efeito, se $s(x) = 0$ a afirmação é imediata; se $s(x) > 0$, então $cs(x) < s(x) \leq f(x)$ e $s_n(x) \uparrow f(x)$. Como s é simples, a continuidade por baixo de μ , aplicada a cada conjunto de nível de s , dá

$$\alpha \geq \lim_n c \int_{E_n} s \, d\mu = c \int s \, d\mu.$$

Fazendo $c \uparrow 1$ e tomando o supremo sobre todas as funções simples $0 \leq s \leq f$, obtemos $\alpha \geq \int f \, d\mu$. □

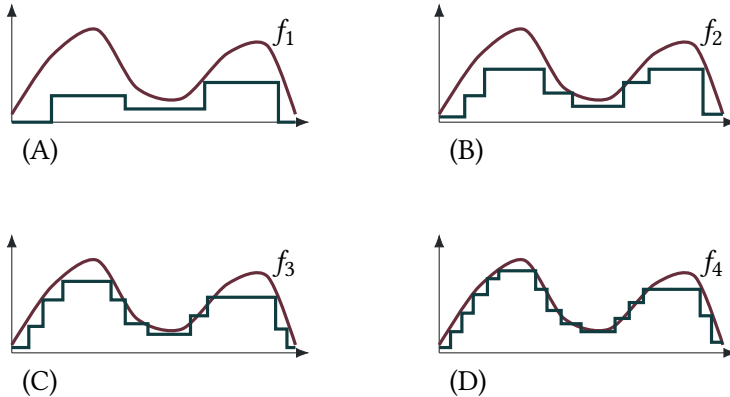


Figura 5.2.: Refinamentos sucessivos de aproximações simples por baixo. Adaptado de [26].

Definição 5.5 — Integrável

- 1** Dizemos que f é **integrável** se $|f|$ é integrável. Equivalentemente, se ambas as funções f^+ e f^- são integráveis. E nesse caso definimos

$$\int f \, d\mu = \int f^+ \, d\mu - \int f^- \, d\mu.$$

- 2** Dizemos que f é **quase integrável** se pelo menos uma das funções f^+ e f^- é integrável. E nesse caso também definimos

$$\int f \, d\mu = \int f^+ \, d\mu - \int f^- \, d\mu.$$

- 3** A integral em um conjunto mensurável $A \subseteq \Omega$ é definida como

$$\int_A f \, d\mu = \int_{\Omega} f \cdot \mathbf{1}_A \, d\mu.$$

5.2. Propriedades da integral de Lebesgue

As propriedades da integral seguem a mesma arquitetura de sua definição:

Estratégia dos três passos

- ☐ primeiro, demonstramos a propriedade para funções simples;
- ☐ depois, passamos às funções mensuráveis não negativas por aproximação monótona;
- ☐ por fim, tratamos funções integráveis separando suas partes positiva e negativa.

Proposição 5.6 — Propriedades da integral

Sejam f e g funções integráveis e $a, b \in \mathbb{R}$. Então:

- 1 (Monotonicidade.) Se $f \leq g$ em quase todo ponto, então $\int f \, d\mu \leq \int g \, d\mu$.
- 2 Se $f = g$ em quase todo ponto, então as integrais são iguais.
- 3 (Linearidade.) $\int (af + bg) \, d\mu = a \int f \, d\mu + b \int g \, d\mu$.
- 4 $\left| \int f \, d\mu \right| \leq \int |f| \, d\mu$.

Demonstração. Começemos pela linearidade para funções não negativas. Para funções simples ela segue do refinamento comum das partições. De fato, se $\varphi = \sum_i a_i \mathbb{1}_{A_i}$, $\psi = \sum_j b_j \mathbb{1}_{B_j}$, então

$$\int_{\Omega} \varphi \, d\mu + \int_{\Omega} \psi \, d\mu = \sum_i a_i \mu(A_i) + \sum_j b_j \mu(B_j) \quad (5.4)$$

$$= \sum_i \sum_j a_i \mu(A_i \cap B_j) + \sum_j \sum_i b_j \mu(B_j \cap A_i) \quad (5.5)$$

$$= \sum_i \sum_j (a_i + b_j) \mu(A_i \cap B_j) \quad (5.6)$$

$$= \int_{\Omega} (\varphi + \psi) \, d\mu. \quad (5.7)$$

Se $f, g \geq 0$ e $a, b \geq 0$, tome funções simples $f_n \uparrow f$ e $g_n \uparrow g$. Então $af_n + bg_n \uparrow af + bg$. Pelo lema de passagem ao limite para aproximações simples,

$$\int (af + bg) \, d\mu = \lim_n \int (af_n + bg_n) \, d\mu = a \int f \, d\mu + b \int g \, d\mu.$$

Agora sejam f e g integráveis e escreva

$$af + bg = u - v,$$

onde

$$u = a^+ f^+ + a^- f^- + b^+ g^+ + b^- g^-, \quad v = a^+ f^- + a^- f^+ + b^+ g^- + b^- g^+.$$

As funções u e v são não negativas e integráveis. Como $(af + bg)^+ + v = (af + bg)^- + u$, a linearidade já provada para funções não negativas fornece

$$\int (af + bg) \, d\mu = \int u \, d\mu - \int v \, d\mu = a \int f \, d\mu + b \int g \, d\mu.$$

Precisamos ainda registrar por que conjuntos nulos não afetam a integral. Se $N \in \mathcal{F}$ e $\mu(N) = 0$, então, para toda função não negativa h ,

$$\int h \mathbb{1}_N \, d\mu = 0.$$

Com efeito, toda função simples abaixo de $h \mathbb{1}_N$ se anula fora de N e, portanto, tem integral zero. Aplicando isso a h^+ e h^- , vemos que uma função integrável e qualquer modificação mensurável sua em N têm a mesma integral.

Para a monotonicidade, podemos assim modificar f e g no conjunto nulo excepcional e supor que $f \leq g$ em todo ponto. A função $g - f$ é não negativa e integrável; pela linearidade,

$$0 \leq \int (g - f) \, d\mu = \int g \, d\mu - \int f \, d\mu.$$

A igualdade quase certa implica as duas desigualdades e, portanto, a igualdade das integrais. Finalmente, de $-|f| \leq f \leq |f|$ e da monotonicidade obtemos

$$-\int |f| \, d\mu \leq \int f \, d\mu \leq \int |f| \, d\mu,$$

o que prova a última afirmação. □

Teorema 5.7 —

Sejam $(\Omega, \mathcal{F}, \mu)$ um espaço de medida e $f, g : \Omega \rightarrow [0, \infty]$ funções mensuráveis. Então:

1 $\int_{\Omega} f \, d\mu < \infty$ então $f < \infty$ μ -q.c.

2 $\int_{\Omega} f \, d\mu = 0 \iff f = 0$ μ -q.c.

3 Se $f = g$ μ -q.c. então $\int_{\Omega} f \, d\mu = \int_{\Omega} g \, d\mu$.

Demonstração. Seja $I = \{f = \infty\}$. Para todo $n \geq 1$, $n\mathbb{1}_I \leq f$ e, portanto,

$$n\mu(I) \leq \int f \, d\mu.$$

Se a integral é finita, segue que $\mu(I) = 0$, o que prova o primeiro item.

Se $f = 0$ quase certamente, toda função simples $0 \leq s \leq f$ se anula fora de um conjunto nulo e tem integral zero. Pela definição por supremo, $\int f \, d\mu = 0$. Reciprocamente, se a integral é zero, então, para $A_n = \{f \geq 1/n\}$,

$$0 \leq \frac{1}{n}\mu(A_n) = \int \frac{1}{n}\mathbb{1}_{A_n} \, d\mu \leq \int f \, d\mu = 0.$$

Como $\{f > 0\} = \bigcup_{n \geq 1} A_n$, temos $f = 0$ quase certamente.

Por fim, se $f = g$ fora de um conjunto nulo N , então $f\mathbb{1}_{N^c} = g\mathbb{1}_{N^c}$. As integrais de $f\mathbb{1}_N$ e $g\mathbb{1}_N$ são zero pelo argumento do parágrafo anterior. A aditividade para funções não negativas conclui que $\int f \, d\mu = \int g \, d\mu$, inclusive quando o valor comum é infinito. $\square$

Observação. Uma consequência do Teorema ??, item [2](#), aplicada às partes positiva e negativa, é que, para qualquer função $f : \Omega \rightarrow \mathbb{R}$ cuja integral esteja definida, temos a liberdade de mudar os valores de $f(\omega)$ num conjunto μ -negligenciável sem alterar o valor da integral, desde que a função resultante seja mensurável.

Definição 5.8 – Integrável

Seja $(\Omega, \mathcal{F}, \mu)$ um espaço de medida. Então $\mathcal{L}_1(\mu)$ denota o conjunto de funções $f : \Omega \rightarrow \mathbb{R}$ que são mensuráveis $\mathcal{F}$ com : $\int_{\Omega} f^+ \, d\mu < \infty$ e $\int_{\Omega} f^- \, d\mu < \infty$;

Em conjuntos de medida finita, a mensurabilidade é a única condição adicional necessária para que uma função limitada seja integrável.

Teorema 5.9 —

Sejam $A \in \mathcal{F}$, com $\mu(A) < \infty$, e $f : A \rightarrow \mathbb{R}$ uma função limitada. Então f é Lebesgue integrável sobre A se e somente se f é mensurável para a σ -álgebra traço em A .

Demonstração. Pela definição adotada, toda função integrável é mensurável. Para a recíproca, suponha que f seja mensurável e escolha $M < \infty$ tal que $|f| \leq M$ em A . Então

$$\int_A |f| \, d\mu \leq M\mu(A) < \infty,$$

e, portanto, f é integrável. É exatamente nesta estimativa que se usa a finitude de $\mu(A)$; não é necessário supor que o espaço de medida seja completo. $\square$

5.3. Comparando a Integral de Lebesgue e de Riemann

As integrais de Riemann e Lebesgue não competem pelo mesmo território: a segunda amplia a primeira e se comporta melhor sob limites. Para tornar essa relação precisa, começamos recordando quais funções limitadas são integráveis no sentido de Riemann.

Teorema 5.10 — Critério de Lebesgue

Uma função limitada num intervalo compacto $[a, b]$ é Riemann integrável se e somente se o conjunto de seus pontos de descontinuidade tem medida nula.

Demonstração. Para $x \in [a, b]$, defina a oscilação de f em x por

$$\omega_f(x) = \inf_{\delta > 0} \left(\sup_{\substack{y \in [a, b] \\ |y-x| < \delta}} f(y) - \inf_{\substack{y \in [a, b] \\ |y-x| < \delta}} f(y) \right).$$

A função f é descontínua em x exatamente quando $\omega_f(x) > 0$. Além disso, cada conjunto $D_m = \{x : \omega_f(x) \geq 1/m\}$ é fechado: se $\omega_f(x) < 1/m$, alguma vizinhança de x tem oscilação menor que $1/m$, e o mesmo vale numa vizinhança menor de cada ponto próximo de x . Logo o conjunto de descontinuidades é $D = \bigcup_{m \geq 1} D_m$.

Suponha primeiro que f seja Riemann integrável. Dado $\eta > 0$, existe uma partição P para a qual a diferença entre as somas superior e inferior é menor que η/m . Salvo pelos finitos pontos da partição, D_m está contido na união dos subintervalos J de P nos quais a oscilação de f é pelo menos $1/m$. Portanto,

$$\lambda(D_m) \leq \sum_{\text{osc}_J(f) \geq 1/m} |J| \leq m \sum_{J \in P} \text{osc}_J(f) |J| < \eta.$$

Assim, $\lambda(D_m) = 0$ para todo m e, por união enumerável, $\lambda(D) = 0$.

Reciprocamente, suponha que $\lambda(D) = 0$. Seja $|f| \leq M$ e fixe $\eta > 0$. O caso $a = b$ é imediato; suponha $a < b$. Pela regularidade exterior da medida de Lebesgue, existe um aberto $G \supset D$ tal que

$$\lambda(G) < \frac{\eta}{4(M+1)}.$$

O compacto $K = [a, b] \setminus G$ é formado apenas por pontos de continuidade. Para cada $x \in K$, escolha $r_x > 0$ de modo que $|f(y) - f(x)| < \eta/(8(b-a))$ quando $|y-x| < 2r_x$. Extraia uma subcobertura finita dos intervalos de centro x e raio r_x , e escolha $\delta > 0$ menor que todos esses raios. Assim, todo intervalo de comprimento menor que δ que encontra K tem oscilação menor que $\eta/(4(b-a))$. Escolha uma partição de malha menor que δ . Nos subintervalos que encontram K , a contribuição à diferença entre as somas de Darboux é menor que $\eta/4$. Os demais estão contidos em G e contribuem, ao todo, menos que $2M\lambda(G) < \eta/2$. Logo a diferença entre as somas superior e inferior é menor que η . Pelo critério de Darboux, f é Riemann integrável. □

Teorema 5.11 —

Dada uma função limitada $f : [a,b] \rightarrow \mathbb{R}$, se f é Riemann integrável, então f é Lebesgue integrável, e ambas as integrais coincidem.

Demonstração. Denote por D o conjunto de descontinuidades de f . Pelo critério anterior, $\lambda(D) = 0$. A restrição de f a $[a,b] \setminus D$ é contínua e, portanto, mensurável na σ -álgebra traço. Como a medida de Lebesgue é completa, qualquer subconjunto de D é mensurável. Segue que f é Lebesgue mensurável. Por ser limitada e estar definida num intervalo de medida finita, ela é Lebesgue integrável pelo Teorema 5.9.

Se I é sua integral de Riemann, para cada n existem funções escada φ_n e ψ_n tais que

$$\varphi_n \leq f \leq \psi_n, \quad \int_a^b \varphi_n(x) \, dx \longrightarrow I, \quad \int_a^b \psi_n(x) \, dx \longrightarrow I.$$

Para funções escada, as integrais de Riemann e de Lebesgue coincidem. Pela monotonicidade da integral de Lebesgue,

$$\int_a^b \varphi_n \, d\lambda \leq \int_a^b f \, d\lambda \leq \int_a^b \psi_n \, d\lambda.$$

Fazendo $n \rightarrow \infty$ e usando o teorema do confronto, obtemos $\int_a^b f \, d\lambda = I$. □

Exemplo 5.12 — .

Considere em $[0,1]$ a função indicadora dos racionais $f = \mathbb{1}_{\mathbb{Q} \cap [0,1]}$. Como $\mathbb{Q}$ é enumerável, $\lambda(\mathbb{Q} \cap [0,1]) = 0$. Logo $\int f \, d\lambda = 0$ e f é Lebesgue integrável.

Mas f não é Riemann integrável, pois o supremo em cada subintervalo de qualquer partição é 1 e o ínfimo é 0.

Lema 5.13 — Regularidade da medida de Lebesgue completada

Se $E \subset \mathbb{R}$ é Lebesgue mensurável e $\lambda(E) < \infty$, então, para todo $\varepsilon > 0$, existe um compacto $K \subset E$ tal que $\lambda(E \setminus K) < \varepsilon$.

Demonstração. Pela definição da completação, podemos escrever $E = B \cup N$, onde B é boreliano, $B \subset E$, e N está contido em um boreliano de medida zero. Logo $\lambda(B) = \lambda(E)$. Se $\lambda(B) = 0$, basta tomar $K = \emptyset$. Caso contrário, a medida

$$\nu(C) = \frac{\lambda(C \cap B)}{\lambda(B)}, \quad C \in \mathcal{B}(\mathbb{R}),$$

é uma probabilidade boreliana em $\mathbb{R}$. Pela regularidade das probabilidades borelianas em espaços poloneses, demonstrada no Capítulo 2, existe compacto $K \subset B \subset E$ com $\nu(B \setminus K) < \varepsilon/\lambda(B)$, isto é, $\lambda(E \setminus K) < \varepsilon$. □

Teorema 5.14 — Teorema de Lusin

Sejam $A \subset \mathbb{R}$ um conjunto Lebesgue mensurável, com $\lambda(A) < \infty$, e $f: A \rightarrow \mathbb{R}$ uma função Lebesgue mensurável. Então, para todo $\varepsilon > 0$, existe um conjunto compacto $K \subset A$ com $\lambda(A \setminus K) < \varepsilon$, tal que a restrição de f a K é contínua.

Demonstração. Seja $\{V_n\}_{n \in \mathbb{N}}$ uma enumeração dos intervalos abertos com extremos racionais. Para cada n , os conjuntos mensuráveis $f^{-1}(V_n)$ e $A \setminus f^{-1}(V_n)$ têm medida finita. Pelo lema anterior, existem compactos

$$K_n \subset f^{-1}(V_n) \quad \text{e} \quad K'_n \subset A \setminus f^{-1}(V_n)$$

tais que

$$\lambda(f^{-1}(V_n) \setminus K_n) < \frac{\varepsilon}{2^{n+2}}, \quad \lambda((A \setminus f^{-1}(V_n)) \setminus K'_n) < \frac{\varepsilon}{2^{n+2}}.$$

Seja

$$K = \bigcap_{n=1}^{\infty} (K_n \cup K'_n).$$

Cada $K_n \cup K'_n$ é compacto; logo K , sendo um fechado contido no primeiro deles, também é compacto. Além disso,

$$\lambda(A \setminus K) \leq \sum_{n=1}^{\infty} \lambda(A \setminus (K_n \cup K'_n)) < \varepsilon.$$

Resta provar a continuidade. Fixe $x \in K$ e um aberto V que contenha $f(x)$. Escolha V_n tal que $f(x) \in V_n \subset V$. Como $x \in f^{-1}(V_n)$, temos $x \notin K'_n$ e, pois $x \in K \subset K_n \cup K'_n$, necessariamente $x \in K_n$. O conjunto $U = K \setminus K'_n$ é uma vizinhança aberta de x em K . Se $y \in U$, então $y \in K_n$, novamente porque $K \subset K_n \cup K'_n$; portanto, $f(y) \in V_n \subset V$. Assim $f|_K$ é contínua. $\square$

5.4. Integração e Limites

Uma das razões para introduzir a integral de Lebesgue é sua estabilidade sob passagens ao limite. Os três teoremas desta seção distinguem as hipóteses que permitem trocar limite e integral. A ordem importa: cada resultado prepara o seguinte.

Teorema 5.15 — Teorema da Convergência Monótona

Se $f_n : \Omega \rightarrow [0, \infty]$ são mensuráveis e $f_n \uparrow f$, então $\int f_n \, d\mu \rightarrow \int f \, d\mu$.

Demonstração. Como $\int f_n \, d\mu \leq \int f_{n+1} \, d\mu$, existe $\alpha \in [0, \infty]$ tal que

$$\lim_{n \rightarrow \infty} \int f_n \, d\mu = \alpha$$

Como $f_n \leq f$ para todo n , temos que $\int f_n \, d\mu \leq \int f \, d\mu$ para todo n . Logo $\alpha \leq \int f \, d\mu$. Seja s uma função simples mensurável tal que $0 \leq s \leq f$ e seja c uma constante tal que $0 < c < 1$ e defina

$$E_n = \{x : f_n(x) \geq cs(x)\}.$$

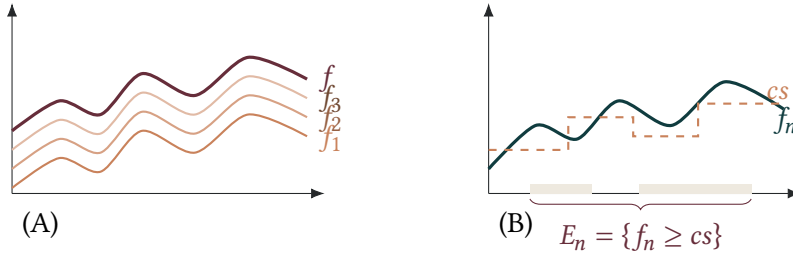


Figura 5.3.: À esquerda, uma sequência crescente aproximando f ; à direita, os conjuntos usados na demonstração do Teorema da Convergência Monótona. Adaptado de [26].

Cada conjunto E_n é mensurável, $E_1 \subset E_2 \subset \dots$ e $\Omega = \bigcup_n E_n$. Para ver a última igualdade, tome $x \in \Omega$. Se $s(x) = 0$, então $x \in E_1$. Se $s(x) > 0$, então $cs(x) < s(x) \leq f(x)$ e, como $f_n(x) \uparrow f(x)$, temos $x \in E_n$ para algum n . Além disso,

$$\int_{\Omega} f_n \, d\mu \geq \int_{E_n} f_n \, d\mu \geq c \int_{E_n} s \, d\mu$$

para todo n . Como s assume apenas um número finito de valores, a continuidade por baixo da medida em cada conjunto de nível de s mostra que $\int_{E_n} s \, d\mu \uparrow \int s \, d\mu$. Fazendo $n \rightarrow \infty$, obtemos $\alpha \geq c \int_{\Omega} s \, d\mu$. Como isso vale para todo $c < 1$, temos

$$\alpha \geq \int s \, d\mu$$

para toda função simples $s \leq f$. Logo, temos

$$\alpha \geq \int_{\Omega} f \, d\mu$$

Isso prova a desigualdade inversa. □

Proposição 5.16 —

Seja $f : \Omega \rightarrow [0, \infty]$ uma aplicação mensurável. Então:

1 $f = 0$ quase certamente se e somente se $\int_{\Omega} f \, d\mu = 0$.

2 Se $\int_{\Omega} f \, d\mu < \infty$, então $f < \infty$ quase certamente.

Demonstração. As duas afirmações já foram provadas, respectivamente, nos itens **2** e **1** do Teorema ??; note que aquele argumento vale para toda medida, sem hipótese de finitude. $\square$

Lema 5.17 — Lema de Fatou

Se $f_n : \Omega \rightarrow [0, \infty]$ são mensuráveis, então

$$\int \liminf_n f_n \, d\mu \leq \liminf_n \int f_n \, d\mu.$$

Demonstração. Defina $g_k(x) = \inf_{i \geq k} f_i(x)$. Então $g_k \leq f_i$ para todo $i \geq k$ e, portanto,

$$\int g_k \, d\mu \leq \inf_{i \geq k} \int f_i \, d\mu.$$

Além disso, $0 \leq g_1 \leq g_2 \leq \dots$, cada g_k é mensurável e $g_k \uparrow \liminf_n f_n$. Pelo Teorema da Convergência Monótona,

$$\int \liminf_n f_n \, d\mu = \lim_{k \rightarrow \infty} \int g_k \, d\mu \leq \lim_{k \rightarrow \infty} \inf_{i \geq k} \int f_i \, d\mu = \liminf_{n \rightarrow \infty} \int f_n \, d\mu.$$

$\square$

Definição 5.18 – Convergência quase todo ponto

Dizemos que $f_n \rightarrow f$ em quase todo ponto ou quase certamente (q.c.) se $\mu(\{\omega : f_n(\omega) \nrightarrow f(\omega)\}) = 0$.

Teorema 5.19 – Teorema da Convergência Dominada

Se f_n e f são mensuráveis, $f_n \rightarrow f$ q.c., $|f_n| \leq g$ q.c. para todo n , e $g \geq 0$ é integrável, então

$$\int f_n \, d\mu \rightarrow \int f \, d\mu.$$

Demonstração. Alterando as funções num único conjunto nulo, podemos supor que a convergência e todas as dominações valham em todo ponto. Assim $|f| \leq g$, de modo que $f \in L^1(\mu)$, e $|f_n - f| \leq 2g$. Podemos aplicar o Lema de Fatou para a função $2g - |f_n - f|$ e obter que

$$\begin{aligned} \int_{\Omega} 2g \, d\mu &\leq \liminf_{n \rightarrow \infty} \int_{\Omega} (2g - |f_n - f|) \, d\mu \\ &= \int_{\Omega} 2g \, d\mu + \liminf_{n \rightarrow \infty} \left(- \int_{\Omega} |f_n - f| \, d\mu \right) \\ &= \int_{\Omega} 2g \, d\mu - \limsup_{n \rightarrow \infty} \int_{\Omega} |f_n - f| \, d\mu \end{aligned}$$

Como $\int_{\Omega} 2g \, d\mu$ é finita podemos subtraí-la

$$\limsup_{n \rightarrow \infty} \int_{\Omega} |f_n - f| \, d\mu \leq 0$$

E assim

$$\lim_{n \rightarrow \infty} \int_{\Omega} |f_n - f| \, d\mu = 0$$

Finalmente,

$$\left| \int f_n \, d\mu - \int f \, d\mu \right| \leq \int |f_n - f| \, d\mu \rightarrow 0,$$

e, portanto,

$$\lim_{n \rightarrow \infty} \int_{\Omega} f_n \, d\mu = \int_{\Omega} f \, d\mu$$



5.5. Espaços $\mathcal{L}^p$

Estudaremos agora os espaços de funções integráveis e as normas que permitem medir a qualidade das aproximações.

Definição 5.20 —

- Dado um espaço de medida $(\Omega, \mathcal{F}, \mu)$ e $1 \leq p < \infty$, definimos $\mathcal{L}^p(\Omega, \mathcal{F}, \mu)$ como o espaço das classes de equivalência das funções mensuráveis tais que

$$\int |f|^p \, d(\mu) < \infty$$

sob a relação de equivalência de igualdade quase certa.

- Para qualquer elemento $f \in \mathcal{L}^p(\Omega, \mathcal{F}, \mu)$ definimos a p -norma como

$$\|f\|_p = \left(\int |f|^p \, d(\mu) \right)^{1/p}.$$

- Definimos $\mathcal{L}^\infty(\Omega, \mathcal{F}, \mu)$ como o espaço das classes de equivalência das funções mensuráveis essencialmente limitadas. Sua norma é

$$\|f\|_\infty := \operatorname{ess\,sup}_{\omega \in \Omega} |f(\omega)| = \inf\{M \geq 0 : |f| \leq M \text{ quase certamente}\}.$$

Mostraremos que $\mathcal{L}^p(\Omega, \mathcal{F}, \mu)$ é um espaço de Banach: sua norma satisfaz a desigualdade triangular e toda sequência de Cauchy converge. Começamos pela desigualdade triangular.

Lema 5.21 — Desigualdade de Hölder

Se $1 < p < \infty$, $q = p/(p-1)$, $f \in L^p(\mu)$ e $h \in L^q(\mu)$, então $fh \in L^1(\mu)$ e

$$\int |fh| \, d\mu \leq \|f\|_p \|h\|_q.$$

Demonstração. A desigualdade de Young

$$uv \leq \frac{u^p}{p} + \frac{v^q}{q}, \quad u, v \geq 0,$$

segue da convexidade de $t \mapsto t^p/p$ ou, equivalentemente, do cálculo elementar do mínimo do lado direito menos uv . Se uma das normas é zero, a conclusão segue do Teorema ???. Caso contrário, aplique Young a

$$u = \frac{|f|}{\|f\|_p}, \quad v = \frac{|h|}{\|h\|_q}$$

e integre. Como $1/p + 1/q = 1$, obtemos

$$\frac{\int |fh| \, d\mu}{\|f\|_p \|h\|_q} \leq \frac{1}{p} \int \frac{|f|^p}{\|f\|_p^p} \, d\mu + \frac{1}{q} \int \frac{|h|^q}{\|h\|_q^q} \, d\mu = \frac{1}{p} + \frac{1}{q} = 1.$$

□

Proposição 5.22 — Desigualdade de Minkowski

Dadas $f, g \in \mathcal{L}^p(\Omega, \mathcal{F}, \mu)$ então

- $f + g \in \mathcal{L}^p(\Omega, \mathcal{F}, \mu)$
- $\|f + g\|_p \leq \|f\|_p + \|g\|_p.$

Demonstração. Como $|f+g| \leq |f|+|g|$, basta provar a desigualdade para funções não negativas. De fato,

$$\|f+g\|_p \leq \| |f|+|g| \|_p.$$

O caso $p = 1$ segue imediatamente da linearidade da integral.

Para $1 < p < \infty$, primeiro demonstraremos que $f+g \in \mathcal{L}^p(\Omega, \mathcal{F}, \mu)$:

$$(f+g)^p \leq 2^{p-1}(f^p + g^p).$$

e, portanto, $\|f+g\|_p^p \leq 2^{p-1}(\|f\|_p^p + \|g\|_p^p) < \infty$

Para provar a desigualdade triangular, note que podemos supor que $\|f+g\|_p > 0$ caso contrário, a desigualdade triangular segue da positividade da norma. Escrevemos

$$\|f+g\|_p^p = \int (f+g)^p d(\mu) = \int f(f+g)^{p-1} d(\mu) + \int g(f+g)^{p-1} d(\mu)$$

Agora podemos aplicar a desigualdade de Hölder, provada no Lema 5.21, a cada termo do lado direito e usar o fato de que $\frac{1}{p} + \frac{1}{q} = 1$ é equivalente a $p = (p-1)q$ para ver que

$$\int f(f+g)^{p-1} d(\mu) \leq \left(\int f^p d(\mu) \right)^{\frac{1}{p}} \left(\int (f+g)^{(p-1)q} d(\mu) \right)^{\frac{1}{q}} = \|f\|_p \|f+g\|_p^{p/q}$$

Aplicando este argumento ao termo $\int g(f+g)^{p-1} d(\mu)$ temos

$$\|f+g\|_p^p \leq (\|f\|_p + \|g\|_p) \cdot \|f+g\|_p^{p/q}$$

dividindo por $\|f+g\|_p^{p/q}$ e usando que $p - \frac{p}{q} = 1$ podemos concluir que $\|f+g\|_p \leq \|f\|_p + \|g\|_p$. □

As demais propriedades de norma são imediatas: a homogeneidade vem da linearidade da integral, e $\|f\|_p = 0$ implica $f = 0$ quase certamente pelo Teorema ???. Juntamente com Minkowski, isso mostra que L^p é um espaço vetorial normado.

Proposição 5.23 – Completude de $\mathcal{L}^p$

Para $1 \leq p < \infty$, o espaço vetorial normado $\mathcal{L}^p(\Omega, \mathcal{F}, \mu)$ é completo.

Demonstração. Seja (f_n) uma sequência de Cauchy em $\mathcal{L}^p(\Omega, \mathcal{F}, \mu)$. Escolha recursivamente $1 \leq n_1 < n_2 < \dots$ de modo que, para cada $j \geq 1$,

$$\|f_m - f_n\|_p \leq 2^{-j} \quad \text{sempre que } m, n \geq n_j.$$

Em particular, $\|f_{n_{j+1}} - f_{n_j}\|_p \leq 2^{-j}$. Para

$$G_N = \sum_{j=1}^N |f_{n_{j+1}} - f_{n_j}|,$$

a desigualdade de Minkowski fornece

$$\|G_N\|_p \leq \sum_{j=1}^N \|f_{n_{j+1}} - f_{n_j}\|_p \leq \sum_{j=1}^N 2^{-j} \leq 1.$$

Como $G_N \uparrow G := \sum_{j=1}^{\infty} |f_{n_{j+1}} - f_{n_j}|$, o Teorema da Convergência Monótona implica $\|G\|_p \leq 1$. Em particular, $G < \infty$ quase certamente. Fora de um conjunto nulo, a série dos incrementos converge absolutamente; logo (f_{n_j}) é uma sequência de Cauchy em $\mathbb{R}$. Defina f como seu limite nesse conjunto e como zero no complementar. A função f é mensurável.

Resta voltar da subsequência à sequência original. Dado $\varepsilon > 0$, escolha N tal que $\|f_m - f_n\|_p < \varepsilon$ sempre que $m, n \geq N$. Para $n \geq N$, o lema de Fatou e a convergência $f_{n_j} \rightarrow f$ quase certamente dão

$$\|f - f_n\|_p^p \leq \liminf_{j \rightarrow \infty} \|f_{n_j} - f_n\|_p^p \leq \varepsilon^p.$$

Assim $f - f_n \in \mathcal{L}^p$, donde $f \in \mathcal{L}^p$, e a mesma estimativa mostra que $\|f_n - f\|_p \rightarrow 0$. Portanto $\mathcal{L}^p$ é completo. $\square$

Sabemos que funções mensuráveis podem ser aproximadas pontualmente por funções simples. Para os espaços $\mathcal{L}^p$, essa aproximação pode ser feita também na norma.

Lema 5.24 —

Para $1 \leq p < \infty$, as funções simples integráveis são densas em $\mathcal{L}^p(\Omega, \mathcal{F}, \mu)$.

Demonstração. Seja primeiro $f \geq 0$ em $\mathcal{L}^p$. Pelo teorema de aproximação por funções simples, existem funções simples $0 \leq s_n \uparrow f$. Como $0 \leq f - s_n \leq f$,

$$|f - s_n|^p \leq |f|^p,$$

e o lado direito é integrável. Pelo Teorema da Convergência Dominada, $\|f - s_n\|_p \rightarrow 0$.

Para f real arbitrária, aplique o argumento a f^+ e f^- : escolha funções simples g_n e h_n tais que $\|f^+ - g_n\|_p \rightarrow 0$ e $\|f^- - h_n\|_p \rightarrow 0$. Então $g_n - h_n$ é simples e, pela desigualdade de Minkowski,

$$\|f - (g_n - h_n)\|_p \leq \|f^+ - g_n\|_p + \|f^- - h_n\|_p \rightarrow 0.$$

□

A construção a partir de funções simples forneceu teoremas de passagem ao limite e espaços L^p completos. Essa combinação explica o papel da integral de Lebesgue na probabilidade: os limites não entram como um remendo posterior; participam da construção desde o início.

EXERCÍCIOS

Ex. 5.1 — Seja $C \subseteq [0,1]$ o conjunto de Cantor e seja λ a medida de Lebesgue. Calcule

$$\int_C x^2 \, d\lambda(x).$$

Indique qual propriedade de C torna o cálculo imediato.

Ex. 5.2 — Seja $u \in L^1(\mathbb{R})$ e defina

$$U(t) = \int_{\mathbb{R}} u(x) \cos(xt) \, dx, \quad t \in \mathbb{R}.$$

Use o Teorema da Convergência Dominada para provar que U é contínua em todo ponto de $\mathbb{R}$.

Ex. 5.3 — Seja $(X, \mathcal{F}, \mu)$ um espaço de medida e sejam $A_1, A_2, \dots \in \mathcal{F}$. Prove que

$$\mu\left(\bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} A_n\right) \leq \liminf_{n \rightarrow \infty} \mu(A_n).$$

Reescreva o conjunto do lado esquerdo como $\liminf A_n$ e identifique o teorema de convergência usado na prova.

Ex. 5.4 — Seja $f : X \rightarrow [0, \infty]$ mensurável. Para $n \geq 1$, defina

$$f_n = (2^{-n} \lfloor 2^n f \rfloor) \wedge n,$$

adotando a convenção $f_n(x) = n$ quando $f(x) = \infty$. Prove que cada f_n é simples e mensurável e que

$$0 \leq f_n \uparrow f \quad \text{pontualmente.}$$

Explique por que essa construção fornece a aproximação por funções simples usada na definição da integral de Lebesgue.

Distribuições

6.1. Distribuição de uma Variável Aleatória

Em muitos problemas, o espaço amostral serve de bastidor. O que observamos é o valor de uma variável aleatória e as probabilidades com que ela visita diferentes regiões da reta. Num modelo epidêmico, por exemplo, interessam o número de infectados e o número de recuperados; os detalhes de Ω podem permanecer invisíveis.

A **distribuição** registra exatamente essa informação e permite, quando possível, deixar o espaço de probabilidade subjacente em segundo plano. Seja ξ uma variável aleatória definida em $(\Omega, \mathcal{F}, \mathbf{P})$. Então

1 $\mathbf{P}(\xi \in \mathbb{R}) = \mathbf{P}(\Omega) = 1;$

2 se $A_1, A_2, \dots \in \mathcal{B}(\mathbb{R})$ são disjuntos então $\xi^{-1}(A_1), \xi^{-1}(A_2), \dots \in \mathcal{F}$ também são disjuntos, pois as pré-imagens preservam interseções, e

$$\mathbf{P}\left(\xi^{-1}\left(\bigcup_{n=1}^{\infty} A_n\right)\right) = \mathbf{P}\left(\bigcup_{n=1}^{\infty} \xi^{-1}(A_n)\right) = \sum_{n=1}^{\infty} \mathbf{P}(\xi^{-1}(A_n)).$$

Essas propriedades mostram que $A \mapsto \mathbf{P}(\xi \in A)$ é uma probabilidade na reta.

Definição 6.1 — Distribuição de uma Variável Aleatória

Se ξ for uma variável aleatória definida em um espaço de probabilidade $(\Omega, \mathcal{F}, \mathbf{P})$, então ξ induz uma medida de probabilidade $\mathbf{P}_\xi$ em $\mathbb{R}$ definida por

$$\mathbf{P}_\xi(A) = \mathbf{P}(\xi^{-1}(A)) = \mathbf{P}(\xi \in A), \quad (6.1)$$

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

Esta medida de probabilidade $\mathbf{P}_\xi$ é denominada **distribuição** de ξ .

Na Equação (6.1), toda a estrutura específica de Ω desaparece. Permanece uma probabilidade na reta, suficiente para responder a qualquer pergunta que dependa apenas do valor de ξ .

De maneira análoga podemos definir a distribuição de um vetor aleatório

Definição 6.2 — Distribuição de Vetores Aleatórios

Se ξ for um vetor aleatório, então ξ induz uma medida de probabilidade $\mathbf{P}_\xi$ em $\mathbb{R}^n$ definida tomando a pré-imagem de cada conjunto de Borel $A \subset \mathbb{R}^n$:

$$\mathbf{P}_\xi(A) = \mathbf{P}(\xi^{-1}(A)) = \mathbf{P}(\xi \in A). \quad (6.2)$$

Esta medida de probabilidade $\mathbf{P}_\xi$ é denominada **distribuição** do vetor ξ .

Se $\xi = (\xi_1, \dots, \xi_n)$ é um vetor aleatório, então as distribuições de cada uma das variáveis aleatórias ξ_i , para $i = 1, \dots, n$, são chamadas de **distribuições marginais** de ξ .

A distribuição não determina a variável aleatória ponto a ponto. Duas variáveis podem estar relacionadas de maneiras muito diferentes e ainda ter a mesma lei.

Exemplo 6.3 — .

Considere o experimento de lançar uma moeda não viciada, e tome como espaço amostral o conjunto $\Omega = \{H, T\}$, onde H representa cara e T representa coroa. Tomamos $\mathbf{P}$ como a medida de probabilidade equiprovável em Ω .

Podemos agora definir a variável aleatória $\xi(H) = 1, \xi(T) = 0$.

Note que $\mathbf{P}(\xi = 1) = \mathbf{P}(\xi = 0) = \frac{1}{2}$, e para $A \in \mathcal{B}(\mathbb{R})$ temos

$$\mathbf{P}(\xi \in A) = \frac{1}{2}\mathbb{1}_A(0) + \frac{1}{2}\mathbb{1}_A(1).$$

Em particular, essa probabilidade é $1/2$ se exatamente um dos pontos $0, 1$ pertence a A , é 1 se ambos pertencem a A e é 0 se nenhum deles pertence a A .

Por outro lado poderíamos definir a variável aleatória $\eta(H) = 0, \eta(T) = 1$.

As variáveis nunca coincidem (de fato $\xi = 1 - \eta$), mas ξ e η possuem a mesma distribuição.



O exemplo distingue duas noções de igualdade que usaremos repetidamente.

Definição 6.4 — Equivalências de Variáveis Aleatórias

- 1** Dizemos que as variáveis aleatórias ξ e η são iguais **quase certamente**, fato que denotaremos por

$$\xi \stackrel{\text{q.c.}}{=} \eta,$$

se $\mathbf{P}(\xi \neq \eta) = 0$.

- 2** Dizemos que as variáveis aleatórias ξ e η são iguais em **distribuição**, fato que denotaremos por

$$\xi \stackrel{\text{d}}{=} \eta,$$

se eles tiverem a mesma distribuição. Equivalentemente,

$$\mathbf{P}(\xi \in A) = \mathbf{P}(\eta \in A) \text{ para todo } A \text{ de Borel.}$$

Observe que se $\xi \stackrel{\text{q.c.}}{=} \eta$, então $\xi \stackrel{\text{d}}{=} \eta$, mas a recíproca não é verdadeira. De fato, se $\xi \stackrel{\text{q.c.}}{=} \eta$, então

$$\mathbf{P}(\xi \in A) = \mathbf{P}(\xi \in A; \xi = \eta) = \mathbf{P}(\eta \in A; \xi = \eta) = \mathbf{P}(\eta \in A).$$

Por outro lado, podemos ter $\xi \stackrel{\text{d}}{=} \eta$ sem que ξ e η estejam definidas em um mesmo espaço de probabilidade.

Exemplo 6.5 — .

Seja agora $\mathbf{P}$ a medida de Lebesgue em $\Omega = [0,1]$, e defina $\xi(\omega) = \omega$ e $\eta(\omega) = 1 - \omega$. Note que para $x \in \mathbb{R}$

$$\{\xi \leq x\} = \{\omega \in [0,1] : \omega \leq x\} = \begin{cases} \emptyset & ; \text{ se } x < 0; \\ [0, x] & ; \text{ se } 0 \leq x \leq 1; , \\ [0,1] & ; \text{ se } x \geq 1 \end{cases}$$

e

$$\{\eta \leq x\} = \{\omega \in [0,1] : 1 - \omega \leq x\} = \begin{cases} \emptyset & ; \text{ se } x < 0; \\ [1 - x, 1] & ; \text{ se } 0 \leq x \leq 1; . \\ [0,1] & ; \text{ se } x \geq 1 \end{cases}$$

Segue que para $x \in [0,1]$

$$\mathbf{P}_\xi((-\infty, x]) = \mathbf{P}(\xi \leq x) = x = \mathbf{P}(\eta \leq x) = \mathbf{P}_\eta((-\infty, x]),$$

e portanto

$$\mathbf{P}_\xi((-\infty, x]) = \mathbf{P}_\eta((-\infty, x]),$$

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

Agora, como $\{(-\infty, x] : x \in \mathbb{R}\}$ é um π -sistema que gera $\mathcal{B}(\mathbb{R})$, então $\mathbf{P}_\xi = \mathbf{P}_\eta$.

◀

No exemplo anterior comparamos a distribuição de duas variáveis aleatórias olhando apenas para as probabilidades em intervalos do tipo $(-\infty, x]$, $x \in \mathbb{R}$. Para tanto argumentamos que

1 a classe $\mathcal{P} = \{(-\infty, x] : x \in \mathbb{R}\}$ é π -sistema;

2 $\mathcal{B}(\mathbb{R})$ é gerada por tais intervalos.

Com isso pudemos concluir que duas medidas de probabilidade em $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ que coincidam em $\mathcal{P}$ são de fato a mesma medida.

Isso demonstra o seguinte resultado.

Teorema 6.6 —

Sejam ξ uma variável aleatória definida em $(\Omega_1, \mathcal{F}_1, \mathbf{P}_1)$ e η uma variável aleatória definida em $(\Omega_2, \mathcal{F}_2, \mathbf{P}_2)$. Então $\xi \stackrel{d}{=} \eta$ se, e somente se,

$$\mathbf{P}_1(\xi \leq x) = \mathbf{P}_2(\eta \leq x) \quad \text{para todo } x \in \mathbb{R}.$$

De maneira equivalente, a distribuição $\mathbf{P}_\xi$ de ξ está completamente determinada pelos valores de $\mathbf{P}_1(\xi \leq x)$, $x \in \mathbb{R}$.

Demonstração. Se ξ e η têm a mesma distribuição, a igualdade nos intervalos $(-\infty, x]$ é imediata. Reciprocamente, suponha que essas igualdades valham para todo $x \in \mathbb{R}$. As probabilidades induzidas $\mathbf{P}_\xi$ e $\mathbf{P}_\eta$ coincidem no π -sistema

$$\mathcal{P} = \{(-\infty, x] : x \in \mathbb{R}\},$$

que gera $\mathcal{B}(\mathbb{R})$. Pelo Teorema de Unicidade de Medidas (Teorema 2.14), elas coincidem em $\mathcal{B}(\mathbb{R})$. Portanto, $\xi \stackrel{d}{=} \eta$. □

Esse resultado permite registrar uma distribuição por meio da função que introduziremos a seguir.

6.2. A Função Distribuição

Definição 6.7 — Função de distribuição

A **função de distribuição** $F : \mathbb{R} \rightarrow [0,1]$ de uma variável aleatória ξ é definida como

$$F(x) := \mathbf{P}(\xi \leq x), \quad x \in \mathbb{R}.$$

O teorema 6.6 nos garante que a função de distribuição determina unicamente a distribuição de ξ .

A função de distribuição comprime a lei de ξ numa única função real. O espaço Ω desaparece da notação, mas não a informação probabilística: F determina completamente a distribuição.

Exemplo 6.8 — .

Seja ξ = número de H em dois lançamentos independentes de uma moeda justa. Então temos

$$\xi = \begin{cases} 0, & \text{com probabilidade } \frac{1}{4} \\ 1, & \text{com probabilidade } \frac{1}{2} \\ 2, & \text{com probabilidade } \frac{1}{4} \end{cases}$$

Então, podemos criar a função de distribuição. Esta função apresentará descontinuidades do tipo salto nos pontos 0, 1 e 2.

$$F(x) = \mathbf{P}(\xi \leq x) = \begin{cases} 0, & x < 0 \\ \frac{1}{4}, & x \in [0, 1) \\ \frac{1}{4} + \frac{1}{2}, & x \in [1, 2) \\ \frac{1}{4} + \frac{1}{2} + \frac{1}{4}, & x \in [2, +\infty) \end{cases}$$

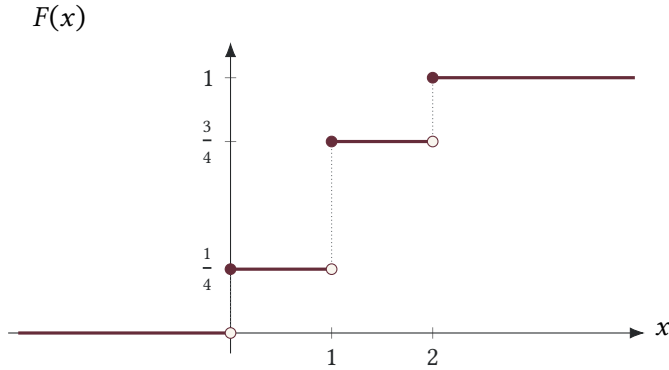


Figura 6.1.: Distribuição do Exemplo 6.8.

Exemplo 6.9 – .

Considere o experimento de sortear ao acaso um ponto na bola unitária $B = \{(a, b) : a^2 + b^2 \leq 1\}$ em $\mathbb{R}^2$, e defina ξ como a distância ao centro da bola.

Para modelar este problema podemos considerar $\mathbf{P}(\cdot) = \mu(\cdot)/\pi$, onde μ é a medida de Lebesgue em B . Ou seja, a probabilidade de sortearmos um ponto em uma região $A \in \mathcal{B}(B)$ é proporcional à área de A . Mais especificamente

$$\mathbf{P}(A) = \frac{\mu(A)}{\pi}.$$

Agora, note que $\xi(a, b) = \sqrt{a^2 + b^2}$ e portanto para $x \in [0, 1]$

$$\mathbf{P}(\xi \leq x) = \mathbf{P}(\{(a, b) : a^2 + b^2 \leq x^2\}) = \frac{\pi x^2}{\pi} = x^2,$$

e portanto

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

Teorema 6.10 — Propriedades da Função de Distribuição

A função de distribuição $F(x)$ de uma variável aleatória ξ possui as seguintes propriedades:

1 F é não decrescente e $0 \leq F(x) \leq 1$.

2 $F(x) \rightarrow 0$ quando $x \rightarrow -\infty$ e $F(x) \rightarrow 1$ quando $x \rightarrow +\infty$.

3 F é contínua à direita, ou seja

$$\lim_{y \rightarrow x_+} F(y) = F(x).$$

4 F possui limites à esquerda em todos os pontos. Além disso,

$$F(x^-) := \lim_{y \rightarrow x^-} F(y) = \mathbf{P}(\xi < x).$$

Em particular,

$$\mathbf{P}(\xi = x) = F(x) - F(x^-).$$

5 F possui no máximo número enumerável de descontinuidades.

Uma função F que é contínua à direita e possui limites à esquerda em todos os pontos é dita **càdlàg**, do francês *continue à droite, limites à gauche*.

Demonstração.

1 Se $x < y$, então $\{\xi \leq x\} \subseteq \{\xi \leq y\}$, de onde $F(x) \leq F(y)$. As cotas $0 \leq F \leq 1$ decorrem dos axiomas da probabilidade.

2 Como $\{\xi \leq -n\} \downarrow \emptyset$ e $\{\xi \leq n\} \uparrow \Omega$, a continuidade da probabilidade por cima e por baixo dá

$$F(-n) \rightarrow 0, \quad F(n) \rightarrow 1.$$

A monotonicidade estende esses limites de sequências inteiras para $x \rightarrow -\infty$ e $x \rightarrow +\infty$.

3 Se $x_n \downarrow x$, então $\{\xi \leq x_n\} \downarrow \{\xi \leq x\}$. Pela continuidade por cima da probabilidade,

$F(x_n) \rightarrow F(x)$. Para uma sequência arbitrária $y_n > x$ com $y_n \rightarrow x$, ponha $z_n = \sup_{k \geq n} y_k$. Então $z_n \downarrow x$ e, pela monotonicidade,

$$F(x) \leq F(y_n) \leq F(z_n).$$

Como $F(z_n) \rightarrow F(x)$, segue que $F(y_n) \rightarrow F(x)$. Logo F é contínua à direita.

4 Se $x_n \uparrow x$, com $x_n < x$, então $\{\xi \leq x_n\} \uparrow \{\xi < x\}$. Pela continuidade por baixo,

$$F(x^-) = \lim_{n \rightarrow \infty} F(x_n) = \mathbf{P}(\xi < x).$$

A monotonicidade mostra que o valor independe da sequência escolhida. Como $\{\xi \leq x\}$ é a união disjunta de $\{\xi < x\}$ e $\{\xi = x\}$,

$$\mathbf{P}(\xi = x) = F(x) - F(x^-).$$

5 Pela continuidade à direita, toda descontinuidade de F é um salto positivo. Para $m \geq 1$, ponha

$$D_m = \{x : F(x) - F(x^-) \geq 1/m\}.$$

Se $x_1 < \dots < x_r$ pertencem a D_m , então, pela monotonicidade,

$$\frac{r}{m} \leq \sum_{j=1}^r (F(x_j) - F(x_j^-)) \leq 1.$$

Portanto, D_m possui no máximo m pontos. O conjunto de descontinuidades é $\bigcup_{m \geq 1} D_m$ e, assim, é no máximo enumerável.

□

Proposição 6.11 —

Se $a < b$, então $\mathbf{P}(\xi \in (a, b]) = F(b) - F(a)$.

Demonstração. Como $\{\xi \leq b\}$ é a união disjunta de $\{\xi \leq a\}$ e $\{a < \xi \leq b\}$,

$$F(b) = F(a) + \mathbf{P}(a < \xi \leq b),$$

o que prova a identidade. □

Definição 6.12 —

Uma função $F : \mathbb{R} \rightarrow [0, 1]$ tal que

- 1 F é não-decrescente;
- 2 F é contínua pela direita;
- 3 $\lim_{x \rightarrow a^-} F(x)$ existe para todo $a \in \mathbb{R}$ (possui limite pela esquerda);
- 4 $\lim_{x \rightarrow +\infty} F(x) = 1$ e $\lim_{x \rightarrow -\infty} F(x) = 0$.

é chamada de candidata a função de distribuição de probabilidade.

As condições anteriores também são suficientes. A demonstração fará mais do que garantir existência: permitirá construir variáveis com distribuições prescritas num mesmo espaço de probabilidade. Essa escolha comum será útil para comparar sequências de variáveis.

Dada uma função F com as propriedades acima, construiremos uma variável aleatória X num espaço de probabilidade $(\Omega, \mathcal{F}, \mathbf{P})$ tal que $\mathbf{P}(X \leq x) = F(x)$, para todo $x \in \mathbb{R}$.

Uma maneira interessante seria definir a variável X como uma função de alguma variável aleatória cuja distribuição seja conhecida. Tomemos, por exemplo, uma variável $U \sim \text{Uniforme}(0,1)$, e tentemos escrever $X = F^*(U)$ para alguma função $F^* : (0,1) \rightarrow \mathbb{R}$.

Para escolher F^* , lembre que o que gostaríamos é que $\mathbf{P}(X \leq x) = \mathbf{P}(F^*(U) \leq x) = F(x)$.

Para isso note primeiro que, como $F(x) \in [0,1]$, então $\mathbf{P}(U \leq F(x)) = F(x)$. Ou

seja, queremos F^* tal que

$$\mathbf{P}(F^*(U) \leq x) = \mathbf{P}(U \leq F(x)).$$

Ou ainda, que $F^*(u) \leq x \Leftrightarrow u \leq F(x)$.

Se F fosse inversível, poderíamos fazer $F^* = F^{-1}$ e, como F é crescente teríamos que

$$F^*(u) \leq x \Leftrightarrow F(F^*(u)) \leq F(x) \Leftrightarrow u \leq F(x).$$

Em geral F não é inversível, mas as considerações acima nos mostram que a chave para conseguir o que queremos ainda está em, de algum modo, “*inverter*” a função distribuição $F : \mathbb{R} \rightarrow [0,1]$.

Como F é não decrescente, as dificuldades para uma inversão clássica se concentram em dois tipos de comportamento:

- 1 intervalos (a,b) onde a função é constante;
- 2 pontos de descontinuidade de F .

Não precisamos, porém, de uma inversa clássica. Basta uma função $F^* : (0,1) \rightarrow \mathbb{R}$ que preserve a equivalência entre as desigualdades: $F^*(\omega) \leq x$ se, e só se, $F(x) \geq \omega$.

Uma maneira de resolver esse problema é tentar definir F^* de modo que

- 1 se F é inversível em (a,b) então queremos que $F^*(\omega) = F^{-1}(\omega)$ em $F((a,b))$;
- 2 se $F(x) = \omega$ para todo $a < x < b$, então F^* é descontínua em ω ;
- 3 se $F(x)$ é descontínua em $x = x_0$ com $\omega_1 = \lim_{x \rightarrow x_0^-} F(x) < F(x_0) = \omega_2$, então $F^*(\omega) = x_0$ para $\omega \in (\omega_1, \omega_2)$.

Essas ideias são melhor visualizadas na Figura 6.2.

Existem diversas formas de definir uma função F^* com tais propriedades, e a seguir apresentaremos uma das mais comuns.

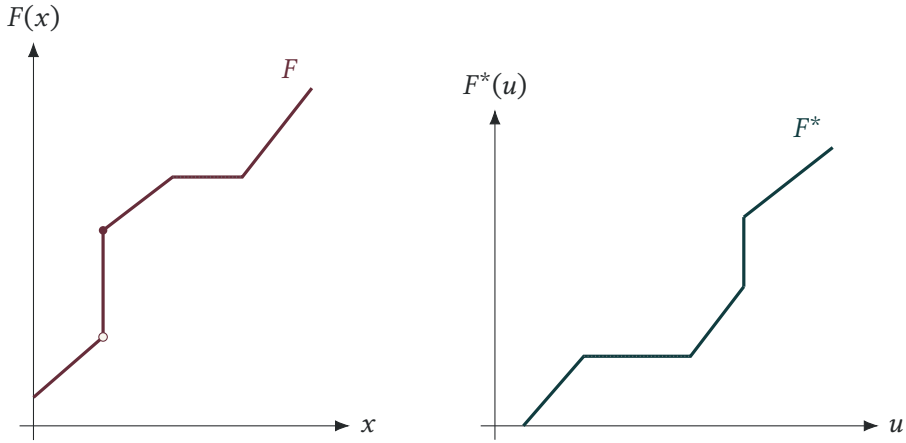


Figura 6.2.: Uma função de distribuição e sua inversa generalizada. Um salto de F produz um trecho constante em F^* , enquanto um trecho constante de F produz um salto em F^* .

Definição 6.13 —

Dada uma candidata a função de distribuição $F : \mathbb{R} \rightarrow [0,1]$, definimos a **inversa generalizada** de F como a função $F^* : (0,1) \rightarrow \mathbb{R}$ dada por

$$F^*(\omega) = \inf\{x \in \mathbb{R} : F(x) \geq \omega\}.$$

Apesar de não ser a inversa de F , F^* ainda herda algumas propriedades de F .

Proposição 6.14 — Propriedades da Inversa Generalizada

Se F é uma candidata a função de distribuição e $u \in (0,1)$, define

$$F^*(u) := \inf\{x \in \mathbb{R} : F(x) \geq u\}.$$

A inversa generalizada satisfaz:

1 $F^*(u) = \sup\{x \in \mathbb{R} : F(x) < u\};$

2 F^* é não decrescente;

3 se $0 < F(x) < 1$, então $F^*(F(x)) \leq x$;

4 $F(F^*(u)) \geq u$;

5 $F^*(u) \leq x$ se e somente se $u \leq F(x)$.

Demonstração. Para $u \in (0,1)$, o conjunto $A_u = \{x : F(x) \geq u\}$ é não vazio e limitado inferiormente, em virtude dos limites de F em $\pm\infty$. Portanto, $F^*(u)$ é um número real.

1 Pela monotonicidade de F , o complemento de A_u é $B_u = \{x : F(x) < u\}$, e todo elemento de B_u é menor que todo elemento de A_u . Se $a = \inf A_u$, então $x < a$ implica $x \in B_u$, enquanto $x > a$ implica que existe $z \in A_u$ com $z < x$ e, pela monotonicidade, $x \in A_u$. Logo $\sup B_u = a = \inf A_u$.

2 Se $u_1 < u_2$, então $A_{u_2} \subseteq A_{u_1}$, de modo que $F^*(u_1) \leq F^*(u_2)$.

3 O ponto x pertence a $A_{F(x)}$; portanto, o ínfimo desse conjunto é menor ou igual a x .

4 Ponha $a = F^*(u)$. Para cada n , escolha $y_n \in A_u$ com $a \leq y_n < a + 1/n$ e defina $x_n = \min\{y_1, \dots, y_n\}$. Então $x_n \downarrow a$, pois $a \leq x_n \leq y_n < a + 1/n$. Pela continuidade de F à direita,

$$F(a) = \lim_n F(x_n) \geq u.$$

5 Se $F^*(u) \leq x$, o item anterior e a monotonicidade dão $u \leq F(F^*(u)) \leq F(x)$. Reciprocamente, se $u \leq F(x)$, então $x \in A_u$ e, por definição, $F^*(u) \leq x$.

□

Com isso, dada uma variável aleatória U definida em um espaço de probabilidade $(\Omega, \mathcal{F}, \mathbf{P})$, e uniformemente distribuída no intervalo $(0,1)$, podemos definir a variável aleatória

$$X = F^*(U),$$

e perceber que para $x \in \mathbb{R}$ temos $F(x) \in [0,1]$, e portanto

$$\mathbf{P}(X \leq x) = \mathbf{P}(F^*(U) \leq x) = \mathbf{P}(U \leq F(x)) = F(x)$$

onde na última igualdade usamos que a probabilidade que a uniforme seja menor que y é y , i.e., $\mathbf{P}(U \leq y) = y$

Mostramos assim o seguinte resultado.

Proposição 6.15 — Distribuição prescrita

Dada uma candidata a função de distribuição $F : \mathbb{R} \rightarrow [0,1]$, existe uma variável aleatória X , definida em algum espaço de probabilidade $(\Omega, \mathcal{F}, \mathbf{P})$, tal que

$$\mathbf{P}(X \leq x) = F(x) \quad \text{para todo } x \in \mathbb{R}.$$

Demonstração. Tome $\Omega = (0,1)$, $\mathcal{F} = \mathcal{B}((0,1))$, $\mathbf{P} = \lambda|_{(0,1)}$ e $U(\omega) = \omega$. Pelo item 5 da Proposição 6.14,

$$\{u \in (0,1) : F^*(u) \leq x\} = (0, F(x)] \cap (0,1),$$

que é boreliano. Portanto, F^* é Borel mensurável e $X = F^*(U)$ é uma variável aleatória. Novamente pelo mesmo item,

$$\{X \leq x\} = \{F^*(U) \leq x\} = \{U \leq F(x)\}.$$

Como U é uniforme em $(0,1)$,

$$\mathbf{P}(X \leq x) = \mathbf{P}(U \leq F(x)) = F(x)$$

para todo $x \in \mathbb{R}$. □

Portanto, toda candidata a função de distribuição é efetivamente a função de distribuição de alguma variável aleatória. A partir deste ponto, podemos usar simplesmente a expressão *função de distribuição* para funções com essas propriedades.

Resta examinar as descontinuidades de funções não decrescentes, como F e F^* .

Proposição 6.16 —

Se $f : I \rightarrow \mathbb{R}$ é uma função não decrescente definida em um intervalo $I \subseteq \mathbb{R}$, então o conjunto D_f formado pelos pontos de descontinuidade de f é no máximo enumerável.

O mesmo resultado vale para f não-crescente.

Demonstração. Suponha primeiro que f seja não decrescente. Em todo ponto interior a de I , os limites laterais

$$f(a-) = \lim_{x \uparrow a} f(x) \quad \text{e} \quad f(a+) = \lim_{x \downarrow a} f(x)$$

existem. Se $a \in D_f$ é interior a I , então $f(a-) < f(a+)$; logo o intervalo $I_a = (f(a-), f(a+))$ é não vazio.

Se $a_1 < a_2$, então $f(a_1+) \leq f(a_2-)$, de modo que I_{a_1} e I_{a_2} são disjuntos. Escolha um racional $r(a) \in I_a$ para cada ponto de descontinuidade interior. O mapa $a \mapsto r(a)$ é injetivo e toma valores em $\mathbb{Q}$; portanto há no máximo um número enumerável desses pontos. Os extremos de I , quando pertencem a I , acrescentam no máximo dois pontos.

Para uma função não crescente, aplique o resultado a $-f$. □

Corolário 6.17 —

Dadas uma função de distribuição F e sua inversa generalizada F^* , os conjuntos de pontos de descontinuidade de F e de F^* são, no máximo, enumeráveis.

Demonstração. As funções F e F^* são não decrescentes; basta aplicar a proposição anterior a cada uma delas. □

Encerramos a seção fixando a terminologia para variáveis com a mesma distribuição.

Definição 6.18 — Variáveis Aleatórias Identicamente Distribuídas

Duas variáveis aleatórias ξ e η são ditas identicamente distribuídas se

$$\mathbf{P}(\xi \leq x) = \mathbf{P}(\eta \leq x) \quad \text{para todo } x \in \mathbb{R}.$$

6.3. Tipos de Distribuições

Definição 6.19 —

Uma função de distribuição F é dita

- 1 discreta** se existem um conjunto enumerável de números reais $\{x_j\}$ e massas pontuais $p_j \geq 0$, com $\sum_j p_j = 1$, tais que

$$F(x) = \sum_{x_j \leq x} p_j \quad \text{para todo } x \in \mathbb{R}$$

- 2 contínua** se for contínua para todo x .

- 3 absolutamente contínua** se existe uma função f não negativa Lebesgue integrável tal que

$$F(b) - F(a) = \int_a^b f(x) dx \quad \text{para todo } a < b$$

- 4 singular contínua** se F é contínua e F' existe e é igual a 0 em quase todo ponto, com respeito à medida de Lebesgue.

Temos o seguinte Teorema de decomposição

Teorema 6.20 — Decomposição de Distribuição

Toda função de distribuição pode ser decomposta em uma combinação convexa de três tipos puros, discretas, absolutamente contínuas, e contínuas singulares. Assim, se F é uma função de distribuição, então

$$F = \alpha F_{ac} + \beta F_d + \gamma F_s,$$

onde $\alpha, \beta, \gamma \geq 0$ e $\alpha + \beta + \gamma = 1$.

- $F_{ac}(x) = \int_{-\infty}^x f(y) dy$, com $f = F'_{ac}$ em quase todo ponto;
- F_d é uma função de saltos com no máximo um número enumerável de saltos;
- F_s é singular contínua.

Componentes com coeficiente zero podem ser omitidas.

Demonstração. Pelo Teorema 6.21, demonstrado a seguir, existem $\delta \in [0,1]$ e funções de distribuição F_d e F_c , respectivamente discreta e contínua, tais que

$$F = \delta F_d + (1 - \delta) F_c.$$

Pelo Teorema 6.22, também demonstrado abaixo, existem $r, s \geq 0$, com $r + s = 1$, e funções de distribuição F_{ac} e F_s , respectivamente absolutamente contínua e singular contínua, tais que $F_c = r F_{ac} + s F_s$. Portanto,

$$F = (1 - \delta) r F_{ac} + \delta F_d + (1 - \delta) s F_s.$$

Os três coeficientes são não negativos e somam 1, como desejado. □

Começaremos provando que toda distribuição pode ser decomposta como soma de uma discreta e uma contínua.

Teorema 6.21 —

Toda função de distribuição pode ser decomposta em uma combinação convexa de uma discreta e uma contínua. Assim, se F é uma função de distribuição, então

$$F = \alpha F_c + \beta F_d$$

onde $\alpha, \beta \geq 0$ e $\alpha + \beta = 1$.

Demonstração. Seja ν a medida de probabilidade cuja função de distribuição é F . Já provamos que F possui no máximo um número enumerável de saltos; enumere seus pontos como $\{x_j\}_{j \geq 1}$ e defina

$$p_j = F(x_j) - F(x_j^-) = \nu(\{x_j\}).$$

A série $\sum_j p_j$ é no máximo 1. A fórmula

$$\nu_d(A) = \sum_{j: x_j \in A} p_j, \quad A \in \mathcal{B}(\mathbb{R}),$$

define uma medida finita e discreta. Além disso, $\nu_d \leq \nu$: para um conjunto A e um subconjunto finito J dos índices com $x_j \in A$,

$$\sum_{j \in J} p_j = \nu\left(\bigcup_{j \in J} \{x_j\}\right) \leq \nu(A),$$

e basta fazer os subconjuntos finitos crescerem. Logo $\nu_c := \nu - \nu_d$ é também uma medida finita e positiva.

Ponha $\beta = \nu_d(\mathbb{R})$ e $\alpha = \nu_c(\mathbb{R}) = 1 - \beta$. Se $\beta > 0$, normalize ν_d/β e denote sua função de distribuição por F_d ; se $\beta = 0$, escolha qualquer função de distribuição discreta para F_d . Analogamente, se $\alpha > 0$, normalize ν_c/α e denote sua função de distribuição por F_c . A medida ν_c não possui átomos, pois retiramos precisamente todas as massas pontuais de ν . Assim,

$$F_c(x) - F_c(x^-) = \frac{\nu_c(\{x\})}{\alpha} = 0$$

quando $\alpha > 0$; pela continuidade à direita, F_c é contínua. Se $\alpha = 0$, escolha qualquer função de distribuição contínua para F_c . Finalmente,

$$F(x) = \nu((-\infty, x]) = \alpha F_c(x) + \beta F_d(x),$$

e $\alpha + \beta = 1$. □

Teorema 6.22 —

Toda função de distribuição contínua pode ser decomposta em uma combinação convexa de uma absolutamente contínua e uma contínua singular. Assim, se F é uma função de distribuição contínua, então

$$F = \alpha F_{ac} + \beta F_s$$

onde $\alpha, \beta \geq 0$ e $\alpha + \beta = 1$. Componentes com coeficiente zero podem ser omitidas.

Para a demonstração desse teorema usaremos a seguinte ferramenta de análise real.

Teorema de diferenciação para funções monótonas. Se G é não decrescente em $[a, b]$, então G' existe e é não negativa quase certamente, e

$$\int_a^b G'(x) \, dx \leq G(b) - G(a).$$

Além disso, se g é integrável em $\mathbb{R}$, a função $x \mapsto \int_{-\infty}^x g(y) \, dy$ é absolutamente contínua em intervalos compactos e tem derivada g quase certamente.

Esse é um resultado avançado de análise real, baseado no lema de cobertura de Vitali e no Teorema Fundamental do Cálculo de Lebesgue. Usaremos esse resultado sem demonstração.

Demonstração. Aplique o teorema de diferenciação a F em cada intervalo compacto e defina $g = F'$ nos pontos em que a derivada existe, tomando $g = 0$ no conjunto excepcional. Então $g \geq 0$ e, para $a < b$,

$$\int_a^b g(y) \, dy \leq F(b) - F(a).$$

Em particular, fazendo $a = -n$, $b = n$ e usando o Teorema da Convergência Monótona, vemos que g é integrável em $\mathbb{R}$ e $\alpha := \int_{\mathbb{R}} g \, d\lambda \leq 1$. Ponha

$$G(x) = \int_{-\infty}^x g(y) \, dy, \quad H(x) = F(x) - G(x), \quad \beta = 1 - \alpha.$$

A função G é contínua e não decrescente. A desigualdade acima mostra que, se $a < b$,

$$H(b) - H(a) = F(b) - F(a) - \int_a^b g(y) \, dy \geq 0.$$

Como F é contínua, H também é contínua. Além disso,

$$G(-\infty) = H(-\infty) = 0, \quad G(+\infty) = \alpha, \quad H(+\infty) = \beta.$$

Se $\alpha > 0$, a função $F_{ac} = G/\alpha$ é uma função de distribuição absolutamente contínua.

Se $\beta > 0$, a função $F_s = H/\beta$ é uma função de distribuição contínua e

$$F'_s = \frac{F' - G'}{\beta} = \frac{g - g}{\beta} = 0 \quad \text{quase certamente;}$$

logo ela é singular contínua. Desse modo, $F = \alpha F_{ac} + \beta F_s$. Quando um coeficiente é zero, a componente correspondente é simplesmente omitida. □

6.3.1. Distribuições Discretas

Seja ξ uma variável aleatória com a função de distribuição F (com probabilidade induzida $\mathbf{P}$).

Definição 6.23 —

A distribuição de ξ (e ξ em si) é dita **discreta** se existir um conjunto enumerável S tal que

$$\mathbf{P}(\xi \in S) = 1.$$

Podemos enumerar o conjunto S , escrevendo $S = \{x_1, x_2, \dots\}$, e definimos $p_k := \mathbf{P}(\xi = x_k)$. Então temos

$$F(x) = \sum_{k: x_k \leq x} p_k,$$

e esta é uma soma sobre um conjunto enumerável. Estes p_k às vezes são denominados **massas pontuais**.

1 Massa de um ponto em zero:

$$\mathbf{P}(\xi = 0) = 1.$$

Portanto, a função de distribuição F é:

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

A distribuição correspondente é a **medida de Dirac** no 0.

- 2** Qualquer variável aleatória simples é discreta. Uma variável aleatória simples é uma variável aleatória que toma um número finito de valores.
- 3** Uma distribuição discreta não precisa manter seus átomos afastados: podemos tomar $S = \mathbb{Q}$ e atribuir massas positivas que somem 1, por exemplo, $p_k = \frac{1}{2^k}$, após enumerar os racionais. A distribuição é discreta; o conjunto de seus átomos não tem a mesma obrigação.

Algumas Distribuições Discretas

Distribuição	Notação	Função de Probabilidade	Domínio
Um ponto	δ_a	$p(a) = 1$	$\{a\}$
Uniforme Discreta	$\text{Unif}\{x_1, \dots, x_n\}$	$p(x_i) = \frac{1}{n}$	$\{x_1, \dots, x_n\}$
Bernoulli	$\text{Be}(p)$	$p(0) = q, p(1) = p$	$\{0, 1\}$
Binomial	$\text{Bin}(n, p)$	$p(k) = \binom{n}{k} p^k q^{n-k}$	$k = 0, 1, \dots, n$
Geométrica	$\text{Ge}(p)$	$p(k) = pq^k$	$k \in \mathbb{N}$
Primeiro Sucesso	$\text{Po}(p)$	$p(k) = pq^{k-1}$	$k \in \mathbb{N}^*$
Poisson	$\text{Po}(m)$	$p(k) = e^{-m} \frac{m^k}{k!}$	$k \in \mathbb{N}$

A distribuição uniforme discreta assume os valores $x_1, \dots, x_n$ e $p(x_i) = \frac{1}{n}$. Aqui e nas linhas seguintes, $q = 1 - p$. A distribuição $\text{Be}(p)$ descreve o resultado de um experimento de lançamento de moeda (não necessariamente honesta), e a distribuição $\text{Bin}(n, p)$ descreve o número de sucessos em n lançamentos. A distribuição $\text{Ge}(p)$ descreve o número de fracassos antes do primeiro sucesso e a distribuição $\text{Po}(p)$ o número de lançamentos necessários para ter sucesso uma vez.

A distribuição de Poisson surge como o limite da distribuição binomial quando $n \rightarrow \infty$ e $p \rightarrow 0$ e $np \rightarrow \lambda$.

Teorema 6.24 — Teorema Limite de Poisson

Se $p_n \in [0, 1]$ e $np_n \rightarrow \lambda \geq 0$, então, para cada $k \in \mathbb{N}$ fixo,

$$\binom{n}{k} p_n^k (1 - p_n)^{n-k} \rightarrow e^{-\lambda} \frac{\lambda^k}{k!}.$$

Demonstração. Para n suficientemente grande, temos $n \geq k$ e $p_n < 1$, e então

$$\binom{n}{k} p_n^k (1 - p_n)^{n-k} = \frac{n(n-1) \cdots (n-k+1)}{n^k} \frac{(np_n)^k}{k!} (1 - p_n)^n (1 - p_n)^{-k}.$$

Os dois primeiros fatores convergem, respectivamente, para 1 e $\lambda^k/k!$, e o último converge para 1. Como $p_n \rightarrow 0$,

$$n \log(1 - p_n) = -(np_n) \frac{-\log(1 - p_n)}{p_n} \rightarrow -\lambda,$$

com a razão interpretada por continuidade quando $p_n = 0$. Logo $(1 - p_n)^n \rightarrow e^{-\lambda}$, e o produto dos quatro limites dá a conclusão. Quando $k = 0$, usa-se a convenção $\lambda^0 = 1$, inclusive para $\lambda = 0$. $\square$

6.3.2. Distribuições Absolutamente Contínuas

A distribuição de ξ (e ξ) é **absolutamente contínua** se existir uma função f , denominada **densidade de ξ** , tal que

$$F(x) = \int_{-\infty}^x f(y) dy \quad \text{para cada } x.$$

Claramente, é necessário que $f \geq 0$ e que

$$\int_{-\infty}^{\infty} f(x) dx = 1.$$

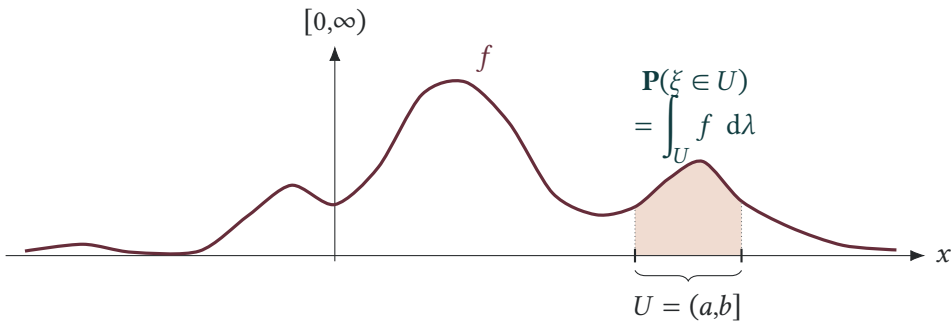


Figura 6.3.: A probabilidade de um intervalo é a área sob a densidade nesse intervalo.

$\square$ Em particular, se f for contínua, então

$$F'(x) = f(x),$$

ou seja, é a derivada de uma função de distribuição.

$\square$ Então, podemos dizer

$$\mathbf{P}(\xi \in (a, b]) = F(b) - F(a) = \int_a^b f(x) dx.$$

□ Alternativamente, podemos começar com uma função f tal que

$$f \geq 0 \quad \text{e} \quad \int_{-\infty}^{\infty} f(x) dx = 1,$$

e definir a função de distribuição pela fórmula

$$F(x) := \int_{-\infty}^x f(y) dy.$$

Exemplo 6.25 — XKCD.

Alice secretamente escolhe dois números reais diferentes por um processo desconhecido e os coloca em dois envelopes. Bob escolhe um dos dois envelopes aleatoriamente (com um lançamento de uma moeda justa) e mostra o número desse envelope. Agora você deve adivinhar se o número no outro envelope fechado é maior ou menor que o que você viu.

Existe uma estratégia que lhe dê uma chance melhor do que 50% de adivinhar corretamente, não importando o procedimento que Alice usasse para escolher seus números?

Solução:

Antes de abrir qualquer envelope, você escolhe um número r de forma aleatória, independentemente da escolha de Bob, usando uma distribuição absolutamente contínua que dê probabilidade positiva a todo intervalo aberto; a distribuição normal padrão é um exemplo. Vamos chamar de x o número do envelope aberto por Bob.

A **estratégia** agora é: se $r < x$, então você prevê que o número oculto y é menor que x ; Caso contrário, você adivinha que y é maior do que x .

Para ver por que a estratégia é vencedora, denote os dois números de Alice por $a < b$ e ponha $\varepsilon = \mathbf{P}(a < r < b) > 0$. Existem três casos:

- Se $r < a$, a regra sempre prevê que o número oculto é menor. Ela acerta exatamente quando Bob mostra b , o que ocorre com probabilidade $1/2$.
- Se $r > b$, a regra sempre prevê que o número oculto é maior. Ela acerta exatamente quando Bob mostra a , novamente com probabilidade $1/2$.

□ Se $a < r < b$, a regra acerta qualquer que seja o envelope mostrado.

Portanto, a probabilidade de vitória é

$$\frac{1}{2}(1 - \varepsilon) + \varepsilon = \frac{1 + \varepsilon}{2} > \frac{1}{2}.$$

O número auxiliar não precisa adivinhar os de Alice: basta ter uma chance positiva de cair entre eles.

◁

Distribuição Uniforme Contínua Uma variável aleatória contínua ξ é **uniforme** no intervalo $[a, b]$, para $a < b$, se sua f.d.p. é

$$f(x) = \begin{cases} \frac{1}{b-a}, & \text{se } x \in [a, b] \\ 0 & \text{caso contrário} \end{cases}$$

e denotamos esse fato por $\xi \sim \text{Uniforme}([a, b])$.

Nesse caso, a probabilidade de ξ estar num subintervalo de $[a, b]$ é proporcional ao comprimento de tal subintervalo; de fato, para $a \leq y < z \leq b$,

$$\mathbf{P}(y \leq \xi \leq z) = \int_y^z \frac{1}{b-a} dx = \frac{z-y}{b-a}.$$

Distribuição Exponencial Uma variável aleatória contínua ξ é **exponencial** com parâmetro $\lambda > 0$, e denotamos esse fato por

$$\xi \sim \text{Exponencial}(\lambda)$$

se sua função de densidade é

$$f_{\xi}(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{se } x \geq 0 \\ 0 & \text{caso contrário.} \end{cases}$$

A função de distribuição é

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

Portanto, $P(\xi > a) = e^{-\lambda a}$ para $a \geq 0$.

Distribuição Normal A variável aleatória ξ possui distribuição **normal** com parâmetros μ e σ^2 , abreviado por $\xi \sim \mathcal{N}(\mu; \sigma^2)$, se sua função densidade de probabilidade é dada por

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

Em particular, a distribuição normal padrão $\mathcal{N}(0; 1)$ tem densidade

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

6.3.3. Distribuições Singulares

Distribuição uniforme no conjunto de Cantor Seja $\mathcal{K}$ a família das uniões finitas de intervalos compactos disjuntos de $[0, 1]$. Para $A = \bigcup_{i=1}^m [a_i, b_i] \in \mathcal{K}$, defina

$$\Phi(A) = \bigcup_{i=1}^m \left(\left[a_i, \frac{2a_i + b_i}{3} \right] \cup \left[\frac{a_i + 2b_i}{3}, b_i \right] \right).$$

Em cada componente, Φ remove o terço central. Portanto,

$$\lambda(\Phi^n(A)) = \left(\frac{2}{3}\right)^n \lambda(A). \quad (6.3)$$

Para $n \geq 0$, defina $C_n = \Phi^n([0, 1])$. Os conjuntos C_n são compactos, $C_{n+1} \subset C_n$ e $\lambda(C_n) = (2/3)^n$. O **conjunto de Cantor** é

$$C = \bigcap_{n=0}^{\infty} C_n.$$

Pela continuidade por cima da medida de Lebesgue, $\lambda(C) = \lim_n \lambda(C_n) = 0$.

Observação 6.26 —

Geometricamente, começamos removendo $(1/3, 2/3)$ de $[0, 1]$ e, em seguida, removemos o terço central de cada intervalo remanescente. Assim,

$$C_0 = [0, 1],$$

$$C_1 = [0, 1/3] \cup [2/3, 1],$$

$$C_2 = [0, 1/9] \cup [2/9, 1/3] \cup [2/3, 7/9] \cup [8/9, 1].$$

Equivalentemente, usando uma das duas expansões ternárias nos pontos que as possuem,

$$C = \left\{ x \in [0, 1] : x = \sum_{i=1}^{\infty} \frac{a_i}{3^i}, a_i \in \{0, 2\} \right\}.$$



Figura 6.4.: As quatro primeiras etapas da construção do conjunto de Cantor.

Agora construiremos uma função de distribuição que cresce sobre o conjunto de Cantor e permanece constante nos intervalos removidos. Defina

$$f_n(x) = \left(\frac{3}{2}\right)^n \mathbb{1}_{C_n}(x), \quad F_n(x) = \int_{-\infty}^x f_n(t) \, dt,$$

e seja ν_n a probabilidade com densidade f_n . Se $\mathcal{J}_n$ denota a coleção dos 2^n intervalos componentes de C_n , então cada $J \in \mathcal{J}_n$ tem comprimento 3^{-n} e

$$\nu_n(J) = \left(\frac{3}{2}\right)^n 3^{-n} = 2^{-n}.$$

Mais geralmente, para $k \geq 0$,

$$\nu_{n+k}(J) = \left(\frac{3}{2}\right)^{n+k} \lambda(C_{n+k} \cap J) = \left(\frac{3}{2}\right)^{n+k} 3^{-n} \left(\frac{2}{3}\right)^k = 2^{-n}.$$

Fixados $x \in \mathbb{R}$ e $k \geq 0$, todos os componentes de $\mathcal{J}_n$ inteiramente à esquerda de x dão a mesma contribuição a $F_n(x)$ e $F_{n+k}(x)$. Há no máximo um componente de nível n cortado pelo ponto x , e a massa que cada uma das duas probabilidades atribui a esse componente é no máximo 2^{-n} . Consequentemente,

$$\sup_{x \in \mathbb{R}} |F_{n+k}(x) - F_n(x)| \leq 2^{1-n}.$$

Logo (F_n) é uniformemente de Cauchy. Seu limite uniforme F é contínuo, não decrescente, vale 0 à esquerda de 0 e vale 1 à direita de 1; portanto, é uma função de distribuição.

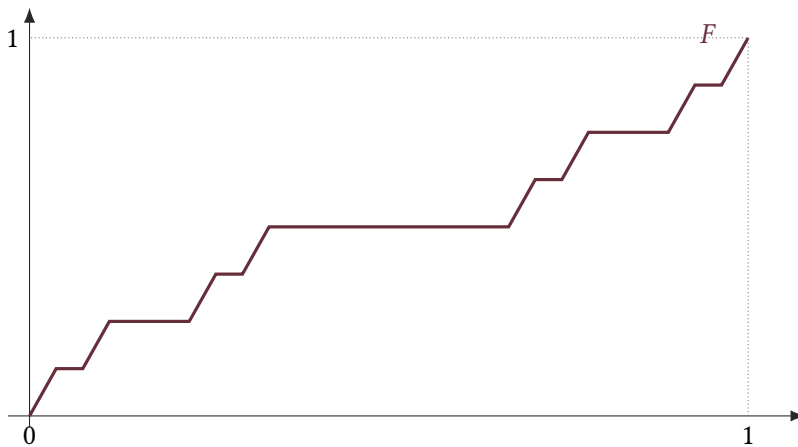


Figura 6.5.: Aproximação de terceira ordem da função de Cantor; os refinamentos convergem uniformemente para a função singular contínua.

Cada componente de $[0,1] \setminus C$ foi removida em alguma etapa. A partir dessa etapa, todas as funções F_n são constantes nessa componente, e o mesmo vale para F . Como $\lambda(C) = 0$, segue que $F' = 0$ quase certamente. Assim, F é uma função de distribuição singular contínua. Ela não pode ter densidade: se $F(x) = \int_{-\infty}^x h(t) \, dt$, então $h = F' = 0$ quase certamente, o que contradiz $\int_{\mathbb{R}} h \, d\lambda = 1$.

A construção probabilística correspondente, por lançamentos de moedas, é desenvolvida no Exercício 6.6, ao final do capítulo.

Outros exemplos selvagens Exemplo 6.27 — Uma densidade uniforme nos irracionais.

A função

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

é uma densidade. Como ela coincide quase certamente com $\mathbb{1}_{[0,1]}$, a distribuição correspondente é simplesmente a uniforme em $[0,1]$. Este exemplo ilustra que uma densidade só é determinada quase certamente.



Exemplo 6.28 — .

Dada uma enumeração $\{r_k\}_{k \geq 1}$ dos racionais, defina a função de massa

$$p(x) = \begin{cases} \frac{6}{\pi^2 k^2}, & x = r_k, \\ 0, & x \notin \mathbb{Q}. \end{cases}$$

Como $\sum_{k=1}^{\infty} 1/k^2 = \pi^2/6$, temos $\sum_{x \in \mathbb{R}} p(x) = 1$; portanto, p define uma distribuição discreta concentrada em $\mathbb{Q}$.



EXERCÍCIOS

Ex. 6.1 — Variáveis aleatórias discretas independentes. Sejam X e Y variáveis aleatórias independentes com valores inteiros.

1. Prove que, para todo $n \in \mathbb{Z}$,

$$\mathbf{P}(X + Y = n) = \sum_{k \in \mathbb{Z}} \mathbf{P}(X = k) \mathbf{P}(Y = n - k).$$

2. Se $X \sim \text{Poisson}(\lambda)$ e $Y \sim \text{Poisson}(\mu)$, determine a distribuição de $X + Y$.
3. Se $X \sim \text{Binomial}(n, p)$ e $Y \sim \text{Binomial}(m, p)$, determine a distribuição de $X + Y$.

Ex. 6.2 — A lei de uma variável aleatória. Seja $X : (\Omega, \mathcal{F}, \mathbf{P}) \rightarrow \mathbb{R}$ uma variável aleatória. Para $A \in \mathcal{B}(\mathbb{R})$, defina

$$P_X(A) = \mathbf{P}(X \in A) = \mathbf{P}(\{\omega \in \Omega : X(\omega) \in A\}).$$

Prove diretamente que P_X é uma medida de probabilidade em $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Ex. 6.3 — Seja $F : \mathbb{R} \rightarrow [0, 1]$ uma função não decrescente, contínua à direita e tal que

$$\lim_{x \rightarrow -\infty} F(x) = 0, \quad \lim_{x \rightarrow +\infty} F(x) = 1.$$

Construa explicitamente um espaço de probabilidade e uma variável aleatória X cuja função de distribuição seja F .

Dica. Tome uma variável uniforme U em $(0, 1)$ e use uma inversa generalizada de F .

Ex. 6.4 — Sejam $Y_1, Y_2, \dots$ variáveis aleatórias independentes, com

$$\mathbf{P}(Y_k = 0) = \mathbf{P}(Y_k = 1) = \frac{1}{2}, \quad k \geq 1,$$

e defina

$$U = \sum_{k=1}^{\infty} 2^{-k} Y_k.$$

1. Mostre que a série converge para um valor em $[0, 1]$.
2. Para $U_n = \sum_{k=1}^n 2^{-k} Y_k$, calcule a função de distribuição de U_n nos pontos diádicos.
3. Passando ao limite, prove que U possui distribuição uniforme em $[0, 1]$. Explique por que a dupla representação binária dos racionais diádicos não altera a conclusão.

Ex. 6.5 — Escreva

$$X = X^+ - X^-, \quad Y = Y^+ - Y^-.$$

Prove, usando apenas transformações mensuráveis, que

$$X \stackrel{d}{=} Y \implies X^+ \stackrel{d}{=} Y^+ \text{ e } X^- \stackrel{d}{=} Y^-.$$

***Ex. 6.6 — Outra construção da distribuição de Cantor.** Sejam $Y_1, Y_2, \dots$ como no exercício da expansão binária e defina

$$Y = \sum_{n=1}^{\infty} \frac{2Y_n}{3^n}.$$

1. Prove que Y toma valores no conjunto de Cantor e que sua função de distribuição F_Y é contínua.
2. Mostre que F_Y é constante em cada componente conexa de $[0,1] \setminus C$, onde C é o conjunto de Cantor. Conclua que $F'_Y(x) = 0$ para Lebesgue-quase todo x .
3. Seja μ_Y a lei de Y e λ a medida de Lebesgue. Prove que μ_Y e λ são mutuamente singulares, isto é, encontre um boreliano A tal que $\lambda(A) = 0$ e $\mu_Y(A^c) = 0$.

Ex. 6.7 — Recorde que uma variável aleatória Y é *estocasticamente maior* do que X quando

$$\mathbf{P}(X \leq x) \geq \mathbf{P}(Y \leq x) \quad \text{para todo } x \in \mathbb{R}.$$

Prove que essa relação vale se, e somente se, existem variáveis X^* e Y^* , definidas no mesmo espaço de probabilidade, tais que

$$X^* \stackrel{d}{=} X, \quad Y^* \stackrel{d}{=} Y, \quad X^* \leq Y^* \quad \text{quase certamente.}$$

Dica. Para uma das implicações, use o mesmo uniforme nas inversas generalizadas das duas funções de distribuição.

Ex. 6.8 — Convenções para a inversa generalizada. Seja F uma função de distribuição e, para $t \in (0,1)$, defina

$$\begin{aligned} F_1^{-1}(t) &= \sup\{y \in \mathbb{R} : F(y) < t\}, & F_2^{-1}(t) &= \inf\{y \in \mathbb{R} : F(y) \geq t\}, \\ F_3^{-1}(t) &= \inf\{y \in \mathbb{R} : F(y) > t\}, & F_4^{-1}(t) &= \sup\{y \in \mathbb{R} : F(y) \leq t\}. \end{aligned}$$

Prove que $F_1^{-1} = F_2^{-1}$ e que $F_3^{-1} = F_4^{-1}$. Dê um exemplo de função de distribuição e de $t \in (0,1)$ para os quais $F_1^{-1}(t) \neq F_3^{-1}(t)$.

Valor Esperado

Uma distribuição contém toda a informação probabilística sobre uma variável, mas frequentemente precisamos resumi-la. A **esperança** realiza a redução mais elementar: associa à variável uma média ponderada de seus valores ou, geometricamente, o centro de massa de sua distribuição. Como todo resumo, ela é útil justamente porque esquece quase tudo.

Considere, para motivar a definição, uma variável simples

$$X = \sum_{k=1}^n a_k \cdot \mathbb{1}_{A_k},$$

com $A_1, A_2, \dots, A_n$ disjuntos. Então a média ponderada $m(X)$ dos valores que X assume é

$$m(X) = \sum_{k=1}^n a_k \mathbf{P}(A_k) = \int X \, d\mathbf{P}.$$

Isso conduz à definição geral.

Definição 7.1 — Valor Esperado ou Esperança

Se ξ é uma variável aleatória definida em $(\Omega, \mathcal{F}, \mathbf{P})$, seu **valor esperado**, **esperança** ou **média** é definido como

$$\mathbf{E}[\xi] := \int_{\Omega} \xi d\mathbf{P}$$

Usaremos aqui a mesma nomenclatura das integrais de Lebesgue. Diremos que X é **integrável** se $\mathbf{E}[X] < \infty$, o que equivale a $\mathbf{E}[X^+] < \infty$ e $\mathbf{E}[X^-] < \infty$. Mais geralmente, a esperança de X está bem definida, possivelmente como $+\infty$ ou $-\infty$, se ao menos uma das quantidades $\mathbf{E}[X^+]$ e $\mathbf{E}[X^-]$ for finita; nesse caso, dizemos que X é **quase integrável**.

A definição recupera imediatamente o cálculo para variáveis simples.

Exemplo 7.2 — Variáveis Aleatórias Simples.

Seja X uma variável aleatória dada por

$$X = \sum_{k=1}^n a_k \cdot \mathbb{1}_{A_k},$$

com $A_1, A_2, \dots, A_n$ disjuntos. Então

$$\mathbf{E}[X] = \int X d\mathbf{P} = \sum_{k=1}^n a_k \mathbf{P}(A_k).$$

Se além disso, os valores $a_1, \dots, a_k \in \mathbb{R}$ são distintos, então $A_k = X^{-1}(\{a_k\}) = \{X = a_k\}$, e

$$\mathbf{E}[X] = \sum_{k=1}^n a_k \mathbf{P}(X = a_k).$$



O caso particular das indicadoras estabelece uma ligação direta entre esperança e probabilidade.

Exemplo 7.3 — A Esperança da Função Indicadora.

Seja 1_A a indicadora de um evento $A \in \mathcal{F}$. Como 1_A é uma variável aleatória simples, seu valor esperado é dado por

$$\mathbf{E}[1_A] = \mathbf{P}(A).$$

◀

Definida pela integral de Lebesgue, a esperança herda as propriedades seguintes.

Teorema 7.4 —

Suponha que ξ, η sejam variáveis aleatórias integráveis. Então

- 1 (Linearidade) $\mathbf{E}[a\xi + b\eta] = a\mathbf{E}[\xi] + b\mathbf{E}[\eta]$ para quaisquer números reais a, b .
- 2 Se $X \equiv b \in \mathbb{R}$, constante, então $\mathbf{E}[X] = b$ (escrevemos também que $\mathbf{E}[b] = b$).
- 3 Se $\xi \geq 0$ então $\mathbf{E}[\xi] \geq 0$.
- 4 (Monotonicidade) Se $\xi \geq \eta$ então $\mathbf{E}[\xi] \geq \mathbf{E}[\eta]$.
- 5 (Desigualdade triangular) $|\mathbf{E}[\xi]| \leq \mathbf{E}[|\xi|]$
- 6 Se $0 \leq \xi_n \uparrow \xi$, então $\mathbf{E}[\xi] = \lim_{n \rightarrow \infty} \mathbf{E}[\xi_n]$, admitindo-se o valor $+\infty$.

Demonstração. Como $\mathbf{E}[\xi] = \int_{\Omega} \xi \, d\mathbf{P}$, a linearidade para funções integráveis, a positividade, a monotonicidade, a desigualdade triangular e a convergência monótona são exatamente as propriedades da integral de Lebesgue demonstradas no capítulo anterior. Para explicitar o único ponto específico de uma probabilidade, se $X \equiv b$, então $X = b1_{\Omega}$ e

$$\mathbf{E}[b] = b\mathbf{P}(\Omega) = b.$$

□

Exemplo 7.5 — .

Suponha que uma esfera seja colorida de modo que 90% de sua área seja vermelha e 10% seja azul, com a região vermelha mensurável. Fixe um cubo inscrito e escolha ao acaso uma rotação do cubo segundo a probabilidade uniforme — a medida de Haar normalizada — em $SO(3)$. Para cada vértice v_i do cubo, sua imagem pela rotação é então uniformemente distribuída na esfera.

Seja ξ_i a indicadora de que o i -ésimo vértice cai na região vermelha. Então $\mathbf{E}[\xi_i] = 0,9$ para todo i e $N = \xi_1 + \dots + \xi_8$ é o número de vértices vermelhos. Por linearidade da esperança,

$$\mathbf{E}[N] = \sum_{i=1}^8 \mathbf{E}[\xi_i] = 8 \cdot 0,9 = 7,2.$$

Se toda rotação tivesse no máximo sete vértices vermelhos, teríamos $N \leq 7$ quase certamente e, portanto, $\mathbf{E}[N] \leq 7$, uma contradição. Logo existe uma rotação para a qual os oito vértices estão em pontos vermelhos da esfera.

Esse é um exemplo da utilização de técnicas probabilísticas para provar fatos determinísticos. Observe que não exigimos que as variáveis ξ_i sejam independentes.

◀

Temos também a seguinte caracterização do operador esperança.

Teorema 7.6 — Caracterização da Esperança

A esperança é o único operador linear $T : \mathcal{L}^1(\Omega, \mathcal{F}, \mathbf{P}) \rightarrow \mathbb{R}$ que satisfaz:

- continuidade na norma de $\mathcal{L}^1$: se $\|\xi_n - \xi\|_1 \rightarrow 0$, então $T(\xi_n) \rightarrow T(\xi)$;
- relação com a probabilidade: $T(\mathbb{1}_A) = \mathbf{P}(A)$ para todo $A \in \mathcal{F}$.

Demonstração. O operador $\xi \mapsto \int_{\Omega} \xi \, d\mathbf{P}$ é linear, $|\mathbf{E}[\xi]| \leq \|\xi\|_1$ e $\mathbf{E}[\mathbb{1}_A] = \mathbf{P}(A)$; portanto, possui as propriedades anunciadas.

Reciprocamente, seja T um operador com essas propriedades. Se $s = \sum_{j=1}^m a_j \mathbb{1}_{A_j}$ é simples, a linearidade e a condição sobre indicadoras dão

$$T(s) = \sum_{j=1}^m a_j \mathbf{P}(A_j) = \mathbf{E}[s].$$

As funções simples são densas em $\mathcal{L}^1$: para $\xi \in \mathcal{L}^1$, escolha simples s_n com $\|s_n - \xi\|_1 \rightarrow 0$. Pela continuidade de T e pela estimativa $|\mathbf{E}[s_n - \xi]| \leq \|s_n - \xi\|_1$,

$$T(\xi) = \lim_n T(s_n) = \lim_n \mathbf{E}[s_n] = \mathbf{E}[\xi].$$

Isso prova a unicidade. □

Fixaremos agora a terminologia para propriedades que valem fora de um evento de probabilidade zero.

Definição 7.7 —

Dado um espaço de probabilidade $(\Omega, \mathcal{F}, \mathbf{P})$, diremos que uma afirmativa A ocorre **quase certamente** se

$$\mathbf{P}(\{\omega \in \Omega : \omega \text{ satisfaz a afirmativa } A\}) = 1.$$

Denotaremos isso por “ A ocorre q.c.”.

Em particular, dadas duas variáveis aleatórias X, Y definidas em Ω , temos que $X = Y$ q.c., se $\mathbf{P}(X = Y) = 1$. Ou ainda que $X \leq Y$ q.c se $\mathbf{P}(X \leq Y) = 1$.

Integrais de Lebesgue ignoram eventos de medida nula, e isso é mostrado no seguinte resultado.

Proposição 7.8 —

Dadas variáveis aleatórias X e Y , vale que

- 1 Se $X = 0$ q.c então $\mathbf{E}[X] = 0$;
- 2 Se $X = Y$ q.c e X é integrável ou $\mathbf{E}[X] = \pm\infty$, então $\mathbf{E}[Y] = \mathbf{E}[X]$;
- 3 Se $X \leq Y$ q.c. e Y é integrável, então $\mathbf{E}[X] \leq \mathbf{E}[Y]$ ($\mathbf{E}[X]$ pode ser $-\infty$).
- 4 Se $X \geq 0$ e $\mathbf{E}[X] = 0$, então $X = 0$ q.c..

Demonstração.

- 1 Como $X^+ = X^- = 0$ quase certamente, ambas as integrais são nulas e, portanto, $\mathbf{E}[X] = 0$.
- 2 Da igualdade quase certa seguem $X^+ = Y^+$ e $X^- = Y^-$ quase certamente. As integrais das partes positiva e negativa coincidem; logo a existência e o valor da esperança de Y são os mesmos que os de X .
- 3 Defina

$$X^* = X\mathbf{1}_{\{X \leq Y\}} + Y\mathbf{1}_{\{X > Y\}}.$$

Então $X^* = X$ quase certamente e $X^* \leq Y$ em todo ponto. Além disso, $(X^*)^+ \leq Y^+$, de modo que a parte positiva de X^* é integrável. Por monotonicidade, $\mathbf{E}[X] = \mathbf{E}[X^*] \leq \mathbf{E}[Y]$, admitindo-se o valor $-\infty$.

- 4 Dado $\epsilon > 0$, defina $Y_\epsilon = \epsilon\mathbf{1}_{\{X > \epsilon\}}$ e observe que $Y_\epsilon \leq X$ para todo $\epsilon > 0$. Segue que

$$\epsilon\mathbf{P}(X > \epsilon) = \mathbf{E}[Y_\epsilon] \leq \mathbf{E}[X] = 0,$$

e portanto $\mathbf{P}(X > \epsilon) = 0$ para todo $\epsilon > 0$.

Como $\{X > 0\} = \bigcup_{n \geq 1} \{X > 1/n\}$, concluímos que $\mathbf{P}(X > 0) = 0$. Como $X \geq 0$, segue que $X = 0$ quase certamente.

□

7.1. Calculando a Esperança

A esperança depende apenas da distribuição da variável. Demonstraremos esse fato seguindo novamente a construção em três passos: funções simples, funções não negativas e funções integráveis.

Tome agora uma variável aleatória X qualquer, e $\phi : \mathbb{R} \rightarrow \mathbb{R}$ uma função mensurável e simples e não-negativa, dada por

$$\phi(x) = \sum_{k=1}^n a_k \mathbb{1}_{A_k}(x),$$

$A_1, \dots, A_n \in \mathcal{B}(\mathbb{R})$ disjuntos.

Temos assim que

$$\phi(X) = \sum_{k=1}^n a_k \mathbb{1}_{A_k}(X) = \sum_{k=1}^n a_k \mathbb{1}_{X \in A_k},$$

que é uma variável aleatória simples.

Segue que

$$\mathbf{E}[\phi(X)] = \sum_{k=1}^n a_k \mathbf{P}(X \in A_k) = \sum_{k=1}^n a_k \mathbf{P}_X(A_k) = \int_{\mathbb{R}} \phi(x) d\mathbf{P}_X(x).$$

Dada agora uma função $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ mensurável e não negativa, tome uma sequência não-decrescente de funções simples e mensuráveis $\phi_n : \mathbb{R} \rightarrow \mathbb{R}, n \geq 1$, não-negativas e tais que $\phi_n(x) \uparrow \varphi(x)$, para todo $x \in \mathbb{R}$.

Pelos mesmos argumentos anteriores, $\phi_n(X), n \geq 1$ é uma sequência de variáveis aleatórias simples e não-negativas, tais que $\phi_n(X) \uparrow \varphi(X)$.

Segue do Teorema da Convergência Monótona, que

$$\mathbf{E}[\varphi(X)] = \lim_{n \rightarrow \infty} \mathbf{E}[\phi_n(X)] = \lim_{n \rightarrow \infty} \int_{\mathbb{R}} \phi_n(x) d\mathbf{P}_X(x) = \int_{\mathbb{R}} \varphi(x) d\mathbf{P}_X(x).$$

Dada agora uma função $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ mensurável, podemos escrever $\varphi = \varphi^+ - \varphi^-$, e aplicando o que descobrimos até agora, vemos que

$$\mathbf{E}[\varphi(X)] = \mathbf{E}[\varphi^+(X)] - \mathbf{E}[\varphi^-(X)] = \int_{\mathbb{R}} \varphi^+ d\mathbf{P}_X - \int_{\mathbb{R}} \varphi^- d\mathbf{P}_X = \int_{\mathbb{R}} \varphi d\mathbf{P}_X.$$

Com isso mostramos que

Teorema 7.9 — Teorema da Mudança de Variáveis

Dada uma variável aleatória X definida em um espaço de probabilidade $(\Omega, \mathcal{F}, \mathbf{P})$ com distribuição $\mathbf{P}_X$, e uma função mensurável $\varphi : \mathbb{R} \rightarrow \mathbb{R}$, então

$$\mathbf{E}[\varphi(X)] = \int_{\mathbb{R}} \varphi(x) \, d\mathbf{P}_X(x).$$

Ou seja, se qualquer um dos lados da equação acima estiver bem definido, o outro também está e a igualdade vale.

Em particular

$$\mathbf{E}[X] = \int_{\mathbb{R}} x \, d\mathbf{P}_X(x)$$

Demonstração. Para uma função simples não negativa $\varphi = \sum_{k=1}^m a_k \mathbb{1}_{A_k}$, a definição da distribuição induzida dá

$$\mathbf{E}[\varphi(X)] = \sum_{k=1}^m a_k \mathbf{P}(X \in A_k) = \sum_{k=1}^m a_k \mathbf{P}_X(A_k) = \int_{\mathbb{R}} \varphi \, d\mathbf{P}_X.$$

Se $\varphi \geq 0$ é mensurável, escolha funções simples $\varphi_n \uparrow \varphi$ e aplique a Convergência Monótona aos dois lados. Para uma função mensurável arbitrária, aplique o caso não negativo a φ^+ e φ^- . Isso também mostra que um lado está bem definido exatamente quando o outro está. A escolha $\varphi(x) = x$ fornece a última fórmula. $\square$

Uma consequência direta do resultado anterior é a seguinte.

Corolário 7.10 — Distribuição determina a esperança

Se ξ é integrável e $\xi \stackrel{d}{=} \eta$, então η é integrável e $\mathbf{E}[\xi] = \mathbf{E}[\eta]$.

Demonstração. As distribuições induzidas são iguais. Aplicando o Teorema 7.9 às funções $x \mapsto |x|$ e $x \mapsto x$, obtemos, respectivamente, a integrabilidade de η e a igualdade das esperanças. $\square$

A dependência apenas da distribuição explica as fórmulas usuais para variáveis discretas e absolutamente contínuas. Vamos deduzi-las.

7.1.1. Variáveis Aleatórias Discretas

Uma variável aleatória discreta concentra sua distribuição num conjunto enumerável de valores. Assim, existe um conjunto $S = \{x_1, x_2, \dots\}$ tal que $\mathbf{P}(X \in S) = 1$.

Com isso, temos que a distribuição $\mathbf{P}_X$ é totalmente definida pelos valores de $\mathbf{P}(\{x_k\}) = \mathbf{P}(X = x_k)$.

De fato, para qualquer $A \in \mathcal{B}(\mathbb{R})$

$$\mathbf{P}_X(A) = \sum_{k: x_k \in A} \mathbf{P}_X(\{x_k\}).$$

Perceba também que

$$X = \sum_{k=1}^{\infty} x_k \mathbb{1}_{A_k},$$

onde $A_k = \{X = x_k\} \in \mathcal{F}$.

Comecemos pelo caso $X \geq 0$, com $x_k \geq 0$ para todo $k \geq 1$. As variáveis

$$X_n = \sum_{k=1}^n x_k \mathbb{1}_{A_k},$$

são simples e portanto

$$\mathbf{E}[X_n] = \sum_{k=1}^n x_k \mathbf{P}(A_k) = \sum_{k=1}^n x_k \mathbf{P}(X = x_k).$$

Além disso, temos que $X_n \uparrow X$ e o Teorema da Convergência Monótona nos diz que

$$\mathbf{E}[X] = \lim_{n \rightarrow \infty} \mathbf{E}[X_n] = \sum_{k=1}^{\infty} x_k \mathbf{P}(X = x_k).$$

O caso com sinal será obtido separando as partes positiva e negativa.

Proposição 7.11 —

Para uma variável aleatória discreta X não negativa ou integrável, assumindo valores em $\{x_1, x_2, \dots\}$, temos

$$\mathbf{E}[X] = \sum_{k=1}^{\infty} x_k \mathbf{P}(X = x_k), \quad (7.1)$$

de modo que a série à direita é absolutamente convergente sempre que X for integrável.

Demonstração. Se $X \geq 0$, as variáveis simples $X_n = \sum_{k=1}^n x_k \mathbf{1}_{\{X=x_k\}}$ crescem para X quase certamente; a Convergência Monótona fornece a fórmula. Se X é integrável, aplique esse caso a X^+ e X^- e subtraia. Além disso, aplicá-lo a $|X|$ mostra que

$$\sum_{k=1}^{\infty} |x_k| \mathbf{P}(X = x_k) = \mathbf{E}|X| < \infty.$$

□

7.1.2. Variáveis Absolutamente Contínuas

Vamos recordar que uma variável aleatória é absolutamente contínua se existe uma função mensurável $f : \mathbb{R} \rightarrow \mathbb{R}^+$, conhecida como **função densidade de probabilidade de X** (ou simplesmente densidade de X) tal que

$$\mathbf{P}_X(A) = \mathbf{P}(X \in A) = \int_A f(x) \, d\mu,$$

onde a integral à direita é a integral de Lebesgue (feita em relação à medida de Lebesgue da reta).

Para calcular $\mathbf{E}[\varphi(X)]$, precisamos primeiro avaliar

$$\int_{\mathbb{R}} \varphi(x) \, d\mathbf{P}_X(x).$$

Para isso, tome primeiro uma função simples

$$\phi(x) = \sum_{k=1}^n a_k \mathbb{1}_{A_k},$$

e note que

$$\begin{aligned} \int_{\mathbb{R}} \phi(x) \, d\mathbf{P}_X(x) &= \sum_{k=1}^n a_k \mathbf{P}_X(A_k) \\ &= \sum_{k=1}^n a_k \int_{A_k} f(x) \, d\mu \\ &= \sum_{k=1}^n a_k \int_{\mathbb{R}} \mathbb{1}_{A_k}(x) f(x) \, d\mu \\ &= \int_{\mathbb{R}} \sum_{k=1}^n a_k \mathbb{1}_{A_k}(x) f(x) \, d\mu \\ &= \int_{\mathbb{R}} \phi(x) f(x) \, dx. \end{aligned}$$

Tomando uma sequência de funções simples $\phi_n, n \geq 1$ não negativas, com $\phi_n \uparrow \varphi \geq 0$, concluímos pelo Teorema da Convergência Monótona que

$$\int_{\mathbb{R}} \varphi(x) \, d\mathbf{P}_X(x) = \lim_n \int_{\mathbb{R}} \phi_n(x) \, d\mathbf{P}_X(x) = \lim_n \int_{\mathbb{R}} \phi_n(x) f(x) \, dx = \int_{\mathbb{R}} \varphi(x) f(x) \, dx.$$

E separando uma função mensurável φ em parte positiva e negativa, concluímos o seguinte resultado.

Proposição 7.12 —

Para uma variável aleatória absolutamente contínua X com função densidade de probabilidade $f : \mathbb{R} \rightarrow \mathbb{R}^+$ e uma função mensurável $\varphi : \mathbb{R} \rightarrow \mathbb{R}$, vale

$$\mathbf{E}[\varphi(X)] = \int_{\mathbb{R}} \varphi(x) f(x) \, dx, \quad (7.2)$$

onde um lado da equação está bem definido se, e somente se, o outro também está.

Demonstração. Pelo Teorema 7.9, $E[\varphi(X)] = \int_{\mathbb{R}} \varphi \, d\mathbf{P}_X$. Para funções simples não negativas, a identidade $\int \varphi \, d\mathbf{P}_X = \int \varphi f \, d\lambda$ segue diretamente de $\mathbf{P}_X(A) = \int_A f \, d\lambda$. A aproximação monótona estende a identidade a toda $\varphi \geq 0$ mensurável; a decomposição em partes positiva e negativa conclui o caso geral. $\square$

7.2. Funções de Vetores Aleatórios

O mesmo princípio vale em dimensão maior. Dado um vetor aleatório $X = (X_1, \dots, X_n)$ e uma função mensurável $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$, queremos calcular $E[\varphi(X)]$ usando apenas a distribuição conjunta de X .

Vamos olhar primeiro para a distribuição $\mathbf{P}_X$ do vetor aleatório X . Lembre-se que, de modo análogo a variáveis aleatórias, $\mathbf{P}_X$ é uma medida de probabilidade em $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ dada por

$$\mathbf{P}_X(A) = \mathbf{P}((X_1, \dots, X_n) \in A),$$

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

De posse da distribuição $\mathbf{P}_X$ podemos recuperar as **distribuições marginais** das variáveis X_k , fazendo

$$\begin{aligned} \mathbf{P}_{X_k}(B) &= \mathbf{P}(X_k \in B) \\ &= \mathbf{P}(X_k \in B; X_j \in \mathbb{R}; j \neq k) \\ &= \mathbf{P}((X_1, \dots, X_n) \in \mathbb{R}^{k-1} \times B \times \mathbb{R}^{n-k}) \\ &= \mathbf{P}_X(\mathbb{R}^{k-1} \times B \times \mathbb{R}^{n-k}). \end{aligned}$$

A relação entre $\mathbf{P}_X$ e as marginais $\mathbf{P}_{X_k}$ pode ser complicada, e em geral precisamos de informações adicionais sobre o vetor para descrevê-la com mais precisão. Uma situação em que essa relação assume uma forma simples é aquela em que $X_1, \dots, X_n$ são independentes. Deixaremos os detalhes para a lista de exercícios.

O argumento do Teorema 7.9 fornece sua versão vetorial.

Teorema 7.13 —

Dado um vetor aleatório $X = (X_1, \dots, X_n)$ definido em um espaço de probabilidade $(\Omega, \mathcal{F}, \mathbf{P})$ com distribuição $\mathbf{P}_X$, e uma função mensurável $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$, então

$$\mathbf{E}[\varphi(X_1, \dots, X_n)] = \int_{\mathbb{R}^n} \varphi(x_1, \dots, x_n) d\mathbf{P}_X(x_1, \dots, x_n).$$

Ou seja, se qualquer um dos lados da equação acima estiver bem definido, o outro também está e a igualdade vale.

Demonstração. O argumento é o mesmo do Teorema 7.9, agora para a medida imagem $\mathbf{P}_X$ em $\mathbb{R}^n$. Para $\varphi = \sum_{j=1}^m a_j \mathbb{1}_{A_j}$ simples e não negativa,

$$\mathbf{E}[\varphi(X)] = \sum_{j=1}^m a_j \mathbf{P}(X \in A_j) = \sum_{j=1}^m a_j \mathbf{P}_X(A_j) = \int_{\mathbb{R}^n} \varphi d\mathbf{P}_X.$$

A Convergência Monótona trata funções não negativas, e a decomposição em partes positiva e negativa trata o caso geral. □

No caso em que as variáveis $X_1, \dots, X_n$ são discretas, assumindo valores em conjuntos enumeráveis $S_1, \dots, S_n$, o vetor também é discreto e assume valores em $S_1 \times \dots \times S_n$. A variável $\varphi(X_1, \dots, X_n)$ assume o valor $\varphi(x_1, \dots, x_n)$ no evento $\{X_1 = x_1, \dots, X_n = x_n\}$.

Vale então o seguinte resultado.

Proposição 7.14 —

Dadas variáveis aleatórias $X_1, \dots, X_n$ discretas assumindo valores em $S_1, \dots, S_n$, respectivamente, e uma função mensurável $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$, então

$$\mathbf{E}[\varphi(X_1, \dots, X_n)] = \sum_{x_1 \in S_1} \dots \sum_{x_n \in S_n} \varphi(x_1, \dots, x_n) \mathbf{P}(X_1 = x_1, \dots, X_n = x_n),$$

sempre que um dos lados estiver bem definido.

Demonstração. Aplique o Teorema 7.13 à distribuição discreta do vetor X . Sua massa no ponto $(x_1, \dots, x_n)$ é precisamente $\mathbf{P}(X_1 = x_1, \dots, X_n = x_n)$. A fórmula segue diretamente para funções não negativas por aproximação monótona das somas finitas e, depois, para funções com sinal pela decomposição positiva e negativa. $\square$

O caso onde $X_1, \dots, X_n$ são absolutamente contínuas é um pouco mais complicado, pois não garante que o vetor X seja absolutamente contínuo. Ou seja, a existência de densidades $f_1, \dots, f_n : \mathbb{R} \rightarrow \mathbb{R}$ não garante a existência de uma função densidade $f : \mathbb{R}^n \rightarrow \mathbb{R}^+$ tal que

$$\begin{aligned} \mathbf{P}_X(A) &= \mathbf{P}((X_1, \dots, X_n) \in A) \\ &= \int_A f(x_1, \dots, x_n) \, d\lambda_n(x_1, \dots, x_n) \\ &= \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} \mathbf{1}_A(x_1, \dots, x_n) f(x_1, \dots, x_n) \, dx_1 \cdots dx_n, \end{aligned}$$

para $A \in \mathcal{B}(\mathbb{R}^n)$, onde λ_n é a medida de Lebesgue em $\mathbb{R}^n$.

Observação 7.15 —

Quando $X = (X_1, \dots, X_n)$ é um vetor aleatório absolutamente contínuo, também é usual dizer que as variáveis $X_1, \dots, X_n$ são conjuntamente absolutamente contínuas.

Não vamos entrar em maiores detalhes sobre esse tipo de vetor, mas podemos mostrar que

Proposição 7.16 —

Para um vetor aleatório absolutamente contínuo X com função densidade de probabilidade $f : \mathbb{R}^n \rightarrow \mathbb{R}^+$ e uma função mensurável $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$, vale

$$\mathbf{E}[\varphi(X)] = \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} \varphi(x_1, \dots, x_n) f(x_1, \dots, x_n) \, dx_1 \cdots dx_n, \quad (7.3)$$

onde um lado da equação está bem definido se, e somente se, o outro também está.

Demonstração. Como $\mathbf{P}_X(A) = \int_A f \, d\lambda_n$, a mesma prova do caso unidimensional dá

$$\int_{\mathbb{R}^n} \varphi \, d\mathbf{P}_X = \int_{\mathbb{R}^n} \varphi(x)f(x) \, d\lambda_n(x)$$

primeiro para funções simples não negativas, depois para funções mensuráveis não negativas e, por fim, para funções com sinal. Combine essa identidade com o Teorema 7.13. □

7.3. Variância, Covariância e Momentos de uma Variável Aleatória

A esperança localiza o centro de uma distribuição, mas não informa como a massa se espalha ao redor dele. O exemplo seguinte torna visível essa perda.

Exemplo 7.17 — .

Tome X e Y variáveis aleatórias discretas tais que

$$\mathbf{P}(X = -1) = \mathbf{P}(X = 1) = \frac{1}{2},$$

e

$$\mathbf{P}(Y = -100) = \mathbf{P}(Y = -50) = \mathbf{P}(Y = 50) = \mathbf{P}(Y = 100) = \frac{1}{4}.$$

Calculando, encontramos que $\mathbf{E}[X] = \mathbf{E}[Y] = 0$, mas a variável aleatória Y assume valores muito mais dispersos em torno da média. ◀

A variância mede essa dispersão por meio da distância quadrática à média. Os momentos e a covariância ampliam a mesma ideia.

Definição 7.18 — Momentos, variância e covariância

Se ξ é uma variável aleatória em $(\Omega, \mathcal{F}, \mathbf{P})$:

- Dado $n \in \mathbb{N}^*$ e $\xi \in \mathcal{L}^n(\mathbf{P})$,

$$m_k := \mathbf{E}[\xi^k], \quad M_k := \mathbf{E}[|\xi|^k], \quad k = 1, \dots, n,$$

são denominados, respectivamente, o k -ésimo **momento** e o k -ésimo **momento absoluto** de ξ .

- Se $\xi \in \mathcal{L}^2(\mathbf{P})$, então

$$\text{Var}[\xi] := \mathbf{E}[(\xi - \mathbf{E}[\xi])^2]$$

é a variância de ξ . O número $\sigma := \sqrt{\text{Var}[\xi]}$ é denominado **desvio padrão** de ξ .

- Se $\xi, \eta \in \mathcal{L}^2(\mathbf{P})$, definimos a covariância por

$$\text{Cov}[\xi, \eta] := \mathbf{E}[(\xi - \mathbf{E}[\xi])(\eta - \mathbf{E}[\eta])]$$

As variáveis ξ e η são ditas não correlacionadas se $\text{Cov}[\xi, \eta] = 0$.

- Se ξ e η não são quase certamente constantes, seu **coeficiente de correlação** é

$$\rho(\xi, \eta) = \frac{\text{Cov}(\xi, \eta)}{\sigma_\xi \sigma_\eta}.$$

A expressão da variância deve ser lida como uma distância quadrática média à esperança. Para cada resultado ω , o desvio $\xi(\omega) - \mathbf{E}[\xi]$ indica quanto o valor observado ficou acima ou abaixo do centro. Se simplesmente fizéssemos a média desses desvios, os positivos e negativos se cancelariam. Ao elevá-los ao quadrado, todos passam a contribuir positivamente e os afastamentos maiores recebem peso maior.

Assim, variância pequena significa que a distribuição está fortemente concentrada em torno da média, enquanto variância grande indica maior dispersão. Como

o quadrado também eleva ao quadrado a unidade de medida, o desvio padrão $\sqrt{\text{Var}[\xi]}$ costuma ser mais fácil de interpretar: ele mede a dispersão na mesma unidade de ξ .

Proposição 7.19 — Propriedades da Variância e Covariância

Dadas $\xi, \eta \in \mathcal{L}^2(\mathbf{P})$, temos:

- 1 $\text{Var}[\xi] \geq 0$
- 2 $\text{Var}[\xi] = \mathbf{E}[\xi^2] - \mathbf{E}[\xi]^2$.
- 3 $\text{Var}[\xi] = 0 \iff \xi = \mathbf{E}[\xi]$ quase certamente.
- 4 A aplicação $f : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto \mathbf{E}[(\xi - x)^2]$ tem um mínimo em $\mathbf{E}[\xi]$ com $f(\mathbf{E}[\xi]) = \text{Var}[\xi]$.
- 5 $\text{Var}[a\xi + b] = a^2 \text{Var}[\xi]$ para $a, b \in \mathbb{R}$.
- 6 $\text{Cov}[\xi, \eta] = \mathbf{E}[\xi\eta] - \mathbf{E}[\xi]\mathbf{E}[\eta]$.
- 7 $\text{Var}[\xi + \eta] = \text{Var}[\xi] + \text{Var}[\eta] + 2 \text{Cov}[\xi, \eta]$.
- 8 $\text{Cov}[\xi, \xi] = \text{Var}[\xi]$.

Demonstração.

- 1 Segue da definição, pois $(\xi - \mathbf{E}[\xi])^2 \geq 0$.
- 2 Segue da linearidade da esperança. De fato

$$\begin{aligned} \text{Var}[\xi] &= \mathbf{E}[(\xi - \mathbf{E}[\xi])^2] \\ &= \mathbf{E}[\xi^2] - 2\mathbf{E}[\xi]^2 + \mathbf{E}[\xi]^2 \\ &= \mathbf{E}[\xi^2] - \mathbf{E}[\xi]^2. \end{aligned}$$

- 3 Se $\text{Var}[\xi] = 0$, a variável não negativa $(\xi - \mathbf{E}[\xi])^2$ tem esperança nula; logo é nula quase certamente. A recíproca é imediata.

4 A identidade

$$\mathbf{E}[(\xi - x)^2] = \text{Var}[\xi] + (x - \mathbf{E}[\xi])^2$$

mostra que o mínimo é atingido em $x = \mathbf{E}[\xi]$ e vale $\text{Var}[\xi]$.

5 Como $a\xi + b - \mathbf{E}[a\xi + b] = a(\xi - \mathbf{E}[\xi])$, temos $\text{Var}[a\xi + b] = a^2 \text{Var}[\xi]$.

6 Segue da linearidade da esperança que

$$\begin{aligned} \text{Cov}[\xi, \eta] &= \mathbf{E}[(\xi - \mathbf{E}[\xi])(\eta - \mathbf{E}[\eta])] \\ &= \mathbf{E}[\xi\eta] - \mathbf{E}[\xi]\mathbf{E}[\eta]. \end{aligned}$$

7 Segue da linearidade da esperança e dos itens anteriores que

$$\begin{aligned} \text{Var}(\xi + \eta) &= \mathbf{E}[\xi^2] + \mathbf{E}[\eta^2] + 2\mathbf{E}[\xi\eta] - \mathbf{E}[\xi]^2 - \mathbf{E}[\eta]^2 - 2\mathbf{E}[\xi]\mathbf{E}[\eta] \\ &= \text{Var}[\xi] + \text{Var}[\eta] + 2 \text{Cov}[\xi, \eta]. \end{aligned}$$

8 Segue direto da definição. □

O item **4** nos fornece uma descrição geométrica: imagine que queremos minimizar o erro quadrático médio de uma estimativa x da variável X . Nesse caso a estimativa que minimiza esse erro é a esperança e a variância é o menor valor do erro quadrático médio.

Exemplo 7.20 — .

Seja $X \in \mathcal{L}^2(\mathbf{P})$ com $\text{Var}[X] > 0$ e seja $Y = aX + b$, com $a \neq 0$. Temos então $\mathbf{E}[Y] = a\mathbf{E}[X] + b$ e $\text{Var}[Y] = a^2 \text{Var}[X]$.

Da mesma forma temos que $\text{Cov}[X, Y] = \mathbf{E}[(X - \mathbf{E}[X])(aX + b - (a\mathbf{E}[X] + b))] = a \text{Var}[X]$.

Segue que

$$\rho(X, Y) = \frac{a \text{Var}[X]}{\sqrt{a^2 (\text{Var}[X])^2}} = \text{sinal}(a) = \begin{cases} 1; & a > 0 \\ -1; & a < 0 \end{cases}.$$

A covariância e correlação servem para medir o “*nível de correlação*” entre as variáveis. Nesse exemplo vimos que se $Y = aX + b$ então $\text{Cov}[X, Y]$ tem o mesmo sinal de a e $\rho(X, Y)$ é 1 ou -1 , de acordo com o sinal de a .

<

Teorema 7.21 — Independência e produto de esperanças

Sejam ξ, η variáveis aleatórias independentes e integráveis. Então $\xi\eta$ é integrável e

$$\mathbf{E}[\xi\eta] = \mathbf{E}[\xi]\mathbf{E}[\eta].$$

Em particular, se $\xi, \eta \in \mathcal{L}^2(\mathbf{P})$, então elas não são correlacionadas.

Demonstração. Assuma primeiro que ξ e η sejam simples, com valores $x_1, \dots, x_n$ e $y_1, \dots, y_m$. Então

$$\begin{aligned} \mathbf{E}[\xi\eta] &= \sum_{i=1}^n \sum_{j=1}^m x_i y_j \mathbf{P}[\xi = x_i, \eta = y_j] \\ &= \sum_{i=1}^n \sum_{j=1}^m x_i y_j \mathbf{P}[\xi = x_i] \mathbf{P}[\eta = y_j] \quad (\text{pela independência}) \\ &= \left(\sum_{i=1}^n x_i \mathbf{P}[\xi = x_i] \right) \left(\sum_{j=1}^m y_j \mathbf{P}[\eta = y_j] \right) \\ &= \mathbf{E}[\xi] \mathbf{E}[\eta]. \end{aligned}$$

Se $\xi, \eta \geq 0$, escolha funções simples $\xi_n \uparrow \xi$ e $\eta_n \uparrow \eta$, obtidas por transformações mensuráveis separadas. A independência é preservada e $\xi_n \eta_n \uparrow \xi\eta$. Pela Convergência Monótona,

$$\mathbf{E}[\xi\eta] = \lim_n \mathbf{E}[\xi_n \eta_n] = \lim_n \mathbf{E}[\xi_n] \mathbf{E}[\eta_n] = \mathbf{E}[\xi] \mathbf{E}[\eta].$$

Aplicando esse caso a $|\xi|$ e $|\eta|$, obtemos $\mathbf{E}|\xi\eta| = \mathbf{E}|\xi| \mathbf{E}|\eta| < \infty$. Por fim, decomponha ξ e η em partes positiva e negativa e use a linearidade. Se ambas pertencem a $\mathcal{L}^2$, a fórmula da covariância dá $\text{Cov}[\xi, \eta] = 0$. □

Vimos assim que variáveis aleatórias independentes têm covariância nula, quando possuem segundo momento finito. A recíproca não é verdadeira, como mostra o Exercício 7.20, ao final do capítulo.

Uma outra consequência importante é sobre a variância da soma de variáveis independentes.

Corolário 7.22 —

Se $X_1, \dots, X_n \in \mathcal{L}^2(\mathbf{P})$ são variáveis aleatórias independentes, então

$$\text{Var}[X_1 + \dots + X_n] = \text{Var}[X_1] + \dots + \text{Var}[X_n].$$

Demonstração. Para $n = 2$ temos

$$\text{Var}[X_1 + X_2] = \text{Var}[X_1] + \text{Var}[X_2] + 2 \text{Cov}[X_1, X_2] = \text{Var}[X_1] + \text{Var}[X_2].$$

Os demais casos seguem por indução, pois $X_1 + \dots + X_{n-1}$ é independente de X_n . □

Proposição 7.23 —

A aplicação $\text{Cov} : \mathcal{L}^2(\mathbf{P}) \times \mathcal{L}^2(\mathbf{P}) \rightarrow \mathbb{R}$ é uma forma simétrica bilinear positiva semidefinida e $\text{Cov}[\xi, \eta] = 0$ se η é constante quase certamente.

Demonstração. Se $\tilde{X} = X - \mathbf{E}[X]$ e $\tilde{Y} = Y - \mathbf{E}[Y]$, então $\text{Cov}[X, Y] = \mathbf{E}[\tilde{X}\tilde{Y}]$. A linearidade em cada entrada e a simetria seguem da linearidade da esperança e do produto. Além disso,

$$\text{Cov}[X, X] = \mathbf{E}[\tilde{X}^2] = \text{Var}[X] \geq 0,$$

de modo que a forma é positiva semidefinida. Se $Y = c$ quase certamente, então $\tilde{Y} = 0$ quase certamente e $\text{Cov}[X, Y] = 0$. □

Proposição 7.24 — Desigualdade de Cauchy–Schwarz

Se $\xi, \eta \in \mathcal{L}^2(\mathbf{P})$, então

$$\text{Cov}[\xi, \eta]^2 \leq \text{Var}[\xi] \text{Var}[\eta].$$

A igualdade é válida se, e somente se, existirem $a, b, c \in \mathbb{R}$, não todas nulas, e tais que $a\xi + b\eta + c = 0$ q.c.

Demonstração. A desigualdade de Cauchy-Schwarz é verdadeira para qualquer forma bilinear positiva semidefinita e, portanto, em particular para a covariância. Ainda assim, faremos uma demonstração aqui para deixar o texto mais completo e auto-contido.

Se $\text{Var}[\eta] = 0$, então η é constante quase certamente e ambos os lados da desigualdade são zero. Além disso, $\eta - \mathbf{E}[\eta] = 0$ já fornece a dependência afim anunciada.

Suponha agora $\text{Var}[\eta] > 0$ e ponha $\theta = -\text{Cov}[\xi, \eta] / \text{Var}[\eta]$.

Temos assim $\theta \text{Var}[\eta] = -\text{Cov}[\xi, \eta]$, e portanto

$$\begin{aligned} 0 &\leq \text{Var}[\xi + \theta\eta] \text{Var}[\eta] \\ &= (\text{Var}[\xi] + 2\theta \text{Cov}[\xi, \eta] + \theta^2 \text{Var}[\eta]) \text{Var}[\eta] \\ &= \text{Var}[\xi] \text{Var}[\eta] - \text{Cov}[\xi, \eta]^2. \end{aligned}$$

A igualdade ocorre se, e somente se, $\text{Var}[\xi + \theta\eta] = 0$, isto é, se $\xi + \theta\eta$ é constante quase certamente. Isso equivale à existência de $a, b, c \in \mathbb{R}$, não todos nulos, tais que $a\xi + b\eta + c = 0$ quase certamente. A recíproca também segue dessa caracterização, trocando as variáveis se necessário. □

Corolário 7.25 —

Dadas variáveis não constantes $X, Y \in \mathcal{L}^2(\mathbf{P})$, então

$$|\rho(X, Y)| \leq 1,$$

com igualdade se, e somente se, existirem constantes $a, b, c \in \mathbb{R}$, não todas

nulas, tais que $aX + bY + c = 0$ quase certamente.

Demonstração. Divida a desigualdade de Cauchy–Schwarz por $\text{Var}[X] \text{Var}[Y] > 0$ e use sua caracterização do caso de igualdade. $\square$

7.4. Integração com respeito à Distribuição

Uma função de distribuição determina uma probabilidade em $\mathbb{R}$. Integrar em relação a essa probabilidade conduz às integrais de **Lebesgue–Stieltjes**; aqui, elas serão sobretudo uma linguagem para calcular esperanças sem retornar ao espaço amostral.

Proposição 7.26 — Medida de Lebesgue–Stieltjes

Dada uma função de distribuição F em $\mathbb{R}$, existe uma única probabilidade $\mathbf{P}_F$ em $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ tal que

$$\mathbf{P}_F((a, b]) = F(b) - F(a)$$

para todo $a < b$.

Demonstração. Pela Proposição 6.15, existe uma variável aleatória X cuja função de distribuição é F . Sua distribuição $\mathbf{P}_X$ satisfaz

$$\mathbf{P}_X((a, b]) = \mathbf{P}(a < X \leq b) = F(b) - F(a),$$

o que prova a existência. Se duas probabilidades possuem essa propriedade, elas coincidem no π -sistema dos intervalos $(-\infty, b]$ (tomando o limite $a \rightarrow -\infty$) e, pelo teorema de unicidade de medidas, coincidem em $\mathcal{B}(\mathbb{R})$. $\square$

Definição 7.27 – Integração com respeito à distribuição

Dada F uma função de distribuição em $\mathbb{R}$. Para toda função $g \in \mathcal{L}^1(\mathbf{P}_F)$ definimos a integral de g com respeito a F

$$\mathbf{E}_F[g] := \int_{\mathbb{R}} g(x) dF(x)$$

onde estamos considerando g como uma variável aleatória no espaço de probabilidade $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbf{P}_F)$.

Proposição 7.28 –

Se $(p_i)_{i \geq 1}$ são números não negativos com $\sum_i p_i = 1$ e

$$F(t) = \sum_{i \geq 1} p_i \mathbb{1}_{\{t_i \leq t\}},$$

então, para toda função mensurável $g \geq 0$,

$$\mathbf{E}_F[g] = \int_{\mathbb{R}} g dF = \sum_i p_i g(t_i).$$

Demonstração. Primeiro suponha que $g = \sum_{j=1}^m b_j \mathbb{1}_{B_j}$ é simples e não negativa. Então

$$\mathbf{E}_F[g] = \sum_{j=1}^m b_j \mathbf{P}_F(B_j) = \sum_{j=1}^m b_j \sum_i p_i \mathbb{1}_{B_j}(t_i) \quad (7.4)$$

$$= \sum_i p_i \sum_{j=1}^m b_j \mathbb{1}_{B_j}(t_i) \quad (7.5)$$

$$= \sum_i p_i g(t_i) \quad (7.6)$$

Para $g \geq 0$ mensurável, tome funções simples $g_n \uparrow g$ e aplique duas vezes o Teorema da Convergência Monótona:

$$\mathbf{E}_F[g] = \lim_{n \rightarrow \infty} \mathbf{E}_F[g_n] = \lim_{n \rightarrow \infty} \sum_i p_i g_n(t_i) = \sum_i p_i g(t_i)$$


Proposição 7.29 —

Suponha que F é absolutamente contínua com densidade f . Então, para toda função mensurável $g \geq 0$,

$$\mathbf{E}_F[g] = \int_{\mathbb{R}} g dF = \int_{-\infty}^{\infty} g(x) f(x) dx$$

Demonstração. Primeiro suponha que $g = \sum_{j=1}^m b_j \mathbb{1}_{B_j}$ é simples e não negativa. Então

$$\mathbf{E}_F[g] = \sum_{j=1}^m b_j \mathbf{P}_F(B_j) = \sum_{j=1}^m b_j \int_{B_j} f(x) \, dx \quad (7.7)$$

$$= \sum_{j=1}^m b_j \int_{-\infty}^{\infty} f(x) \mathbb{1}_{B_j}(x) \, dx \quad (7.8)$$

$$= \int_{-\infty}^{\infty} f(x) \left[\sum_{j=1}^m b_j \mathbb{1}_{B_j}(x) \right] \, dx \quad (7.9)$$

$$= \int_{-\infty}^{\infty} f(x) g(x) dx \quad (7.10)$$

Para $g \geq 0$ mensurável, aproxime g por funções simples não negativas e use o Teorema da Convergência Monótona dos dois lados. □

Utilizando as Proposições 7.28 e 7.29 podemos integrar com respeito a uma distribuição que seja combinação de uma discreta e uma absolutamente contínua.

Teorema 7.30 — Mudança de Variáveis

Dada uma variável aleatória ξ e sua função de distribuição F_ξ , então

$$\mathbf{E}[\xi] = \int_{\mathbb{R}} x \, dF_\xi(x)$$

e de modo mais geral,

$$\mathbf{E}[g(\xi)] = \int_{\mathbb{R}} g(x) \, dF_\xi(x)$$

para toda função mensurável $g : \mathbb{R} \rightarrow \mathbb{R}$, sempre que qualquer um dos lados esteja bem definido.

Demonstração. A medida de Lebesgue–Stieltjes $\mathbf{P}_{F_\xi}$ é a distribuição $\mathbf{P}_\xi$, pois ambas têm função de distribuição F_ξ . Portanto, o Teorema 7.9 dá

$$\mathbf{E}[g(\xi)] = \int_{\mathbb{R}} g \, d\mathbf{P}_\xi = \int_{\mathbb{R}} g(x) \, dF_\xi(x).$$

Tomando $g(x) = x$, obtemos a primeira identidade. □

Exemplo 7.31 — .

Se $\xi \sim \mathcal{N}(0; 1)$ então $\mathbf{E}(\xi) = 0$.

De fato,

$$\mathbf{E}[\xi] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-x^2/2} \, dx = 0,$$

pois o integrando é uma função ímpar e integrável. ◀

Exemplo 7.32 — .

Se $\xi \sim \text{Exp}(\lambda)$ então

$$\mathbf{E}[\xi^n] = \int_0^\infty x^n \lambda e^{-\lambda x} dx = \frac{n!}{\lambda^n},$$

onde a última igualdade segue por n integrações por partes.



7.5. Desigualdades de Markov e Chebyshev

As desigualdades de Markov e Chebyshev convertem informação sobre momentos em controle das caudas de uma distribuição. São estimativas simples e universais: pedem pouco, e por isso também prometem pouco.

Teorema 7.33 — Desigualdade de Markov

Dada uma variável aleatória $\xi \geq 0$, vale que

$$\mathbf{P}(\xi \geq t) \leq \frac{\mathbf{E}[\xi]}{t} \quad \text{para todo } t > 0. \quad (7.11)$$

Em particular, dada uma variável aleatória ξ qualquer

$$\mathbf{P}(|\xi| \geq t) \leq \frac{\mathbf{E}[|\xi|]}{t}.$$

Demonstração. Começamos definindo $\eta := t\mathbf{1}_{\{\xi \geq t\}}$, e observando que $\eta \leq \xi$ pontualmente.

Assim sendo temos

$$\mathbf{E}[\xi] \geq \mathbf{E}[\eta] = \mathbf{E}[t\mathbf{1}_{\{\xi \geq t\}}] = t\mathbf{P}(\xi \geq t),$$

concluindo o resultado. □

A simplicidade da cota tem um preço: ela pode estar longe do valor exato. Aplicar Markov a outras funções da variável permite refiná-la.

Teorema 7.34 — Desigualdade de Chebyshev

Dadas uma variável aleatória $\xi \geq 0$ e uma constante $p > 0$, vale

$$\mathbf{P}(\xi \geq t) \leq \frac{\mathbf{E}[\xi^p]}{t^p} \quad \text{para todo } t > 0. \quad (7.12)$$

Em particular

$$\mathbf{P}(|\xi| \geq t) \leq \frac{\mathbf{E}[|\xi|^p]}{t^p},$$

para uma variável aleatória ξ qualquer.

Demonstração. Basta reduzir o enunciado à desigualdade de Markov: $\mathbf{P}(\xi \geq t) = \mathbf{P}(\xi^p \geq t^p) \leq \frac{\mathbf{E}[\xi^p]}{t^p}$, pela desigualdade de Markov. $\square$

Para que a cota seja informativa, precisamos de $\xi \in \mathcal{L}^p(\mathbf{P})$. Ainda assim, ela evidencia um fato mais forte sobre a cauda.

Se $X \in \mathcal{L}^p(\mathbf{P})$ para todo $p > 0$, então $t^p \mathbf{P}(|X| > t) \rightarrow 0$ para cada $p > 0$: basta aplicar a desigualdade com um expoente $q > p$.

Assim, nessa forma da desigualdade, exigimos um momento de ordem p de ξ , mas obtemos um controle mais forte para a cauda.

Uma formulação ainda mais flexível é obtida substituindo a potência por uma função não decrescente.

Teorema 7.35 — Desigualdade de Markov estendida

Se $\varphi : [0, \infty) \rightarrow [0, \infty)$ é não decrescente, ξ é uma variável aleatória e $t \geq 0$ satisfaz $\varphi(t) > 0$, então

$$\mathbf{P}(|\xi| \geq t) \leq \frac{\mathbf{E}[\varphi(|\xi|)]}{\varphi(t)}.$$

Demonstração. No evento $\{|\xi| \geq t\}$, a monotonicidade dá $\varphi(|\xi|) \geq \varphi(t)$. Assim,

$$\varphi(t)\mathbb{1}_{\{|\xi| \geq t\}} \leq \varphi(|\xi|).$$

Tomando esperanças e dividindo por $\varphi(t) > 0$, obtemos o resultado. $\square$

Outra aplicação interessante da desigualdade de Markov é no estudo de como os valores de uma variável aleatória se dispersam ao redor da média, e como os diversos momentos centrais, e em particular a variância, informam esse comportamento.

O corolário abaixo é geralmente apresentado na literatura como sendo a própria desigualdade de Chebyshev, e trata exatamente deste assunto.

Corolário 7.36 – Variância

Suponha que ξ tenha média μ e variância $0 < \sigma^2 < \infty$. Então

$$\mathbf{P}(|\xi - \mu| \geq t\sigma) \leq \frac{1}{t^2} \text{ para qualquer } t > 0, \quad (7.13)$$

ou ainda

$$\mathbf{P}(|\xi - \mu| \geq t) \leq \frac{\sigma^2}{t^2} \text{ para qualquer } t > 0. \quad (7.14)$$

Demonstração. Use a desigualdade de Chebyshev para a variável aleatória $|\xi - \mu| \geq 0$, com expoente $p = 2$. Como $\mathbf{E}|\xi - \mu|^2 = \sigma^2$,

$$\mathbf{P}(|\xi - \mu| \geq t\sigma) \leq \frac{\sigma^2}{(t\sigma)^2} = \frac{1}{t^2}.$$

$\square$

Exemplo 7.37 – .

Dada uma variável aleatória ξ com variância $0 < \sigma^2 < \infty$ e média $\mu \in \mathbb{R}$, então

$$\mathbf{P}(\mu - \sqrt{2} \cdot \sigma < \xi < \mu + \sqrt{2} \cdot \sigma) = 1 - \mathbf{P}(|\xi - \mu| \geq \sqrt{2} \cdot \sigma) \geq 1 - \frac{1}{2} = \frac{1}{2}.$$

Ou seja, ξ está a uma distância menor que $\sqrt{2} \cdot \sigma$ de sua média, com probabilidade $\geq \frac{1}{2}$.

◁

Exemplo 7.38 — .

Considere o experimento de lançar uma moeda de maneira sequencial e independente. Para cada $n \geq 1$, defina X_n como a indicadora de que observamos cara no n -ésimo lançamento. Suponha que em um lançamento da moeda, a probabilidade de observar cara é $p \in (0,1)$.

Com isso temos que as variáveis $X_1, X_2, \dots$ são independentes com $\mathbf{E}[X_n] = p$ e $\text{Var}[X_n] = p(1-p)$.

Queremos medir a proporção de caras observadas em n lançamentos, e para isso definiremos $S_n = X_1 + \dots + X_n$. Isto é, S_n é o total de caras observadas em n lançamentos.

Segue assim que

$$\mathbf{E}[S_n] = n p,$$

e, como $X_1, \dots, X_n$ são independentes,

$$\text{Var}(S_n) = n p(1-p).$$

Tome agora $\epsilon > 0$ pequeno e note que

$$\mathbf{P}\left(\left|\frac{S_n}{n} - p\right| > \epsilon\right) = \mathbf{P}(|S_n - np| > n\epsilon) \leq \frac{\text{Var}(S_n)}{n^2\epsilon^2} = \frac{p(1-p)}{n\epsilon^2}.$$

Concluimos assim que

$$\mathbf{P}\left(\left|\frac{S_n}{n} - p\right| > \epsilon\right) \rightarrow 0,$$

quando $n \rightarrow \infty$, para todo $\epsilon > 0$.

◁

7.6. Desigualdade Maximal de Kolmogorov

Para somas de variáveis independentes, não basta controlar apenas o valor final: muitas vezes queremos saber se alguma soma parcial se afastou demais da origem. A desigualdade maximal de Kolmogorov fornece esse controle sem pagar um fator adicional pelo número de instantes observados.

Teorema 7.39 —

Seja $(\xi_n)_{n \geq 1}$ uma sequência de variáveis aleatórias independentes, definidas no mesmo espaço de probabilidade, com esperança 0 e variância finita. Seja $S_n = \sum_{k=1}^n \xi_k$. Então para $\lambda > 0$,

$$\mathbf{P}\left(\max_{1 \leq k \leq n} |S_k| \geq \lambda\right) \leq \frac{\text{Var}(S_n)}{\lambda^2}$$

Demonstração.

Comece observando que o evento $\{\max_{1 \leq k \leq n} |S_k| \geq \lambda\}$ é a união disjunta dos eventos

$$A_k = \{|S_k| \geq \lambda, |S_j| < \lambda \text{ para } j < k\}, k = 1, \dots, n.$$

Observe também que, para cada $k = 1, \dots, n$, temos $S_n - S_k = \xi_{k+1} + \dots + \xi_n$, que é independente da σ -álgebra gerada por $\xi_1, \dots, \xi_k$. Em particular, é independente de $S_k \mathbb{1}_{A_k}$.

Além disso, $\mathbf{E}[S_n - S_k] = 0$ e, pela independência,

$$\mathbf{E}[S_k(S_n - S_k) \mathbb{1}_{A_k}] = \mathbf{E}[S_k \mathbb{1}_{A_k}] \mathbf{E}[S_n - S_k] = 0.$$

Além disso, dentro do evento A_k vale que $\lambda^{-2} S_k^2 \geq 1$, de modo que $\mathbb{1}_{A_k} \leq \lambda^{-2} S_k^2 \mathbb{1}_{A_k}$, e

$$\begin{aligned} \mathbf{P}(A_k) &\leq \lambda^{-2} \mathbf{E}[S_k^2 \mathbb{1}_{A_k}] \\ &\leq \lambda^{-2} \mathbf{E}[(S_k^2 + (S_n - S_k)^2) \mathbb{1}_{A_k}] \\ &= \lambda^{-2} \mathbf{E}[(S_k + S_n - S_k)^2 \mathbb{1}_{A_k}], \end{aligned}$$

onde a última igualdade decorre do termo misto nulo calculado acima.

Portanto,

$$\mathbf{P}(A_k) \leq \lambda^{-2} \mathbf{E}[S_n^2 \mathbb{1}_{A_k}].$$

Somando sobre os eventos disjuntos A_k , encontramos

$$\mathbf{P}\left(\max_{1 \leq k \leq n} |S_k| \geq \lambda\right) \leq \lambda^{-2} \mathbf{E}\left[S_n^2 \mathbb{1}_{\{\max_{1 \leq k \leq n} |S_k| \geq \lambda\}}\right] \leq \lambda^{-2} \mathbf{E}[S_n^2].$$

O resultado é concluído quando vemos que $\text{Var}(S_n) = \mathbf{E}[S_n^2] - \mathbf{E}[S_n]^2 = \mathbf{E}[S_n^2]$.

□

7.7. Desigualdade de Jensen, Hölder e Minkowski

Nesta seção apresentaremos três desigualdades fundamentais. A desigualdade de Jensen relaciona funções convexas e esperança; as desigualdades de Hölder e Minkowski controlam produtos e somas em espaços L^p . Antes, recordaremos alguns conceitos de análise real.

Definição 7.40 —

Uma função $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ é **convexa** se para todo $a, b \in \mathbb{R}$ e $\alpha \in [0, 1]$ vale que

$$\varphi(\alpha a + (1 - \alpha)b) \leq \alpha \varphi(a) + (1 - \alpha)\varphi(b).$$

Geometricamente, isso se traduz como o gráfico da função, no intervalo $[a, b]$, estar abaixo da reta secante ao gráfico nos pontos $(a, \varphi(a))$ e $(b, \varphi(b))$.

Funções como $|x|$, x^2 e $e^{\lambda x}$ são exemplos clássicos de funções convexas.

Um resultado importante sobre funções convexas, e que utilizaremos na demonstração do resultado principal, é o seguinte.

Proposição 7.41 —

Se $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ é convexa, então, dado $t_0 \in \mathbb{R}$, o mapa

$$t \mapsto \frac{\varphi(t) - \varphi(t_0)}{t - t_0},$$

definido em $\mathbb{R} \setminus \{t_0\}$, é não decrescente.

Demonstração. Se $t_0 < s < t$, escreva $s = \alpha t + (1 - \alpha)t_0$, com $\alpha = (s - t_0)/(t - t_0)$. Pela convexidade,

$$\frac{\varphi(s) - \varphi(t_0)}{s - t_0} \leq \frac{\varphi(t) - \varphi(t_0)}{t - t_0}.$$

O caso $s < t < t_0$ é obtido escrevendo t como combinação convexa de s e t_0 . Finalmente, se $s < t_0 < t$, escreva t_0 como combinação convexa de s e t ; a desigualdade resultante é

$$\frac{\varphi(s) - \varphi(t_0)}{s - t_0} \leq \frac{\varphi(t) - \varphi(t_0)}{t - t_0}.$$

Isso cobre todas as posições possíveis. □

Proposição 7.42 — Desigualdade de Jensen

Se ξ é integrável e $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ é convexa, então $\mathbf{E}[\varphi(\xi)]$ está bem definida em $(-\infty, \infty]$ e

$$\varphi(\mathbf{E}[\xi]) \leq \mathbf{E}[\varphi(\xi)].$$

Demonstração. Ponha $t_0 = \mathbf{E}[\xi]$. Pela proposição anterior, existe $m \in \mathbb{R}$ entre as inclinações secantes à esquerda e à direita de t_0 . Assim,

$$\varphi(t) \geq \varphi(t_0) + m(t - t_0) \quad \text{para todo } t \in \mathbb{R}.$$

Essa reta de suporte mostra também que a parte negativa de $\varphi(\xi)$ é dominada por uma função integrável; logo sua esperança está bem definida. Tomando esperanças na desigualdade,

$$\mathbf{E}[\varphi(\xi)] \geq \varphi(t_0) + m(\mathbf{E}[\xi] - t_0) = \varphi(\mathbf{E}[\xi]).$$



Terminamos com uma aplicação bastante importante da desigualdade de Jensen.

Exemplo 7.43 — .

Seja X uma variável aleatória e sejam $0 < p < q$. Vamos mostrar que $X \in \mathcal{L}^q(\mathbf{P})$ implica $X \in \mathcal{L}^p(\mathbf{P})$.

Como $q/p > 1$, a função $\varphi(x) = x^{q/p}$ é convexa em $[0, \infty)$. Não podemos aplicar Jensen diretamente a $|X|^p$, pois sua integrabilidade é precisamente o que queremos provar. Para $M > 0$, defina

$$Y_M = |X|^p \wedge M.$$

A variável Y_M é limitada e, portanto, integrável. Pela desigualdade de Jensen,

$$\mathbf{E}[Y_M]^{q/p} \leq \mathbf{E}[Y_M^{q/p}] \leq \mathbf{E}[|X|^q].$$

Portanto,

$$\mathbf{E}[Y_M] \leq \mathbf{E}[|X|^q]^{p/q}.$$

Como $Y_M \uparrow |X|^p$ quando $M \rightarrow \infty$, o Teorema da Convergência Monótona fornece

$$\mathbf{E}[|X|^p] \leq \mathbf{E}[|X|^q]^{p/q} < \infty.$$

Logo $X \in \mathcal{L}^p(\mathbf{P})$.



Proposição 7.44 — Desigualdade de Hölder

Se $p, q \in [1, \infty]$ são expoentes conjugados, $1/p + 1/q = 1$, $\xi \in \mathcal{L}^p(\mathbf{P})$ e $\eta \in \mathcal{L}^q(\mathbf{P})$, então

$$\mathbf{E}[|\xi\eta|] \leq \|\xi\|_p \|\eta\|_q.$$

Demonstração. Nos casos $(p, q) = (1, \infty)$ e $(\infty, 1)$, a desigualdade segue diretamente de $|\xi\eta| \leq \|\eta\|_\infty|\xi|$ e de sua versão simétrica. Suponha $1 < p, q < \infty$. Se uma das normas é zero, o resultado é imediato. Caso contrário, ponha $U = |\xi|/\|\xi\|_p$ e $V = |\eta|/\|\eta\|_q$. Pela desigualdade de Young, $uv \leq u^p/p + v^q/q$, temos

$$\mathbf{E}[UV] \leq \frac{1}{p}\mathbf{E}[U^p] + \frac{1}{q}\mathbf{E}[V^q] = \frac{1}{p} + \frac{1}{q} = 1.$$

Multiplicando pelas duas normas, obtemos a conclusão. □

Proposição 7.45 — Desigualdade de Minkowski

Se $f, g \in \mathcal{L}^p(\mathbf{P})$ e $1 \leq p \leq \infty$, então

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p.$$

Demonstração. Para $p = 1$, integre $|f + g| \leq |f| + |g|$. Para $p = \infty$, tome o supremo essencial nessa mesma desigualdade. Suponha $1 < p < \infty$. A estimativa $|f + g|^p \leq 2^{p-1}(|f|^p + |g|^p)$ mostra primeiro que $f + g \in \mathcal{L}^p$. Se $\|f + g\|_p = 0$, não há o que provar. Caso contrário, Hölder, com conjugado $q = p/(p - 1)$, fornece

$$\begin{aligned} \|f + g\|_p^p &\leq \int_{\Omega} |f| |f + g|^{p-1} d\mathbf{P} + \int_{\Omega} |g| |f + g|^{p-1} d\mathbf{P} \\ &\leq (\|f\|_p + \|g\|_p) \|f + g\|_q^{p-1} \\ &= (\|f\|_p + \|g\|_p) \|f + g\|_p^{p-1}. \end{aligned}$$

Dividindo pelo último fator, obtemos a desigualdade. □

7.8. Funções Geradoras de Probabilidade

Considere uma variável aleatória ξ , tomando valores nos naturais. Seja $p_r = \mathbf{P}(\xi = r)$.

Definição 7.46 – Função Geradora de Probabilidade

A **função geradora de probabilidade** de ξ é definida como

$$p(z) = \mathbf{E}[z^\xi] = \sum_{r=0}^{\infty} \mathbf{P}(\xi = r)z^r = p_0 + p_1z + p_2z^2 + \cdots, \quad |z| \leq 1.$$

Observamos que função geradora de probabilidade é uma série de potência e converge absolutamente para $|z| \leq 1$, pois

$$|p(z)| \leq \sum_r p_r |z|^r \leq \sum_r p_r = 1.$$

A função geradora reúne as probabilidades nos coeficientes de uma série. Sua utilidade aparece quando operações sobre variáveis se traduzem em operações algébricas sobre essas funções.

Exemplo 7.47 – .

Considere um dado justo. Então $p_r = 1/6$ para $r = 1, \dots, 6$. assim

$$p(z) = \mathbf{E}[z^\xi] = \frac{1}{6}(z + z^2 + \cdots + z^6) = \frac{1}{6}z \left(\frac{1 - z^6}{1 - z} \right).$$

◀

Exemplo 7.48 – .

Neste exemplo, calculamos a função geradora de probabilidade para a distribuição de Poisson de parâmetro α . Com base na definição, temos:

$$p(z) = \sum_{j=0}^{\infty} \frac{e^{-\alpha} \alpha^j}{j!} z^j \tag{7.15}$$

$$= \sum_{j=0}^{\infty} \frac{e^{-\alpha} (\alpha z)^j}{j!} \tag{7.16}$$

$$= \frac{e^{-\alpha}}{e^{-\alpha z}} \sum_{j=0}^{\infty} \frac{e^{-\alpha z} (\alpha z)^j}{j!} \tag{7.17}$$

$$= e^{\alpha(z-1)} \tag{7.18}$$


Teorema 7.49 —

A distribuição de ξ é determinada exclusivamente pela sua função geradora de probabilidade.

Demonstração. No interior do intervalo de convergência, podemos diferenciar a série termo a termo. Assim,

$$\left. \frac{d^n}{dz^n} p(z) \right|_{z=0} = n! p_n.$$

Logo $p_n = p^{(n)}(0)/n!$ para todo $n \geq 0$, e recuperamos toda a distribuição. □

Teorema 7.50 — Lema de Abel

$$\mathbf{E}[\xi] = \lim_{z \uparrow 1} p'(z),$$

admitindo-se o valor $+\infty$. Se p' possui extensão contínua a 1, então $\mathbf{E}[\xi] = p'(1)$.

Demonstração. Para $0 < z < 1$, a diferenciação termo a termo dá

$$p'(z) = \sum_{r=1}^{\infty} r p_r z^{r-1}.$$

Quando $z \uparrow 1$, cada parcela cresce para $r p_r$. Pelo Teorema da Convergência Monótona aplicado à medida de contagem em $\mathbb{N}$,

$$\lim_{z \uparrow 1} p'(z) = \sum_{r=1}^{\infty} r p_r = \mathbf{E}[\xi].$$

A última afirmação é imediata da continuidade em 1. □

Teorema 7.51 —

$$\mathbf{E}[\xi(\xi - 1)] = \lim_{z \uparrow 1} p''(z),$$

admitindo-se o valor $+\infty$.

Demonstração. Para $0 < z < 1$,

$$p''(z) = \sum_{r=2}^{\infty} r(r-1)p_r z^{r-2}.$$

As parcelas são não negativas e crescem quando $z \uparrow 1$. A Convergência Monótona fornece

$$\lim_{z \uparrow 1} p''(z) = \sum_{r=2}^{\infty} r(r-1)p_r = \mathbf{E}[\xi(\xi - 1)].$$

□

Exemplo 7.52 — .

Considere a distribuição de Poisson. Então

$$p_r = \mathbf{P}(\xi = r) = \frac{1}{r!} \lambda^r e^{-\lambda}.$$

Então

$$p(z) = \mathbf{E}[z^\xi] = \sum_0^{\infty} z^r \frac{1}{r!} \lambda^r e^{-\lambda} = e^{\lambda z} e^{-\lambda} = e^{\lambda(z-1)}.$$

Temos

$$\mathbf{E}[\xi] = \left. \frac{d}{dz} e^{\lambda(z-1)} \right|_{z=1} = \lambda,$$

e

$$\mathbf{E}[\xi(\xi - 1)] = \left. \frac{d^2}{dz^2} e^{\lambda(z-1)} \right|_{z=1} = \lambda^2$$

assim

$$\text{Var}(\xi) = \mathbf{E}[\xi^2] - \mathbf{E}[\xi]^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

◀

Teorema 7.53 —

Suponha que $\xi_1, \xi_2, \dots, \xi_n$ sejam variáveis aleatórias independentes com funções geradoras de probabilidade $p_1, p_2, \dots, p_n$. Então, a função geradora de probabilidade de $\xi_1 + \xi_2 + \dots + \xi_n$ é $p_1(z)p_2(z) \cdots p_n(z)$.

Demonstração.

$$\mathbf{E}[z^{\xi_1 + \dots + \xi_n}] = \mathbf{E}[z^{\xi_1} \cdots z^{\xi_n}] = \mathbf{E}[z^{\xi_1}] \cdots \mathbf{E}[z^{\xi_n}] = p_1(z) \cdots p_n(z).$$

□

Exemplo 7.54 — .

Se ξ e η forem variáveis aleatórias de Poisson independentes com parâmetros λ, μ respectivamente, então

$$\mathbf{E}[t^{\xi + \eta}] = \mathbf{E}[t^{\xi}] \mathbf{E}[t^{\eta}] = e^{\lambda(t-1)} e^{\mu(t-1)} = e^{(\lambda + \mu)(t-1)}$$

Portanto, $\xi + \eta \sim \mathbf{P}(\lambda + \mu)$.

Também podemos fazê-lo diretamente:

$$\mathbf{P}(\xi + \eta = r) = \sum_{i=0}^r \mathbf{P}(\xi = i, \eta = r - i) = \sum_{i=0}^r \mathbf{P}(\xi = i) \mathbf{P}(\eta = r - i),$$

mas é muito mais complicado.

◀

Somas de um número aleatório de termos

As funções geradoras também permitem estudar somas cujo número de parcelas é aleatório. Numa seguradora, por exemplo, tanto o número de pedidos de indenização quanto o valor de cada pedido podem variar. A soma reúne as duas fontes de aleatoriedade.

Exemplo 7.55 — .

Sejam $\xi_1, \xi_2, \dots$ variáveis aleatórias independentes, identicamente distribuídas e com valores em $\mathbb{N}$, com função geradora $p(z) = \mathbf{E}[z^{\xi_1}]$. Seja N , também com valores em $\mathbb{N}$, independente de todas as ξ_i e com função geradora h . Para $S = \xi_1 + \dots + \xi_N$ (com $S = 0$ quando $N = 0$), temos

$$\begin{aligned}
 \mathbf{E}[z^S] &= \mathbf{E}[z^{\xi_1 + \dots + \xi_N}] \\
 &= \mathbf{E}_N[\underbrace{\mathbf{E}_{\xi_i}[z^{\xi_1 + \dots + \xi_N} \mid N]}_{\text{assumindo fixo } N}] \\
 &= \sum_{n=0}^{\infty} \mathbf{P}(N = n) \mathbf{E}[z^{\xi_1 + \xi_2 + \dots + \xi_n}] \\
 &= \sum_{n=0}^{\infty} \mathbf{P}(N = n) \mathbf{E}[z^{\xi_1}] \mathbf{E}[z^{\xi_2}] \dots \mathbf{E}[z^{\xi_n}] \\
 &= \sum_{n=0}^{\infty} \mathbf{P}(N = n) (\mathbf{E}[z^{\xi_1}])^n \\
 &= \sum_{n=0}^{\infty} \mathbf{P}(N = n) p(z)^n \\
 &= h(p(z))
 \end{aligned}$$

como $h(x) = \sum_{n=0}^{\infty} \mathbf{P}(N = n)x^n$.

assim

$$\begin{aligned}
 \mathbf{E}[S] &= \left. \frac{d}{dz} h(p(z)) \right|_{z=1} \\
 &= h'(p(1))p'(1) \\
 &= \mathbf{E}[N] \mathbf{E}[\xi_1]
 \end{aligned}$$

Para calcular a variância, use o fato de que

$$\mathbf{E}[S(S-1)] = \left. \frac{d^2}{dz^2} h(p(z)) \right|_{z=1}.$$

Se N e ξ_1 possuem segundos momentos finitos, a diferenciação em $z = 1$ fornece

$$\text{Var}(S) = \mathbf{E}[N] \text{Var}(\xi_1) + \mathbf{E}[\xi_1]^2 \text{Var}(N).$$

7.9. Processos de Ramificação

Originalmente, processos de ramificações foram considerados por Galton e Watson nos anos de 1870 quando estes procuravam um modelo quantitativo para o fenômeno do desaparecimento de sobrenomes, mesmo no cenário de uma população crescente. O modelo é construído sob a suposição de que cada homem em uma dada família tem a probabilidade p_k de ter k filhos e que o sobrenome de família é passado aos filhos, então desejamos determinar a probabilidade que após n gerações um indivíduo não tenha descendentes homens.

Vamos fazer perguntas como o número esperado de indivíduos em uma geração específica e a probabilidade de extinção.

Considere $\xi_0, \xi_1, \dots$, onde ξ_n é o número de indivíduos na geração n -ésima. Assumimos o seguinte:

- 1 $\xi_0 = 1$
- 2 Cada indivíduo vive por uma unidade de tempo e produz k descendentes com probabilidade p_k .
- 3 Os números de descendentes de indivíduos distintos e de gerações distintas são independentes. Assim,

$$\xi_{n+1} = \eta_1^n + \eta_2^n + \dots + \eta_{\xi_n}^n,$$

onde as variáveis η_i^n são independentes, identicamente distribuídas como ξ_1 e independentes da história até a geração n .

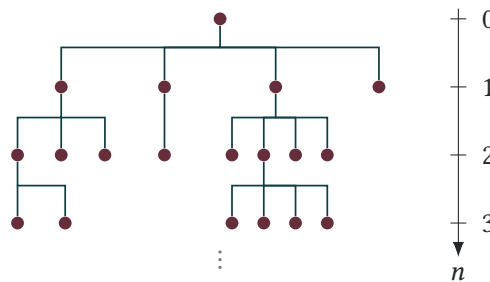


Figura 7.1.: Realização de um processo de ramificação; esquema baseado em [14].

A função geradora transforma a recorrência do processo de ramificação numa relação de composição. Seja $F(z)$ a função geradora de probabilidade de η_i^n . Então

$$F(z) = \mathbf{E}[z^{\eta_i^n}] = \mathbf{E}[z^{\xi_1}] = \sum_{k=0}^{\infty} p_k z^k.$$

Definimos

$$F_n(z) = \mathbf{E}[z^{\xi_n}], \quad F_0(z) = z.$$

O principal teorema dos processos de ramificação que provaremos nesta seção é

Teorema 7.56 —

Para $0 \leq z \leq 1$,

$$F_{n+1}(z) = F_n(F(z)) = F(F_n(z)) = F^{\circ(n+1)}(z).$$

Demonstração. Condicionando no tamanho da geração n e usando a independência das descendências,

$$\begin{aligned} F_{n+1}(z) &= \sum_{k=0}^{\infty} \mathbf{P}(\xi_n = k) \mathbf{E}[z^{\eta_1^n + \dots + \eta_k^n} \mid \xi_n = k] \\ &= \sum_{k=0}^{\infty} \mathbf{P}(\xi_n = k) F(z)^k = F_n(F(z)). \end{aligned}$$

Como F_0 é a identidade, a recorrência mostra por indução que $F_n = F^{\circ n}$. Portanto, $F_n \circ F = F^{\circ(n+1)} = F \circ F_n$. □

Teorema 7.57 —

Suponha que

$$\mathbf{E}[\xi_1] = \sum k p_k = \mu$$

e

$$\text{Var}(\xi_1) = \sum_{k=0}^{\infty} (k - \mu)^2 p_k = \sigma^2 < \infty.$$

Então

$$\mathbf{E}[\xi_n] = \mu^n, \quad \text{Var}(\xi_n) = \sigma^2 \mu^{n-1} (1 + \mu + \dots + \mu^{n-1}).$$

Demonstração. Condicionamente a $\xi_{n-1} = k$, a variável ξ_n é soma de k cópias independentes de ξ_1 . Logo sua média condicional é $k\mu$ e sua variância condicional é $k\sigma^2$. Em particular,

$$\mathbf{E}[\xi_n] = \mu \mathbf{E}[\xi_{n-1}],$$

e a condição inicial $\xi_0 = 1$ dá $\mathbf{E}[\xi_n] = \mu^n$.

Usando $\mathbf{E}[\xi_n^2 \mid \xi_{n-1} = k] = k\sigma^2 + k^2\mu^2$ e somando sobre k , obtemos

$$\mathbf{E}[\xi_n^2] = \sigma^2 \mathbf{E}[\xi_{n-1}] + \mu^2 \mathbf{E}[\xi_{n-1}^2].$$

Portanto, escrevendo $V_n = \text{Var}(\xi_n)$,

$$V_n = \mu^2 V_{n-1} + \sigma^2 \mu^{n-1}, \quad V_0 = 0.$$

Uma indução nessa recorrência fornece

$$V_n = \sigma^2 \mu^{n-1} (1 + \mu + \dots + \mu^{n-1}).$$

□

Probabilidade de Extinção

Considere $A_n = \{\xi_n = 0\}$, o evento de que a população já esteja extinta na n -ésima geração. Seja q a probabilidade de extinção eventual e

$$A = \bigcup_{n=1}^{\infty} A_n = \{\text{a extinção ocorre eventualmente}\}.$$

Como $A_1 \subseteq A_2 \subseteq A_3 \subseteq \dots$, sabemos que

$$q = \mathbf{P}(A) = \lim_{n \rightarrow \infty} \mathbf{P}(A_n) = \lim_{n \rightarrow \infty} \mathbf{P}(\xi_n = 0).$$

Como

$$\mathbf{P}(\xi_n = 0) = F_n(0),$$

e F é contínua em $[0,1]$, temos

$$F(q) = F\left(\lim_{n \rightarrow \infty} F_n(0)\right) = \lim_{n \rightarrow \infty} F(F_n(0)) = \lim_{n \rightarrow \infty} F_{n+1}(0) = q.$$

assim

$$F(q) = q.$$

Alternativamente, usando a lei da probabilidade total

$$q = \sum_k \mathbf{P}(\xi_1 = k) \mathbf{P}(\text{extinção} \mid \xi_1 = k) = \sum_k p_k q^k = F(q),$$

onde a segunda igualdade vem do fato de que, para que toda a população se extinga, cada população individual deve se extinguir.

Isso significa que para encontrar a probabilidade de extinção, precisamos encontrar um ponto fixo de F .

Teorema 7.58 —

A probabilidade de extinção q é a menor solução em $[0,1]$ da equação $F(s) = s$. Escreva $\mu = \mathbf{E}[\xi_1]$.

1 Se $\mu \leq 1$ e $\mathbf{P}(\xi_1 = 1) < 1$, então $q = 1$.

2 Se $\mu > 1$, então $q < 1$.

3 No caso degenerado $\mathbf{P}(\xi_1 = 1) = 1$, temos $q = 0$.

Demonstração. Já vimos que $F(q) = q$. Se $\alpha \in [0,1]$ é qualquer outro ponto fixo, então $F_0(0) = 0 \leq \alpha$ e, pela monotonicidade de F , $F_{n+1}(0) = F(F_n(0)) \leq F(\alpha) = \alpha$. Portanto,

$$q = \lim_{n \rightarrow \infty} F_n(0) \leq \alpha,$$

e q é o menor ponto fixo.

A função $G(s) = F(s) - s$ é convexa, $G(1) = 0$ e $F'(1-) = \mu$. Se $\mu \leq 1$, a convexidade dá, para $s < 1$,

$$\frac{F(1) - F(s)}{1 - s} \leq F'(1-) = \mu \leq 1,$$

de modo que $F(s) \geq s$. Se houvesse igualdade para algum $s < 1$, a convexidade e a igualdade das inclinações forçariam $F(t) = t$ em $[s, 1]$; pela identidade de séries de potências, isso significaria $F(t) = t$ em todo $(-1, 1)$, isto é, $\mathbf{P}(\xi_1 = 1) = 1$. Fora desse caso degenerado, não há ponto fixo abaixo de 1, e portanto $q = 1$.

Se $\mu > 1$, então $G'(1-) > 0$ e, conseqüentemente, $G(s) < 0$ para $s < 1$ suficientemente próximo de 1. Como $G(0) = F(0) \geq 0$, a continuidade fornece um ponto fixo em $[0, 1]$; logo o menor deles satisfaz $q < 1$. Finalmente, no caso degenerado $\mathbf{P}(\xi_1 = 1) = 1$, temos $F(s) = s$ para todo s ; cada indivíduo tem exatamente um descendente e a população nunca se extingue, de modo que $q = 0$. $\square$

7.10. Funções Geradoras de Momento

Definição 7.59 —

A **função geradora de momentos** de ξ é

$$M_\xi(t) = \mathbf{E}[e^{t\xi}],$$

desde que essa esperança seja finita para todo t em algum intervalo $(-h, h)$, com $h > 0$.

Para variáveis discretas ou absolutamente contínuas temos que

$$M_\xi(t) = \mathbf{E}[e^{t\xi}] = \begin{cases} \sum_i e^{tx_i} p(x_i) & \text{se } \xi \text{ é discreta,} \\ \int_{\mathbb{R}} e^{tx} f(x) dx & \text{se } \xi \text{ é absolutamente contínua.} \end{cases}$$

Exemplo 7.60 — .

A **distribuição degenerada de probabilidade** é a distribuição associada a uma variável aleatória discreta exibindo certeza do resultado. Se ξ for um variável aleatório degenerada, então $\xi = c$ com probabilidade 1. A função geradora de momento da variável aleatória degenerada é particularmente simples:

$$\sum_{x_i=c} e^{x_i t} = e^{ct}.$$

**Teorema 7.61 —**

Se M_ξ é finita em $(-h, h)$, então é infinitamente diferenciável nesse intervalo e, para $n \geq 1$ e $|t| < h$,

$$M_\xi^{(n)}(t) = \frac{d^n}{dt^n} \mathbb{E}(e^{t\xi}) = \mathbb{E}\left(\frac{d^n}{dt^n} e^{t\xi}\right) = \mathbb{E}(\xi^n e^{t\xi}).$$

Demonstração. Fixe t com $|t| < h$ e escolha $\delta > 0$ tal que $|t| + 2\delta < h$. Existe $C_{n,\delta}$ tal que $|x|^n \leq C_{n,\delta} e^{\delta|x|}$. Para s numa vizinhança suficientemente pequena de t ,

$$|x|^n e^{sx} \leq C_{n,\delta} (e^{(t+2\delta)x} + e^{(t-2\delta)x}).$$

O lado direito é integrável pela hipótese sobre M_ξ . O teorema de diferenciação sob o sinal de integral, obtido da Convergência Dominada, pode então ser aplicado sucessivamente e fornece a fórmula. □

Teorema 7.62 —

Se ξ e η são independentes e suas funções geradoras de momentos são finitas num intervalo comum em torno de 0, então

$$M_{\xi+\eta}(t) = M_\xi(t)M_\eta(t)$$

nesse intervalo.

Demonstração.

$$\begin{aligned} M_{\xi+\eta}(t) &= \mathbb{E}[e^{t(\xi+\eta)}] \\ &= \mathbb{E}[e^{t\xi} e^{t\eta}] \\ &= \mathbb{E}[e^{t\xi}] \mathbb{E}[e^{t\eta}] \\ &= M_\xi(t) M_\eta(t). \end{aligned}$$



Teorema 7.63 —

A função geradora de momento determina de forma única a distribuição de probabilidade.

Demonstração. Suponha que duas distribuições μ e ν tenham a mesma função geradora num intervalo $(-h, h)$. Para $z \in \mathbb{C}$ com $|\operatorname{Re} z| < h$, as funções

$$L_\mu(z) = \int_{\mathbb{R}} e^{zx} d\mu(x), \quad L_\nu(z) = \int_{\mathbb{R}} e^{zx} d\nu(x)$$

são analíticas. A justificativa é a mesma dominação exponencial usada no teorema anterior. Como coincidem no intervalo real $(-h, h)$, o teorema da identidade para funções analíticas mostra que coincidem em toda a faixa $|\operatorname{Re} z| < h$. Em particular, $L_\mu(it) = L_\nu(it)$ para todo $t \in \mathbb{R}$: as funções características são iguais. Pelo Teorema da Unicidade de Fourier (Teorema 12.13), $\mu = \nu$. □

Teorema 7.64 — FGM da distribuição normal

Se $\zeta \sim \mathcal{N}(\mu, \sigma^2)$, então

$$M_\zeta(t) = \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right).$$

Demonstração.

$$\begin{aligned} M_\zeta(t) &= \mathbb{E}[e^{t\zeta}] \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{tx} e^{-(x-\mu)^2/(2\sigma^2)} dx \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \exp\left(\frac{-(x^2 - 2\mu x + \mu^2 - 2\sigma^2 tx)}{2\sigma^2}\right) dx \end{aligned}$$

Completando o quadrado:

$$\begin{aligned} x^2 - 2\mu x + \mu^2 - 2\sigma^2 tx &= x^2 - 2(\mu + \sigma^2 t)x + \mu^2 \\ &= (x - (\mu + \sigma^2 t))^2 - (\mu + \sigma^2 t)^2 + \mu^2 \\ &= (x - (\mu + \sigma^2 t))^2 - \sigma^4 t^2 - 2\mu\sigma^2 t. \end{aligned}$$

Então voltando ao cálculo da função geradora de momento

$$\begin{aligned}
 M_{\zeta}(t) &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \exp\left(\frac{-((x - (\mu + \sigma^2 t))^2 - \sigma^4 t^2 - 2\mu\sigma^2 t)}{2\sigma^2}\right) dx \\
 &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{\sigma^4 t^2 + 2\mu\sigma^2 t}{2\sigma^2}\right) \int_{-\infty}^{\infty} \exp\left(\frac{-(x - (\mu + \sigma^2 t))^2}{2\sigma^2}\right) dx \\
 &= \exp\left(\frac{\sigma^4 t^2 + 2\mu\sigma^2 t}{2\sigma^2}\right) \\
 &= \exp(\mu t + \sigma^2 t^2 / 2)
 \end{aligned}$$

□

Teorema 7.65 —

Se $\zeta_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$, e $\zeta_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$ e ζ_1 e ζ_2 são independentes, então $\zeta_1 + \zeta_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$. Ou seja, a soma de variáveis aleatórias normais independentes é uma variável aleatória normal cuja média é a soma dos médias e cuja variância é a soma das variâncias.

Demonstração. Calculamos a função geradora de momento da soma usando o teorema sobre somas de variáveis aleatórias independentes.

$$\begin{aligned}
 M_{\zeta_1 + \zeta_2}(t) &= M_{\zeta_1}(t) M_{\zeta_2}(t) \\
 &= \exp(\mu_1 t + \sigma_1^2 t^2 / 2) \exp(\mu_2 t + \sigma_2^2 t^2 / 2) \\
 &= \exp((\mu_1 + \mu_2)t + (\sigma_1^2 + \sigma_2^2)t^2 / 2)
 \end{aligned}$$

Essa é a função geradora de momentos de $\mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$; pela unicidade, as distribuições coincidem.

□

7.11. Entropia

Informação

Gostaríamos de desenvolver uma medida da informação que obtemos de observar a ocorrência de um evento de probabilidade p . Nossa primeira redução será ignorar quaisquer características específicas do evento e apenas observar se ou não aconteceu. Assim, pensamos no evento como a observação de um símbolo cuja probabilidade de ocorrência é p .

A abordagem que vamos adotar é axiomática: Queremos que a medida de informação $I(p)$ tenha as seguintes propriedades

Definição 7.66 – Informação

Uma medida de informação é uma função $I : (0,1] \rightarrow [0,\infty)$ com as seguintes propriedades:

- 1 A informação é uma quantidade não negativa: $I(p) \geq 0$.
- 2 Se um evento tiver probabilidade 1, não obtemos informações da ocorrência do evento: $I(1) = 0$.
- 3 Se ocorrerem dois eventos independentes (cuja probabilidade conjunta é o produto de suas probabilidades individuais), então a informação que obtemos ao observar os eventos é a soma das duas informações: $I(p_1 \cdot p_2) = I(p_1) + I(p_2)$. (Esta é a propriedade fundamental)
- 4 A medida de informação depende continuamente da probabilidade.

Essas propriedades determinam a forma de I . De fato, defina $J(x) = I(e^{-x})$ para $x \geq 0$. Então J é contínua, não negativa e

$$J(x + y) = I(e^{-x}e^{-y}) = J(x) + J(y).$$

A equação aditiva contínua em $[0,\infty)$ implica $J(x) = cx$ para alguma constante

$c \geq 0$. Logo

$$I(p) = -c \log p.$$

Se a informação não for identicamente nula, então $c > 0$ e podemos escolher uma base $b > 1$ de modo que

$$I(p) = \log_b(1/p) = -\log_b p.$$

A escolha da base determina apenas a unidade de informação.

Normalmente, pensamos em termos de $\log_2(p)$.

Exemplo 7.67 — .

Por exemplo, lançar uma moeda justa uma vez nos dará eventos H e T com probabilidade $1/2$ e, portanto, um único lançamento de uma moeda nos dá $-\log_2(1/2) = 1$ bit de informação.

Lançando uma moeda justa n vezes temos $-\log_2((1/2)^n) = \log_2(2^n) = n \cdot \log_2(2) = n$ bits de informação.

Assim, obtemos que n lançamentos de uma moeda justa nos fornece n bits de informação, e leva n dígitos binários para especificar. Que estas duas medidas são as mesmas nos assegura que fizemos uma boa escolha de definição de informação.

◀

Entropia

Suponha agora que temos n símbolos $\{a_1, a_2, \dots, a_n\}$ e que alguma fonte está nos fornecendo um fluxo desses símbolos. Suponha ainda que a fonte emite os símbolos com probabilidades $\{p_1, p_2, \dots, p_n\}$, respectivamente. Por enquanto, também assumiremos que os símbolos são emitidos de forma independente

Qual é a quantidade média de informação que obtemos de cada símbolo que vemos no fluxo?

O que realmente queremos aqui é uma média ponderada. Se observarmos o símbolo a_i , nós obteremos $\log(1/p_i)$ informação dessa observação particular. A longo

prazo, digamos N de observações, veremos aproximadamente $N \cdot p_i$ ocorrências do símbolo a_i . Assim, após N observações obteremos a informação total I de

$$I = \sum_{i=1}^n (N \cdot p_i) \cdot \log(1/p_i).$$

Mas, então, a média de informações que obtemos por símbolo observado será

$$\begin{aligned} I/N &= (1/N) \sum_{i=1}^n (N \cdot p_i) \cdot \log(1/p_i) \\ &= \sum_{i=1}^n p_i \cdot \log(1/p_i) \end{aligned}$$

Adotamos a convenção de que $0 \cdot \log(1/0)$ vale 0, correspondente ao limite da expressão quando a probabilidade tende a zero.

Isso nos leva a seguinte definição.

Definição 7.68 —

Seja uma distribuição de probabilidade $P = \{p_1, p_2, \dots, p_n\}$. Definimos **entropia** da distribuição P por:

$$H(P) = \sum_{i=1}^n p_i \cdot \log(1/p_i).$$

Para uma variável discreta X com $\mathbf{P}(X = a_i) = p_i$, escreva $p(a_i) = p_i$. Então essa definição pode ser escrita como

$$H(X) = \mathbf{E} \left[\log \frac{1}{p(X)} \right].$$

Há uma noção análoga para uma distribuição absolutamente contínua com densidade f , chamada **entropia diferencial**:

$$h(f) = \int_{\mathbb{R}} f(x) \log \frac{1}{f(x)} \, dx,$$

quando a integral está bem definida. Apesar da semelhança formal, a entropia diferencial pode ser negativa e muda sob transformações de escala; por isso não deve ser confundida com a entropia discreta.

Em outras palavras, a entropia de uma distribuição de probabilidade é o valor esperado da informação da distribuição.

Teorema 7.69 — Desigualdade de Gibbs

Considere as distribuições de probabilidade, $P = \{p_1, p_2, \dots, p_n\}$ e $Q = \{q_1, q_2, \dots, q_n\}$, onde $p_i, q_i \geq 0$ e $\sum_i p_i = \sum_i q_i = 1$. Então

$$-\sum p_i \log q_i \geq -\sum p_i \log p_i$$

com as convenções $0 \log 0 = 0$ e $-p_i \log q_i = +\infty$ se $p_i > 0 = q_i$. A igualdade ocorre se, e somente se, $p_i = q_i$ para todo i .

Demonstração. Se existe i com $p_i > 0 = q_i$, o lado esquerdo é infinito e a desigualdade é imediata. Caso contrário, omitimos os índices com $p_i = 0$ e usamos $\log x \leq x - 1$:

$$\begin{aligned} \sum_{i: p_i > 0} p_i \log \left(\frac{q_i}{p_i} \right) &\leq \sum_{i: p_i > 0} p_i \left(\frac{q_i}{p_i} - 1 \right) \\ &= \sum_{i: p_i > 0} q_i - 1 \leq 0. \end{aligned}$$

Isso equivale à desigualdade anunciada. A igualdade em $\log x \leq x - 1$ ocorre apenas quando $x = 1$; portanto, ela exige $q_i = p_i$ para todo índice com $p_i > 0$. Como ambas as famílias somam 1, também não pode haver massa de Q fora desse suporte, e $p_i = q_i$ para todo i . □

A desigualdade de Gibbs permite identificar a distribuição de maior entropia

entre n resultados possíveis. Para uma distribuição $P = \{p_1, p_2, \dots, p_n\}$, temos

$$\begin{aligned}
 H(P) - \log(n) &= \sum_{i=1}^n p_i \log(1/p_i) - \log(n) \\
 &= \sum_{i=1}^n p_i \log(1/p_i) - \log(n) \sum_{i=1}^n p_i \\
 &= \sum_{i=1}^n p_i \log(1/p_i) - \sum_{i=1}^n p_i \log(n) \\
 &= \sum_{i=1}^n p_i (\log(1/p_i) - \log(n)) \\
 &= \sum_{i=1}^n p_i (\log(1/p_i) + \log(1/n)) \\
 &= \sum_{i=1}^n p_i \log\left(\frac{1/n}{p_i}\right) \\
 &\leq 0,
 \end{aligned}$$

com igualdade somente quando $p_i = \frac{1}{n}$ para todos i . A distribuição uniforme maximiza, portanto, a informação média por observação: sem resultados favorecidos, a surpresa fica mais bem distribuída.

EXERCÍCIOS

Ex. 7.1 — Seja X uma variável aleatória discreta com valores distintos $x_1, x_2, \dots$. Suponha que

$$\sum_{k=1}^{\infty} |x_k| \mathbf{P}(X = x_k) < \infty.$$

Prove, a partir da definição da integral de Lebesgue, que

$$\mathbf{E}[X] = \sum_{k=1}^{\infty} x_k \mathbf{P}(X = x_k). \quad (7.19)$$

Ex. 7.2 — Construa variáveis aleatórias $X, Y : [0,1] \rightarrow \mathbb{R}$, no espaço com medida de Lebesgue, tais que

$$\mathbf{P}(X > Y) > \frac{1}{2}, \quad \mathbf{E}[X] < \mathbf{E}[Y].$$

Calcule explicitamente as duas probabilidades e as duas esperanças.

Ex. 7.3 — Prove que, para toda função boreliana $g : \mathbb{R} \rightarrow \mathbb{R}$,

$$\xi \stackrel{d}{=} \eta \implies g(\xi) \stackrel{d}{=} g(\eta).$$

Compare com os casos concretos $g(x) = x^+$ e $g(x) = x^-$ estudados no Capítulo 6.

Ex. 7.4 — Seja $X_1, X_2, \dots$ uma sequência de variáveis aleatórias mutuamente independentes. Prove que as σ -álgebras

$$\sigma(X_1, X_3, X_5, \dots) \quad \text{e} \quad \sigma(X_2, X_4, X_6, \dots)$$

são independentes.

Ex. 7.5 — Sejam $X_1, X_2, \dots$ variáveis aleatórias i.i.d., com média μ e variância $\sigma^2 < \infty$. Seja N uma variável aleatória com valores em $\mathbb{N}_0$, independente da sequência, com

$$\mathbf{E}[N] = m, \quad \text{Var}(N) = v < \infty.$$

Defina

$$S = \sum_{i=1}^N X_i = \sum_{i=1}^{\infty} X_i \mathbf{1}_{\{N \geq i\}},$$

com a convenção de que a soma vazia vale zero. Calcule $\mathbf{E}[S]$ e $\text{Var}(S)$.

Ex. 7.6 — Prove a desigualdade de Cauchy–Schwarz: se $X, Y \in L^2(\mathbf{P})$, então

$$|\mathbf{E}[XY]| \leq \sqrt{\mathbf{E}[X^2]\mathbf{E}[Y^2]}.$$

Prove também que há igualdade se, e somente se, X e Y são linearmente dependentes como elementos de $L^2(\mathbf{P})$.

Ex. 7.7 — Sejam X e Z variáveis aleatórias independentes, ambas com distribuição normal padrão. Para $a, b \in \mathbb{R}$, não ambos nulos, defina

$$Y = aX + bZ.$$

1. Calcule $\text{Corr}(X, Y)$.
2. Verifique diretamente que $|\text{Corr}(X, Y)| \leq 1$.
3. Determine as condições necessárias e suficientes sobre a e b para que $\text{Corr}(X, Y) = 1$.

4. Faça o mesmo para $\text{Corr}(X, Y) = -1$.

Ex. 7.8 — Problema do Pareamento Suponha que $n \geq 2$ pessoas deixem seus chapéus no vestiário e, ao final, os chapéus sejam devolvidos segundo uma permutação uniforme. Seja X o número de pessoas que recebem o próprio chapéu. Calcule $E[X]$ e $\text{Var}(X)$.

Ex. 7.9 — Seja X uma variável aleatória real. Prove que

$$E|X| = \infty \implies \sum_{n=0}^{\infty} P(|X| > n) = \infty.$$

Mostre, de fato, as desigualdades que relacionam a série acima com $E|X|$.

Ex. 7.10 — Seja X uma variável aleatória real. Prove que, para todo $\varepsilon > 0$, existe uma variável aleatória limitada X_ε tal que

$$P(X \neq X_\varepsilon) < \varepsilon.$$

Faça a construção por truncamento e indique uma cota determinística para $|X_\varepsilon|$.

Ex. 7.11 — Coletor de cupons. Cada pacote contém, de maneira independente, uma figurinha escolhida uniformemente entre n tipos. Determine o número esperado de pacotes necessários para obter ao menos uma figurinha de cada tipo.

Ex. 7.12 — Seja X uma variável aleatória com valores em $\mathbb{N}_0$ e segundo momento finito. Sua função geradora de probabilidades é

$$G_X(s) = E[s^X] = \sum_{i=0}^{\infty} P(X = i)s^i, \quad 0 \leq s \leq 1.$$

Justifique as derivações em $s = 1$ pela esquerda e prove que

$$\text{Var}(X) = G_X''(1) + G_X'(1) - G_X'(1)^2.$$

Desigualdades

Ex. 7.13 — Uma moeda justa é lançada independentemente n vezes, e S_n denota o número de caras. Use a desigualdade de Chebyshev para provar que, para todo

$\varepsilon > 0$,

$$\lim_{n \rightarrow \infty} \mathbf{P} \left(\left| \frac{S_n}{n} - \frac{1}{2} \right| < \varepsilon \right) = 1.$$

Ex. 7.14 — Dígitos binários e função geradora exponencial. Para $\omega \in [0,1)$, seja $d_k(\omega) \in \{0,1\}$ o k -ésimo dígito de sua expansão binária; nos racionais diádicos, escolha a expansão que termina em zeros. Defina

$$z_k(\omega) = 2d_k(\omega) - 1, \quad s_n(\omega) = \sum_{k=1}^n z_k(\omega).$$

Com respeito à medida de Lebesgue, prove que

$$M_n(t) := \int_0^1 e^{ts_n(\omega)} d\omega = \left(\frac{e^t + e^{-t}}{2} \right)^n, \quad t \in \mathbb{R}. \quad (7.20)$$

Diferenciando em $t = 0$, conclua que

$$\int_0^1 s_n(\omega) d\omega = 0, \quad \int_0^1 s_n(\omega)^2 d\omega = n.$$

Dica. Divida $[0,1)$ nos intervalos diádicos $[k/2^n, (k+1)/2^n)$, $0 \leq k < 2^n$.

Ex. 7.15 — Cota exponencial para os dígitos binários. Com a notação do exercício anterior, prove que, para todo $\varepsilon > 0$,

$$\mathbf{P} \left(\left| \frac{S_n}{n} \right| \geq \varepsilon \right) \leq 2e^{-n\varepsilon^2/2}. \quad (7.21)$$

Dica. Use (7.20), o argumento exponencial de Markov e

$$\frac{e^x + e^{-x}}{2} \leq e^{x^2/2}, \quad x \in \mathbb{R}.$$

Ex. 7.16 — Demonstração probabilística do Teorema de Weierstrass. Se $f : [0,1] \rightarrow \mathbb{R}$ é contínua, use a desigualdade de Chebyshev e a continuidade uniforme de f para provar que

$$\sum_{k=0}^n f\left(\frac{k}{n}\right) \binom{n}{k} x^k (1-x)^{n-k} \longrightarrow f(x)$$

uniformemente em $x \in [0,1]$.

Dica. Para cada x , tome $S_n \sim \text{Binomial}(n, x)$ e escreva o polinômio acima como $\mathbf{E}[f(S_n/n)]$. Separe os eventos $\{|S_n/n - x| < \delta\}$ e $\{|S_n/n - x| \geq \delta\}$ usando uma escolha de δ que não dependa de x .

Ex. 7.17 — Otimalidade da desigualdade de Chebyshev. Fixe $t \geq 1$, $\mu \in \mathbb{R}$ e $\sigma > 0$. Construa uma variável aleatória X com média μ e variância σ^2 tal que

$$\mathbf{P}(|X - \mu| \geq t\sigma) = \frac{1}{t^2}.$$

Ex. 7.18 — Seja X uma variável aleatória não negativa com $\mathbf{E}[X] < \infty$, e defina $X_n = X \wedge n$.

1. Dê um exemplo para o qual $\mathbf{E}[X^2] = \infty$.
2. Prove que, para todo $p > 2$,

$$\sum_{n=1}^{\infty} n^{-p} \mathbf{E}[X_n^2] < \infty.$$

3. Prove que a conclusão continua válida no caso-limite $p = 2$.

Dica para o último item. Para cada $x \geq 0$, estime $\sum_{n \geq 1} n^{-2} (x \wedge n)^2$ separando os índices $n \leq x$ e $n > x$.

Ex. 7.19 — Sejam $X_1, \dots, X_n$ variáveis aleatórias reais independentes e $S_k = X_1 + \dots + X_k$. Prove a desigualdade de Etemadi: para todo $t > 0$,

$$\mathbf{P}\left(\max_{1 \leq k \leq n} |S_k| \geq t\right) \leq 3 \max_{1 \leq k \leq n} \mathbf{P}(|S_k| \geq t/3).$$

Dica. Para $r > 0$, considere os eventos disjuntos de primeira passagem

$$A_j = \left\{ \max_{1 \leq k < j} |S_k| < 3r, |S_j| \geq 3r \right\}, \quad j = 1, \dots, n.$$

Use a independência de A_j e $S_n - S_j$ e trate separadamente o caso em que $\max_k \mathbf{P}(|S_k| \geq r) > 1/3$.

Ex. 7.20 — Dê um exemplo de variáveis aleatórias limitadas e dependentes X e Y tais que $\text{Cov}(X, Y) = 0$. Calcule a covariância e mostre diretamente que as variáveis não são independentes.

Dica. Tome uma variável X centrada e procure Y que seja nula sempre que $X \neq 0$.

Medida Produto

Para estudar duas ou mais grandezas simultaneamente, precisamos medir subconjuntos de um produto de espaços. A medida produto realiza essa construção, e os teoremas de Tonelli e Fubini mostram quando uma integral em várias variáveis pode ser calculada por integrais sucessivas. A ordem de integração parece apenas tipográfica; sem as hipóteses corretas, não é.

Definição 8.1 —

Dados $(\Omega_1, \mathcal{F}_1)$ e $(\Omega_2, \mathcal{F}_2)$ espaços mensuráveis, definimos a σ -álgebra produto $\mathcal{F}_1 \otimes \mathcal{F}_2$ como a σ -álgebra gerada pela coleção dos retângulos mensuráveis:

$$\mathcal{F}_1 \otimes \mathcal{F}_2 = \sigma \langle \{A \times B : A \in \mathcal{F}_1, B \in \mathcal{F}_2\} \rangle$$

Nesse caso também diremos que o produto dos espaços mensuráveis $(\Omega_1, \mathcal{F}_1)$ e $(\Omega_2, \mathcal{F}_2)$ é o espaço mensurável $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2)$.

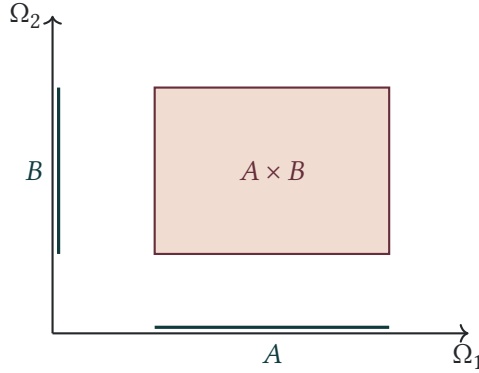


Figura 8.1.: Retângulo mensurável em $\mathcal{F}_1 \otimes \mathcal{F}_2$.

As projeções oferecem outra descrição da σ -álgebra produto. Ao produto cartesiano $\Omega_1 \times \Omega_2$ dos conjuntos Ω_1 e Ω_2 associamos as aplicações

$$\pi_1 : \Omega_1 \times \Omega_2 \longrightarrow \Omega_1, \quad \pi_2 : \Omega_1 \times \Omega_2 \longrightarrow \Omega_2.$$

Podemos fazer uso dessas aplicações para construir a σ -álgebra produto. Dados dois espaços mensuráveis $(\Omega_1, \mathcal{F}_1)$ e $(\Omega_2, \mathcal{F}_2)$, podemos formar uma coleção de subconjuntos em $\Omega_1 \times \Omega_2$ puxando de $\mathcal{F}_1$:

$$\pi_1^*(\mathcal{F}_1) := \{\pi_1^{-1}(E) : E \in \mathcal{F}_1\} = \{E \times \Omega_2 : E \in \mathcal{F}_1\}.$$

A família $\pi_1^*(\mathcal{F}_1)$ é uma σ -álgebra, pois a imagem inversa preserva complementos e uniões enumeráveis.

Podemos fazer a mesma construção para a projeção no segundo espaço.

Teorema 8.2 — Produto finito de σ -álgebras

A σ -álgebra produto é a σ -álgebra gerada pelas imagens inversas das duas projeções:

$$\mathcal{F}_1 \otimes \mathcal{F}_2 = \sigma\langle \pi_1^*(\mathcal{F}_1) \cup \pi_2^*(\mathcal{F}_2) \rangle.$$

Demonstração. Para $A \in \mathcal{F}_1$ e $B \in \mathcal{F}_2$,

$$A \times B = \pi_1^{-1}(A) \cap \pi_2^{-1}(B).$$

Logo todo retângulo mensurável pertence à σ -álgebra do lado direito. Reciprocamente, $\pi_1^{-1}(A) = A \times \Omega_2$ e $\pi_2^{-1}(B) = \Omega_1 \times B$ são retângulos mensuráveis. Portanto, cada uma das duas famílias geradoras está contida em $\mathcal{F}_1 \otimes \mathcal{F}_2$, o que prova as duas inclusões. $\square$

Suponha que $E \subset \Omega_1 \times \Omega_2$. Dado $x \in \Omega_1$ e $y \in \Omega_2$, definimos a x -seção $E^x \subset \Omega_2$ e a y -seção $E^y \subset \Omega_1$ de E como

$$E^y = \{x \in \Omega_1 : (x, y) \in E\}$$

$$E^x = \{y \in \Omega_2 : (x, y) \in E\}$$

Conforme indicado na próxima proposição, todas as seções de um conjunto mensurável são mensuráveis.

Proposição 8.3 —

Se $(\Omega_1, \mathcal{F}_1)$ e $(\Omega_2, \mathcal{F}_2)$ são espaços mensuráveis e $E \in \mathcal{F}_1 \otimes \mathcal{F}_2$, então $E^x \in \mathcal{F}_2$ para todo $x \in \Omega_1$ e $E^y \in \mathcal{F}_1$ para todo $y \in \Omega_2$.

Demonstração. Provaremos apenas para a seção E^y ; a demonstração para a seção E^x é análoga.

Considere a família $\mathcal{M} = \{E \in \mathcal{F}_1 \otimes \mathcal{F}_2 : E^y \in \mathcal{F}_1\}$. Mostraremos que $\mathcal{M}$ é uma σ -álgebra contendo os retângulos mensuráveis e logo deve ser $\mathcal{F}_1 \otimes \mathcal{F}_2$, e isso é suficiente para demonstrar a proposição.

- $\square$ Suponha que $E = A \times B$ é um retângulo mensurável. Então $E^y = A$ quando $y \in B$ e $E^y = \emptyset$ quando $y \notin B$. Em ambos os casos $E^y \in \mathcal{F}_1$, logo $E \in \mathcal{M}$.
- $\square$ $\emptyset^y = \{x \in \Omega_1 : (x, y) \in \emptyset\} = \emptyset \in \mathcal{F}_1$, assim $\mathcal{M}$ contém $\emptyset$.
- $\square$ Se $E_1, E_2, \dots \in \mathcal{M}$, então $(\bigcup E_n)^y = \bigcup E_n^y \in \mathcal{F}_1$. Logo $\mathcal{M}$ é fechado em relação a união enumerável.

□ Se $E \in \mathcal{M}$, e $F = E^c$, então $F^y = \{x \in \Omega_1 : (x, y) \notin E\} = \Omega_1 \setminus E^y \in \mathcal{F}_1$. Logo $\mathcal{M}$ é fechada sob complementação.

□

Proposição 8.4 —

Suponha que $\mathbb{R}^m, \mathbb{R}^n$ estejam equipados com suas σ -álgebras Borel $\mathcal{B}(\mathbb{R}^m), \mathcal{B}(\mathbb{R}^n)$ e seja $\mathbb{R}^{m+n} = \mathbb{R}^m \times \mathbb{R}^n$. Então

$$\mathcal{B}(\mathbb{R}^{m+n}) = \mathcal{B}(\mathbb{R}^m) \otimes \mathcal{B}(\mathbb{R}^n).$$

Demonstração. Cada retângulo $A \times B$, com $A \in \mathcal{B}(\mathbb{R}^m)$ e $B \in \mathcal{B}(\mathbb{R}^n)$, é boreliano em $\mathbb{R}^{m+n}$, pois

$$A \times B = \pi_1^{-1}(A) \cap \pi_2^{-1}(B)$$

e as projeções coordenadas são contínuas. Isso prova $\mathcal{B}(\mathbb{R}^m) \otimes \mathcal{B}(\mathbb{R}^n) \subseteq \mathcal{B}(\mathbb{R}^{m+n})$.

Na direção contrária, escolha bases enumeráveis de abertos em $\mathbb{R}^m$ e $\mathbb{R}^n$. Os produtos de elementos dessas bases formam uma base enumerável da topologia de $\mathbb{R}^{m+n}$. Todo aberto de $\mathbb{R}^{m+n}$ é, portanto, união enumerável desses retângulos e pertence a $\mathcal{B}(\mathbb{R}^m) \otimes \mathcal{B}(\mathbb{R}^n)$. Como os abertos geram a σ -álgebra de Borel, segue a inclusão oposta. □

Suponha que $(\Omega_1, \mathcal{F}_1)$ e $(\Omega_2, \mathcal{F}_2)$ sejam espaços mensuráveis. A interseção de retângulos mensuráveis é um retângulo mensurável

$$(A \times B) \cap (C \times D) = (A \cap C) \times (B \cap D),$$

e o complemento de um retângulo mensurável é uma união finita de retângulos

$$(A \times B)^c = (A^c \times B) \cup (A \times B^c) \cup (A^c \times B^c).$$

Assim, a família de uniões finitas de retângulos mensuráveis em $\Omega_1 \times \Omega_2$ forma uma álgebra, que denotamos por $\mathcal{F}_0$. Esta álgebra não é, em geral, uma σ -álgebra, mas gera a mesma σ -álgebra produto que os retângulos mensuráveis.

Teorema 8.5 — Existência da medida produto

Dados espaços de medida σ -finitos $(\Omega_1, \mathcal{F}_1, \mu_1)$ e $(\Omega_2, \mathcal{F}_2, \mu_2)$, existe uma única medida μ em $\mathcal{F}_1 \otimes \mathcal{F}_2$ satisfazendo

$$\mu(A \times B) = \mu_1(A)\mu_2(B), \quad A \in \mathcal{F}_1, B \in \mathcal{F}_2.$$

Diremos que μ é a medida produto de μ_1 e μ_2 . E será denotada por $\mu_1 \otimes \mu_2$.

Demonstração. Seja $\mathcal{F}_0$ a álgebra das uniões finitas disjuntas de retângulos mensuráveis. Para $E \in \mathcal{F}_0$, defina

$$\mu_0(E) := \int_{\Omega_1} \mu_2(E^x) \, d\mu_1(x).$$

Como E é união finita de retângulos, a função $x \mapsto \mu_2(E^x)$ é mensurável e simples; em particular, a definição faz sentido. Para um retângulo,

$$\mu_0(A \times B) = \int_{\Omega_1} \mathbb{1}_A(x) \mu_2(B) \, d\mu_1(x) = \mu_1(A)\mu_2(B).$$

Se (E_k) é uma sequência disjunta em $\mathcal{F}_0$ cuja união E ainda pertence a $\mathcal{F}_0$, então, para cada x , as seções E_k^x são disjuntas e $E^x = \bigcup_k E_k^x$. Pela aditividade de μ_2 e pela Convergência Monótona,

$$\mu_0(E) = \int \sum_{k=1}^{\infty} \mu_2(E_k^x) \, d\mu_1(x) = \sum_{k=1}^{\infty} \mu_0(E_k).$$

Logo μ_0 é uma pré-medida, e o Teorema de Extensão de Carathéodory fornece uma medida μ em $\sigma\langle \mathcal{F}_0 \rangle = \mathcal{F}_1 \otimes \mathcal{F}_2$.

Para a unicidade, escolha $U_j \uparrow \Omega_1$ e $V_j \uparrow \Omega_2$ com $\mu_1(U_j), \mu_2(V_j) < \infty$. Os retângulos $U_j \times V_j$ crescem para o espaço produto e têm medida finita. Assim, o segundo caso do Teorema de Unicidade de Medidas 2.14, aplicado ao π -sistema dos retângulos, mostra que a extensão é única. $\square$

Corolário 8.6 — Produto finito de medidas

Sejam $(\Omega_j, \mathcal{F}_j, \mu_j)$, $1 \leq j \leq n$, espaços de medida σ -finitos. Existe uma única medida σ -finita

$$\mu_1 \otimes \cdots \otimes \mu_n$$

na σ -álgebra produto $\mathcal{F}_1 \otimes \cdots \otimes \mathcal{F}_n$, definida por iteração, tal que

$$(\mu_1 \otimes \cdots \otimes \mu_n)(A_1 \times \cdots \times A_n) = \prod_{j=1}^n \mu_j(A_j)$$

para todo retângulo mensurável $A_1 \times \cdots \times A_n$. Sob a identificação natural dos produtos cartesianos, essa medida não depende da parentização dos fatores.

Demonstração. Itere o teorema anterior. O produto de duas medidas σ -finitas é σ -finito, pois produtos dos conjuntos de uma exaustão por conjuntos de medida finita fornecem uma exaustão do espaço produto. Assim a construção pode ser repetida um número finito de vezes.

As diferentes parentizações produzem medidas na mesma σ -álgebra: todas as σ -álgebras iteradas são geradas pelos retângulos $A_1 \times \cdots \times A_n$. Além disso, todas as medidas obtidas atribuem a cada retângulo o valor $\prod_j \mu_j(A_j)$. A unicidade no π -sistema dos retângulos, usando a σ -finitude, mostra que essas medidas coincidem. □

Teorema 8.7 — Mensurabilidade de Funções com Variável Fixa

Sejam $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2)$ e $(Z, \mathcal{M})$ espaços mensuráveis.

Se $f : \Omega_1 \times \Omega_2 \rightarrow Z$ é mensurável, então as funções $f^y : \Omega_1 \rightarrow Z$, $f^x : \Omega_2 \rightarrow Z$ obtidas fixando uma variável também são mensuráveis.

Demonstração. Provaremos apenas para f^y . Seja $I^y : \Omega_1 \rightarrow \Omega_1 \times \Omega_2$ a aplicação de inclusão $I^y(x) = (x, y)$. Dado $E \in \mathcal{F}_1 \otimes \mathcal{F}_2$, por 8.3 $(I^y)^{-1}(E) = E^y \in \mathcal{F}_1$, logo I^y é uma função mensurável. Mas $f^y = f \circ I^y$. $\square$

8.1. Áreas de Seções Transversais

Um conjunto no espaço produto pode ser examinado fixando uma coordenada. As medidas dessas seções desempenham o papel abstrato de áreas transversais e serão o integrando na fórmula da medida produto.

Teorema 8.8 — Mensurabilidade das áreas das seções

Sejam $(\Omega_1, \mathcal{F}_1, \mu)$, $(\Omega_2, \mathcal{F}_2, \nu)$ espaços de medida σ -finitos. Se $E \in \mathcal{F}_1 \otimes \mathcal{F}_2$,

- $\square \nu(E^x)$ é uma função mensurável de $x \in \Omega_1$.
- $\square \mu(E^y)$ é uma função mensurável de $y \in \Omega_2$.

Demonstração. Provaremos a mensurabilidade de $y \mapsto \mu(E^y)$; a outra afirmação segue trocando os fatores. Suponha primeiro que $\mu(\Omega_1) < \infty$ e ponha

$$\mathcal{M} = \{E \in \mathcal{F}_1 \otimes \mathcal{F}_2 : \mu(E^y) \text{ é uma função mensurável de } y\}.$$

$\mathcal{M}$ é igual a $\mathcal{F}_1 \otimes \mathcal{F}_2$, pois:

- $\square$ Se $E = A \times B$ é um retângulo mensurável, então $\mu(E^y) = \mu(A)\mathbb{1}_B(y)$, que é mensurável. Se E é uma união finita disjunta de retângulos mensuráveis E_j , então $\mu(E^y) = \sum_j \mu(E_j^y)$, que também é mensurável.

Então $\mathcal{M}$ contém a álgebra das uniões finita disjuntas de retângulos mensuráveis.

- $\square$ Se E_n são conjuntos crescentes em $\mathcal{M}$, então $\mu((\bigcup_n E_n)^y) = \mu(\bigcup_n E_n^y) = \lim_{n \rightarrow \infty} \mu(E_n^y)$ é mensurável, e logo $\bigcup_n E_n \in \mathcal{M}$.

De modo análogo, se E_n são conjuntos decrescentes em $\mathcal{M}$, então a continu-

idade por cima dá

$$\mu\left(\left(\bigcap_n E_n\right)^y\right) = \lim_n \mu(E_n^y).$$

Aqui se usa a finitude de $\mu(\Omega_1)$.

- Assim $\mathcal{M}$ é uma classe monótona contendo a álgebra das uniões finitas de retângulos mensuráveis. Pelo Teorema da Classe Monótona (1.29), $\mathcal{M} = \mathcal{F}_1 \otimes \mathcal{F}_2$.

No caso σ -finito, escolha $\Omega_{1,n} \uparrow \Omega_1$ com $\mu(\Omega_{1,n}) < \infty$ e reaplique o argumento à medida finita $\mu_n(F) = \mu(F \cap \Omega_{1,n})$. Então

$$\mu(E^y) = \lim_{n \rightarrow \infty} \mu_n(E^y),$$

que é mensurável como limite de funções mensuráveis. □

8.2. Integrais Iteradas

Construímos a medida produto por extensão de Carathéodory e sabemos quanto ela vale nos retângulos. As seções permitem agora calculá-la em conjuntos mensuráveis gerais por uma fórmula integral explícita.

Teorema 8.9 — Fórmula integral da medida produto

Sejam $(\Omega_1, \mathcal{F}_1, \mu)$ e $(\Omega_2, \mathcal{F}_2, \nu)$ espaços de medida σ -finitos. Para todo $E \in \mathcal{F}_1 \otimes \mathcal{F}_2$,

$$(\mu \otimes \nu)(E) = \int_{x \in \Omega_1} \underbrace{\int_{y \in \Omega_2} \mathbb{1}_E \, d\nu}_{\nu(E^x)} \, d\mu = \int_{y \in \Omega_2} \underbrace{\int_{x \in \Omega_1} \mathbb{1}_E \, d\mu}_{\mu(E^y)} \, d\nu.$$

Demonstração. Denote por $\lambda_1(E)$ a primeira integral iterada e por $\lambda_2(E)$ a segunda. Essas integrais existem pelo Teorema 8.8. As medidas λ_1 e λ_2 são σ -aditivas pelo Teorema da Convergência Monótona logo ambas são medidas em $\mathcal{F}_1 \otimes \mathcal{F}_2$. Mais ainda se $E = A \times B$, então expandindo as duas integrais temos $\lambda_1(E) = \mu(A)\nu(B) = \lambda_2(E)$.

As três medidas são σ -finitas e coincidem no π -sistema dos retângulos. Pelo Teorema de Unicidade de Medidas 2.14, $\lambda_1 = \lambda_2 = \mu \otimes \nu$ em toda $\mathcal{F}_1 \otimes \mathcal{F}_2$. $\square$

Teorema 8.10 — Fubini

Sejam $(\Omega_1, \mathcal{F}_1, \mu)$ e $(\Omega_2, \mathcal{F}_2, \nu)$ espaços de medida σ -finitos. Se $\xi : \Omega_1 \times \Omega_2 \rightarrow \mathbb{R}$ é $\mu \otimes \nu$ -integrável, então as seções são integráveis quase em toda parte e

$$\int_{\Omega_1 \times \Omega_2} \xi \, d(\mu \otimes \nu) = \int_{x \in \Omega_1} \left[\int_{y \in \Omega_2} \xi(x, y) \, d\nu \right] d\mu = \int_{y \in \Omega_2} \left[\int_{x \in \Omega_1} \xi(x, y) \, d\mu \right] d\nu.$$

Demonstração. Para $\xi = 1_E$, a afirmação é o Teorema 8.9. Por linearidade, ela vale para funções simples não negativas. Se $\xi \geq 0$ é mensurável, escolha $s_n \uparrow \xi$ simples. Para cada variável fixa, $s_n(x, \cdot) \uparrow \xi(x, \cdot)$; duas aplicações da Convergência Monótona dão

$$\int \xi \, d(\mu \otimes \nu) = \int_{\Omega_1} \left(\int_{\Omega_2} \xi(x, y) \, d\nu(y) \right) d\mu(x),$$

e, por simetria, a igualdade na outra ordem. Esse é o conteúdo de Tonelli para funções não negativas.

Agora aplique o caso não negativo a $|\xi|$. Como $\int |\xi| \, d(\mu \otimes \nu) < \infty$, a função $x \mapsto \int |\xi(x, y)| \, d\nu(y)$ é finita para μ -quase todo x . Logo as seções de ξ são integráveis fora de um conjunto nulo. Aplicando a fórmula já provada a ξ^+ e ξ^- e subtraindo, obtemos a primeira integral iterada. O argumento simétrico fornece a segunda. $\square$

Teorema 8.11 — Tonelli

Sejam $(\Omega_1, \mathcal{F}_1, \mu)$ e $(\Omega_2, \mathcal{F}_2, \nu)$ espaços de medida σ -finitos.

1 Se $f : \Omega_1 \times \Omega_2 \rightarrow [0, \infty]$ é $\mathcal{F}_1 \otimes \mathcal{F}_2$ -mensurável, então as funções

$$x \mapsto \int_{\Omega_2} f(x, y) \, d\nu(y), \quad y \mapsto \int_{\Omega_1} f(x, y) \, d\mu(x)$$

são mensuráveis e

$$\int_{\Omega_1 \times \Omega_2} f \, d(\mu \otimes \nu) = \int_{\Omega_1} \left(\int_{\Omega_2} f(x, y) \, d\nu(y) \right) d\mu(x) = \int_{\Omega_2} \left(\int_{\Omega_1} f(x, y) \, d\mu(x) \right) d\nu(y),$$

com valores em $[0, \infty]$.

2 Se $g : \Omega_1 \times \Omega_2 \rightarrow \mathbb{R}$ é mensurável, então g é $\mu \otimes \nu$ -integrável se e somente se

$$\int_{\Omega_1} \left(\int_{\Omega_2} |g(x, y)| \, d\nu(y) \right) d\mu(x) < \infty.$$

Equivalentemente, a mesma condição pode ser escrita com a ordem das integrais invertida.

Demonstração. Para funções simples não negativas, a mensurabilidade das integrais parciais segue do Teorema 8.8, e as igualdades seguem da linearidade e do Teorema 8.9. Para $f \geq 0$ mensurável, tome funções simples $s_n \uparrow f$. As integrais parciais de s_n crescem para as integrais parciais de f ; logo estas são mensuráveis. Aplicando a Convergência Monótona no espaço produto e em cada integral iterada, obtemos as duas igualdades do primeiro item.

Aplicando o primeiro item a $|g|$, temos

$$\int_{\Omega_1 \times \Omega_2} |g| \, d(\mu \otimes \nu) = \int_{\Omega_1} \left(\int_{\Omega_2} |g(x, y)| \, d\nu(y) \right) d\mu(x),$$

e analogamente na ordem inversa. Portanto, qualquer uma dessas quantidades é finita se e somente se g é integrável. $\square$

Corolário 8.12 — Representação por camadas

Se $(E, \mathcal{E}, \mu)$ é um espaço de medida σ -finito, $f : E \rightarrow [0, \infty]$ é mensurável e $r > 0$, então

$$\int_E f^r \, d\mu = r \int_0^\infty t^{r-1} \mu(f > t) \, dt,$$

com a igualdade entendida em $[0, \infty]$. Em particular,

$$\int_E f \, d\mu = \int_0^\infty \mu(f > t) \, dt,$$

e, se $0 \leq f \leq 1$, a integral externa pode ser restrita a $[0, 1]$.

Demonstração. Para todo $x \in E$,

$$f(x)^r = \int_0^\infty r t^{r-1} \mathbb{1}_{\{f(x) > t\}} \, dt.$$

O conjunto $\{(x, t) : f(x) > t\}$ é mensurável, pois

$$\{(x, t) : f(x) > t\} = \bigcup_{q \in \mathbb{Q}_{>0}} \{f > q\} \times (0, q).$$

Logo o integrando no lado direito é mensurável e não negativo. Tonelli permite integrar primeiro em x ou em t , produzindo a fórmula. $\square$

Resumidamente, se uma integral dupla é absolutamente convergente, então é válido mudar a ordem de integração.

Exemplo 8.13 — .

Se ξ não é não-negativa, a hipótese de convergência absoluta é crucial para mudar a ordem de integração.

Um contraexemplo elementar é a sequência duplamente indexada $a_{m,n} = (-m)^n / n!$, para $m \geq 1$ e $n \geq 0$. Somando primeiro em n , obtemos $\sum_{n \geq 0} a_{m,n} = e^{-m}$ e, portanto,

$$\sum_{m \geq 1} \sum_{n \geq 0} a_{m,n} = \sum_{m \geq 1} e^{-m} = \frac{1}{e-1}.$$

Para cada n fixo, entretanto, os termos $a_{m,n}$ nem sequer tendem a zero quando $m \rightarrow \infty$, de modo que a soma na ordem inversa não existe. A série dupla não é absolutamente convergente.

<

8.3. Distribuição Conjunta

A medida produto volta agora à probabilidade. Para um vetor aleatório, a distribuição conjunta registra simultaneamente as coordenadas e, ao contrário das marginais isoladas, conserva a dependência entre elas.

Definição 8.14 — Distribuição conjunta

Se $\xi = (\xi_1, \dots, \xi_n)$ for um vetor aleatório definido em $(\Omega, \mathcal{F}, \mathbf{P})$, então ξ induz uma medida de probabilidade $\mathbf{P}_\xi$ em $\mathbb{R}^n$, dada, para $A \in \mathcal{B}(\mathbb{R}^n)$, por

$$\mathbf{P}_\xi(A) = \mathbf{P}(\xi^{-1}(A)) = \mathbf{P}(\xi \in A). \quad (8.1)$$

Essa medida é denominada **distribuição conjunta** de ξ .

Definição 8.15 — Função de Distribuição Conjunta

Se $\xi = (\xi_1, \dots, \xi_n)$ for um vetor aleatório, sua função de distribuição conjunta é

$$F_\xi(x_1, \dots, x_n) = \mathbf{P}(\xi_1 \leq x_1, \dots, \xi_n \leq x_n).$$

A partir da definição da independência vemos que a função de distribuição conjunta fatora como o produto de funções de distribuição.

$$F_{(\xi_1, \dots, \xi_n)}(x_1, \dots, x_n) = F_{\xi_1}(x_1) \times \dots \times F_{\xi_n}(x_n),$$

Teorema 8.16 —

As variáveis aleatórias $\xi_1, \dots, \xi_n$ são independentes se e somente se sua distribuição conjunta satisfaz

$$\mathbf{P}_\xi = \mathbf{P}_{\xi_1} \otimes \dots \otimes \mathbf{P}_{\xi_n}.$$

Demonstração. Se as variáveis são independentes, então, para quaisquer $A_k \in \mathcal{B}(\mathbb{R})$,

$$\mathbf{P}_\xi(A_1 \times \dots \times A_n) = \mathbf{P}(\xi_1 \in A_1, \dots, \xi_n \in A_n) = \prod_{k=1}^n \mathbf{P}_{\xi_k}(A_k).$$

As duas probabilidades coincidem no π -sistema dos retângulos; pelo Teorema de Unicidade de Medidas 2.14, coincidem em toda $\mathcal{B}(\mathbb{R}^n)$.

Reciprocamente, se a distribuição conjunta é a medida produto, a mesma identidade vale para todo retângulo boreliano. Essa identidade é exatamente a definição de independência de $\xi_1, \dots, \xi_n$. $\square$

Assim, a independência equivale a dizer que a lei conjunta é o produto das leis marginais; as variáveis não precisam estar originalmente definidas num espaço produto.

8.4. Extensão de Kolmogorov em $\mathbb{R}^{\mathbb{N}}$

Para construir conjuntamente uma sequência de variáveis aleatórias, precisamos de um espaço no qual todas possam ser observadas. Passamos, por isso, dos produtos finitos aos produtos infinitos, que também servirão de espaço amostral para processos estocásticos.

Nessa seção nos concentraremos na construção de medidas no espaço $\mathbb{R}^{\mathbb{N}}$, onde todas as ideias centrais são apresentadas. Posteriormente, na Seção 8.8, o argumento que apresentaremos será generalizado para produtos mais gerais.

Para provar o Teorema de Extensão de Kolmogorov, seguiremos os passos de Bochner[10, 2]. Antes de entrarmos diretamente nos espaços produto, apresenta-

remos algumas definições e resultados que preparam essa construção.

Considere então um conjunto Ω e uma sequência crescente $\{\mathcal{F}_n\}_{n=1}^{\infty}$ de σ -álgebras em Ω . E para cada n , seja μ_n uma medida de probabilidade na σ -álgebra $\mathcal{F}_n$.

Definição 8.17 —

Diremos que a sequência $\{(\mathcal{F}_n, \mu_n)\}_{n=1}^{\infty}$ é **Kolmogorov consistente** se sempre que $m \leq n$ temos $\mathcal{F}_m \subset \mathcal{F}_n$ e $\mu_n|_{\mathcal{F}_m} = \mu_m$.

Dado $\mathcal{A} = \bigcup_{n=1}^{\infty} \mathcal{F}_n$, uma **extensão** da sequência de probabilidades $\{\mu_n\}_{n=1}^{\infty}$ é uma medida μ em $\sigma\langle\mathcal{A}\rangle$ satisfazendo $\mu_n = \mu|_{\mathcal{F}_n}$.

Teorema 8.18 — Extensão de Bochner

Dado $\{(\mathcal{F}_n, \mu_n)\}_{n=1}^{\infty}$ uma sequência Kolmogorov consistente. Suponha que existe uma classe compacta $\mathcal{K} \subseteq \bigcup_{n=1}^{\infty} \mathcal{F}_n$ de Ω tal que

$$\mu_n(A) = \sup\{\mu_n(C) : C \in \mathcal{K} \cap \mathcal{F}_n \text{ e } C \subset A\}.$$

Então existe uma única extensão de Kolmogorov para a σ -álgebra $\sigma(\bigcup_{n=1}^{\infty} \mathcal{F}_n)$.

Demonstração.

Defina μ na álgebra $\mathcal{A} = \bigcup_{n=1}^{\infty} \mathcal{F}_n$ por $\mu(E) = \mu_n(E)$ para $E \in \mathcal{F}_n$. A condição de consistência de Kolmogorov garante que μ está bem definida. Também é simples ver que $\mu(\emptyset) = 0$ e $\mu(\Omega) = 1$.

A aditividade finita de μ é direta. Se $A_i \in \mathcal{A}$, para $i = 1, \dots, r$, existe m tal que $A_i \in \mathcal{F}_m$ para todos esses índices; basta então usar a aditividade de μ_m .

Para concluir a demonstração, observamos que se $A \in \mathcal{A}$, então existe n tal que

$A \in \mathcal{F}_n$ e

$$\begin{aligned}\mu(A) &= \mu_n(A) \\ &= \sup\{\mu(C) : C \in \mathcal{K} \cap \mathcal{F}_n \text{ e } C \subset A\} \\ &\leq \sup\{\mu(C) : C \in \mathcal{K} \text{ e } C \subset A\} \\ &\leq \mu(A).\end{aligned}$$

Assim, pelo Teorema de Extensão Compacta 2.29, μ é uma pré-medida. O Teorema de Extensão de Carathéodory fornece então uma extensão a uma probabilidade em $\sigma\langle\mathcal{A}\rangle$; como a pré-medida é finita, essa extensão é única. $\square$

Com essas ferramentas, voltamos à construção de medidas de probabilidade em

$$\mathbb{R}^{\mathbb{N}} = \{(\omega_1, \omega_2, \dots) : \omega_i \in \mathbb{R}, i \in \mathbb{N}\}.$$

O primeiro passo é eleger uma σ -álgebra onde definiremos nossa medida. Consi-

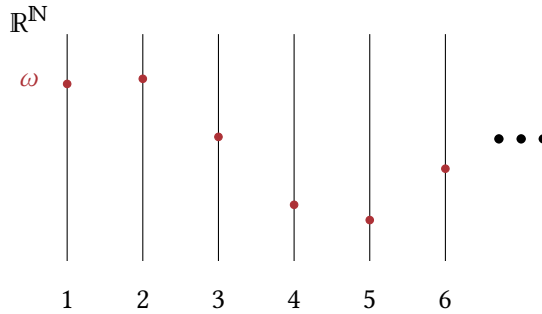


Figura 8.2.: Representação visual de $\mathbb{R}^{\mathbb{N}}$.

dere a classe dos cilindros retangulares de dimensão finita:

Definição 8.19 —

Definimos

$$\mathcal{C} = \{\{\omega : \omega_i \in (a_i, b_i] \text{ para } i = 1, \dots, k\} : k \in \mathbb{N}, -\infty \leq a_i < b_i \leq \infty\}.$$

A σ -álgebra produto em $\mathbb{R}^{\mathbb{N}}$ é

$$\mathcal{B}(\mathbb{R})^{\otimes \mathbb{N}} := \sigma\langle \mathcal{C} \rangle.$$

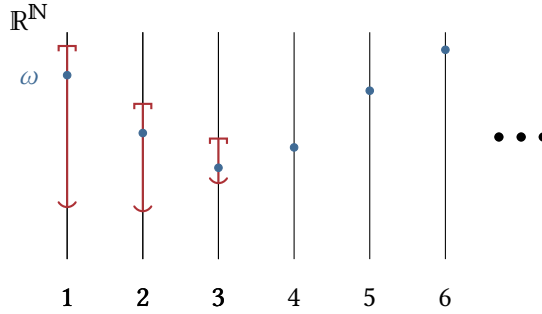


Figura 8.3.: Representação visual de um cilindro em $\mathbb{R}^{\mathbb{N}}$. Em azul um ponto do cilindro.

Observação 8.20 —

Esta é, de fato, a σ -álgebra com a qual trabalhamos intuitivamente nos cursos mais básicos de probabilidade. Lembre, por exemplo, que quando queremos estudar o “lançamento de infinitas moedas”, sempre descrevemos os eventos de interesse a partir de eventos que dependem do resultado de finitos lançamentos. Mesmo eventos do tipo $A = \{\text{observamos cara infinitas vezes}\}$ é escrito como

$$A = \bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n,$$

com $A_n = \{\text{observamos cara na } n\text{-ésima jogada}\}$.

Para aplicar o Teorema 8.18, fixaremos uma notação para os cilindros. Dado $B \subset \mathbb{R}^n$, denotaremos por $B \times \mathbb{R}^{\mathbb{N}-n}$ o conjunto:

$$B \times \mathbb{R}^{\mathbb{N}-n} := \{(\omega_1, \omega_2, \dots) \text{ tais que } (\omega_1, \dots, \omega_n) \in B\}$$

Se $A \in \mathcal{C}$, então existem $n \in \mathbb{N}$ e $B \in \mathcal{B}(\mathbb{R}^n)$ tais que $A = B \times \mathbb{R}^{\mathbb{N}-n}$.

Considere agora a sequência encaixante de σ -álgebras:

$$\mathcal{F}_n = \{A \times \mathbb{R}^{\mathbb{N}-n} : A \in \mathcal{B}(\mathbb{R}^n)\}.$$

Se $m < n$ e $B \in \mathcal{B}(\mathbb{R}^m)$, então

$$B \times \mathbb{R}^{\mathbb{N}-m} = (B \times \mathbb{R}^{n-m}) \times \mathbb{R}^{\mathbb{N}-n} \in \mathcal{F}_n,$$

de modo que $\mathcal{F}_m \subset \mathcal{F}_n$, e $\mathcal{A} = \bigcup_{n=1}^{\infty} \mathcal{F}_n$ é uma álgebra, com $\mathcal{C} \subseteq \mathcal{A}$.

Proposição 8.21 —

$$\sigma\langle \mathcal{A} \rangle = \sigma\langle \mathcal{C} \rangle.$$

Demonstração. Como $\mathcal{C} \subset \mathcal{A}$, temos imediatamente $\sigma\langle \mathcal{C} \rangle \subset \sigma\langle \mathcal{A} \rangle$. Para ver a inclusão contrária, dado $n \geq 1$, faça $\mathcal{G}_n = \{A \in \mathcal{B}(\mathbb{R}^n) : A \times \mathbb{R}^{\mathbb{N}-n} \in \sigma\langle \mathcal{C} \rangle\}$. Como $\mathcal{G}_n$ é uma σ -álgebra contendo os retângulos do tipo

$$\prod_{k=1}^n (a_k, b_k] \subset \mathbb{R}^n,$$

então $\mathcal{G}_n = \mathcal{B}(\mathbb{R}^n)$. Segue que $\mathcal{F}_n \subseteq \sigma\langle \mathcal{C} \rangle$ para todo $n \geq 1$ e

$$\sigma\langle \mathcal{A} \rangle \subseteq \sigma\langle \mathcal{C} \rangle.$$

□

Teorema 8.22 — Extensão de Kolmogorov em $\mathbb{R}^{\mathbb{N}}$

Dadas medidas de probabilidade $\mathbf{P}_n$ em $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ tais que

$$\mathbf{P}_{n+1}((a_1, b_1] \times \cdots \times (a_n, b_n] \times \mathbb{R}) = \mathbf{P}_n((a_1, b_1] \times \cdots \times (a_n, b_n]),$$

para todo $n \geq 1$ e todos os intervalos indicados, existe uma única medida de probabilidade $\mathbf{P}$ em $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}(\mathbb{R})^{\otimes \mathbb{N}})$ com

$$\mathbf{P}\{\omega : \omega_i \in (a_i, b_i], 1 \leq i \leq n\} = \mathbf{P}_n((a_1, b_1] \times \cdots \times (a_n, b_n]).$$

Demonstração.

Para começar, dado $n \geq 1$ defina a medida de probabilidade Q_n em $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ por

$$Q_n(A) = \mathbf{P}_{n+1}(A \times \mathbb{R}).$$

Como Q_n coincide com $\mathbf{P}_n$ no π -sistema dos retângulos $\prod_{k=1}^n (a_k, b_k]$, o Teorema de Unicidade de Medidas 2.14 mostra que ambas coincidem em toda $\mathcal{B}(\mathbb{R}^n)$. Logo, para todo $n \geq 1$ e $A \in \mathcal{B}(\mathbb{R}^n)$,

$$\mathbf{P}_{n+1}(A \times \mathbb{R}) = \mathbf{P}_n(A),$$

e por indução

$$\mathbf{P}_{n+m}(A \times \mathbb{R}^m) = \mathbf{P}_n(A).$$

Com a notação introduzida anteriormente, defina a medida μ_n em $(\mathbb{R}^{\mathbb{N}}, \mathcal{F}_n)$ por

$$\mu_n(B \times \mathbb{R}^{\mathbb{N}-n}) = \mathbf{P}_n(B).$$

Se $m < n$ e $D \in \mathcal{B}(\mathbb{R}^m)$, então

$$\mu_n(D \times \mathbb{R}^{\mathbb{N}-m}) = \mathbf{P}_n(D \times \mathbb{R}^{n-m}) = \mathbf{P}_m(D) = \mu_m(D \times \mathbb{R}^{\mathbb{N}-m}).$$

Concluimos que $\{(\mathcal{F}_n, \mu_n)\}_{n \geq 1}$ é Kolmogorov consistente.

Considere agora as classes de cilindros

$$\mathcal{K}_n = \{K \times \mathbb{R}^{\mathbb{N}-n}, K \in \mathcal{B}(\mathbb{R}^n) \text{ compacto}\} \quad \text{e} \quad \mathcal{K} = \bigcup_{n \geq 1} \mathcal{K}_n.$$

Observe que, se $C \in \mathcal{K}_n$ e $D \in \mathcal{K}_m$, com $m < n$, então $C \cap D \in \mathcal{K}_n$.

Afirmamos que classe $\mathcal{K}$ é compacta, o que provaremos no lema a seguir.

Dado agora um conjunto $A \in \mathcal{A}$ então $A = D \times \mathbb{R}^{\mathbb{N}-n}$ com $D \in \mathcal{B}(\mathbb{R}^n)$, para algum $n \geq 1$. Pela regularidade interior das probabilidades de Borel em $\mathbb{R}^n$, para todo $\varepsilon > 0$ existe um compacto $K \subseteq D$ tal que

$$\mu_n(A \setminus (K \times \mathbb{R}^{\mathbb{N}-n})) = \mathbf{P}_n(D \setminus K) < \varepsilon.$$

Em particular,

$$\mu_n(A) \leq \mu_n(K \times \mathbb{R}^{\mathbb{N}-n}) + \varepsilon.$$

Consequentemente

$$\mu_n(A) = \sup\{\mu_n(C) : C \in \mathcal{K} \cap \mathcal{F}_n \text{ e } C \subset A\}.$$

e o teorema segue do Teorema 8.18. □

Dado $A \in \mathcal{A}$ não vazio, definimos $\dim(A) := \min\{n \geq 1 : A \in \mathcal{F}_n\}$, e $\dim(\mathbb{R}^N) = 0$.

Se $A_1, A_2 \in \mathcal{A}$ e $A_1 \cap A_2 \neq \emptyset$, então $\dim(A_1 \cap A_2) \leq \max\{\dim(A_1), \dim(A_2)\}$. De fato, se $m < n$ e $A_1 = B_1 \times \mathbb{R}^{N-m}$, $A_2 = B_2 \times \mathbb{R}^{N-n}$, então

$$(B_1 \times \mathbb{R}^{N-m}) \cap (B_2 \times \mathbb{R}^{N-n}) = ((B_1 \times \mathbb{R}^{n-m}) \cap B_2) \times \mathbb{R}^{N-n} \in \mathcal{F}_n.$$

Lema 8.23 —

A classe $\mathcal{K}$ definida na demonstração anterior como

$$\mathcal{K}_n = \{K \times \mathbb{R}^{N-n}, K \in \mathcal{B}(\mathbb{R}^n) \text{ compacto}\} \quad \text{e} \quad \mathcal{K} = \bigcup_{n \geq 1} \mathcal{K}_n.$$

é compacta.

Demonstração. Considere uma sequência $(C_i)_{i \geq 1}$ em $\mathcal{K}$ e suponha que

$$B_m := \bigcap_{i=1}^m C_i \neq \emptyset \quad \text{para todo } m \geq 1.$$

Para cada i , escolha r_i tal que $C_i \in \mathcal{K}_{r_i}$ e defina $d_m = \max\{r_1, \dots, r_m\}$. Como os cilindros de menor nível são fechados e algum dos cilindros que formam B_m tem base compacta em $\mathbb{R}^{d_m}$, existe um compacto não vazio $L_m \subseteq \mathbb{R}^{d_m}$ tal que

$$B_m = L_m \times \mathbb{R}^{N-d_m}.$$

A sequência $(d_m)_{m \geq 1}$ é não decrescente. Distinguímos os mesmos dois casos sugeridos pelos níveis dos cilindros.

Se (d_m) é limitada, existe d e existe m_0 tais que $d_m = d$ para $m \geq m_0$. Nesse caso os compactos não vazios $L_m \subseteq \mathbb{R}^d$ são decrescentes. A compacidade de L_{m_0} implica $\bigcap_{m \geq m_0} L_m \neq \emptyset$, e portanto

$$\bigcap_{i=1}^{\infty} C_i = \bigcap_{m=m_0}^{\infty} B_m = \left(\bigcap_{m=m_0}^{\infty} L_m \right) \times \mathbb{R}^{N-d} \neq \emptyset.$$

Suponha agora que $d_m \rightarrow \infty$. Escolha $x^{(m)} \in B_m$ para cada m . Fixada a coordenada j , tome m_j tal que $d_{m_j} \geq j$. Para $m \geq m_j$, temos $x^{(m)} \in B_m \subseteq B_{m_j}$; assim, a j -ésima coordenada de $x^{(m)}$ pertence, a partir de m_j , à projeção compacta $\pi_j(L_{m_j})$. Extraímos sucessivamente uma subsequência na qual a primeira coordenada converge, depois uma subsequência desta na qual a segunda coordenada converge, e assim por diante. O procedimento diagonal fornece uma subsequência $(y^{(\ell)})_{\ell \geq 1}$ de $(x^{(m)})_{m \geq 1}$ que converge coordenada a coordenada para algum $y \in \mathbb{R}^N$.

Fixe $p \geq 1$. A partir de certo índice, $y^{(\ell)} \in B_p$. Escrevendo $B_p = L_p \times \mathbb{R}^{N-d_p}$, as primeiras d_p coordenadas de $y^{(\ell)}$ convergem para as de y . Como L_p é fechado, $y \in B_p$. Isso vale para todo p , logo

$$y \in \bigcap_{p=1}^{\infty} B_p = \bigcap_{i=1}^{\infty} C_i.$$

Em ambos os casos, a propriedade da interseção finita implica que a interseção enumerável é não vazia. Pela contrapositiva, $\mathcal{K}$ é uma classe compacta.

□

8.5. Existência de Variáveis Aleatórias Independentes

O produto de medidas não apenas descreve independência: ele permite construí-la. As projeções coordenadas do espaço produto terão as leis prescritas e serão independentes por definição da medida.

Teorema 8.24 —

Dadas probabilidades $\mathbf{P}_1, \dots, \mathbf{P}_n$ em $\mathbb{R}$, então existem variáveis aleatórias independentes $\xi_1, \dots, \xi_n$ com distribuições $\mathbf{P}_1, \dots, \mathbf{P}_n$.

Demonstração. Considere o espaço de probabilidade

$$(\Omega, \mathcal{F}, \mathbf{P}) = (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mathbf{P}_1 \otimes \dots \otimes \mathbf{P}_n)$$

e defina $\xi_k(x_1, \dots, x_n) = x_k$. As projeções coordenadas são mensuráveis. Para $A_1, \dots, A_n \in \mathcal{B}(\mathbb{R})$,

$$\mathbf{P}(\xi_1 \in A_1, \dots, \xi_n \in A_n) = (\mathbf{P}_1 \otimes \dots \otimes \mathbf{P}_n)(A_1 \times \dots \times A_n) = \prod_{k=1}^n \mathbf{P}_k(A_k).$$

Logo $\xi_1, \dots, \xi_n$ são independentes e a distribuição de ξ_k é $\mathbf{P}_k$. □

Podemos generalizar o resultado anterior para sequências de variáveis aleatórias i.i.d.

Teorema 8.25 — Sequências de variáveis i.i.d.

Seja F uma distribuição em $\mathbb{R}$, então existe uma sequência de variáveis aleatórias independentes $\xi_1, \dots, \xi_n, \dots$ com distribuição F .

Demonstração. Para cada $n \geq 1$, seja $\mathbf{P}_n = F^{\otimes n}$ em $\mathbb{R}^n$. Essas probabilidades são consistentes, pois

$$\mathbf{P}_{n+1}(A \times \mathbb{R}) = F^{\otimes n}(A)F(\mathbb{R}) = \mathbf{P}_n(A)$$

para todo $A \in \mathcal{B}(\mathbb{R}^n)$. Pelo Teorema de Extensão de Kolmogorov, existe uma probabilidade $\mathbf{P}$ em $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}(\mathbb{R})^{\otimes \mathbb{N}})$ com essas marginais. Defina $\xi_k(\omega) = \omega_k$. As projeções são mensuráveis e, para $A_1, \dots, A_n \in \mathcal{B}(\mathbb{R})$,

$$\mathbf{P}(\xi_1 \in A_1, \dots, \xi_n \in A_n) = \prod_{k=1}^n F(A_k).$$

Logo as coordenadas são independentes e cada uma tem distribuição F . □

Agora vamos traduzir a teoria desenvolvida até aqui para densidades. Uma função mensurável $f : \mathbb{R} \rightarrow [0, \infty)$ é uma densidade de ξ se

$$\mathbf{P}(\xi \in A) = \int_A f(x) \, dx \quad \text{para todo } A \in \mathcal{B}(\mathbb{R}).$$

Necessariamente $\int_{\mathbb{R}} f \, dx = 1$, e a densidade é determinada apenas quase em toda parte. Do mesmo modo, uma densidade conjunta de $\xi = (\xi_1, \dots, \xi_n)$ é uma função mensurável $\mathbf{f} : \mathbb{R}^n \rightarrow [0, \infty)$ tal que

$$\mathbf{P}(\xi \in A) = \int_A \mathbf{f}(x) \, dx \quad \text{para todo } A \in \mathcal{B}(\mathbb{R}^n).$$

Teorema 8.26 —

Sejam $\xi_1, \dots, \xi_n$ variáveis aleatórias com densidades marginais $f_1, \dots, f_n$ e densidade conjunta $\mathbf{f}$. Então $\xi_1, \dots, \xi_n$ são independentes se e somente se

$$\mathbf{f}(x_1, \dots, x_n) = \prod_{i=1}^n f_i(x_i)$$

para quase todo $(x_1, \dots, x_n) \in \mathbb{R}^n$.

Para passar de identidades de integrais a identidades pontuais entre densidades, usaremos a seguinte ferramenta de análise real.

Teorema 8.27 — Teorema de Diferenciação de Lebesgue

Se $f \in L^1(\mathbb{R})$, então

$$\lim_{\epsilon \downarrow 0} \frac{1}{\epsilon} \int_x^{x+\epsilon} f(y) \, dy = f(x)$$

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

Demonstração. É o caso $n = 1$ da versão multidimensional enunciada a seguir. $\square$

Teorema 8.28 — Diferenciação de Lebesgue em $\mathbb{R}^n$

Se $f \in L^1(\mathbb{R}^n)$ e $Q_\epsilon(x) = \prod_{i=1}^n [x_i, x_i + \epsilon]$, então

$$\lim_{\epsilon \downarrow 0} \frac{1}{\epsilon^n} \int_{Q_\epsilon(x)} f(y) \, dy = f(x)$$

para quase todo $x \in \mathbb{R}^n$.

Demonstração. A prova analítica parte da desigualdade maximal fraca (1,1) de Hardy–Littlewood. Para $g \in C_c(\mathbb{R}^n)$, a média sobre $Q_\epsilon(x)$ converge uniformemente a $g(x)$ em compactos. Dado $f \in L^1$, aproxime-o em L^1 por tais funções g . A desigualdade maximal controla, por uma constante vezes $\|f - g\|_1/t$, a medida do conjunto onde o limite superior das médias de $|f - g|$ excede t . Fazendo primeiro $\|f - g\|_1 \rightarrow 0$ e depois $t \downarrow 0$, obtém-se a convergência quase em toda parte. Essa ferramenta clássica, cuja prova completa pertence à análise real, pode ser consultada em [17]. $\square$

Prova do Teorema 8.26. Suponha primeiro que as variáveis sejam independentes. Para $x = (x_1, \dots, x_n)$ e $\epsilon > 0$, a independência dá

$$\begin{aligned} \frac{1}{\epsilon^n} \int_{Q_\epsilon(x)} f(y) \, dy &= \frac{1}{\epsilon^n} \prod_{i=1}^n \mathbf{P}(x_i \leq \xi_i \leq x_i + \epsilon) \\ &= \prod_{i=1}^n \left(\frac{1}{\epsilon} \int_{x_i}^{x_i + \epsilon} f_i(y_i) \, dy_i \right). \end{aligned}$$

Pelos dois teoremas de diferenciação, o lado esquerdo converge a $f(x)$ para quase todo x , e cada fator do lado direito converge a $f_i(x_i)$ fora de um conjunto nulo. Pelo Teorema de Fubini, a união dos conjuntos excepcionais coordenados ainda tem medida de Lebesgue nula em $\mathbb{R}^n$. Logo $f(x) = \prod_i f_i(x_i)$ quase em toda parte.

Reciprocamente, suponha essa fatoração. Para $A_1, \dots, A_n \in \mathcal{B}(\mathbb{R})$, Fubini–Tonelli

fornece

$$\begin{aligned} \mathbf{P}(\xi_1 \in A_1, \dots, \xi_n \in A_n) &= \int_{A_1 \times \dots \times A_n} \prod_{i=1}^n f_i(x_i) \, dx \\ &= \prod_{i=1}^n \int_{A_i} f_i(x_i) \, dx_i = \prod_{i=1}^n \mathbf{P}(\xi_i \in A_i). \end{aligned}$$

Essa é a definição de independência. □

Vejamos alguns exemplos.

Exemplo 8.29 — Distribuição Uniforme.

Sejam ξ e η variáveis aleatórias independentes, distribuídas uniformemente em $[0,1]$. Assim, a densidade de ξ e η é dada pela função de densidade

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

Portanto, a densidade conjunta de ξ e η é

$$\varphi(x, y) = \begin{cases} 1, & \text{se } (x, y) \in [0,1] \times [0,1] \\ 0, & \text{caso contrário.} \end{cases}$$

◁

Exemplo 8.30 — Distribuição Gaussiana em $\mathbb{R}^n$.

Também denominada de distribuição normal multivariada. A densidade de cada coordenada ξ_k é dada pela gaussiana padrão unidimensional

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

e as coordenadas são independentes, então a densidade de $(\xi_1, \dots, \xi_n)$ é

$$\varphi(x_1, \dots, x_n) = \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{\|x\|^2}{2}\right),$$

onde $\|x\|^2 = x_1^2 + \dots + x_n^2$.

◁

Faremos agora uma demonstração alternativa da afirmação de que para variáveis independentes a esperança do produto é o produto das esperanças. Compare essa demonstração com a demonstração do Teorema 7.21.

Teorema 8.31 —

Sejam $\xi, \eta \in \mathcal{L}^1(\mathbf{P})$ independentes. Então $(\xi, \eta) \in \mathcal{L}^1(\mathbf{P})$ e

$$\mathbf{E}[\xi\eta] = \mathbf{E}[\xi] \times \mathbf{E}[\eta].$$

Demonstração. A distribuição conjunta de (ξ, η) é $\mathbf{P}_\xi \otimes \mathbf{P}_\eta$. Por Tonelli,

$$\mathbf{E}|\xi\eta| = \int_{\mathbb{R}^2} |xy| \, d(\mathbf{P}_\xi \otimes \mathbf{P}_\eta)(x, y) = \left(\int_{\mathbb{R}} |x| \, d\mathbf{P}_\xi(x) \right) \left(\int_{\mathbb{R}} |y| \, d\mathbf{P}_\eta(y) \right) < \infty.$$

Portanto $\xi\eta$ é integrável e Fubini pode ser aplicado:

$$\mathbf{E}[\xi\eta] = \int_{\mathbb{R}^2} xy \, d(\mathbf{P}_\xi \otimes \mathbf{P}_\eta)(x, y) = \left(\int_{\mathbb{R}} x \, d\mathbf{P}_\xi(x) \right) \left(\int_{\mathbb{R}} y \, d\mathbf{P}_\eta(y) \right) = \mathbf{E}[\xi]\mathbf{E}[\eta].$$

□

8.6. Convolução e Soma de Variáveis Aleatórias Independentes

Se conhecemos apenas as distribuições de ξ e η , a distribuição de $\xi + \eta$ não está determinada, embora sua esperança esteja, quando existe. Para ver isso, tome uma variável simétrica ξ , isto é, $\xi \stackrel{d}{=} -\xi$. Um exemplo é o resultado de uma moeda honesta, com cara representada por 1 e coroa por -1 .

As somas $\xi + \xi$ e $\xi + (-\xi)$ têm parcelas com as mesmas distribuições, mas a primeira é 2ξ e a segunda é identicamente nula. Portanto, é preciso conhecer a distribuição conjunta dos somandos. Sob independência, ela é o produto das marginais e podemos calcular a lei da soma.

Teorema 8.32 —

Se ξ e η são variáveis aleatórias independentes com funções de distribuição F e G , respectivamente, então a função de distribuição de $\xi + \eta$ é

$$H(z) = \int_{\mathbb{R}} F(z - y) \, dG(y) = \int_{\mathbb{R}} G(z - x) \, dF(x).$$

Demonstração. Pela independência, a distribuição conjunta de (ξ, η) é $\mathbf{P}_{\xi} \otimes \mathbf{P}_{\eta}$. Tonelli aplicado à indicadora do semiplano $\{(x, y) : x + y \leq z\}$ dá

$$\begin{aligned} H(z) &= \int_{\mathbb{R}^2} \mathbb{1}_{\{x+y \leq z\}} \, d(\mathbf{P}_{\xi} \otimes \mathbf{P}_{\eta})(x, y) \\ &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \mathbb{1}_{\{x \leq z-y\}} \, d\mathbf{P}_{\xi}(x) \right) d\mathbf{P}_{\eta}(y) \\ &= \int_{\mathbb{R}} F(z - y) \, d\mathbf{P}_{\eta}(y) = \int_{\mathbb{R}} F(z - y) \, dG(y). \end{aligned}$$

Trocando a ordem de integração, obtemos $H(z) = \int_{\mathbb{R}} G(z - x) \, dF(x)$. □

Corolário 8.33 —

Se ξ e η são independentes e têm densidades f e g , respectivamente, então $\xi + \eta$ tem densidade

$$h(z) = \int_{\mathbb{R}} f(z - y)g(y) \, dy.$$

A integral anterior é denotada por $h = f \star g$ e é conhecida como **convolução** de f e g .

Demonstração. A função h é mensurável e não negativa, por Tonelli. Para todo $A \in \mathcal{B}(\mathbb{R})$, novamente por Tonelli e pela mudança de variável $z = x + y$,

$$\begin{aligned} \int_A h(z) \, dz &= \int_{\mathbb{R}} \int_{\mathbb{R}} \mathbb{1}_A(z) f(z-y) g(y) \, dz \, dy \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} \mathbb{1}_A(x+y) f(x) g(y) \, dx \, dy \\ &= \mathbf{P}(\xi + \eta \in A). \end{aligned}$$

Logo h é uma densidade de $\xi + \eta$. Tomando $A = \mathbb{R}$, vemos também que $\int_{\mathbb{R}} h = 1$. $\square$

Exemplo 8.34 — .

Sejam ξ e η variáveis aleatórias uniformes independentes em $[0,1]$. A densidade de $\xi + \eta$ é

$$h(x) = \int_{\mathbb{R}} \mathbb{1}_{[0,1]}(x-y) \mathbb{1}_{[0,1]}(y) \, dy.$$

O integrando vale 1 exatamente quando $\max(0, x-1) \leq y \leq \min(1, x)$. Portanto,

$$h(x) = \begin{cases} x, & 0 \leq x \leq 1, \\ 2-x, & 1 < x \leq 2, \\ 0, & \text{caso contrário.} \end{cases}$$

$\triangleleft$

A convolução também permite suavizar funções. Veremos como usá-la para aproximar funções contínuas por funções suaves.

Definição 8.35 — Funções Teste

Uma função é dita **função teste** ou suave de suporte compacto, denotada por $f \in C_c^\infty(\mathbb{R})$ se f é de classe C^∞ e seu suporte

$$\text{supp}(f) := \overline{\{x \in \mathbb{R} : f(x) \neq 0\}} = \overline{f^{-1}(\mathbb{R} \setminus \{0\})}.$$

é um conjunto compacto.

Para começar, estabelecemos a existência de uma função teste em $[-1,1]$.

Lema 8.36 —

A função

$$f(x) = \begin{cases} e^{\frac{-1}{1-x^2}} & |x| < 1 \\ 0 & |x| \geq 1 \end{cases}$$

pertence a $C_c^\infty(\mathbb{R})$ e tem suporte $[-1,1]$.

Demonstração. No intervalo $(-1,1)$, a função é suave. Uma indução usando as regras do produto e da cadeia mostra que, para cada $m \geq 0$, existem um polinômio P_m e um inteiro $N_m \geq 0$ tais que

$$f^{(m)}(x) = \frac{P_m(x)}{(1-x^2)^{N_m}} \exp\left(-\frac{1}{1-x^2}\right), \quad |x| < 1.$$

Para todo $N \geq 0$, $t^{-N}e^{-1/t} \rightarrow 0$ quando $t \downarrow 0$. Tomando $t = 1 - x^2$, concluímos que $f^{(m)}(x) \rightarrow 0$ quando $x \rightarrow \pm 1$ pelo interior. Como f é identicamente nula fora $[-1,1]$, todas as derivadas se prolongam continuamente por zero nos dois extremos. Logo $f \in C_c^\infty(\mathbb{R})$; como $f > 0$ em $(-1,1)$, seu suporte é $[-1,1]$. $\square$

Normalizando a função do lema anterior por sua integral, obtemos uma função $\rho \geq 0$ em $C_c^\infty(\mathbb{R})$, com suporte em $[-1,1]$ e integral igual a 1.

Lema 8.37 —

Seja $\rho \geq 0$ uma função em $C_c^\infty(\mathbb{R})$, com suporte em $[-1,1]$ e $\int_{\mathbb{R}} \rho = 1$. Se $f : \mathbb{R} \rightarrow \mathbb{R}$ é contínua, defina

$$f_n(x) = n \int_{-n}^n \rho(n(x-y))f(y) \, dy.$$

Então $f_n \in C_c^\infty(\mathbb{R})$,

$$f_n^{(m)}(x) = n^{m+1} \int_{-n}^n \rho^{(m)}(n(x-y))f(y) \, dy,$$

e $f_n \rightarrow f$ uniformemente em conjuntos compactos. Se f é limitada, então $\|f_n\|_\infty \leq \|f\|_\infty$.

Demonstração. Ponha $\rho_n(u) = n\rho(nu)$. Seu suporte está contido em $[-1/n, 1/n]$, e

$$f_n(x) = \int_{-n}^n \rho_n(x-y)f(y) \, dy.$$

Daí $\text{supp}(f_n) \subseteq [-n-1/n, n+1/n]$. Como f é integrável em intervalos compactos, a Convergência Dominada permite diferenciar sob o sinal de integral, para cada n fixo. Iterando,

$$f_n^{(m)}(x) = \int_{-n}^n \rho_n^{(m)}(x-y)f(y) \, dy = n^{m+1} \int_{-n}^n \rho^{(m)}(n(x-y))f(y) \, dy.$$

O mesmo argumento mostra que essas derivadas são contínuas; portanto $f_n \in C_c^\infty(\mathbb{R})$.

Seja $K \subset \mathbb{R}$ compacto. A função f é uniformemente contínua no compacto $\{y : d(y, K) \leq 1\}$. Dado $\varepsilon > 0$, escolha $\delta > 0$ tal que $|f(y) - f(x)| < \varepsilon$ quando $x \in K$, $d(y, K) \leq 1$ e $|x-y| < \delta$. Para n suficientemente grande, $1/n < \delta$ e $[x-1/n, x+1/n] \subset [-n, n]$ para todo $x \in K$. Usando $\int \rho_n = 1$, obtemos, uniformemente em $x \in K$,

$$|f_n(x) - f(x)| \leq \int \rho_n(x-y)|f(y) - f(x)| \, dy \leq \varepsilon.$$

Por fim, se f é limitada,

$$|f_n(x)| \leq \|f\|_\infty \int_{-n}^n \rho_n(x-y) \, dy \leq \|f\|_\infty,$$

o que conclui a prova. □

8.6.1. Testando Convergência com Funções Suaves

Lema 8.38 —

Sejam $(\xi_n)_{n \geq 1}$ e ξ variáveis aleatórias reais. Então $\xi_n \xrightarrow{d} \xi$ se e somente se

$$\lim_{n \rightarrow \infty} \mathbf{E}[f(\xi_n)] = \mathbf{E}[f(\xi)]$$

para toda $f \in C_c^\infty(\mathbb{R})$.

Demonstração. Se $\xi_n \xrightarrow{d} \xi$ e F_n, F são as respectivas funções de distribuição, então $F_n(t) \rightarrow F(t)$ em todo ponto de continuidade de F , portanto quase em toda parte. Para $f \in C_c^\infty(\mathbb{R})$,

$$f(x) = - \int_x^\infty f'(t) \, dt.$$

Por Fubini,

$$\mathbf{E}[f(\xi_n)] = - \int_{\mathbb{R}} f'(t) F_n(t) \, dt.$$

Como $|F_n| \leq 1$ e f' tem suporte compacto, a Convergência Dominada dá $\mathbf{E}[f(\xi_n)] \rightarrow \mathbf{E}[f(\xi)]$.

Reciprocamente, suponha a convergência das esperanças para toda função teste. Primeiro obteremos controle uniforme das caudas. Para $R > 0$, escolha $\psi_R \in C_c^\infty(\mathbb{R})$ tal que

$$\mathbb{1}_{[-R/2, R/2]} \leq \psi_R \leq \mathbb{1}_{[-R, R]}.$$

Então

$$\limsup_{n \rightarrow \infty} \mathbf{P}(|\xi_n| > R) \leq 1 - \mathbf{E}[\psi_R(\xi)] \leq \mathbf{P}(|\xi| > R/2).$$

O último termo tende a zero quando $R \rightarrow \infty$.

Seja agora g uma função contínua limitada e ponha $M = \|g\|_\infty$. O caso $M = 0$ é trivial. Dado $\varepsilon > 0$, escolha R de modo que

$$\mathbf{P}(|\xi| > R) < \frac{\varepsilon}{16M}, \quad \mathbf{P}(|\xi_n| > R) < \frac{\varepsilon}{16M}$$

para todo n suficientemente grande. Pelo lema dos molificadores, existe $g_k \in C_c^\infty(\mathbb{R})$ com $\|g_k\|_\infty \leq M$ e $\sup_{|x| \leq R} |g_k(x) - g(x)| < \varepsilon/8$ para k grande. Fixado tal k e tomando depois n grande,

$$\begin{aligned} |\mathbf{E}[g(\xi_n)] - \mathbf{E}[g(\xi)]| &\leq \mathbf{E}|g - g_k|(\xi_n) + |\mathbf{E}[g_k(\xi_n)] - \mathbf{E}[g_k(\xi)]| + \mathbf{E}|g_k - g|(\xi) \\ &< \varepsilon. \end{aligned}$$

Logo as esperanças convergem para toda g contínua e limitada.

Por fim, seja x um ponto de continuidade de F . Para $\delta > 0$, escolha funções contínuas $g_\delta^-, g_\delta^+ : \mathbb{R} \rightarrow [0, 1]$ com

$$g_\delta^- \leq \mathbb{1}_{(-\infty, x]} \leq g_\delta^+,$$

tais que $g_\delta^- = 1$ em $(-\infty, x - \delta]$ e zero em $[x, \infty)$, enquanto $g_\delta^+ = 1$ em $(-\infty, x]$ e zero em $[x + \delta, \infty)$. A convergência já provada para funções contínuas limitadas implica

$$\mathbf{E}[g_\delta^-(\xi)] \leq \liminf_n F_n(x) \leq \limsup_n F_n(x) \leq \mathbf{E}[g_\delta^+(\xi)].$$

Fazendo $\delta \downarrow 0$ e usando a continuidade de F em x , os extremos convergem a $F(x)$. Portanto $F_n(x) \rightarrow F(x)$ e $\xi_n \xrightarrow{d} \xi$. $\square$

8.7. Aplicações

Encerramos o capítulo com duas aplicações em que a estrutura produto é menos visível, mas decisiva: simetrias de sequências independentes e passagem ao limite em distribuições.

8.7.1. Lei 0 – 1 de Hewitt-Savage

Definição 8.39 –

Uma bijeção $\pi : \mathbb{N} \rightarrow \mathbb{N}$ é uma **permutação finita** se existe N_0 tal que $\pi(n) = n$ para todo $n > N_0$.

Se $\xi = (\xi_1, \xi_2, \dots)$ é uma sequência de variáveis aleatórias, escrevemos

$$T_\pi \xi = (\xi_{\pi(1)}, \xi_{\pi(2)}, \dots).$$

Um evento $A = \{\xi \in B\} \in \sigma(\xi_1, \xi_2, \dots)$ é **simétrico** se $\{\xi \in B\} = \{T_\pi \xi \in B\}$ para toda permutação finita π . Os eventos simétricos formam a σ -álgebra permutável $\mathcal{E}$.

Teorema 8.40 – Lei 0–1 de Hewitt-Savage

Se $(\xi_n)_{n \geq 1}$ é uma sequência de variáveis aleatórias independentes e identicamente distribuídas, então

$$\mathbf{P}(A) \in \{0, 1\} \quad \text{para todo } A \in \mathcal{E}.$$

Demonstração. Fixe $A \in \mathcal{G}$ e $\varepsilon > 0$. A álgebra dos eventos que dependem de um número finito de coordenadas gera $\sigma(\xi_1, \xi_2, \dots)$. Pelo Lema 2.13, existe um evento $C \in \sigma(\xi_1, \dots, \xi_n)$ tal que

$$\mathbf{P}(A \triangle C) < \varepsilon$$

para algum n . Seja π a permutação que troca os blocos $\{1, \dots, n\}$ e $\{n+1, \dots, 2n\}$ e fixa os demais índices, e ponha $C' = \{T_\pi \xi \in D\}$ quando $C = \{\xi \in D\}$.

Como a sequência é i.i.d., sua lei é invariante por π ; como A é simétrico,

$$\mathbf{P}(A \triangle C') = \mathbf{P}(A \triangle C) < \varepsilon \quad \text{e} \quad \mathbf{P}(C') = \mathbf{P}(C).$$

Além disso, C depende das primeiras n coordenadas e C' das n coordenadas seguintes, logo são independentes. Portanto

$$\mathbf{P}(C \cap C') = \mathbf{P}(C)^2.$$

Como

$$\mathbf{P}(A \triangle (C \cap C')) \leq \mathbf{P}(A \triangle C) + \mathbf{P}(A \triangle C') < 2\varepsilon,$$

temos $|\mathbf{P}(A) - \mathbf{P}(C)^2| < 2\varepsilon$ e $|\mathbf{P}(C) - \mathbf{P}(A)| < \varepsilon$. Fazendo $\varepsilon \downarrow 0$, obtemos $\mathbf{P}(A) = \mathbf{P}(A)^2$, isto é, $\mathbf{P}(A) \in \{0, 1\}$. □

8.7.2. Passeios Aleatórios

Se $(\xi_i)_{i \geq 1}$ é uma sequência de vetores aleatórios i.i.d. em $\mathbb{R}^d$, então

$$S_n = \sum_{i=1}^n \xi_i$$

é um passeio aleatório em $\mathbb{R}^d$.

Teorema 8.41 —

Se S_n é um passeio aleatório em $\mathbb{R}$, então, quase certamente, exatamente uma das seguintes alternativas ocorre:

- ☐ $S_n = 0$ para todo n ;
- ☐ $S_n \rightarrow \infty$;
- ☐ $S_n \rightarrow -\infty$;
- ☐ $\liminf_n S_n = -\infty$ e $\limsup_n S_n = \infty$.

Demonstração. Uma permutação finita dos incrementos não altera S_n para todo n suficientemente grande. Logo os eventos $\{\limsup_n S_n \leq t\}$ e $\{\liminf_n S_n \leq t\}$ são simétricos. Pela Lei 0–1 de Hewitt–Savage, as variáveis aleatórias estendidas

$$L = \limsup_n S_n, \quad \ell = \liminf_n S_n$$

são constantes quase certamente; denote essas constantes também por $L, \ell \in [-\infty, \infty]$.

Considere o passeio deslocado $S'_n = \xi_2 + \dots + \xi_{n+1}$. Ele tem a mesma lei que (S_n) , de modo que $\limsup_n S'_n = L$ e $\liminf_n S'_n = \ell$ quase certamente. Como $S_{n+1} = \xi_1 + S'_n$, se L é finito então $L = \xi_1 + L$, e se ℓ é finito então $\ell = \xi_1 + \ell$. Em qualquer desses casos, $\xi_1 = 0$ quase certamente e, pela distribuição idêntica, $S_n = 0$ para todo n quase certamente.

Fora desse caso degenerado, L e ℓ pertencem a $\{-\infty, \infty\}$. Como $\ell \leq L$, restam somente $(\ell, L) = (\infty, \infty)$, $(-\infty, -\infty)$ ou $(-\infty, \infty)$, que correspondem às três alternativas restantes. □

8.8. Extensão de Kolmogorov geral

Nesta seção generalizamos o Teorema de Extensão de Kolmogorov para produtos arbitrários. Queremos construir uma medida no espaço

$$\Omega_A := \prod_{\alpha \in A} \Omega_\alpha$$

associado a uma família de espaços $\{(\Omega_\alpha, \mathcal{B}_\alpha)\}_{\alpha \in A}$, todos não vazios. Seus elementos são $x_A = (x_\alpha)_{\alpha \in A}$.

De modo correspondente, temos as aplicações de projeção nas coordenadas $\pi_\beta : \Omega_A \rightarrow \Omega_\beta$, que leva x_A para x_β . Também podemos definir a composta de tais aplicações de projeção de coordenadas.

Por exemplo, dados $B \subset A$, podemos definir a projeção parcial $\pi_B : \Omega_A \rightarrow \Omega_B$ de modo natural. Mais geralmente se $C \subset B \subset A$, podemos definir $\pi_{C \leftarrow B} : \Omega_B \rightarrow \Omega_C$. Então temos as leis de composição

$$\pi_{D \leftarrow C} \circ \pi_{C \leftarrow B} = \pi_{D \leftarrow B}$$

se $D \subset C \subset B \subset A$.

Podemos definir a σ -álgebra puxada por cada coordenada:

$$\pi_\beta^*(\mathcal{B}_\beta) := \{\pi_\beta^{-1}(E) : E \in \mathcal{B}_\beta\},$$

e então a σ -álgebra produto

$$\mathcal{B}_A = \bigotimes_{\beta \in A} \mathcal{B}_\beta := \sigma \left\langle \bigcup_{\beta \in A} \pi_\beta^*(\mathcal{B}_\beta) \right\rangle.$$

Esta é a menor σ -álgebra que torna mensuráveis todas as projeções π_β e, mais geralmente, todas as projeções π_B para B finito.

Todo evento do produto depende, de fato, de no máximo uma quantidade enumerável de coordenadas. Para justificar essa afirmação, considere

$$\mathcal{M} = \{\pi_B^{-1}(E_B) : B \subset A \text{ enumerável}, E_B \in \mathcal{B}_B\}.$$

A família $\mathcal{M}$ contém os cilindros coordenados que geram $\mathcal{B}_A$ e é fechada por complementos. Se $E_n = \pi_{B_n}^{-1}(D_n) \in \mathcal{M}$, então $B = \bigcup_n B_n$ é enumerável. Escrevendo $\rho_n : \Omega_B \rightarrow \Omega_{B_n}$ para a projeção natural,

$$E_n = \pi_B^{-1}(\rho_n^{-1}(D_n)),$$

e, portanto,

$$\bigcup_n E_n = \pi_B^{-1}\left(\bigcup_n \rho_n^{-1}(D_n)\right) \in \mathcal{M}.$$

Logo $\mathcal{M}$ é uma σ -álgebra e $\mathcal{B}_A \subseteq \mathcal{M}$. A inclusão oposta vale porque cada π_B é mensurável para a σ -álgebra produto. Assim $\mathcal{M} = \mathcal{B}_A$.

Em particular, se $E \in \mathcal{B}_A$, existem um conjunto enumerável $B \subset A$ e $E_B \in \mathcal{B}_B$ tais que $E = \pi_B^{-1}(E_B)$. Se $f : \Omega_A \rightarrow [0, \infty]$ é mensurável, aplique esse fato aos eventos $\{f > q\}$, $q \in \mathbb{Q}_{\geq 0}$, e tome a união dos conjuntos de coordenadas assim obtidos. Há, portanto, um único conjunto enumerável $B \subset A$ que determina simultaneamente todos esses eventos. Se dois pontos de Ω_A têm a mesma projeção em Ω_B , eles pertencem aos mesmos eventos $\{f > q\}$ para todo racional $q \geq 0$ e, por isso, têm o mesmo valor de f . Como os fatores são não vazios, π_B é sobrejetora; podemos então definir $f_B : \Omega_B \rightarrow [0, \infty]$ por $f_B(\pi_B(x)) = f(x)$. As pré-imagens $\{f_B > q\}$ são mensuráveis para todo $q \in \mathbb{Q}_{\geq 0}$, logo f_B é mensurável e

$$f = f_B \circ \pi_B.$$

As projeções coordenadas tornam-se, assim, uma família de variáveis aleatórias definida num único espaço mensurável. Se μ_A é uma probabilidade em $\mathcal{B}_A$, sua marginal em Ω_B , para $B \subset A$, é

$$\mu_B(E_B) := (\pi_B)_* \mu_A(E_B) = \mu_A(\pi_B^{-1}(E_B)),$$

para $E_B \in \mathcal{B}_B$. Essas marginais satisfazem

$$(\pi_{C \leftarrow B})_* \mu_B = \mu_C$$

sempre que $C \subset B \subset A$.

Teorema 8.42 — Extensão de Kolmogorov

Seja $\{(\Omega_\alpha, \mathcal{B}_\alpha)\}_{\alpha \in A}$ uma família de espaços de Borel Hausdorff não vazios. Para cada subconjunto finito $B \subset A$, seja μ_B uma probabilidade internamente regular em Ω_B , equipado com a topologia produto. Suponha que

$$(\pi_{C \leftarrow B})_* \mu_B = \mu_C$$

sempre que $C \subset B \subset A$, com B e C finitos. Então existe uma única probabilidade μ_A em $\mathcal{B}_A$ que satisfaz $(\pi_B)_* \mu_A = \mu_B$ para todo $B \subset A$ finito.

Demonstração. Seja $\mathcal{A}_0$ a álgebra dos cilindros de dimensão finita:

$$\mathcal{A}_0 = \{\pi_B^{-1}(E) : B \subset A \text{ finito}, E \in \mathcal{B}_B\}.$$

Defina

$$\mu_0(\pi_B^{-1}(E)) := \mu_B(E).$$

Essa definição independe da representação. De fato, duas representações podem ser levadas ao conjunto finito comum de coordenadas $D = B \cup C$, e a compatibilidade das marginais mostra que ambas recebem o mesmo valor. Pelo mesmo procedimento de levar uma coleção finita de cilindros a um conjunto comum de coordenadas, vemos que μ_0 é uma probabilidade finitamente aditiva em $\mathcal{A}_0$.

Considere a família de cilindros com base compacta

$$\mathcal{K} = \{\pi_B^{-1}(K) : B \subset A \text{ finito}, K \subset \Omega_B \text{ compacto}\}.$$

A regularidade interior de cada μ_B dá, para $E = \pi_B^{-1}(E_B) \in \mathcal{A}_0$,

$$\mu_0(E) = \sup\{\mu_0(C) : C \in \mathcal{K}, C \subset E\}.$$

Verifiquemos cuidadosamente que $\mathcal{K}$ é uma classe compacta. Se $C_n = \pi_{B_n}^{-1}(K_n)$ possui a propriedade da interseção finita, ponha $J = \bigcup_n B_n$, que é enumerável. Para cada $\alpha \in J$, escolha um índice $n(\alpha)$ com $\alpha \in B_{n(\alpha)}$ e defina

$$L_\alpha = \pi_{\{\alpha\} \leftarrow B_{n(\alpha)}}(K_{n(\alpha)}).$$

Cada L_α é compacto e não vazio. Pelo Teorema de Tychonoff, $L = \prod_{\alpha \in J} L_\alpha$ é compacto. Para cada n , o conjunto

$$D_n = \{x \in L : \pi_{B_n \leftarrow J}(x) \in K_n\}$$

é fechado em L , pois K_n é compacto e Ω_{B_n} é Hausdorff. A família (D_n) tem a propriedade da interseção finita: para uma coleção finita de índices, acrescente também, para cada coordenada que ela envolve, os cilindros $C_{n(\alpha)}$; a propriedade da interseção finita dos C_n fornece as coordenadas necessárias em L , e as demais podem ser escolhidas arbitrariamente. Pela compacidade de L , $\bigcap_n D_n \neq \emptyset$. Estendendo um ponto dessa interseção às coordenadas de $A \setminus J$, obtemos

$$\bigcap_{n=1}^{\infty} C_n \neq \emptyset.$$

Portanto $\mathcal{K}$ é uma classe compacta.

Este é o ponto decisivo da construção: a classe compacta converte uma interseção enumerável vazia de aproximantes numa interseção finita já vazia. Assim, a aditividade finita de μ_0 basta para controlar sua continuidade no vazio.

O Teorema de Extensão Compacta 2.29 transforma agora a aditividade finita de μ_0 em continuidade no vazio e, portanto, em σ -aditividade na álgebra $\mathcal{A}_0$: μ_0 é uma pré-medida. O Teorema de Extensão de Carathéodory fornece uma extensão a $\sigma\langle\mathcal{A}_0\rangle = \mathcal{B}_A$. Como $\mu_0(\Omega_A) = 1$, a extensão é única. Por construção, suas marginais finitas são exatamente as probabilidades μ_B . $\square$

Proposição 8.43 — Fórmula das caudas

Se X é uma variável aleatória não negativa, sem hipótese de integrabilidade, então

$$\mathbf{E}[X] = \int_0^{\infty} \mathbf{P}(X > t) \, dt,$$

com a igualdade entendida em $[0, \infty]$.

Demonstração. Para cada ω ,

$$X(\omega) = \int_0^\infty \mathbb{1}_{\{X(\omega) > t\}} dt.$$

O integrando é não negativo. Pelo Teorema de Tonelli,

$$\mathbf{E}X = \int_0^\infty \mathbf{E}[\mathbb{1}_{\{X > t\}}] dt = \int_0^\infty \mathbf{P}(X > t) dt.$$

□

EXERCÍCIOS

Produto

Seja $E \subseteq \Omega_1 \times \Omega_2$. Para $x \in \Omega_1$, defina a x -seção de E por

$$E^x = \{y \in \Omega_2 : (x, y) \in E\}.$$

Ex. 8.1 — Sejam $E, F \subseteq \Omega_1 \times \Omega_2$ e $x \in \Omega_1$. Prove que

1. $(E \cap F)^x = E^x \cap F^x$;
2. $(E^c)^x = (E^x)^c$, tomando os complementos, respectivamente, em $\Omega_1 \times \Omega_2$ e em Ω_2 ;
3. $(\bigcup_{n \geq 1} E_n)^x = \bigcup_{n \geq 1} E_n^x$.

Ex. 8.2 — Se $(\Omega_1, \mathcal{F}_1, \mu_1)$ e $(\Omega_2, \mathcal{F}_2, \mu_2)$ são espaços de medida σ -finitos, prove que a medida produto $\mu_1 \otimes \mu_2$ também é σ -finita. Construa explicitamente uma exaustão do espaço produto por retângulos de medida finita.

Ex. 8.3 — Prove que, para todo inteiro $d \geq 1$,

$$\mathcal{B}(\mathbb{R}^d) = \bigotimes_{i=1}^d \mathcal{B}(\mathbb{R}).$$

Faça as duas inclusões: use a continuidade das projeções numa direção e uma base enumerável de retângulos abertos na outra.

Ex. 8.4 — Uma prova alternativa da existência da medida produto. Sejam μ_1 e μ_2 probabilidades borelianas em $\mathbb{R}^m$ e $\mathbb{R}^n$, respectivamente. Na álgebra das uniões finitas disjuntas de retângulos borelianos, considere a função finitamente aditiva determinada por

$$\mu_0(A \times B) = \mu_1(A)\mu_2(B).$$

Use a regularidade interior das probabilidades borelianas e a classe dos compactos de $\mathbb{R}^{m+n}$ para verificar as hipóteses do Teorema de Extensão Compacta. Conclua, pelo Teorema de Carathéodory, a existência de $\mu_1 \otimes \mu_2$.

Ex. 8.5 — Construção quantílica de uma sequência i.i.d. O teorema de existência de sequências i.i.d. deste capítulo garante que, dada uma função de distribuição F em $\mathbb{R}$, existe uma sequência infinita de variáveis independentes com essa distribuição. Dê uma construção alternativa e explícita usando uma sequência de uniformes e a inversa generalizada de F .

Ex. 8.6 — Sejam ξ e η independentes, com

$$f_\xi(x) = \alpha e^{-\alpha x} \mathbb{1}_{[0, \infty)}(x), \quad \alpha > 0,$$

e $\eta \sim U(0,1)$. Determine a densidade de $Z = \xi + \eta$, apresentando-a separadamente nos intervalos $z < 0$, $0 \leq z < 1$ e $z \geq 1$.

Ex. 8.7 — Independência probabilística e independência linear. Seja V um espaço vetorial de dimensão finita sobre um corpo finito $\mathbb{F}_q$, seja ξ uniforme em V e seja

$$\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbb{F}_q$$

uma forma bilinear não degenerada. Dados vetores não nulos $v_1, \dots, v_n \in V$, prove que as variáveis aleatórias

$$\langle \xi, v_1 \rangle, \dots, \langle \xi, v_n \rangle$$

são independentes se, e somente se, $v_1, \dots, v_n$ são linearmente independentes.

Dica. Estude a imagem da transformação linear

$$T : V \rightarrow \mathbb{F}_q^n, \quad T(x) = (\langle x, v_1 \rangle, \dots, \langle x, v_n \rangle).$$

Ex. 8.8 — Distribuição do produto. Sejam ξ e η variáveis aleatórias independentes e estritamente positivas, com funções de distribuição F e G , respectivamente.

1. Se H é a função de distribuição de $\xi\eta$, prove que, para $z > 0$,

$$H(z) = \int_{(0,\infty)} F\left(\frac{z}{y}\right) dG(y).$$

2. Aplique a fórmula quando ξ e η são independentes e uniformes em $(0,1)$. Determine a função de distribuição e uma densidade de $\xi\eta$.

Ex. 8.9 — Infinitas moedas e o intervalo unitário. Considere

$$\Omega = \{0,1\}^{\mathbb{N}}, \quad \Phi(\omega) = \sum_{k=1}^{\infty} \frac{\omega_k}{2^k},$$

com a σ -álgebra produto e a probabilidade produto de moedas justas. Seja E o conjunto das sequências que são eventualmente iguais a 1.

1. Prove que E é enumerável e tem probabilidade zero.
2. Prove que a restrição $\Phi : \Omega \setminus E \rightarrow [0,1)$ é uma bijeção e que sua inversa tem coordenadas

$$d_k(x) = \lfloor 2^k x \rfloor - 2 \lfloor 2^{k-1} x \rfloor, \quad x \in [0,1).$$

Conclua que a bijeção e sua inversa são mensuráveis.

3. Mostre, primeiro nos intervalos diádicos, que a imagem da medida produto por Φ é a medida de Lebesgue. Conclua que os dois espaços são isomorfos como espaços de probabilidade depois de retirar conjuntos nulos.

Ex. 8.10 — Uma identidade de janelas deslizantes. Seja X uma variável aleatória real e fixe $a > 0$. Use Tonelli para provar que

$$\int_{\mathbb{R}} \mathbf{P}(x < X < x + a) dx = a.$$

Acoplamento

Comparar as distribuições de duas variáveis aleatórias nem sempre permite comparar diretamente seus valores. Se elas vivem em espaços amostrais diferentes, expressões como $\xi(\omega) - \eta(\omega)$ sequer estão definidas: não há, a princípio, um mesmo resultado ω no qual possamos observar as duas variáveis.

O acoplamento contorna essa dificuldade construindo cópias das variáveis num espaço de probabilidade comum. As distribuições marginais permanecem inalteradas, mas a dependência entre as cópias pode ser escolhida de acordo com o problema. Essa liberdade permite ordenar duas variáveis, fazer duas trajetórias se encontrarem ou estimar a diferença entre distribuições pela probabilidade de desacordo entre as cópias.

Começemos com duas moedas cujas probabilidades de sucesso são q e r , com $0 \leq q < r \leq 1$. Se as lançarmos independentemente, a primeira pode produzir sucesso e a segunda, fracasso. Entretanto, escolhendo uma única variável $U \sim \text{Uni}[0,1]$ e definindo

$$\xi = \mathbb{1}_{\{U \leq q\}}, \quad \eta = \mathbb{1}_{\{U \leq r\}},$$

obtemos as distribuições desejadas e, ao mesmo tempo, a relação $\xi \leq \eta$ em todo resultado. Cada moeda conserva sua lei; a novidade está na companhia.

9.1. Construções conjuntas

Definição 9.1 — Acoplamento

Sejam μ e ν medidas de probabilidade num mesmo espaço mensurável $(E, \mathcal{E})$.

- Um **acoplamento de μ e ν** é uma medida de probabilidade $\hat{\mu}$ em $(E \times E, \mathcal{E} \otimes \mathcal{E})$ cujas marginais são μ e ν :

$$\hat{\mu} \circ \pi_1^{-1} = \mu, \quad \hat{\mu} \circ \pi_2^{-1} = \nu,$$

onde $\pi_1(x, y) = x$ e $\pi_2(x, y) = y$.

- Um **acoplamento de variáveis aleatórias ξ e η** , ambas com valores em $(E, \mathcal{E})$, é um par $(\hat{\xi}, \hat{\eta})$ definido num mesmo espaço de probabilidade e tal que

$$\hat{\xi} \stackrel{d}{=} \xi, \quad \hat{\eta} \stackrel{d}{=} \eta.$$

Denotaremos por $\mathcal{C}(\mu, \nu)$ o conjunto dos acoplamentos de μ e ν . A medida produto $\mu \otimes \nu$ sempre pertence a esse conjunto e corresponde ao **acoplamento independente**. Se $\mu = \nu$, há também o acoplamento diagonal: escolha Z com distribuição μ e tome (Z, Z) . Esses dois extremos mostram que as marginais não determinam a distribuição conjunta. A independência é uma escolha, não uma cláusula da definição.

Exemplo 9.2 — Duas moedas, dois acoplamentos.

Sejam ξ e η variáveis de Bernoulli com parâmetros $0 \leq q < r \leq 1$. No acoplamento independente, com as linhas indexadas pelo valor da primeira variável e as colunas pelo valor da segunda, a distribuição conjunta é

$$\begin{pmatrix} (1-q)(1-r) & (1-q)r \\ q(1-r) & qr \end{pmatrix}.$$

Nesse caso, o evento $\{\xi = 1, \eta = 0\}$ tem probabilidade $q(1-r) > 0$ quando $0 < q < r < 1$.

No acoplamento construído na abertura do capítulo,

$$\xi = \mathbb{1}_{\{U \leq q\}}, \quad \eta = \mathbb{1}_{\{U \leq r\}},$$

a distribuição conjunta é

$$\begin{pmatrix} 1-r & r-q \\ 0 & q \end{pmatrix}.$$

Agora $\xi \leq \eta$ sempre. Por isso, essa construção é chamada de **acoplamento monótono**.

A mesma distribuição conjunta admite outra construção. Tome variáveis independentes $\xi_0 \sim \text{Be}(q)$ e $\zeta \sim \text{Be}((r-q)/(1-q))$ e faça $\eta_0 = \xi_0 \vee \zeta$. Então $\eta_0 \sim \text{Be}(r)$ e $\xi_0 \leq \eta_0$. O mecanismo é diferente, mas o acoplamento é o mesmo em lei.

◀

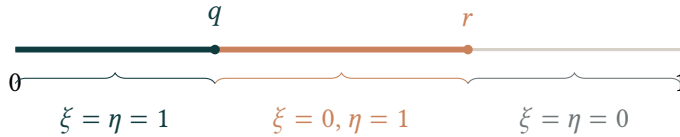


Figura 9.1.: O mesmo uniforme produz duas moedas ordenadas. O desacordo ocorre exatamente quando $q < U \leq r$.

Exemplo 9.3 — Acoplamentos discretos como matrizes.

Se E é enumerável, um acoplamento de probabilidades μ e ν pode ser representado por uma matriz não negativa

$$Q(x, y) = \mathbf{P}(\hat{\xi} = x, \hat{\eta} = y), \quad x, y \in E.$$

As condições sobre as marginais tornam-se

$$\sum_{y \in E} Q(x, y) = \mu(\{x\}), \quad \sum_{x \in E} Q(x, y) = \nu(\{y\}).$$

Reciprocamente, toda matriz não negativa com essas somas marginais define um acoplamento. Assim, construir um acoplamento discreto significa distribuir massa numa tabela sem alterar as somas de suas linhas e colunas.



9.2. Acoplamento de processos

Para comparar processos, é preciso preservar mais do que a distribuição em cada instante: a dependência temporal também faz parte de sua lei.

Definição 9.4 — Acoplamento de processos

Sejam $X = (X_t)_{t \in T}$ e $Y = (Y_t)_{t \in T}$ processos com valores em $(E, \mathcal{E})$, possivelmente definidos em espaços de probabilidade diferentes. Um **acoplamento de X e Y** é um par de processos $(\hat{X}, \hat{Y})$ definido num mesmo espaço de probabilidade tal que

$$\mathcal{L}((\hat{X}_t)_{t \in T}) = \mathcal{L}((X_t)_{t \in T}), \quad \mathcal{L}((\hat{Y}_t)_{t \in T}) = \mathcal{L}((Y_t)_{t \in T})$$

no espaço de trajetórias $(E^T, \mathcal{E}^{\otimes T})$. Equivalentemente, os processos acoplados devem possuir todas as distribuições finito-dimensionais corretas.

Exigir apenas $\hat{X}_t \stackrel{d}{=} X_t$ para cada t não é suficiente. Por exemplo, sejam X_0 e X_1 moedas justas independentes e seja $Y_0 = Y_1 = Z$, onde Z é uma moeda justa. Em cada instante, $X_t \stackrel{d}{=} Y_t$, mas

$$\mathbf{P}(X_0 = X_1) = \frac{1}{2}, \quad \mathbf{P}(Y_0 = Y_1) = 1.$$

As marginais unidimensionais coincidem, mas não registram que um dos processos lança uma nova moeda e o outro repete a primeira. As leis das trajetórias são diferentes.

Exemplo 9.5 — Passeios aleatórios ordenados em $\mathbb{Z}$.

Para $p \in [0,1]$, seja $S(p) = (S_n(p))_{n \geq 0}$ o passeio aleatório iniciado em 0, com probabilidade p de saltar para a direita. Escolha uma sequência i.i.d. $(U_i)_{i \geq 1}$ de variáveis uniformes em $[0,1]$ e, para cada $p \in [0,1]$, defina

$$B_i(p) = \mathbb{1}_{\{U_i \leq p\}}, \quad \hat{S}_n(p) = \sum_{i=1}^n (2B_i(p) - 1).$$

Para cada p , o processo $\hat{S}(p)$ possui a lei do passeio de parâmetro p . Além disso, se $0 \leq q \leq r \leq 1$, então

$$B_i(q) \leq B_i(r) \quad \text{para todo } i, \quad \text{e, portanto,} \quad \hat{S}_n(q) \leq \hat{S}_n(r) \quad \text{para todo } n.$$

Uma única sequência de uniformes acopla simultaneamente toda a família de passeios e ordena suas trajetórias.

<

9.3. Variação total

A escolha da dependência só será útil à comparação se pudermos relacioná-la a uma quantidade que dependa apenas das marginais. A distância de variação total fornece essa ligação.

Definição 9.6 — Distância de variação total

Para probabilidades μ e ν num espaço mensurável $(E, \mathcal{E})$, definimos

$$d_{TV}(\mu, \nu) := \sup_{A \in \mathcal{E}} |\mu(A) - \nu(A)|.$$

Essa normalização faz com que $0 \leq d_{TV}(\mu, \nu) \leq 1$.

No Capítulo 13, veremos que $\mu - \nu$ é uma medida com sinal e que

$$|\mu - \nu|(E) = 2d_{TV}(\mu, \nu).$$

As caracterizações de que precisamos aqui admitem provas diretas, sem recorrer à teoria de medidas com sinal.

Proposição 9.7 — Fórmula discreta

Se E é finito ou enumerável, munido de $\mathcal{G} = 2^E$, e μ, ν são probabilidades em E , então

$$d_{TV}(\mu, \nu) = \frac{1}{2} \sum_{x \in E} |\mu(\{x\}) - \nu(\{x\})|.$$

Demonstração. Escreva $a_x = \mu(\{x\}) - \nu(\{x\})$ e $A_+ = \{x \in E : a_x \geq 0\}$. Como $\sum_x a_x = 0$,

$$\sum_{x \in A_+} a_x = - \sum_{x \notin A_+} a_x = \frac{1}{2} \sum_{x \in E} |a_x|.$$

Para todo $A \subseteq E$,

$$\mu(A) - \nu(A) = \sum_{x \in A} a_x \leq \sum_{x \in A_+} a_x,$$

e a cota é atingida em $A = A_+$. Aplicando o mesmo argumento a $\nu - \mu$, obtemos a fórmula. □

Proposição 9.8 — Caracterização dual

Sejam μ e ν probabilidades em $(E, \mathcal{G})$. Então

$$d_{TV}(\mu, \nu) = \sup_{0 \leq f \leq 1} \left| \int_E f d\mu - \int_E f d\nu \right| = \frac{1}{2} \sup_{|f| \leq 1} \left| \int_E f d\mu - \int_E f d\nu \right|,$$

onde os supremos percorrem as funções mensuráveis sujeitas às restrições indicadas.

Demonstração. Se $0 \leq f \leq 1$, a representação por camadas do Corolário 8.12 dá

$$\int_E f d\mu - \int_E f d\nu = \int_0^1 [\mu(f > t) - \nu(f > t)] dt.$$

O valor absoluto dessa expressão é no máximo $d_{TV}(\mu, \nu)$. Por outro lado, tomando $f = 1_A$ e aproximando o supremo sobre A , obtemos a igualdade no primeiro supremo.

Se $|f| \leq 1$, então $g = (f + 1)/2$ satisfaz $0 \leq g \leq 1$. Como $\mu(E) = \nu(E) = 1$,

$$\int_E f d\mu - \int_E f d\nu = 2 \left(\int_E g d\mu - \int_E g d\nu \right).$$

Isso prova a segunda igualdade. □

A definição por eventos fornece a desigualdade triangular:

$$d_{TV}(\mu, \nu) \leq d_{TV}(\mu, \rho) + d_{TV}(\rho, \nu)$$

para quaisquer probabilidades μ, ν e ρ .

9.4. A desigualdade de acoplamento

Proposição 9.9 — Desigualdade de acoplamento

Sejam μ e ν probabilidades em $(E, \mathcal{E})$ e seja $(\hat{\xi}, \hat{\eta})$ um acoplamento para o qual o evento $\{\hat{\xi} \neq \hat{\eta}\}$ seja mensurável. Então

$$d_{TV}(\mu, \nu) \leq \mathbf{P}(\hat{\xi} \neq \hat{\eta}).$$

Demonstração. Para todo $A \in \mathcal{E}$,

$$\begin{aligned}\mu(A) - \nu(A) &= \mathbf{P}(\hat{\xi} \in A, \hat{\eta} \notin A) - \mathbf{P}(\hat{\eta} \in A, \hat{\xi} \notin A) \\ &\leq \mathbf{P}(\hat{\xi} \neq \hat{\eta}).\end{aligned}$$

Trocando μ e ν , obtemos a mesma cota para $\nu(A) - \mu(A)$. Basta tomar o supremo sobre A . □

O lado direito depende do acoplamento; o esquerdo, apenas das marginais. Podemos, portanto, escolher as cópias de modo a reduzir a probabilidade de desacordo, sem mudar as distribuições que queremos comparar.

Corolário 9.10 —

Nas condições da proposição, se $f : E \rightarrow \mathbb{R}$ é mensurável e limitada, então

$$|\mathbf{E}f(\hat{\xi}) - \mathbf{E}f(\hat{\eta})| \leq 2 \|f\|_{\infty} \mathbf{P}(\hat{\xi} \neq \hat{\eta}).$$

Demonstração. As variáveis $f(\hat{\xi})$ e $f(\hat{\eta})$ coincidem quando $\hat{\xi} = \hat{\eta}$. Logo,

$$|\mathbf{E}f(\hat{\xi}) - \mathbf{E}f(\hat{\eta})| \leq \mathbf{E}|f(\hat{\xi}) - f(\hat{\eta})| \leq 2 \|f\|_{\infty} \mathbf{P}(\hat{\xi} \neq \hat{\eta}).$$

□

Exemplo 9.11 — Perturbação Poisson.

Sejam $\lambda_1 \geq \lambda_2 \geq 0$. Escolha variáveis independentes

$$\hat{\eta} \sim \text{Po}(\lambda_2), \quad Z \sim \text{Po}(\lambda_1 - \lambda_2),$$

e defina $\hat{\xi} = \hat{\eta} + Z$. Então $\hat{\xi} \sim \text{Po}(\lambda_1)$ e

$$d_{\text{TV}}(\text{Po}(\lambda_1), \text{Po}(\lambda_2)) \leq \mathbf{P}(Z > 0) = 1 - e^{-(\lambda_1 - \lambda_2)} \leq \lambda_1 - \lambda_2.$$

Não foi preciso subtrair as duas funções de massa: o desacordo fica restrito à ocorrência de um termo adicional não nulo. A estimativa traduz essa observação em controle da distância de variação total.

◀

9.5. Acoplamentos maximais

A desigualdade de acoplamento fornece uma cota para qualquer construção conjunta. Cabe perguntar se alguma escolha atinge a própria distância de variação total.

Definição 9.12 — Acoplamento maximal

Um acoplamento $(\hat{\xi}, \hat{\eta})$ de μ e ν é **maximal** se

$$\mathbf{P}(\hat{\xi} \neq \hat{\eta}) = d_{\text{TV}}(\mu, \nu).$$

Ele maximiza a probabilidade de as duas cópias coincidirem.

O adjetivo se refere à coincidência: é o desacordo que queremos minimizar.

Teorema 9.13 — Caso discreto

Se E é finito ou enumerável, munido de $\mathcal{G} = 2^E$, e μ, ν são probabilidades em E , existe um acoplamento maximal. Em particular,

$$\inf_{\hat{\mu} \in \mathcal{C}(\mu, \nu)} \hat{\mu}(\{(x, y) : x \neq y\}) = d_{\text{TV}}(\mu, \nu).$$

Demonstração. Escreva $\mu(x) = \mu(\{x\})$, $\nu(x) = \nu(\{x\})$ e ponha

$$a = \sum_{x \in E} (\mu(x) \wedge \nu(x)).$$

Pela Proposição 9.7,

$$d_{\text{TV}}(\mu, \nu) = \frac{1}{2} \sum_{x \in E} |\mu(x) - \nu(x)| = 1 - a.$$

Se $a = 1$, então $\mu = \nu$ e o acoplamento diagonal resolve o problema. Suponha $a < 1$ e defina

$$\gamma_{\mu}(x) = \frac{\mu(x) - \mu(x) \wedge \nu(x)}{1 - a}, \quad \gamma_{\nu}(x) = \frac{\nu(x) - \mu(x) \wedge \nu(x)}{1 - a}.$$

Quando $a > 0$, defina também

$$\gamma_0(x) = \frac{\mu(x) \wedge \nu(x)}{a}.$$

Escolha $B \sim \text{Be}(a)$. Se $B = 1$, escolha W com distribuição γ_0 e faça $\hat{\xi} = \hat{\eta} = W$. Se $B = 0$, escolha $U \sim \gamma_\mu$ e $V \sim \gamma_\nu$, independentemente, e faça $(\hat{\xi}, \hat{\eta}) = (U, V)$. Quando $a = 0$, usamos apenas a segunda etapa.

As distribuições residuais possuem suportes disjuntos: $\gamma_\mu(x) > 0$ implica $\mu(x) > \nu(x)$, enquanto $\gamma_\nu(x) > 0$ implica $\nu(x) > \mu(x)$. Portanto, na segunda etapa, $U \neq V$ quase certamente. Além disso, para todo $x \in E$,

$$a\gamma_0(x) + (1-a)\gamma_\mu(x) = \mu(x), \quad a\gamma_0(x) + (1-a)\gamma_\nu(x) = \nu(x),$$

com a interpretação evidente quando $a = 0$. Assim, as marginais estão corretas e

$$\mathbf{P}(\hat{\xi} \neq \hat{\eta}) = \mathbf{P}(B = 0) = 1 - a = d_{\text{TV}}(\mu, \nu).$$

A desigualdade de acoplamento mostra que nenhuma construção pode produzir uma probabilidade de desacordo menor. □

No exemplo das moedas, o acoplamento monótono da Figura 9.1 satisfaz

$$\mathbf{P}(\xi \neq \eta) = r - q = d_{\text{TV}}(\text{Be}(q), \text{Be}(r)).$$

Portanto, ele é simultaneamente monótono e maximal. Essas propriedades não se confundem: uma ordena as cópias; a outra maximiza a probabilidade de coincidência. Para essas moedas, a mesma construção faz as duas coisas.

Corolário 9.14 — Critério de convergência

Se E é finito ou enumerável, então

$$d_{\text{TV}}(\mu_n, \mu) \longrightarrow 0$$

se, e somente se, para cada n existem variáveis (X_n, Y_n) num mesmo espaço de probabilidade, com leis μ_n e μ , tais que

$$\mathbf{P}(X_n \neq Y_n) \longrightarrow 0.$$

Os espaços de probabilidade podem depender de n .

Demonstração. Se os acoplamentos existem, a desigualdade de acoplamento fornece a convergência em variação total. Reciprocamente, escolha para cada n um acoplamento maximal e aplique o Teorema 9.13. $\square$

9.6. Uma família inteira no mesmo espaço: percolação

A mesma ideia permite comparar, num único espaço, toda uma família de modelos. Retomemos a percolação de elos discutida na Seção 3.3.

Seja $G = (V, E_G)$ um grafo com conjunto enumerável de arestas. No espaço $[0,1]^{E_G}$, escolha variáveis independentes $(U_e)_{e \in E_G}$, todas uniformes em $[0,1]$. Para cada $p \in [0,1]$, defina a configuração

$$\omega_p(e) = \mathbb{1}_{\{U_e \leq p\}}, \quad e \in E_G.$$

Uma aresta está aberta na configuração ω_p precisamente quando sua marca uniforme não excede p .

Uma propriedade $A \subseteq \{0,1\}^{E_G}$ é chamada **crescente** se

$$\omega \in A, \quad \omega \leq \omega' \quad \implies \quad \omega' \in A,$$

onde a ordem entre configurações é coordenada a coordenada.

Proposição 9.15 — Percolação acoplada

Para cada p , a configuração ω_p possui a lei da percolação de elos de parâmetro p . Além disso, para toda realização das marcas,

$$p \leq q \quad \implies \quad \omega_p \leq \omega_q \quad \text{simultaneamente para todos } p, q \in [0,1].$$

Consequentemente, se A é um evento mensurável crescente, então

$$p \mapsto \mathbf{P}_p(A)$$

é uma função não decrescente.

Demonstração. Para p fixado, as variáveis $\omega_p(e)$ são independentes e

$$\mathbf{P}(\omega_p(e) = 1) = \mathbf{P}(U_e \leq p) = p.$$

Logo, ω_p possui a lei produto da percolação de parâmetro p .

Se $p \leq q$, então, para todo $u \in [0,1]$,

$$\mathbb{1}_{\{u \leq p\}} \leq \mathbb{1}_{\{u \leq q\}}.$$

Aplicando essa desigualdade a cada marca U_e , obtemos $\omega_p \leq \omega_q$. A afirmação vale ponto a ponto no espaço das marcas e, portanto, simultaneamente para todos os pares $p \leq q$. Finalmente, se A é crescente,

$$\mathbb{1}_{\{\omega_p \in A\}} \leq \mathbb{1}_{\{\omega_q \in A\}},$$

e a desigualdade entre as probabilidades segue ao tomar esperanças. □



Figura 9.2.: No acoplamento simultâneo, a configuração de parâmetro $q \geq p$ é obtida acrescentando arestas à configuração de parâmetro p ; nenhuma aresta aberta precisa ser fechada.

O evento de que dois vértices estejam ligados por um caminho aberto é crescente. Também é crescente o evento de que o cluster aberto de um vértice x seja infinito. Portanto,

$$\theta_x(p) = \mathbf{P}_p(|C_x| = \infty)$$

é não decrescente em p , como foi usado na construção da probabilidade crítica na Seção 3.3. A comparação entre probabilidades decorre agora de uma inclusão entre configurações construídas no mesmo espaço.

O acoplamento preserva cada modelo e escolhe como eles serão observados juntos. Neste exemplo, essa escolha transforma a ordem entre os parâmetros numa ordem entre as próprias configurações.

EXERCÍCIOS

Acoplamentos

Ex. 9.1 — Sejam $X \sim \text{Be}(q)$ e $Y \sim \text{Be}(r)$, com $0 \leq q \leq r \leq 1$.

1. Mostre que todo acoplamento de X e Y é determinado pelo valor $a = \mathbf{P}(X = 1, Y = 1)$.
2. Determine o intervalo de valores possíveis de a .
3. Identifique nesse intervalo os acoplamentos independente e monótono, este último caracterizado por $X \leq Y$ quase certamente. Qual deles minimiza $\mathbf{P}(X \neq Y)$? Calcule esse mínimo.

Ex. 9.2 — Sejam X_0, X_1 moedas justas independentes e seja $Y_0 = Y_1 = Z$, onde Z é uma moeda justa.

1. Verifique que $X_t \stackrel{d}{=} Y_t$ para $t = 0, 1$.
2. Mostre que os vetores (X_0, X_1) e (Y_0, Y_1) não possuem a mesma lei. Explique por que marginais unidimensionais corretas não bastam na definição de acoplamento de processos.

Ex. 9.3 — Sejam μ e ν probabilidades num espaço mensurável E , seja (X, Y) um acoplamento de μ e ν e seja $f : E \rightarrow S$ mensurável. Mostre que $(f(X), f(Y))$ é um acoplamento das medidas-imagem de μ e ν por f . Deduza que

$$d_{\text{TV}}(\mu \circ f^{-1}, \nu \circ f^{-1}) \leq d_{\text{TV}}(\mu, \nu)$$

quando E é finito ou enumerável. Prove também a desigualdade diretamente, a partir da definição por supremo, sem a hipótese de enumerabilidade.

Ex. 9.4 — Para $i = 1, \dots, n$, seja (X_i, Y_i) um acoplamento de μ_i e ν_i , e suponha que

esses pares sejam independentes.

1. Mostre que

$$\mathbf{P}((X_1, \dots, X_n) \neq (Y_1, \dots, Y_n)) \leq \sum_{i=1}^n \mathbf{P}(X_i \neq Y_i).$$

2. Usando acoplamentos maximais, prove, para espaços finitos ou enumeráveis,

$$d_{\text{TV}}\left(\bigotimes_{i=1}^n \mu_i, \bigotimes_{i=1}^n \nu_i\right) \leq \sum_{i=1}^n d_{\text{TV}}(\mu_i, \nu_i).$$

Ex. 9.5 — Use o acoplamento de variáveis de Poisson construído no capítulo para provar que, se $\lambda_n, \lambda \geq 0$ e $\lambda_n \rightarrow \lambda$, então

$$d_{\text{TV}}(\text{Po}(\lambda_n), \text{Po}(\lambda)) \longrightarrow 0.$$

Ex. 9.6 — No acoplamento simultâneo da percolação de elos num grafo localmente finito G , fixe vértices x e y e denote por C_x o aglomerado aberto de x .

1. Mostre diretamente que o evento $\{x \longleftrightarrow y\}$ é crescente.
2. Para $n \geq 1$, seja A_n o evento de que x esteja ligado por um caminho aberto a algum vértice a distância pelo menos n . Mostre que A_n é crescente e que

$$\{|C_x| = \infty\} = \bigcap_{n \geq 1} A_n$$

num grafo localmente finito.

3. Conclua novamente que $p \mapsto \theta_x(p)$ é não decrescente.

Modos de Convergência

Em probabilidade, dizer que ξ_n se aproxima de ξ ainda não é dizer o bastante: é preciso escolher o que significa aproximar. Neste capítulo apresentaremos e compararemos as noções mais usuais de convergência de variáveis aleatórias. A finitude da medida será usada, por exemplo, na inclusão $\mathcal{L}^q \subseteq \mathcal{L}^p$ para $q \geq p$ e na passagem da convergência quase certa para a convergência em probabilidade; em espaços de medida infinita, essas afirmações exigem hipóteses adicionais.

10.1. Convergências Determinísticas

Começamos pelas noções determinísticas, que comparam funções sem usar a medida. Como já são conhecidas de outros contextos, recordaremos apenas as definições, alguns exemplos e as diferenças que serão importantes adiante.

10.1.1. Convergência Pontual

Seja Ω um conjunto e $\xi_n : \Omega \rightarrow \mathbb{R}$ uma sequência de funções com o mesmo domínio Ω . Dizemos que $\xi_n(x)$ **converge pontualmente** para uma função $\xi : \Omega \rightarrow \mathbb{R}$ se:

$$\lim_{n \rightarrow \infty} \xi_n(x) = \xi(x),$$

para todo $x \in \Omega$.

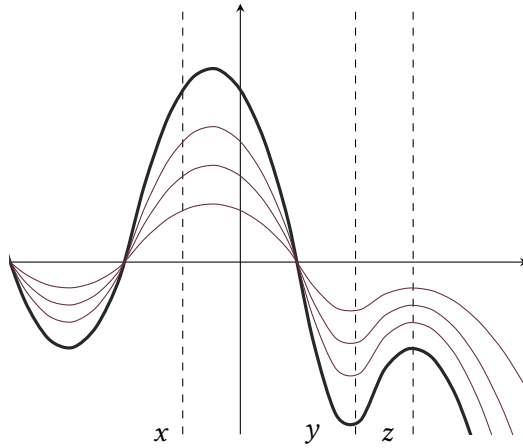


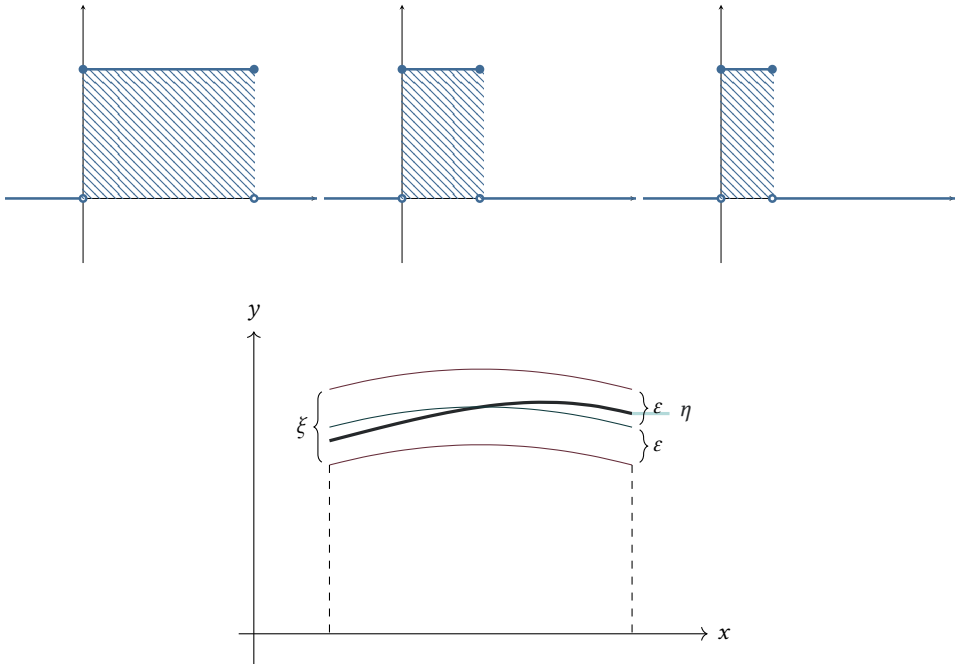
Figura 10.1.: A sequência $\{\xi_n\}_{n=1}^{\infty}$ converge pontualmente para a função constante 0.

Assim, para quaisquer $x, y, z \in \Omega$, temos $\lim_n \xi_n(x) = 0$, $\lim_n \xi_n(y) = 0$ e $\lim_n \xi_n(z) = 0$.

Exemplo 10.1 — .

Para $n \in \mathbb{N}$ definimos $\xi_n(x) = \begin{cases} 1 & \text{se } 0 < x < 1/n \\ 0 & \text{caso contrário} \end{cases}$

Essa sequência converge pontualmente para a função constante zero.



10.1.2. Convergência Uniforme

A **distância uniforme** entre duas funções ξ e η é definida como

$$\|\xi - \eta\|_u = \sup_{x \in \Omega} |\xi(x) - \eta(x)|.$$

Essa distância mede o afastamento máximo de ξ e η ; quando o supremo é infinito, entendemos $\|\xi - \eta\|_u$ como uma distância estendida.

Uma sequência de funções $\{\xi_n\}_{n=1}^\infty$ **converge uniformemente** para ξ se $\lim_{n \rightarrow \infty} \|\xi_n - \xi\|_u = 0$. Isso significa não só que $\lim_n \xi_n(x) = \xi(x)$ para cada $x \in \Omega$, mas também que a aproximação é uniforme em todo o domínio: dado $\varepsilon > 0$, um mesmo índice n_0 serve para todos os pontos $x \in \Omega$.

Exemplo 10.2 — .

Se $\xi_n(x) = 1/n$ para todo $x \in [0, 1]$, então a sequência $\{\xi_n\}_{n=1}^{\infty}$ converge para zero uniformemente em $[0, 1]$.

**Exemplo 10.3 — .**

Para $n \in \mathbb{N}$ definimos $\xi_n(x) = \begin{cases} 1 & \text{se } 0 < x < 1/n \\ 0 & \text{caso contrário} \end{cases}$

Essa sequência converge pontualmente para a função constante zero, mas não uniformemente, pois $\|\xi_n\|_u = 1$ para todo n .



10.2. Convergência em Espaços de Probabilidade

Passamos agora às convergências que usam a estrutura mensurável e a probabilidade. O conjunto excepcional deixa de ser apenas um defeito: seu tamanho passa a fazer parte da definição.

10.2.1. Convergência em Distribuição

A convergência em distribuição é a mais fraca entre as que estudaremos. Ela compara apenas as leis de ξ_n com a lei de ξ , por meio de suas funções de distribuição. O espaço amostral fica fora de cena: as variáveis nem sequer precisam estar definidas no mesmo espaço de probabilidade.

Definição 10.4 — Convergência em Distribuição

Uma sequência de variáveis aleatórias $\{\xi_n\}_{n=1}^{\infty}$, **converge em distribuição** para uma variável aleatória ξ , fato denotado por $\xi_n \xrightarrow{d} \xi$, se

$$\lim_{n \rightarrow \infty} F_{\xi_n}(x) = F_{\xi}(x),$$

para todo x tal que $F_{\xi}(x)$ é contínua.

Exemplo 10.5 — .

A exigência de que apenas os pontos de continuidade de F_X sejam considerados na definição acima é, de fato, essencial.

Para ver isso, tome uma sequência $\{X_n; n \geq 1\}$ de variáveis aleatórias, definidas em um mesmo espaço de probabilidade $(\Omega, \mathcal{F}, \mathbf{P})$, e tais que $X_n(\omega) = \frac{1}{n}$ para todo $\omega \in \Omega$. Tal sequência converge para 0 de todos os modos que vimos até agora — pontualmente e uniformemente — e também dos modos que descreveremos a seguir: quase certamente, em $\mathcal{L}^p$ (para qualquer $p \geq 1$) e em probabilidade.

Além disso, temos

$$F_{X_n}(x) = \mathbf{P}(X_n \leq x) = \begin{cases} 0 & , \text{ se } x < \frac{1}{n}; \\ 1 & , \text{ se } x \geq \frac{1}{n} \end{cases},$$

e para a variável $X = 0$ temos

$$F_X(x) = \mathbf{P}(X \leq x) = \begin{cases} 0 & , \text{ se } x < 0; \\ 1 & , \text{ se } x \geq 0 \end{cases}$$

Observe que para $x < 0$

$$\lim_n F_{X_n}(x) = 0 = F_X(x),$$

e para $x > 0$

$$\lim_n F_{X_n}(x) = 1 = F_X(x),$$

mas para $x = 0$ temos

$$\lim_n F_{X_n}(0) = 0 \neq 1 = F_X(0).$$

Mas como $x = 0$ é ponto de descontinuidade de F_X , então ainda podemos afirmar que $X_n \xrightarrow{d} 0$.



Ao trabalhar com variáveis aleatórias que tomam valores inteiros, o seguinte teorema é frequentemente útil.

Teorema 10.6 —

Considere a sequência $\{\xi_n\}_{n=1}^\infty$ e a variável aleatória ξ . Suponha que ξ e ξ_n sejam não negativas e tomem valores inteiros.

Então $\xi_n \xrightarrow{d} \xi$ se e somente se

$$\lim_{n \rightarrow \infty} \mathbf{P}(\xi_n = k) = \mathbf{P}(\xi = k), \quad k = 0, 1, 2, \dots$$

Demonstração.

Como ξ toma valores inteiros, sua função de distribuição F_ξ é contínua em todo $x \in \mathbb{R} \setminus \{0, 1, 2, \dots\}$. Se $\xi_n \xrightarrow{d} \xi$, então

$$\lim_{n \rightarrow \infty} F_{\xi_n}(x) = F_\xi(x), \quad x \in \mathbb{R} \setminus \{0, 1, 2, \dots\}.$$

Assim, para $k = 0, 1, 2, \dots$, temos

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbf{P}(\xi_n = k) &= \lim_{n \rightarrow \infty} \left[F_{\xi_n} \left(k + \frac{1}{2} \right) - F_{\xi_n} \left(k - \frac{1}{2} \right) \right] \\ &= \lim_{n \rightarrow \infty} F_{\xi_n} \left(k + \frac{1}{2} \right) - \lim_{n \rightarrow \infty} F_{\xi_n} \left(k - \frac{1}{2} \right) \\ &= F_\xi \left(k + \frac{1}{2} \right) - F_\xi \left(k - \frac{1}{2} \right) \quad (\text{pois } \xi_n \xrightarrow{d} \xi) \\ &= \mathbf{P}(\xi = k). \end{aligned}$$

Para provar a recíproca, suponha que

$$\lim_{n \rightarrow \infty} \mathbf{P}(\xi_n = k) = \mathbf{P}(\xi = k), \quad k = 0, 1, 2, \dots$$

Se $x < 0$, então $F_{\xi_n}(x) = F_{\xi}(x) = 0$. Se $x \geq 0$, temos

$$\begin{aligned}\lim_{n \rightarrow \infty} F_{\xi_n}(x) &= \lim_{n \rightarrow \infty} \mathbf{P}(\xi_n \leq x) \\ &= \lim_{n \rightarrow \infty} \sum_{k=0}^{\lfloor x \rfloor} \mathbf{P}(\xi_n = k),\end{aligned}$$

onde $\lfloor x \rfloor$ denota o maior inteiro menor ou igual a x . Como, para x fixo, a soma é finita, podemos trocar o limite com a soma e obtemos

$$\begin{aligned}\lim_{n \rightarrow \infty} F_{\xi_n}(x) &= \sum_{k=0}^{\lfloor x \rfloor} \lim_{n \rightarrow \infty} \mathbf{P}(\xi_n = k) \\ &= \sum_{k=0}^{\lfloor x \rfloor} \mathbf{P}(\xi = k) \\ &= \mathbf{P}(\xi \leq x) = F_{\xi}(x).\end{aligned}$$

□

Teorema 10.7 – Aproximação de Poisson para a binomial

Seja $\{\xi_n\}_{n \geq 1}$ uma sequência de variáveis aleatórias tal que

$$\xi_n \sim \text{Bin}\left(n, \frac{\lambda}{n}\right), \quad \text{para } n \in \mathbb{N}, n > \lambda,$$

onde $\lambda > 0$ é constante. Então ξ_n converge em distribuição para uma variável aleatória com distribuição $\text{Po}(\lambda)$.

Demonstração. Pelo teorema anterior, basta mostrar que, para cada $k = 0, 1, 2, \dots$,

$$\lim_{n \rightarrow \infty} \mathbf{P}(\xi_n = k) = \frac{e^{-\lambda} \lambda^k}{k!}.$$

Temos

$$\begin{aligned}\lim_{n \rightarrow \infty} \mathbf{P}(\xi_n = k) &= \lim_{n \rightarrow \infty} \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \lambda^k \lim_{n \rightarrow \infty} \frac{n!}{k!(n-k)!} \left(\frac{1}{n^k}\right) \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \frac{\lambda^k}{k!} \lim_{n \rightarrow \infty} \left[\frac{n(n-1) \cdots (n-k+1)}{n^k} \right] \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k}.\end{aligned}$$

Para k fixo, temos

$$\begin{aligned}\lim_{n \rightarrow \infty} \frac{n(n-1)(n-2)\dots(n-k+1)}{n^k} &= 1, \\ \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^{-k} &= 1, \\ \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^n &= e^{-\lambda}.\end{aligned}$$

Logo

$$\lim_{n \rightarrow \infty} \mathbf{P}(\xi_n = k) = \frac{e^{-\lambda} \lambda^k}{k!},$$

que é a função de probabilidade de $\text{Po}(\lambda)$. □

10.2.2. Convergência em Probabilidade

Na convergência em probabilidade, não exigimos proximidade em cada ponto; exigimos que a probabilidade de um afastamento fixo desapareça. No contexto da Teoria da Medida, essa noção é chamada de *convergência em medida*.

Definição 10.8 — Convergência em Probabilidade

Uma sequência de variáveis aleatórias $\{\xi_n\}_{n \geq 1}$ **converge em probabilidade** para a variável aleatória ξ se

$$\forall \epsilon > 0 \quad \mathbf{P}(|\xi_n - \xi| > \epsilon) \rightarrow 0 \text{ quando } n \rightarrow \infty.$$

Denotaremos tal fato por $\xi_n \xrightarrow{\mathbf{P}} \xi$.

A sequência pode continuar errando, mas a região em que o erro é maior do que ϵ deve tornar-se improvável.

Exemplo 10.9 – .

Dado $\xi_n \sim \text{Exp}(n)$, então $\xi_n \xrightarrow{\mathbf{P}} 0$.

Pois

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbf{P}(|\xi_n - 0| \geq \epsilon) &= \lim_{n \rightarrow \infty} \mathbf{P}(\xi_n \geq \epsilon) && (\text{ pois } \xi_n \geq 0) \\ &= \lim_{n \rightarrow \infty} e^{-n\epsilon} && (\text{ pois } \xi_n \sim \text{Exp}(n)) \\ &= 0, && \text{ para todo } \epsilon > 0. \end{aligned}$$

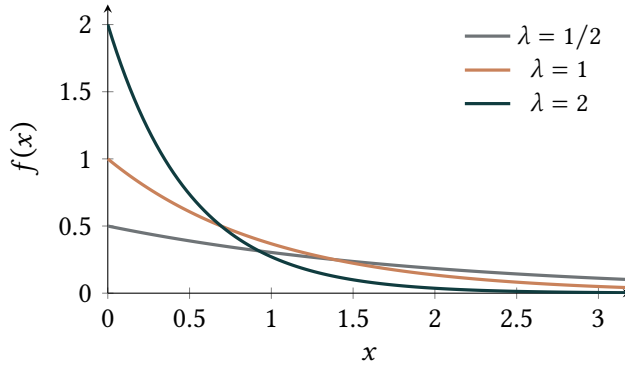


Figura 10.2.: Densidades exponenciais de parâmetros 1/2, 1 e 2.

Exemplo 10.10 – .

Seja ξ uma variável aleatória e $\xi_n = \xi + \eta_n$, com

$$\mathbf{E}[\eta_n] = \frac{1}{n}, \quad \text{Var}[\eta_n] = \frac{\sigma^2}{n},$$

onde $\sigma > 0$ é uma constante. Então $\xi_n \xrightarrow{\mathbf{P}} \xi$.

Pela desigualdade triangular para $\eta_n - \mathbf{E}[\eta_n] + \mathbf{E}[\eta_n]$, temos

$$|\eta_n| \leq |\eta_n - \mathbf{E}[\eta_n]| + \frac{1}{n}.$$

Então, para todo $\epsilon > 0$ e todo $n > 1/\epsilon$, temos

$$\begin{aligned}
 \mathbf{P}(|\xi_n - \xi| \geq \epsilon) &= \mathbf{P}(|\eta_n| \geq \epsilon) \\
 &\leq \mathbf{P}\left(|\eta_n - \mathbf{E}[\eta_n]| + \frac{1}{n} \geq \epsilon\right) \\
 &= \mathbf{P}\left(|\eta_n - \mathbf{E}[\eta_n]| \geq \epsilon - \frac{1}{n}\right) \\
 &\leq \frac{\text{Var}[\eta_n]}{\left(\epsilon - \frac{1}{n}\right)^2} \quad (\text{Chebyshev}) \\
 &= \frac{\sigma^2}{n\left(\epsilon - \frac{1}{n}\right)^2} \rightarrow 0.
 \end{aligned}$$

Logo $\xi_n \xrightarrow{\mathbf{P}} \xi$.

◀

10.2.3. Convergência Quase Certa

Na convergência quase certa, recuperamos a convergência pontual, permitindo apenas que ela falhe em um conjunto de probabilidade nula.

Começaremos com uma leve adaptação do exemplo 10.1, onde tínhamos uma sequência convergindo pontualmente.

Exemplo 10.11 — .

Tome $\Omega = [0,1]$, munido da σ -álgebra de Borel, e seja $\mathbf{P}$ a medida de Lebesgue em $[0,1]$. Para $n \geq 1$, defina

$$\xi_n(\omega) = \begin{cases} 1 & \text{se } 0 \leq \omega < 1/n \\ 0 & \text{caso contrário} \end{cases}$$

Observe que, se $\omega \in (0,1]$, então $\xi_n(\omega) = 0$ para todo $n > 1/\omega$, de modo que $\xi_n(\omega) \rightarrow 0$.

No entanto, vale também que $\xi_n(0) = 1$ para todo $n > 0$, de modo que

$$\{\omega \in [0,1] : \xi_n(\omega) \not\rightarrow 0\} = \{0\},$$

e

$$\mathbf{P}\left(\left\{\omega : \lim_{n \rightarrow \infty} \xi_n(\omega) = 0\right\}\right) = \mathbf{P}((0,1]) = 1.$$

Dizemos assim que $\xi_n \rightarrow 0$ **quase certamente**, quando $n \rightarrow \infty$.

◀

Definição 10.12 – Convergência Quase Certa

Sejam ξ uma variável aleatória e $\{\xi_n\}_{n \geq 1}$ uma sequência de variáveis aleatórias definidas no mesmo espaço de probabilidade. Dizemos que ξ_n **converge quase certamente** para ξ , isto é, $\xi_n \xrightarrow{q.c.} \xi$ se

$$\mathbf{P}\left(\lim_{n \rightarrow \infty} \xi_n = \xi\right) = 1$$

ou, equivalentemente, se

$$\mathbf{P}\left(\left\{\omega \in \Omega : \lim_{n \rightarrow \infty} \xi_n(\omega) = \xi(\omega)\right\}\right) = 1.$$

A convergência quase certa é uma convergência pontual em um conjunto de probabilidade 1: temos $\xi_n(\omega) \rightarrow \xi(\omega)$ para quase todo ω , isto é, exceto possivelmente em um conjunto de probabilidade nula.

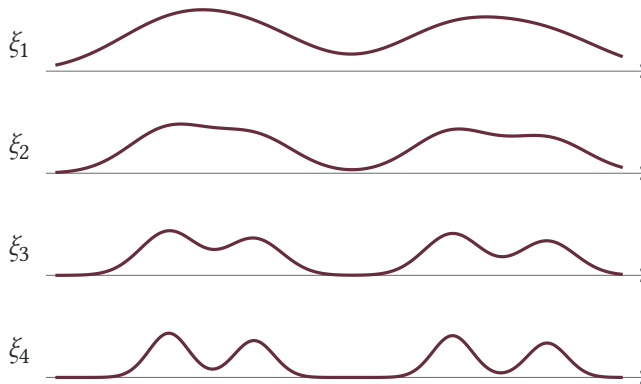


Figura 10.3.: As regiões onde ξ_n se afasta de zero se estreitam em torno de um conjunto finito; fora desse conjunto, $\xi_n(\omega) \rightarrow 0$.

Exemplo 10.13 — .

Em $\Omega = [0,1]$, com a medida de Lebesgue, para $n \geq 1$ definimos $\xi_n(x) = \begin{cases} 1 & \text{se } x \in \mathbb{Q} \\ 1/n & \text{caso contrário} \end{cases}$

Essa sequência não converge pontualmente para a função constante zero, pois vale 1 em todo racional para todo n , mas converge quase certamente para zero porque $\mathbb{Q} \cap [0,1]$ tem medida de Lebesgue nula.

**Heurística 10.14 — math.exchange**

Considere um homem que joga três moedas todas as manhãs. Todas as tardes, ele doa um real para uma instituição de caridade para cada cara que aparecer nos lançamentos. Porém, a primeira vez que o resultado for três coroas ele irá parar de doar permanentemente.

Seja $\{\xi_n\}_{n=1}^{\infty}$ o montante diário que a instituição de caridade recebeu dele. Podemos ter quase certeza de que um dia esse valor será zero e permanecerá zero para sempre depois disso, ou seja, temos que $\xi_n \xrightarrow{q.c.} 0$

Em qualquer prazo finito, porém, ainda há probabilidade positiva de que as doações continuem. A convergência quase certa não fornece uma data marcada para o último pagamento.

Teorema 10.15 — Critério de Convergência Quase Certa

Sejam $(\xi_n)_{n \geq 1}$ e ξ variáveis aleatórias. Então $\xi_n \xrightarrow{q.c.} \xi$ se, e somente se, para todo $\epsilon > 0$,

$$\mathbf{P}(|\xi_n - \xi| \geq \epsilon \text{ i.v.}) = 0.$$

Equivalentemente, basta verificar essa condição para $\epsilon = 1/k$, $k = 1, 2, \dots$

Demonstração. Para $k, n \geq 1$, ponha $A_{k,n} = \{|\xi_n - \xi| \geq 1/k\}$. Pela definição de convergência de seqüências reais,

$$\{\xi_n \not\rightarrow \xi\} = \bigcup_{k=1}^{\infty} \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_{k,m} = \bigcup_{k=1}^{\infty} \{A_{k,n} \text{ i.v.}\}.$$

Se $\xi_n \rightarrow \xi$ quase certamente, o evento à esquerda tem probabilidade zero; logo cada evento $\{A_{k,n} \text{ i.v.}\}$ tem probabilidade zero. Reciprocamente, se todos esses eventos têm probabilidade zero, sua união enumerável também tem probabilidade zero, e portanto $\xi_n \rightarrow \xi$ quase certamente.

Por fim, a condição para todos os $\epsilon > 0$ implica imediatamente a condição para $\epsilon = 1/k$. Para a recíproca, dado $\epsilon > 0$, escolha k com $1/k \leq \epsilon$ e use $\{|\xi_n - \xi| \geq \epsilon\} \subseteq A_{k,n}$. □

Teorema 10.16 —

Considere as variáveis aleatórias $(\xi_n)_{n \geq 1}$ e ξ . Se, para todo $\epsilon > 0$, tivermos

$$\sum_{n=1}^{\infty} \mathbf{P}(|\xi_n - \xi| > \epsilon) < \infty,$$

então $\xi_n \xrightarrow{q.c.} \xi$.

Demonstração. Para cada $k \geq 1$, aplique o primeiro lema de Borel–Cantelli aos eventos $A_n^{(k)} = \{|\xi_n - \xi| > 1/k\}$. A hipótese fornece $\sum_n \mathbf{P}(A_n^{(k)}) < \infty$, logo $\mathbf{P}(A_n^{(k)} \text{ i.v.}) = 0$. Fora da união desses eventos-limite, para todo k temos eventualmente $|\xi_n - \xi| \leq 1/k$; logo $\xi_n \rightarrow \xi$. Portanto, $\xi_n \xrightarrow{q.c.} \xi$. □

Exemplo 10.17 — .

Considere uma seqüência $\{\xi_n, n = 1, 2, 3, \dots\}$ tal que

$$\xi_n = \begin{cases} -\frac{1}{n} & \text{com probabilidade } \frac{1}{2} \\ \frac{1}{n} & \text{com probabilidade } \frac{1}{2} \end{cases}$$

Então $\xi_n \xrightarrow{q.c.} 0$.

Pelo teorema anterior basta mostrar que

$$\sum_{n=1}^{\infty} \mathbf{P}(|\xi_n| > \epsilon) < \infty.$$

Observe que $|\xi_n| = 1/n$. Logo, $|\xi_n| > \epsilon$ se e somente se $n < 1/\epsilon$. Logo temos

$$\begin{aligned} \sum_{n=1}^{\infty} \mathbf{P}(|\xi_n| > \epsilon) &\leq \sum_{n=1}^{\lfloor 1/\epsilon \rfloor} \mathbf{P}(|\xi_n| > \epsilon) \\ &\leq \lfloor 1/\epsilon \rfloor < \infty. \end{aligned}$$

◁

Exemplo 10.18 – Corcovas deslizantes.

Considere $\Omega = [0,1]$ com a medida de Lebesgue. Para cada $m \geq 2$, particione $[0,1]$ em m intervalos mensuráveis disjuntos $I_{m,1}, \dots, I_{m,m}$, cada um com medida $1/m$, e enumere em blocos as funções indicadoras

$$\mathbb{1}_{I_{2,1}}, \mathbb{1}_{I_{2,2}}, \mathbb{1}_{I_{3,1}}, \dots, \mathbb{1}_{I_{3,3}}, \dots$$

Denote a sequência resultante por (ξ_n) . Em cada bloco, temos $\mathbf{P}(\xi_n = 1) = 1/m$; portanto, ao avançarmos pelos blocos, $\mathbf{P}(|\xi_n| > \epsilon) \rightarrow 0$ para todo $0 < \epsilon < 1$. Assim, $\xi_n \rightarrow 0$ em probabilidade.

Por outro lado, em cada bloco e para cada $\omega \in [0,1]$, exatamente uma das indicadoras vale 1 em ω e as demais valem 0. Logo $(\xi_n(\omega))$ assume os valores 0 e 1 infinitas vezes e não converge em nenhum ponto.

◁

10.2.4. Convergência em $\mathcal{L}^p$

A convergência em $\mathcal{L}^p$ controla o tamanho médio do erro, medido pela potência p .

Definição 10.19 —

Fixe $p > 0$. Uma sequência de variáveis aleatórias $\{\xi_n\}_{n=1}^{\infty}$ **converge em $\mathcal{L}^p$** para a variável aleatória ξ se

$$\|\xi_n - \xi\|_p \rightarrow 0 \text{ quando } n \rightarrow \infty.$$

Em outras palavras

$$\mathbb{E}|\xi_n - \xi|^p \rightarrow 0 \text{ quando } n \rightarrow \infty.$$

Para $0 < p < 1$, a expressão $\|X\|_p = (\mathbb{E}|X|^p)^{1/p}$ não é uma norma, mas a mesma notação de convergência continua sendo útil.

Exemplo 10.20 — .

Seja $X_n \sim \text{Uni}\left(0, \frac{1}{n}\right)$. Então $X_n \xrightarrow{\mathcal{L}^p} 0$ para todo $p \geq 1$.

A densidade de X_n é dada por

$$f_{X_n}(x) = \begin{cases} n & 0 \leq x \leq \frac{1}{n} \\ 0 & \text{caso contrário} \end{cases}$$

Então

$$\begin{aligned} \mathbb{E}(|X_n - 0|^p) &= \int_0^{\frac{1}{n}} x^p n \, dx \\ &= \frac{1}{(p+1)n^p} \rightarrow 0, \quad \text{para todo } p \geq 1. \end{aligned}$$

Proposição 10.21 —

Se $0 < p \leq q < \infty$ e $\xi_n \xrightarrow{\mathcal{L}^q} \xi$, então $\xi_n \xrightarrow{\mathcal{L}^p} \xi$.

Demonstração. Se $p = q$, não há o que provar. Suponha $p < q$ e ponha $r = q/p > 1$, com expoente conjugado r' . Aplicando Hölder a $|\xi_n - \xi|^p \cdot 1$ e usando $\mathbf{P}(\Omega) = 1$, obtemos

$$\mathbf{E}|\xi_n - \xi|^p \leq (\mathbf{E}|\xi_n - \xi|^{pr})^{1/r} (\mathbf{E}1^{r'})^{1/r'} = (\mathbf{E}|\xi_n - \xi|^q)^{p/q}.$$

Logo

$$\|\xi_n - \xi\|_p \leq \|\xi_n - \xi\|_q \rightarrow 0.$$

□

Exemplo 10.22 — .

Considere uma sequência $\{X_n\}$ tal que

$$X_n = \begin{cases} n^2 & \text{com probabilidade } \frac{1}{n} \\ 0 & \text{com probabilidade } 1 - \frac{1}{n} \end{cases}$$

Mostre que

- $X_n \xrightarrow{p} 0$.
- X_n não converge em $\mathcal{L}^p$ para qualquer $p \geq 1$.
- Para mostrar que $X_n \xrightarrow{p} 0$, podemos escrever, para qualquer $\epsilon > 0$

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbf{P}(|X_n| \geq \epsilon) &= \lim_{n \rightarrow \infty} \mathbf{P}(X_n = n^2) \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \\ &= 0. \end{aligned}$$

Assim $X_n \xrightarrow{p} 0$.

□ Para qualquer $p \geq 1$, podemos escrever

$$\begin{aligned}\lim_{n \rightarrow \infty} \mathbf{E}(|X_n|^p) &= \lim_{n \rightarrow \infty} \left(n^{2p} \frac{1}{n} \right) \\ &= \lim_{n \rightarrow \infty} n^{2p-1} \\ &= \infty \quad (\text{pois } p \geq 1).\end{aligned}$$

Portanto, X_n não converge em $\mathcal{L}^p$ para qualquer $p \geq 1$. Em particular, embora $X_n \xrightarrow{\mathbf{P}} 0$, o valor esperado de X_n não converge para 0.

◀

10.3. Comparando os Modos de Convergência

As noções anteriores não formam apenas uma lista: há uma hierarquia entre elas, embora algumas implicações dependam de hipóteses adicionais. As relações principais estão resumidas no diagrama:

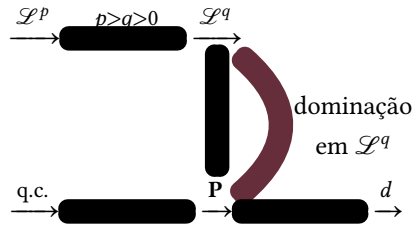


Figura 10.4.: Relação entre os modos de convergência em espaços de probabilidade.

Convergência quase certa implica convergência em probabilidade

Teorema 10.23 – Convergência q.c. e em Probabilidade

Dados ξ_n e ξ variáveis aleatórias.

- 1 $\xi_n \rightarrow \xi$ q.c. implica $\xi_n \rightarrow \xi$ em probabilidade.
- 2 Se $\xi_n \rightarrow \xi$ em probabilidade, então existe uma subsequência $\xi_{n_k} \rightarrow \xi$ quase certamente.

Demonstração.

- 1 Suponha $\xi_n \rightarrow \xi$ q.c. e fixe $\epsilon > 0$. Ponha $A_n = \{|\xi_n - \xi| > \epsilon\}$ e $B_N = \bigcup_{n \geq N} A_n$. Então $B_N \downarrow \limsup_n A_n$, cuja probabilidade é zero pelo Teorema 10.15. Como $\mathbf{P}(B_1) \leq 1 < \infty$, a continuidade da probabilidade por cima dá $\mathbf{P}(B_N) \downarrow 0$. Para $n \geq N$, $A_n \subset B_N$; logo $\mathbf{P}(A_n) \rightarrow 0$.

Concluimos portanto que $\xi_n \rightarrow \xi$ em probabilidade.

- 2 Suponha $\xi_n \rightarrow \xi$ em probabilidade. Queremos encontrar uma subsequência que converge quase certamente.

Pela convergência em probabilidade, podemos escolher recursivamente $n_1 < n_2 < \dots$ de modo que

$$\mathbf{P}\left(|\xi_{n_k} - \xi| > \frac{1}{k}\right) < 2^{-k}, \quad k = 1, 2, \dots$$

Como $\sum_k 2^{-k} < \infty$, o primeiro lema de Borel–Cantelli mostra que $|\xi_{n_k} - \xi| > 1/k$ ocorre apenas finitas vezes, quase certamente. Portanto, $\xi_{n_k} \rightarrow \xi$ quase certamente.

□

Convergência em $\mathcal{L}^p$ implica convergência em Probabilidade

Proposição 10.24 —

$\xi_n \xrightarrow{\mathcal{L}^p} \xi$ implica $\xi_n \xrightarrow{\mathbf{P}} \xi$.

Demonstração. Usando a desigualdade de Markov, temos, para todo $\epsilon > 0$,

$$\mathbf{P}(|\xi_n - \xi| > \epsilon) \leq \mathbf{P}(|\xi_n - \xi|^p \geq \epsilon^p) \leq \frac{\mathbf{E}|\xi_n - \xi|^p}{\epsilon^p} \rightarrow 0.$$

□

Vimos que se a sequência ξ_n converge em $\mathcal{L}^p$ para ξ então a sequência também converge em probabilidade, porém a recíproca nem sempre é verdadeira.

Teorema 10.25 — Convergência dominada em probabilidade

Se $p > 0$, $\xi_n \xrightarrow{\mathbf{P}} \xi$ e existe uma variável aleatória não negativa $\eta \in \mathcal{L}^p$ tal que $|\xi_n| \leq \eta$ quase certamente para todo $n \geq 1$, então $\xi_n \xrightarrow{\mathcal{L}^p} \xi$.

Demonstração.

Toda subsequência (ξ_{n_k}) possui, pela parte anterior, uma subsubsequência $(\xi_{n_{k_j}})$ que converge para ξ quase certamente. Ao passar ao limite nessa subsubsequência, obtemos $|\xi| \leq \eta$ quase certamente. Consequentemente,

$$|\xi_{n_{k_j}} - \xi|^p \leq (|\xi_{n_{k_j}}| + |\xi|)^p \leq 2^p \eta^p.$$

O lado direito é integrável e o lado esquerdo converge quase certamente para zero. Pelo Teorema da Convergência Dominada, $\mathbf{E}|\xi_{n_{k_j}} - \xi|^p \rightarrow 0$.

Assim, toda subsequência da sequência numérica $a_n = \mathbf{E}|\xi_n - \xi|^p$ possui uma subsubsequência convergente a zero. Se a_n não convergisse a zero, haveria uma subsequência limitada inferiormente por alguma constante positiva, uma contradição. Logo $a_n \rightarrow 0$.

□

Teorema 10.26 — Lema de Scheffé para variáveis não negativas

Sejam (ξ_n) e ξ variáveis aleatórias não negativas e integráveis. Se $\xi_n \xrightarrow{q.c.} \xi$ e $\mathbf{E}[\xi_n] \rightarrow \mathbf{E}[\xi]$, então $\xi_n \xrightarrow{\mathcal{L}^1} \xi$.

Demonstração. Decompondo nos conjuntos $\{\xi_n \leq \xi\}$ e $\{\xi_n > \xi\}$,

$$\begin{aligned} \mathbf{E}|\xi_n - \xi| &= \mathbf{E}[(\xi - \xi_n)\mathbb{1}_{\{\xi_n \leq \xi\}}] + \mathbf{E}[(\xi_n - \xi)\mathbb{1}_{\{\xi_n > \xi\}}] \\ &= 2\mathbf{E}[(\xi - \xi_n)\mathbb{1}_{\{\xi_n \leq \xi\}}] + \mathbf{E}[\xi_n] - \mathbf{E}[\xi]. \end{aligned}$$

Além disso, $0 \leq (\xi - \xi_n)\mathbb{1}_{\{\xi_n \leq \xi\}} \leq \xi$, e esse termo converge quase certamente para zero. A convergência dominada e a hipótese sobre as esperanças mostram que os dois termos da última expressão tendem a zero. $\square$

Convergência em Probabilidade implica Convergência em Distribuição

Teorema 10.27 —

Se uma sequência de variáveis aleatórias ξ_n converge em probabilidade para uma variável aleatória ξ , então ξ_n também converge em distribuição para ξ .

Demonstração. Tome $\epsilon > 0$ e note que, como $\xi_n \xrightarrow{\mathbf{P}} \xi$, então

$$\lim_{n \rightarrow \infty} \mathbf{P}(|\xi_n - \xi| > \epsilon) = 0.$$

Fixado também $\delta > 0$, existe $n_0 \geq 1$ tal que

$$\mathbf{P}(|\xi_n - \xi| > \epsilon) < \delta,$$

para todo $n \geq n_0$.

Se $\xi_n \leq x$ e $|\xi_n - \xi| \leq \epsilon$, então $\xi \leq x + \epsilon$. Logo, para $n \geq n_0$,

$$\mathbf{P}(\xi_n \leq x) \leq \mathbf{P}(\xi \leq x + \epsilon) + \delta.$$

Concluimos assim que

$$\limsup_n \mathbf{P}(\xi_n \leq x) \leq \mathbf{P}(\xi \leq x + \epsilon),$$

para todo $\epsilon > 0$.

De modo análogo, se $\xi \leq x - \epsilon$ e $|\xi_n - \xi| \leq \epsilon$, então $\xi_n \leq x$. Portanto,

$$\mathbf{P}(\xi \leq x - \epsilon) \leq \mathbf{P}(\xi_n \leq x) + \delta.$$

Segue então que para todo $\epsilon > 0$,

$$\mathbf{P}(\xi \leq x - \epsilon) \leq \liminf_n \mathbf{P}(\xi_n \leq x) \leq \limsup_n \mathbf{P}(\xi_n \leq x) \leq \mathbf{P}(\xi \leq x + \epsilon).$$

Tomando x como ponto de continuidade de F_ξ e fazendo $\epsilon \downarrow 0$, encontramos

$$\lim_n \mathbf{P}(\xi_n \leq x) = \mathbf{P}(\xi \leq x),$$

concluindo o resultado. □

10.4. Teorema de Representação de Skorokhod

A convergência em distribuição compara leis, não trajetórias, e por isso não exige que as variáveis estejam no mesmo espaço. O Teorema de Representação de Skorokhod mostra até onde podemos inverter essa fraqueza: em um espaço adequado, é possível realizar cópias com as mesmas leis que convergem quase certamente.

Teorema 10.28 — Teorema de Representação de Skorokhod

Suponha que $\xi_n \xrightarrow{d} \xi$. Então existem variáveis aleatórias ξ'_n com a mesma distribuição que ξ_n e uma variável aleatória ξ' com a mesma distribuição que ξ , tais que

$$\xi'_n \xrightarrow{q.c.} \xi'.$$

Demonstração. Sejam F_n e F as funções de distribuição de ξ_n e ξ , respectivamente. Em $(0,1)$, munido da medida de Lebesgue, considere a variável uniforme $U(u) = u$ e defina

$$\xi'_n = F_n^*(U) \quad \text{e} \quad \xi' = F^*(U),$$

onde as estrelas denotam as inversas generalizadas. Pela Proposição 6.15, $\xi'_n \stackrel{d}{=} \xi_n$ e $\xi' \stackrel{d}{=} \xi$.

Resta provar a convergência quase certa. Fixe $u \in (0,1)$ em que F^* é contínua e fixe $\varepsilon > 0$. Como uma função monótona possui no máximo uma quantidade enumerável de descontinuidades, podemos escolher um ponto de continuidade x_- de F tal que

$$F^*(u) - \varepsilon < x_- < F^*(u).$$

Pela caracterização $F^*(u) \leq x \iff u \leq F(x)$, a desigualdade $x_- < F^*(u)$ implica $F(x_-) < u$. Como $F_n(x_-) \rightarrow F(x_-)$, temos $F_n(x_-) < u$ para todo n suficientemente grande e, novamente pela mesma caracterização,

$$F_n^*(u) > x_- > F^*(u) - \varepsilon.$$

Daí, $\liminf_n F_n^*(u) \geq F^*(u)$.

Para obter a desigualdade oposta, a continuidade de F^* em u permite escolher $u' \in (u,1)$ tão próximo de u que $F^*(u') < F^*(u) + \varepsilon/2$. Escolha então um ponto de continuidade x_+ de F no intervalo

$$F^*(u') < x_+ < F^*(u) + \varepsilon.$$

Como $F^*(u') \leq x_+$, temos $u' \leq F(x_+)$. Portanto, $F_n(x_+) \rightarrow F(x_+)$ e, como $u' > u$, vale $F_n(x_+) > u$ para todo n suficientemente grande. Segue que

$$F_n^*(u) \leq x_+ < F^*(u) + \varepsilon,$$

e assim $\limsup_n F_n^*(u) \leq F^*(u)$. Concluimos que $F_n^*(u) \rightarrow F^*(u)$ em todo ponto de continuidade de F^* .

Pelo Corolário 6.17, o conjunto D_{F^*} das descontinuidades de F^* é no máximo enumerável. Como U é uniforme, $\mathbf{P}(U \in D_{F^*}) = 0$. Logo

$$\xi'_n = F_n^*(U) \longrightarrow F^*(U) = \xi' \quad \text{quase certamente,}$$

como queríamos. □

10.5. Princípio de Seleção de Helly

O Princípio de Seleção de Helly fornece uma forma de compacidade para funções de distribuição, mas exige uma ressalva: o limite obtido pode perder massa no infinito e, portanto, não ser ele próprio uma função de distribuição.

Teorema 10.29 — Princípio de Seleção de Helly

Seja (F_n) uma sequência de funções de distribuição. Então existem uma subseqüência (F_{n_k}) e uma função limitada, contínua à direita e não decrescente F tais que

$$F_{n_k}(x) \longrightarrow F(x) \quad \text{em todo ponto de continuidade } x \text{ de } F.$$

Demonstração. Enumere os racionais como $\mathbb{Q} = \{q_1, q_2, \dots\}$. Como cada sequência $(F_n(q_j))_{n \geq 1}$ toma valores em $[0, 1]$, aplicações sucessivas de Bolzano–Weierstrass e o argumento diagonal fornecem uma subseqüência (F_{n_k}) para a qual $F_{n_k}(q)$ converge para todo $q \in \mathbb{Q}$. Escreva

$$G(q) = \lim_{k \rightarrow \infty} F_{n_k}(q), \quad q \in \mathbb{Q}.$$

A monotonicidade das funções F_{n_k} implica que G é não decrescente.

Defina, para $x \in \mathbb{R}$,

$$F(x) = \inf\{G(q) : q \in \mathbb{Q}, q > x\}.$$

É imediato que F toma valores em $[0,1]$ e é não decrescente. Além disso, se $L = \inf_{y>x} F(y)$, então $L \geq F(x)$. Reciprocamente, dado $\varepsilon > 0$, escolha $q \in \mathbb{Q}$, $q > x$, tal que $G(q) < F(x) + \varepsilon$ e depois escolha y com $x < y < q$. Como o racional q participa do ínfimo que define $F(y)$,

$$L \leq F(y) \leq G(q) < F(x) + \varepsilon.$$

Logo $L = F(x)$, o que prova a continuidade à direita de F .

Resta provar a convergência. Fixe $x \in \mathbb{R}$ e $\varepsilon > 0$. Pela definição de $F(x)$, existe $q \in \mathbb{Q}$, $q > x$, tal que $G(q) < F(x) + \varepsilon/2$. Para k suficientemente grande, $F_{n_k}(q) < F(x) + \varepsilon$ e, pela monotonicidade,

$$F_{n_k}(x) \leq F_{n_k}(q) < F(x) + \varepsilon.$$

Portanto, $\limsup_k F_{n_k}(x) \leq F(x)$.

Suponha agora que x seja ponto de continuidade de F . Escolha $r < x$ de modo que $F(r) > F(x) - \varepsilon/2$ e um racional q com $r < q < x$. Da definição de $F(r)$,

$$G(q) \geq F(r) > F(x) - \varepsilon/2.$$

Para k suficientemente grande, $F_{n_k}(q) > F(x) - \varepsilon$ e, novamente pela monotonicidade,

$$F_{n_k}(x) \geq F_{n_k}(q) > F(x) - \varepsilon.$$

Assim, $\liminf_k F_{n_k}(x) \geq F(x)$. As duas estimativas provam que $F_{n_k}(x) \rightarrow F(x)$ em todo ponto de continuidade de F . □

O princípio de seleção de Helly diz que, dada $(F_n)_{n \geq 1}$, alguma subsequência $F_{n_k}(x)$ converge para uma função $F(x)$ em todos os pontos de continuidade de F . A função limite é não decrescente e contínua à direita, mas ainda pode não ser uma função de distribuição.

Falta garantir que

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

O princípio de Helly, sozinho, não garante essas duas igualdades: à medida que n cresce, pode haver massa de probabilidade escapando para $+\infty$ ou para $-\infty$, como mostra o exemplo abaixo.

Exemplo 10.30 — .

Dada uma variável aleatória U uniformemente distribuída em $[0,1]$ e constantes $\delta, \epsilon > 0$ com $\delta + \epsilon < 1$, defina

$$X_n := -n \cdot \mathbb{1}_{\{U < \delta\}} + n \cdot \mathbb{1}_{\{U > 1-\epsilon\}},$$

para $n \geq 1$.

Note que para todo $t > 0$ temos que

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbf{P}(X_n > t) &= \mathbf{P}(U > 1 - \epsilon) = \epsilon, \quad \text{e} \\ \lim_{n \rightarrow \infty} \mathbf{P}(X_n < -t) &= \mathbf{P}(U < \delta) = \delta, \end{aligned}$$

de modo que quando $n \rightarrow \infty$ uma “massa” de probabilidade ϵ está sendo perdida em $+\infty$ e δ em $-\infty$.

Como consequência, temos

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = \begin{cases} \delta & , \text{ para } x < 0; \\ 1 - \epsilon & , \text{ para } x \geq 0, \end{cases}$$

que é uma função não-decrescente e contínua pela direita, mas não é uma função distribuição.

◀

Uma forma de resolver esse problema é garantir que a massa de probabilidade fique quase toda contida em um intervalo limitado, uniformemente ao longo da sequência. Isto é, para todo $\epsilon > 0$ devem existir $a, b \in \mathbb{R}$ tais que $\mathbf{P}(a < X_n \leq b) > 1 - \epsilon$ para todo $n \geq 1$.

Esta propriedade é denominada *tightness* e está precisamente definida abaixo.

Definição 10.31 – Tightness

Diremos que uma família $\{\mu_i : i \in I\}$ de medidas de probabilidade em $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ é **tight** se, para todo $\epsilon > 0$, existe um intervalo compacto $[a, b]$ tal que

$$\mu_i([a, b]) > 1 - \epsilon \quad \text{para todo } i \in I.$$

Analogamente, uma família $\{X_i : i \in I\}$ de variáveis aleatórias é tight se a família de suas distribuições $\{\mathbf{P}_{X_i} : i \in I\}$ é tight. Usaremos a mesma terminologia para a família correspondente de funções de distribuição.

Proposição 10.32 – Convergência em distribuição implica tightness

Se $X_n \xrightarrow{d} X$, então a família $\{X_n : n \geq 1\}$ é tight.

Demonstração. Fixe $\epsilon > 0$ e escreva F_n e F para as funções de distribuição de X_n e X . Escolha pontos de continuidade $a < b$ de F tais que

$$F(a) < \frac{\epsilon}{4}, \quad F(b) > 1 - \frac{\epsilon}{4}.$$

Como $F_n(a) \rightarrow F(a)$ e $F_n(b) \rightarrow F(b)$, para todo n suficientemente grande,

$$\mathbf{P}(a < X_n \leq b) = F_n(b) - F_n(a) > 1 - \epsilon.$$

Restam apenas finitos índices. Para cada um deles, a continuidade por baixo em $\{|X_j| \leq M\} \uparrow \Omega$ permite escolher um intervalo compacto que contenha massa maior que $1 - \epsilon$. Tome então um único intervalo compacto que contenha todos esses intervalos e também $[a, b]$. Esse intervalo funciona simultaneamente para todo n . $\square$

Teorema 10.33 — Teorema de Prokhorov em $\mathbb{R}$

Se $(\mathbf{P}_n)$ é uma sequência tight de medidas de probabilidade em $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, então existem uma subsequência $(\mathbf{P}_{n_k})$ e uma medida de probabilidade $\mathbf{P}$ tais que $\mathbf{P}_{n_k}$ converge fracamente para $\mathbf{P}$.

Demonstração. Seja $F_n(x) = \mathbf{P}_n((-\infty, x])$. Pelo princípio de seleção de Helly, existem uma subsequência (F_{n_k}) e uma função $F : \mathbb{R} \rightarrow [0, 1]$, não decrescente e contínua à direita, tais que $F_{n_k}(x) \rightarrow F(x)$ em todo ponto de continuidade de F .

Fixe $\varepsilon > 0$. Pela tightness, existe $[a_0, b_0]$ tal que $\mathbf{P}_n([a_0, b_0]) > 1 - \varepsilon$ para todo n . Como F possui no máximo uma quantidade enumerável de descontinuidades, podemos escolher pontos de continuidade $a < a_0$ e $b > b_0$. Então

$$F(b) - F(a) = \lim_{k \rightarrow \infty} (F_{n_k}(b) - F_{n_k}(a)) = \lim_{k \rightarrow \infty} \mathbf{P}_{n_k}((a, b]) \geq 1 - \varepsilon.$$

Se $L_- := \lim_{x \rightarrow -\infty} F(x)$ e $L_+ := \lim_{x \rightarrow \infty} F(x)$, segue que $L_+ - L_- \geq 1 - \varepsilon$. Como isso vale para todo $\varepsilon > 0$ e $0 \leq L_- \leq L_+ \leq 1$, necessariamente $L_- = 0$ e $L_+ = 1$.

Portanto, F é uma função de distribuição. Pela Proposição 6.15, existe uma variável aleatória X com função de distribuição F ; seja $\mathbf{P}$ a lei de X . A convergência obtida nos pontos de continuidade de F é precisamente $\mathbf{P}_{n_k} \Rightarrow \mathbf{P}$. $\square$

Assim, tomando X uma variável aleatória com função distribuição F , mostramos o seguinte resultado.

Proposição 10.34 —

Se $(X_n)_{n \geq 1}$ é uma sequência tight de variáveis aleatórias, então existem uma subsequência (n_k) e uma variável aleatória X tais que $X_{n_k} \xrightarrow{d} X$.

Demonstração. A família das distribuições $(\mathbf{P}_{X_n})$ é tight. Pelo teorema anterior, alguma subsequência $(\mathbf{P}_{X_{n_k}})$ converge fracamente para uma medida de probabilidade $\mathbf{P}$. No espaço canônico $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbf{P})$, a aplicação identidade define uma variável aleatória X com lei $\mathbf{P}$. Pela definição de convergência em distribuição, $X_{n_k} \xrightarrow{d} X$. □

Exemplo 10.35 — .

Sejam $X_1, X_2, \dots$ variáveis aleatórias independentes, identicamente distribuídas e integráveis, e ponha $S_n = X_1 + \dots + X_n$. Pela desigualdade triangular,

$$\mathbf{E}|S_n| \leq \sum_{j=1}^n \mathbf{E}|X_j| = n\mathbf{E}|X_1|.$$

A desigualdade de Markov fornece, para todo $t > 0$,

$$\mathbf{P}\left(\left|\frac{S_n}{n}\right| > t\right) \leq \frac{\mathbf{E}|S_n|}{nt} \leq \frac{\mathbf{E}|X_1|}{t}.$$

O último termo tende a zero quando $t \rightarrow \infty$, uniformemente em n . Logo $(S_n/n)_{n \geq 1}$ é tight e, pela proposição anterior, possui uma subsequência convergente em distribuição. ◀

A tightness impede que a massa escape e garante limites ao menos ao longo de subsequências. Para obter a convergência da sequência inteira, resta mostrar que todas as subsequências convergentes conduzem ao mesmo limite.

Proposição 10.36 — Critério de unicidade

Seja $(X_n)_{n \geq 1}$ uma sequência tight. Suponha que existe uma variável aleatória X tal que toda subsequência de (X_n) que converge em distribuição tem como limite X . Então $X_n \xrightarrow{d} X$.

Demonstração. Denote por F a função de distribuição de X e por F_n a de X_n . Se X_n não convergisse em distribuição para X , existiriam um ponto de continuidade x de F , uma constante $\epsilon > 0$ e uma subsequência (n_k) tais que

$$|F_{n_k}(x) - F(x)| \geq \epsilon \quad \text{para todo } k.$$

A subsequência (X_{n_k}) continua tight. Pela Proposição 10.34, ela possui uma subsubsequência convergente em distribuição; pela hipótese, seu limite é X . Logo $F_{n_{k_l}}(x) \rightarrow F(x)$, contradizendo a desigualdade anterior. $\square$

Os modos de convergência não formam uma única cadeia, mas cumprem papéis distintos. A convergência quase certa controla trajetórias; a convergência em probabilidade controla conjuntos excepcionais; a convergência em L^p acrescenta controle de tamanho; e a convergência em distribuição retém apenas as leis. Skorokhod permite, sob hipóteses adequadas, representar a última por convergência quase certa, enquanto Helly e a tightness fornecem os limites por subsequências. Saber que um limite pode ser extraído é metade do argumento; identificá-lo é a outra metade.

EXERCÍCIOS

Ex. 10.1 — Seja $(\xi_n)_{n \geq 1}$ uma sequência de variáveis aleatórias tal que

$$\xi_n \sim \text{Ge}\left(\frac{\lambda}{n}\right), \quad n > \lambda,$$

onde $\lambda > 0$ e $\text{Ge}(p)$ conta o número de fracassos antes do primeiro sucesso. Defina $\eta_n = \xi_n/n$. Calcule $\mathbf{P}(\eta_n > x)$ e mostre que η_n converge em distribuição para uma variável exponencial de taxa λ .

Ex. 10.2 — Sejam (ξ_n) e (η_n) duas sequências de variáveis aleatórias definidas no mesmo espaço de probabilidade. Suponha que

$$\xi_n \xrightarrow{\mathbf{P}} \xi \quad \text{e} \quad \eta_n \xrightarrow{\mathbf{P}} \eta.$$

Prove que $\xi_n + \eta_n \xrightarrow{\mathbf{P}} \xi + \eta$.

Ex. 10.3 — Considere uma sequência (ξ_n) tal que

$$\xi_n = \begin{cases} n, & \text{com probabilidade } 1/n^2, \\ 0, & \text{com probabilidade } 1 - 1/n^2. \end{cases}$$

Mostre que:

1. $\xi_n \xrightarrow{\mathbf{P}} 0$.
2. $\xi_n \xrightarrow{\mathcal{L}^r} 0$ para $0 < r < 2$.
3. ξ_n não converge para zero em $\mathcal{L}^r$ para $r \geq 2$.
4. $\xi_n \xrightarrow{q.c.} 0$.

Ex. 10.4 — Suponha que $\xi_n \xrightarrow{\mathcal{L}^r} \xi$ para $r \geq 1$. Mostre que

$$\mathbf{E}|\xi_n|^r \longrightarrow \mathbf{E}|\xi|^r.$$

Ex. 10.5 — Suponha que $\xi_n \xrightarrow{d} \xi$ e $\eta_n \xrightarrow{\mathbf{P}} c$, onde c é constante. Prove que

$$\eta_n \xi_n \xrightarrow{d} c\xi.$$

Ex. 10.6 — Prove o Teorema da Aplicação Contínua para a convergência em probabilidade: se $g : \mathbb{R} \rightarrow \mathbb{R}$ é contínua, então

$$\xi_n \xrightarrow{\mathbf{P}} \xi \implies g(\xi_n) \xrightarrow{\mathbf{P}} g(\xi).$$

Ex. 10.7 — Seja (ξ_n) uma sequência de variáveis aleatórias mutuamente independentes que converge em probabilidade para ξ . Mostre que ξ é constante quase certamente.

Dica. Escolha duas subsequências com índices disjuntos. Seus pares permanecem independentes, mas convergem em probabilidade para (ξ, ξ) . Conclua que duas cópias independentes com a lei de ξ são iguais quase certamente.

Ex. 10.8 — Prove que, se $\xi_n \xrightarrow{\mathcal{L}^2} \xi$, então

$$\text{Var}(\xi_n) \longrightarrow \text{Var}(\xi).$$

Ex. 10.9 — Mostre diretamente que a convergência em $\mathcal{L}^p$ implica convergência em probabilidade para todo $p > 0$. Dê um contraexemplo para a recíproca, por exemplo quando $p = 1$.

Ex. 10.10 — Considere variáveis aleatórias $\xi_1, \xi_2, \dots$ com funções de distribuição $F_1, F_2, \dots$. Suponha que $\xi_n \xrightarrow{d} \xi$ e que a função de distribuição F de ξ seja contínua. Prove que $F_n \rightarrow F$ uniformemente, isto é,

$$\|F_n - F\|_\infty := \sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \rightarrow 0.$$

Ex. 10.11 — Considere variáveis aleatórias de valores inteiros $\xi_1, \xi_2, \dots$ e uma variável aleatória ξ também de valores inteiros. Mostre que $\xi_n \xrightarrow{d} \xi$ se, e somente se,

$$\mathbf{P}(\xi_n = k) \rightarrow \mathbf{P}(\xi = k) \quad \text{para todo } k \in \mathbb{Z}.$$

Na volta, justifique por que a massa não pode escapar para o infinito.

Ex. 10.12 — Considere variáveis aleatórias independentes $\xi_1, \xi_2, \dots$, uniformemente distribuídas em $[0, 1]$, e defina $Z_n = \max_{k \leq n} \xi_k$.

1. Prove que $Z_n \rightarrow 1$ em probabilidade.
2. Prove que $Z_n \rightarrow 1$ quase certamente.

Ex. 10.13 — Da hipergeométrica à binomial. Para cada $n \geq 5$, uma urna contém n bolas azuis e n bolas vermelhas. Retiram-se dez bolas, sem reposição, e X_n denota o número de bolas azuis retiradas. Determine a lei de X_n e prove que

$$X_n \xrightarrow{d} \text{Binomial}(10, 1/2).$$

Teoremas-limite

Repetir um experimento não elimina o acaso de cada observação, mas pode estabilizar sua média. Esse é o conteúdo das leis dos grandes números. Para somas

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

de variáveis aleatórias independentes com média comum μ , estudaremos condições sob as quais

$$\frac{S_n}{n} \longrightarrow \mu.$$

As flutuações individuais não desaparecem; tornam-se pequenas na escala da média.

As versões fraca e forte diferem pelo modo de convergência e pelas hipóteses exigidas. Começaremos pelo controle da variância e passaremos à convergência quase certa. Duas aplicações encerram o capítulo: a integração de Monte Carlo e a aproximação por polinômios de Bernstein.

11.1. Lei fraca dos grandes números

A lei fraca afirma a convergência de S_n/n em probabilidade. Com variâncias uniformemente limitadas, obtemos primeiro a convergência em $\mathcal{L}^2$; a desigualdade de Chebyshev fornece então a convergência em probabilidade.

Teorema 11.1 — Lei Fraca dos Grandes Números

Sejam $\xi_1, \xi_2, \dots$ variáveis aleatórias independentes com $E[\xi_k] = \mu$ e $\text{Var}(\xi_k) \leq c < \infty$, para todo $k \geq 1$. Nestas condições, se $S_n = \xi_1 + \xi_2 + \dots + \xi_n$, então

$$\frac{S_n}{n} \xrightarrow{\mathcal{L}^2} \mu \quad \text{e} \quad \frac{S_n}{n} \xrightarrow{p} \mu.$$

Demonstração. Basta mostrar que

$$E \left[\left(\frac{S_n}{n} - \mu \right)^2 \right] \rightarrow 0.$$

A convergência em probabilidade seguirá da convergência em $\mathcal{L}^2$, pela desigualdade de Chebyshev.

Pela linearidade da esperança, $E \left[\frac{S_n}{n} \right] = \mu$. A independência permite somar as variâncias:

$$\begin{aligned} E \left[\left(\frac{S_n}{n} - \mu \right)^2 \right] &= \text{Var} \left(\frac{S_n}{n} \right) = \frac{1}{n^2} \text{Var}(S_n) \\ &= \frac{1}{n^2} (\text{Var}(\xi_1) + \dots + \text{Var}(\xi_n)) \\ &\leq \frac{cn}{n^2} \\ &= \frac{c}{n} \rightarrow 0. \end{aligned}$$

□

A demonstração fornece uma taxa $O(n^{-1})$ para o erro quadrático médio e, por Chebyshev, para a probabilidade de um desvio fixo. A estimativa ainda não é somá-

vel em n : para concluir convergência quase certa por Borel–Cantelli, precisaremos de outro argumento.

Observação 11.2 —

A demonstração permanece válida para variáveis aleatórias ξ_k não necessariamente independentes, desde que sejam duas a duas não correlacionadas. De fato, a independência foi usada apenas para escrever a variância da soma como a soma das variâncias.

A mesma ideia se estende a vetores aleatórios de dimensão finita.

Teorema 11.3 — Lei fraca dos grandes números para vetores

Sejam $\xi^{(1)}, \xi^{(2)}, \dots$ vetores aleatórios i.i.d. com valores em $\mathbb{R}^d$ e $\mathbf{E}\|\xi^{(1)}\|^2 < \infty$.

Então

$$\frac{1}{n} \sum_{i=1}^n \xi^{(i)} \xrightarrow{\mathbf{P}} \mathbf{E}[\xi^{(1)}].$$

Demonstração. Escreva $\xi^{(i)} = (\xi_1^{(i)}, \dots, \xi_d^{(i)})$. Cada sequência de coordenadas $\xi_j^{(1)}, \xi_j^{(2)}, \dots$ satisfaz as hipóteses do Teorema 11.1; logo, para todo j ,

$$\frac{1}{n} \sum_{i=1}^n \xi_j^{(i)} \xrightarrow{\mathbf{P}} \mathbf{E}[\xi_j^{(1)}].$$

Para a norma do máximo e qualquer $\varepsilon > 0$, a probabilidade de que o vetor se afaste de sua média por mais de ε é limitada pela soma das d probabilidades correspondentes às coordenadas. Como d é finito, essa soma tende a zero. $\square$

O caso dos ensaios de Bernoulli antecede historicamente as versões gerais. Nele, a média amostral é a frequência relativa de sucessos, e a lei dos grandes números dá uma forma precisa à interpretação frequentista da probabilidade.

Teorema 11.4 — Lei dos Grandes Números de Bernoulli

Seja S_n o número de sucessos em n ensaios independentes, cada um com probabilidade de sucesso p . Então S_n tem distribuição binomial com parâmetros (n, p) e

$$\frac{S_n}{n} \rightarrow p$$

em probabilidade, quando $n \rightarrow \infty$.

Demonstração. Se ξ_k é a indicadora de sucesso no k -ésimo ensaio, então as variáveis ξ_k são independentes, $S_n = \sum_{k=1}^n \xi_k$, $E[\xi_k] = p$ e $\text{Var}(\xi_k) = p(1-p) \leq 1/4$. Basta aplicar o Teorema 11.1. □

Exemplo 11.5 — .

Ao lançar uma moeda equilibrada n vezes, a proporção de caras converge em probabilidade para $1/2$. Em particular,

$$P(0,49n \leq \text{número de caras} \leq 0,51n) \rightarrow 1.$$

Isso não obriga a moeda a compensar um excesso anterior de caras. A moeda não contrai dívidas: a lei controla a proporção acumulada, não a previsão do próximo lançamento.



Podemos dispensar a hipótese de variância finita se, em contrapartida, supusermos que as variáveis aleatórias ξ_k são identicamente distribuídas. A ferramenta será o método de truncamento de Markov: primeiro controlamos variáveis limitadas e depois estimamos a contribuição das caudas.

Mais adiante, um exercício apresenta uma extensão para somas independentes não necessariamente identicamente distribuídas.

Teorema 11.6 — Método de Truncamento de Markov

Considere ξ uma variável aleatória, com média finita. Considere $M > 1$ o nível de truncamento.

Defina

$$[\xi]^M := \begin{cases} \xi, & \text{se } |\xi| \leq M \\ 0, & \text{se } |\xi| > M \end{cases} = \xi \cdot \mathbf{1}_{\{|\xi| \leq M\}}.$$

Então

- $|[\xi]^M| \leq M$ pontualmente, e $|[\xi]^M| \leq |\xi|$
- $[\xi]^M \rightarrow \xi$ pontualmente quando $M \rightarrow \infty$,
- A variável $[\xi]^M$ possui todos os momentos finitos.
- Pelo Teorema da Convergência Dominada $\mathbf{E}[[\xi]^M] \rightarrow \mathbf{E}[\xi]$ quando $M \rightarrow \infty$. Mais ainda,

$$\mathbf{E}|[\xi]^M - \xi| \rightarrow 0 \text{ quando } M \rightarrow \infty. \quad (11.1)$$

Demonstração. As duas desigualdades do primeiro item seguem diretamente da definição. Para cada ponto em que ξ assume um valor real, quando M ultrapassa $|\xi|$ o truncamento coincide com ξ ; daí a convergência pontual. Além disso, para todo $r > 0$,

$$\mathbf{E}[|[\xi]^M|^r] \leq M^r < \infty.$$

Por fim, $|[\xi]^M| \leq |\xi|$ e $[\xi]^M \rightarrow \xi$; como ξ é integrável, o Teorema da Convergência Dominada dá a convergência das esperanças. Aplicando-o também a $|[\xi]^M - \xi| = |\xi| \mathbf{1}_{\{|\xi| > M\}}$, obtemos (11.1). □

Corolário 11.7 —

Dada uma variável aleatória integrável ξ ,

$$\mathbf{E} \left[|\xi| \cdot \mathbf{1}_{\{|\xi| > M\}} \right] \rightarrow 0 \text{ quando } M \rightarrow \infty.$$

Demonstração. Pela definição do truncamento,

$$|\xi - \lfloor \xi \rfloor^M| = |\xi| \mathbf{1}_{\{|\xi| > M\}}.$$

A conclusão é exatamente (11.1). □

Teorema 11.8 — Lei Fraca dos Grandes Números

Sejam $\xi_1, \xi_2, \dots$ variáveis aleatórias independentes identicamente distribuídas com $\mathbf{E}[\xi_k] = \mu$ para todo k .

Considere $S_n = \xi_1 + \xi_2 + \dots + \xi_n$.

Então

$$\frac{S_n}{n} \rightarrow \mu$$

em probabilidade (quando $n \rightarrow \infty$).

Demonstração. Dado $M > 1$, considere

$$\xi_k^{(M)} := \xi_k \mathbf{1}_{\{|\xi_k| \leq M\}}, \quad S_n^{(M)} := \sum_{k=1}^n \xi_k^{(M)}.$$

As variáveis $\xi_k^{(M)}$ continuam i.i.d. e

$$\text{Var}(\xi_k^{(M)}) \leq \mathbf{E}[(\xi_k^{(M)})^2] \leq M^2.$$

Pelo Teorema 11.1,

$$\frac{S_n^{(M)}}{n} \xrightarrow{\mathcal{L}^2} \mathbf{E}[\xi_1^{(M)}].$$

Em particular, por Cauchy–Schwarz,

$$\mathbf{E} \left| \frac{S_n^{(M)}}{n} - \mathbf{E}[\xi_1^{(M)}] \right| \rightarrow 0.$$

Pela desigualdade triangular e pela distribuição comum,

$$\mathbf{E} \left| \frac{S_n}{n} - \mathbf{E}[\xi_1] \right| \leq \mathbf{E} \left| \frac{S_n^{(M)}}{n} - \mathbf{E}[\xi_1^{(M)}] \right| + 2\mathbf{E}|\xi_1 - \xi_1^{(M)}|.$$

Portanto, para todo M fixo,

$$\limsup_{n \rightarrow \infty} \mathbf{E} \left| \frac{S_n}{n} - \mu \right| \leq 2\mathbf{E}|\xi_1 - \xi_1^{(M)}|.$$

Fazendo $M \rightarrow \infty$ e usando (11.1), concluímos na verdade que $S_n/n \rightarrow \mu$ em $\mathcal{L}^1$ e, portanto, em probabilidade. $\square$

11.2. Lei forte dos grandes números

Na lei fraca, o conjunto em que a média se afasta de μ pode mudar com n ; exigimos apenas que sua probabilidade tenda a zero. A lei forte afirma algo sobre a sequência inteira: para variáveis i.i.d. integráveis,

$$\frac{S_n}{n} \rightarrow \mu \text{ q.c.}$$

Começaremos pelo caso de quarto momento finito. A estimativa resultante será somável, o que permitirá aplicar Borel–Cantelli; depois removeremos essa hipótese adicional.

Teorema 11.9 – Lei Forte dos Grandes Números - I

Sejam $\xi_1, \xi_2, \dots$ variáveis aleatórias i.i.d. com esperança μ e assumamos $\mathbf{E}[\xi_k^4] < \infty$. Então, $S_n = \xi_1 + \xi_2 + \dots + \xi_n$ satisfaz

$$\frac{S_n}{n} \rightarrow \mu \text{ quase certamente.}$$

Demonstração. Começemos pelo caso $\mu = 0$. O caso geral segue ao aplicar esse caso às variáveis centradas $\xi_k - \mu$, que ainda são i.i.d. e têm quarto momento finito.

Na expansão

$$\mathbf{E}[S_n^4] = \sum_{1 \leq i, j, k, \ell \leq n} \mathbf{E}[\xi_i \xi_j \xi_k \xi_\ell],$$

todo termo em que algum índice aparece uma única vez é nulo, por independência e pelo fato de $\mathbf{E}[\xi_i] = 0$. Restam os n termos com os quatro índices iguais e, para cada par de índices distintos, as seis ordenações que os fazem aparecer duas vezes cada. Portanto,

$$\mathbf{E}[S_n^4] = n\mathbf{E}[\xi_1^4] + 6\binom{n}{2}\mathbf{E}[\xi_1^2 \xi_2^2].$$

Como a independência dá $\mathbf{E}[\xi_1^2 \xi_2^2] = (\mathbf{E}[\xi_1^2])^2 \leq \mathbf{E}[\xi_1^4]$, existe $C < \infty$ tal que

$$\mathbf{E}[S_n^4] \leq 3Cn^2.$$

Consequentemente, pela desigualdade de Markov aplicada a S_n^4 ,

$$\mathbf{P}\left(\left|\frac{S_n}{n}\right| > \varepsilon\right) \leq \frac{3C}{\varepsilon^4 n^2}.$$

A série do lado direito é convergente. Pelo primeiro lema de Borel–Cantelli, para cada $m \geq 1$,

$$\mathbf{P}\left(\left|\frac{S_n}{n}\right| > \frac{1}{m} \text{ infinitas vezes}\right) = 0.$$

Fora da união enumerável desses eventos nulos, para cada m existe um índice a partir do qual $|S_n/n| \leq 1/m$. Portanto $S_n/n \rightarrow 0$ quase certamente. $\square$

O resultado se estende a vetores de dimensão finita pela aplicação do caso escalar a cada coordenada.

Teorema 11.10 — Lei Forte dos Grandes Números

Sejam $\xi^{(1)}, \xi^{(2)}, \dots$ vetores aleatórios i.i.d. com valores em $\mathbb{R}^d$ e $\mathbf{E}\|\xi^{(1)}\|^4 < \infty$.

Então

$$\frac{1}{n} \sum_{i=1}^n \xi^{(i)} \xrightarrow{\text{q.c.}} \mathbf{E}[\xi^{(1)}].$$

Demonstração. Cada coordenada satisfaz a versão escalar precedente. Como há apenas d coordenadas, a interseção dos d eventos de probabilidade um em que as convergências coordenada a coordenada ocorrem ainda tem probabilidade um. Em dimensão finita, convergência coordenada a coordenada equivale à convergência em qualquer norma. $\square$

Proposição 11.11 — Esperança Finita é Necessária

Sejam $\xi_1, \xi_2, \dots$ variáveis aleatórias independentes e identicamente distribuídas, com $\mathbf{E}|\xi_1| = \infty$. Então $S_n = \xi_1 + \xi_2 + \dots + \xi_n$ satisfaz

$$\mathbf{P}\left(\frac{S_n}{n} \text{ convergir em } \mathbb{R}\right) = 0.$$

Demonstração. A fórmula das caudas para $|\xi_1|$ implica

$$\sum_{n=1}^{\infty} \mathbf{P}(|\xi_1| > n) = \infty.$$

Como as variáveis são identicamente distribuídas, a mesma série é $\sum_n \mathbf{P}(|\xi_n| > n)$. Os eventos $\{|\xi_n| > n\}$ são independentes; portanto, o segundo lema de Borel–Cantelli dá

$$\mathbf{P}(|\xi_n| > n \text{ infinitas vezes}) = 1.$$

Por outro lado, se S_n/n convergisse em um ponto ω , então

$$\frac{\xi_n}{n} = \frac{S_n}{n} - \frac{n-1}{n} \frac{S_{n-1}}{n-1} \longrightarrow 0.$$

Isso é incompatível com $|\xi_n|/n > 1$ infinitas vezes. Logo o evento de convergência está contido no complementar de um evento de probabilidade um e, portanto, tem probabilidade zero. $\square$

Teorema 11.12 – Lei Forte dos Grandes Números de Kolmogorov

Sejam $\xi_1, \xi_2, \dots$ variáveis aleatórias independentes e identicamente distribuídas, com $\mathbf{E}|\xi_1| < \infty$ e $\mu = \mathbf{E}[\xi_1]$. Então, para $S_n = \xi_1 + \dots + \xi_n$,

$$\frac{S_n}{n} \rightarrow \mu \quad \text{quase certamente.}$$

Demonstração. Combinaremos o truncamento com o controle das médias ao longo de uma subsequência geométrica.

Passo 1: redução ao caso não negativo. Suponha primeiro que $\xi_k \geq 0$. Ao final, o caso geral seguirá de $\xi_k = \xi_k^+ - \xi_k^-$, pois as partes positiva e negativa são integráveis e cada uma forma uma sequência i.i.d.

Passo 2: truncamento. Defina

$$Y_k = \xi_k \mathbf{1}_{\{\xi_k \leq k\}}, \quad T_n = \sum_{k=1}^n Y_k.$$

Pela representação por camadas do Corolário 8.12,

$$\sum_{k=1}^{\infty} \mathbf{P}(Y_k \neq \xi_k) = \sum_{k=1}^{\infty} \mathbf{P}(\xi_1 > k) \leq \int_0^{\infty} \mathbf{P}(\xi_1 > t) dt = \mathbf{E}[\xi_1] < \infty.$$

O primeiro lema de Borel–Cantelli mostra que $Y_k = \xi_k$ para todo k suficientemente grande, quase certamente. Assim, $(S_n - T_n)/n \rightarrow 0$ quase certamente.

Os truncamentos Y_k são independentes, mas seus níveis de corte variam com k , de modo que já não são identicamente distribuídos. Suas médias e variâncias devem, por isso, conservar esse índice. Como $0 \leq Y_k \leq k$, a mesma representação por camadas, agora com $r = 2$, dá

$$\begin{aligned} \text{Var}(Y_k) &\leq \mathbf{E}[Y_k^2] = \int_0^{\infty} 2t \mathbf{P}(Y_k > t) dt \\ &\leq \int_0^{\infty} 2t \mathbf{P}(\xi_1 > t) \mathbf{1}_{\{t < k\}} dt. \end{aligned}$$

Por Tonelli,

$$\sum_{k=1}^{\infty} \frac{\text{Var}(Y_k)}{k^2} \leq \int_0^{\infty} 2t \mathbf{P}(\xi_1 > t) \sum_{k>t} \frac{1}{k^2} dt < \infty.$$

De fato, a soma interna é limitada por uma constante para $0 \leq t < 1$ e por $2/t$ para $t \geq 1$; a última integral é, portanto, dominada por uma constante mais $4\mathbf{E}[\xi_1]$.

Passo 3: uma subsequência geométrica. Fixe $\alpha > 1$ e ponha $m_j = \lfloor \alpha^j \rfloor$. Descartar um número finito de termos permite supor que (m_j) é estritamente crescente e que $m_j \geq \alpha^j/2$. Se $R_n = T_n - \mathbf{E}[T_n]$, então Chebyshev, a independência e Tonelli dão

$$\begin{aligned} \sum_j \mathbf{P}(|R_{m_j}| > \varepsilon m_j) &\leq \frac{1}{\varepsilon^2} \sum_j \frac{1}{m_j^2} \sum_{k \leq m_j} \text{Var}(Y_k) \\ &= \frac{1}{\varepsilon^2} \sum_{k=1}^{\infty} \text{Var}(Y_k) \sum_{j: m_j \geq k} \frac{1}{m_j^2} \\ &\leq C_{\alpha, \varepsilon} \sum_{k=1}^{\infty} \frac{\text{Var}(Y_k)}{k^2} < \infty, \end{aligned}$$

pois a soma geométrica interna é no máximo C_α/k^2 . Aplicando Borel–Cantelli a $\varepsilon = 1/r$, $r \in \mathbb{N}$, obtemos

$$\frac{R_{m_j}}{m_j} \longrightarrow 0 \quad \text{quase certamente.}$$

Além disso, $\|Y_k - \xi_k\|_1 = \mathbf{E}[\xi_1 \mathbf{1}_{\{\xi_1 > k\}}] \rightarrow 0$. Como $\mathbf{E}[\xi_k] = \mu$, segue que $\mathbf{E}[Y_k] \rightarrow \mu$ e, pelo lema de Cesàro,

$$\frac{\mathbf{E}[T_{m_j}]}{m_j} \longrightarrow \mu.$$

Concluimos que $T_{m_j}/m_j \rightarrow \mu$ quase certamente.

Passo 4: preenchimento das lacunas. Como T_n é crescente, se $m_j \leq n \leq m_{j+1}$, então

$$\frac{m_j}{m_{j+1}} \frac{T_{m_j}}{m_j} \leq \frac{T_n}{n} \leq \frac{m_{j+1}}{m_j} \frac{T_{m_{j+1}}}{m_{j+1}}.$$

Como $m_{j+1}/m_j \rightarrow \alpha$, segue que, quase certamente,

$$\frac{\mu}{\alpha} \leq \liminf_n \frac{T_n}{n} \leq \limsup_n \frac{T_n}{n} \leq \alpha\mu.$$

Repetindo o argumento para a sequência enumerável $\alpha_r = 1 + 1/r$ e fazendo $r \rightarrow \infty$, obtemos $T_n/n \rightarrow \mu$ quase certamente. Pelo Passo 2, o mesmo vale para S_n/n .

Finalmente, para variáveis de sinal arbitrário, aplicamos o caso não negativo separadamente a ξ_k^+ e ξ_k^- e subtraímos os limites. $\square$

11.3. Integração de Monte Carlo

Considere uma função integrável $f : [0,1] \rightarrow \mathbb{R}$. Queremos aproximar numericamente $\int_0^1 f(x) dx$; a escolha do intervalo $[0,1]$ é apenas uma normalização.

Uma abordagem determinística usa pontos equidistantes $x_1, \dots, x_n$ e a aproximação

$$\int_0^1 f(x) dx \approx \frac{1}{n} \sum_{k=1}^n f(x_k).$$

Para funções regulares, essa é a ideia das somas de Riemann. Para uma função apenas integrável, porém, os valores numa malha prescrita podem não representar bem a integral. O método de Monte Carlo substitui a malha fixa por pontos aleatórios: sejam $U_1, U_2, \dots$ variáveis aleatórias i.i.d. uniformes em $[0,1]$ e defina

$$I_n = \frac{1}{n} \sum_{k=1}^n f(U_k).$$

O acaso entra na escolha dos pontos, não no valor da integral que procuramos.

Teorema 11.13 — Integração de Monte-Carlo

Se $f : [0,1] \rightarrow \mathbb{R}$ é Borel-mensurável e integrável, então

$$I_n \xrightarrow{\text{q.c.}} \int_0^1 f(x) dx.$$

Em particular, a convergência também ocorre em probabilidade.

Demonstração. As variáveis $f(U_k)$ são i.i.d. e integráveis. Como U_k tem distribuição uniforme,

$$\mathbb{E}[f(U_k)] = \int_0^1 f(x) dx.$$

A conclusão segue da Lei Forte dos Grandes Números de Kolmogorov. □

A integrabilidade basta para garantir a convergência desse método de computação científica. Quantificar o erro é outra questão: hipóteses adicionais, como variância finita, permitem obter estimativas para amostras de tamanho finito.

11.4. Polinômios de Bernstein e o teorema de Weierstrass

As leis dos grandes números também permitem demonstrar um resultado inteiramente determinístico. A interpretação dos polinômios de Bernstein como esperanças de funções de variáveis binomiais conduz a uma prova do Teorema de Weierstrass. Esta seção é adaptada de [23].

Polinômios de Bernstein

Os polinômios de Bernstein são definidos por

$$b_{n,j}(x) = \binom{n}{j} x^j (1-x)^{n-j},$$

para $j = 0, 1, \dots, n$.

- Os polinômios $b_{n,j}$ formam uma família de $n + 1$ polinômios de grau no máximo n .
- Para $0 < j < n$, o máximo de $b_{n,j}$ em $[0,1]$ é atingido unicamente em j/n .
- A família $b_{n,0}, \dots, b_{n,n}$ é uma base do espaço dos polinômios de grau no máximo n .
- Eles satisfazem a simetria $b_{n,j}(x) = b_{n,n-j}(1-x)$, a positividade $b_{n,j}(x) \geq 0$ em $[0,1]$ e a normalização $\sum_{j=0}^n b_{n,j}(x) = 1$.

Teorema de Aproximação de Weierstrass

Associemos agora esses polinômios a uma função contínua $f : [0,1]^d \rightarrow \mathbb{C}$, definindo

$$B_n[f](\mathbf{x}) = \sum_{0 \leq i_1, \dots, i_d \leq n} f\left(\frac{i_1}{n}, \dots, \frac{i_d}{n}\right) \prod_{k=1}^d \binom{n}{i_k} x_k^{i_k} (1-x_k)^{n-i_k}.$$

O teorema afirma que esses polinômios aproximam f uniformemente.

A estrutura dos polinômios torna-se mais clara com a interpretação probabilística

$$B_n[f](\mathbf{x}) = \mathbf{E} \left[f \left(\frac{\mathbf{S}_n}{n} \right) \right],$$

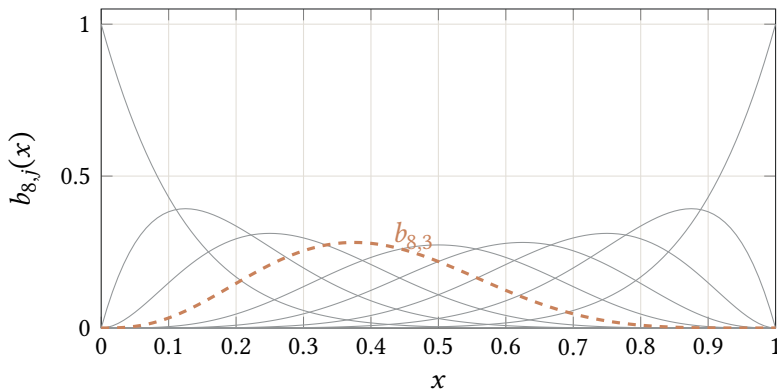


Figura 11.1.: Polinômios de Bernstein $b_{8,j}$, $j = 0, \dots, 8$. A curva tracejada destaca $b_{8,3}$. Para cada x , a soma das nove curvas é 1.

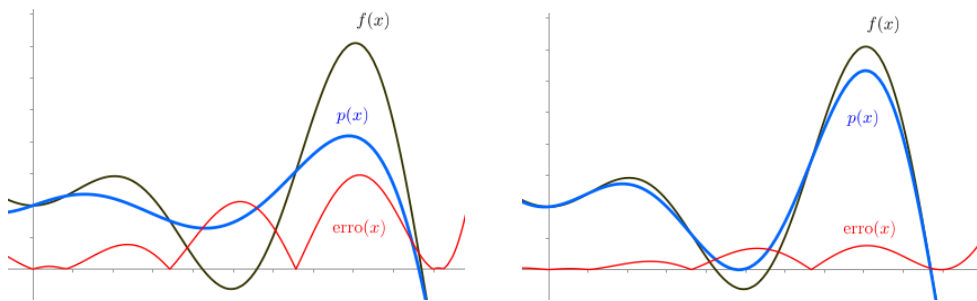


Figura 11.2.: Aproximações de Bernstein com $n = 10$ (à esquerda) e $n = 50$ (à direita), produzidas pelo código de simulação que acompanha as notas.

onde as coordenadas de $\mathbf{S}_n$ são independentes e

$$S_{k,n} \sim \text{Bin}(n, x_k), \quad 1 \leq k \leq d.$$

A lei dos grandes números sugere a concentração de $\mathbf{S}_n/n$ perto de $\mathbf{x}$. Para passar dessa observação à convergência *uniforme*, precisamos de uma estimativa que não dependa de $\mathbf{x}$. A desigualdade de Chebyshev fornecerá justamente esse controle.

Teorema 11.14 – Teorema de Aproximação de Weierstrass

Sejam $d \geq 1$ e $f : [0,1]^d \rightarrow \mathbb{C}$ uma função contínua. Defina

$$B_n[f](\mathbf{x}) = \sum_{0 \leq i_1, \dots, i_d \leq n} f\left(\frac{i_1}{n}, \dots, \frac{i_d}{n}\right) \prod_{k=1}^d \binom{n}{i_k} x_k^{i_k} (1 - x_k)^{n-i_k}$$

para $n \geq 1$ e $\mathbf{x} = (x_1, \dots, x_d) \in [0,1]^d$. Então $B_n[f] \rightarrow f$ uniformemente em $[0,1]^d$.

Demonstração. Fixe $\mathbf{x} = (x_1, \dots, x_d) \in [0,1]^d$. Para $1 \leq k \leq d$ e $1 \leq j \leq n$, sejam $\xi_{k,j}$ variáveis independentes de Bernoulli com parâmetro x_k . Ponha

$$S_{k,n} = \sum_{j=1}^n \xi_{k,j}, \quad \mathbf{S}_n = (S_{1,n}, \dots, S_{d,n}).$$

A independência e a fórmula da distribuição binomial mostram, pela definição da esperança, que

$$B_n[f](\mathbf{x}) = \mathbf{E} \left[f \left(\frac{\mathbf{S}_n}{n} \right) \right].$$

Use a norma do máximo em $\mathbb{R}^d$. Como

$$\mathbf{E} \left[\frac{S_{k,n}}{n} \right] = x_k, \quad \text{Var} \left(\frac{S_{k,n}}{n} \right) = \frac{x_k(1-x_k)}{n} \leq \frac{1}{4n},$$

a desigualdade da união e a desigualdade de Chebyshev dão, para todo $\delta > 0$,

$$\mathbf{P} \left(\left\| \frac{\mathbf{S}_n}{n} - \mathbf{x} \right\|_{\infty} \geq \delta \right) \leq \sum_{k=1}^d \mathbf{P} \left(\left| \frac{S_{k,n}}{n} - x_k \right| \geq \delta \right) \leq \frac{d}{4n\delta^2}.$$

Essa cota não depende de $\mathbf{x}$.

Seja $M = \max_{\mathbf{y} \in [0,1]^d} |f(\mathbf{y})|$. Dado $\varepsilon > 0$, a continuidade uniforme de f fornece $\delta > 0$ tal que $|f(\mathbf{y}) - f(\mathbf{x})| < \varepsilon/2$ sempre que $\|\mathbf{y} - \mathbf{x}\|_{\infty} < \delta$. Separando a esperança conforme esse evento e seu complementar, obtemos

$$\begin{aligned} |B_n[f](\mathbf{x}) - f(\mathbf{x})| &\leq \mathbf{E} \left| f \left(\frac{\mathbf{S}_n}{n} \right) - f(\mathbf{x}) \right| \\ &\leq \frac{\varepsilon}{2} + 2M \mathbf{P} \left(\left\| \frac{\mathbf{S}_n}{n} - \mathbf{x} \right\|_{\infty} \geq \delta \right) \\ &\leq \frac{\varepsilon}{2} + \frac{Md}{2n\delta^2}. \end{aligned}$$

Para n suficientemente grande, o último termo é menor que $\varepsilon/2$, uniformemente em $\mathbf{x}$. Logo $\|B_n[f] - f\|_\infty \rightarrow 0$. $\square$

A prova introduz variáveis aleatórias para construir uma aproximação determinística. Depois de tomada a esperança, o acaso já fez seu trabalho.

Uma extensão da Lei Forte: Kolmogorov e Kronecker

A Lei Forte apresentada no texto trata sobretudo de sequências i.i.d. Há uma extensão simples e útil para somas de variáveis independentes que não precisam ter a mesma distribuição. Ela repousa em três ferramentas que também voltarão a aparecer, em linguagem de martingalas, no Capítulo 15.

Teorema 11.15 — Desigualdade maximal de Kolmogorov

Sejam $X_1, \dots, X_n$ variáveis aleatórias independentes, centradas e com segundo momento finito, e ponha $S_k = X_1 + \dots + X_k$. Então, para todo $r > 0$,

$$\mathbf{P}\left(\max_{1 \leq k \leq n} |S_k| \geq r\right) \leq \frac{\text{Var}(S_n)}{r^2} = \frac{1}{r^2} \sum_{k=1}^n \text{Var}(X_k).$$

Demonstração. Seja A_k o evento em que k é o primeiro índice para o qual $|S_k| \geq r$. Os eventos A_k são disjuntos e A_k depende apenas de $X_1, \dots, X_k$. Como $S_n - S_k$ é independente de A_k e de S_k , e tem média zero,

$$\mathbf{E}[S_n^2 \mathbf{1}_{A_k}] = \mathbf{E}[S_k^2 \mathbf{1}_{A_k}] + \mathbf{P}(A_k) \mathbf{E}[(S_n - S_k)^2] \geq r^2 \mathbf{P}(A_k).$$

Somando em k e usando a disjunção,

$$r^2 \mathbf{P}\left(\max_{k \leq n} |S_k| \geq r\right) \leq \mathbf{E}[S_n^2] = \text{Var}(S_n),$$

pois os incrementos são independentes e centrados. $\square$

Teorema 11.16 – Critério de convergência de Kolmogorov

Se $X_1, X_2, \dots$ são independentes, centradas, têm segundo momento finito e

$$\sum_{k=1}^{\infty} \text{Var}(X_k) < \infty,$$

então a série $\sum_{k \geq 1} X_k$ converge quase certamente.

Demonstração. Escreva $S_n = \sum_{k \leq n} X_k$. Para $m < n$, a desigualdade maximal aplicada à cauda fornece

$$\mathbf{P}\left(\max_{m < j \leq n} |S_j - S_m| \geq r\right) \leq \frac{1}{r^2} \sum_{k=m+1}^n \text{Var}(X_k).$$

Fazendo $n \rightarrow \infty$ e usando continuidade da probabilidade,

$$\mathbf{P}\left(\sup_{j > m} |S_j - S_m| \geq r\right) \leq \frac{1}{r^2} \sum_{k > m} \text{Var}(X_k) \rightarrow 0.$$

Se $C_m(r) = \{\sup_{j, \ell \geq m} |S_j - S_\ell| > r\}$, então $C_m(r) \subseteq \{\sup_{j \geq m} |S_j - S_m| > r/2\}$; portanto $\mathbf{P}(C_m(r)) \rightarrow 0$. Como $C_m(r)$ decresce em m , segue que, para cada $r > 0$, quase toda trajetória é eventualmente r -Cauchy. Intersectando sobre $r \in \mathbb{Q}_{>0}$, (S_n) é Cauchy quase certamente e, portanto, converge. $\square$

Lema 11.17 – Lema de Kronecker, na forma usada aqui

Se a série $\sum_{k \geq 1} x_k/k$ converge, então

$$\frac{1}{n} \sum_{k=1}^n x_k \rightarrow 0.$$

Demonstração. Ponha $T_n = \sum_{k=1}^n x_k/k$ e $T_0 = 0$. Como $x_k = k(T_k - T_{k-1})$, uma soma por partes dá

$$\frac{1}{n} \sum_{k=1}^n x_k = T_n - \frac{1}{n} \sum_{k=1}^{n-1} T_k.$$

Se $T_n \rightarrow T$, a média de Cesàro à direita também converge a T ; a diferença, portanto, tende a zero. □

EXERCÍCIOS

Ex. 11.1 — Truncamento. Seja ξ uma variável aleatória integrável. Para $M > 0$, defina

$$\xi^{(M)} = \xi \mathbb{1}_{\{|\xi| \leq M\}}.$$

Prove que:

- $|\xi^{(M)}| \leq M$ e $|\xi^{(M)}| \leq |\xi|$ pontualmente;
- $\xi^{(M)} \rightarrow \xi$ quase certamente quando $M \rightarrow \infty$;
- para cada M fixo, $\xi^{(M)}$ possui momentos de todas as ordens;
- $\mathbf{E}[\xi^{(M)}] \rightarrow \mathbf{E}[\xi]$ quando $M \rightarrow \infty$.

Indique explicitamente qual função domina a sequência no último item.

Ex. 11.2 — Sejam $\xi_1, \xi_2, \dots$ variáveis aleatórias i.i.d. com distribuição $N(0,1)$. Prove que

$$\frac{\xi_1^2 + \dots + \xi_n^2}{(\xi_1 - 1)^2 + \dots + (\xi_n - 1)^2}$$

converge quase certamente e determine o limite.

Ex. 11.3 — Sejam $\xi_1, \xi_2, \dots$ variáveis aleatórias i.i.d. com distribuição uniforme em $(0,1)$. Prove que a média geométrica

$$\left(\prod_{k=1}^n \xi_k \right)^{1/n}$$

converge quase certamente e determine o limite.

Dica. Aplique a Lei Forte dos Grandes Números a $\log \xi_1, \log \xi_2, \dots$ e verifique antes a integrabilidade.

Ex. 11.4 — Lei Fraca por funções geradoras de momentos. Sejam $\xi_1, \xi_2, \dots$ variáveis aleatórias i.i.d., com $\mathbf{E}|\xi_1| < \infty$, e suponha que

$$M_\xi(s) = \mathbf{E}[e^{s\xi_1}]$$

seja finita para $s \in [-c, c]$, onde $c > 0$. Defina $\bar{\xi}_n = n^{-1} \sum_{k=1}^n \xi_k$ e $\mu = \mathbf{E}[\xi_1]$.

1. Prove que, para todo $s \in [-c, c]$,

$$M_{\bar{\xi}_n}(s) = \left[M_\xi\left(\frac{s}{n}\right) \right]^n \longrightarrow e^{s\mu}.$$

2. Reconheça o limite como a função geradora de momentos da constante μ e conclua que $\bar{\xi}_n \xrightarrow{\mathbf{P}} \mu$.

Dica. Mostre que $M_\xi(u) = 1 + u\mathbf{E}[\xi_1] + o(u)$ quando $u \rightarrow 0$.

Ex. 11.5 — Sejam $\xi_1, \xi_2, \dots$ variáveis aleatórias independentes com

$$\mathbf{E}[\xi_k] = 0, \quad \sigma_k^2 = \text{Var}(\xi_k) < \infty.$$

Para $S_n = \sum_{k=1}^n \xi_k$, prove que

$$\sum_{k=1}^{\infty} \frac{\sigma_k^2}{k^2} < \infty \implies \frac{S_n}{n} \xrightarrow{q.c.} 0.$$

Dica. Aplique o Teorema 11.16 à série $\sum_k \xi_k/k$ e depois o Lema 11.17.

Ex. 11.6 — Seja $\xi_1, \xi_2, \dots$ uma sequência i.i.d. com

$$\mathbf{P}(\xi_k = 0) = \mathbf{P}(\xi_k = 2) = \frac{1}{2},$$

e defina

$$Z = \sum_{k=1}^{\infty} \frac{\xi_k}{3^k}.$$

1. Prove que a série converge e que Z pertence ao conjunto de Cantor quase certamente.
2. Prove que Z possui a distribuição de Cantor.
3. Calcule $\mathbf{E}[Z]$ e $\text{Var}(Z)$ justificando a troca entre soma e esperança.

Ex. 11.7 — Integração de Monte Carlo. Seja $f : [0,1] \rightarrow [0,1]$ contínua. Considere $\xi_1, c_1, \xi_2, c_2, \dots$ independentes e uniformes em $[0,1]$ e defina

$$\eta_i = \mathbb{1}_{\{c_i < f(\xi_i)\}}.$$

1. Mostre que $\eta_1, \eta_2, \dots$ são i.i.d. de Bernoulli e calcule $\mathbf{E}[\eta_i]$.
2. Use a Lei Forte dos Grandes Números para concluir que

$$\frac{1}{n} \sum_{i=1}^n \eta_i \longrightarrow \int_0^1 f(x) \, dx \quad \text{quase certamente.}$$

Teorema do Limite Central

Nos capítulos anteriores vimos que, se $X_1, X_2, \dots$ são variáveis aleatórias independentes e identicamente distribuídas, com média comum $\mu = E[X_1]$ finita, então

$$\frac{1}{n} \sum_{k=1}^n (X_k - \mu) \xrightarrow{q.c.} 0.$$

Escrevendo $S_n = X_1 + \dots + X_n$, a Lei Forte dos Grandes Números afirma que o erro $S_n - n\mu$ é pequeno quando comparado com n . Ela não diz, porém, em qual escala ainda podemos enxergar as flutuações desse erro.

Suponha agora que

$$0 < \sigma^2 = \text{Var}(X_1) < \infty.$$

Pela independência,

$$\text{Var}\left(\frac{S_n - n\mu}{a_n}\right) = \frac{n\sigma^2}{a_n^2}.$$

Se desejamos uma escala na qual a variância não desapareça nem exploda, a escolha natural é $a_n = \sqrt{n}$. A pergunta deixa então de ser se a média converge e passa a ser:

qual é a distribuição das flutuações de ordem $\sqrt{n}$?

O Teorema do Limite Central responde que, depois da normalização adequada, essas flutuações convergem para uma distribuição normal. O aspecto mais surpreendente é a universalidade: a forma do limite não depende da distribuição comum das parcelas, desde que a variância seja finita.

Para demonstrar esse resultado, transformaremos distribuições em funções. Somas de variáveis independentes se tornarão produtos dessas funções, e um problema de convergência de medidas se tornará um problema de convergência pontual. As funções que realizam essa passagem são as funções características.

12.1. Funções características

12.1.1. Por que introduzir números complexos?

Os momentos de uma variável aleatória guardam muita informação sobre sua distribuição. Seria tentador empacotá-los numa série

$$\sum_{m=0}^{\infty} \mathbf{E}[X^m] t^m,$$

mas essa série raramente converge. A função geradora de momentos melhora a situação ao introduzir os fatores $m!$:

$$\mathbf{E}[e^{tX}] = \sum_{m=0}^{\infty} \frac{\mathbf{E}[X^m]}{m!} t^m,$$

sempre que as trocas de soma e esperança são legítimas. Ainda assim, $\mathbf{E}[e^{tX}]$ pode ser infinita para todo $t \neq 0$.

A saída é substituir t por it . Como

$$|e^{itX}| = 1,$$

a nova esperança existe para toda variável aleatória, sem qualquer hipótese sobre momentos. Neste caso, passar aos complexos é justamente o que simplifica as coisas.

12.1.2. Uma breve passagem pelas integrais complexas

Identificaremos $\mathbb{C}$ com $\mathbb{R}^2$, munido de sua σ -álgebra de Borel. Assim, uma função $f : \Omega \rightarrow \mathbb{C}$ é mensurável se, e somente se,

$$f = g + ih,$$

com $g, h : \Omega \rightarrow \mathbb{R}$ mensuráveis. Dizemos que f é integrável quando $|f|$ é integrável e definimos

$$\int_{\Omega} f \, d\mu = \int_{\Omega} g \, d\mu + i \int_{\Omega} h \, d\mu.$$

Proposição 12.1 —

Sejam f e g funções complexas integráveis e sejam $a, b \in \mathbb{C}$. Então:

1 $af + bg$ é integrável e

$$\int (af + bg) \, d\mu = a \int f \, d\mu + b \int g \, d\mu;$$

2

$$\overline{\int f \, d\mu} = \int \bar{f} \, d\mu;$$

3

$$\left| \int f \, d\mu \right| \leq \int |f| \, d\mu.$$

Demonstração. As duas primeiras propriedades seguem diretamente da definição em termos das partes real e imaginária. Para a terceira, ponha $z = \int f \, d\mu$. Se $z = 0$, não há o que provar. Caso contrário, tome $\alpha = \bar{z}/|z|$. Então $|\alpha| = 1$ e

$$|z| = \operatorname{Re}(\alpha z) = \int \operatorname{Re}(\alpha f) \, d\mu \leq \int |\alpha f| \, d\mu = \int |f| \, d\mu.$$

□

Proposição 12.2 — Extensões complexas dos teoremas de integração

Valem as seguintes extensões.

- 1 Convergência Dominada.** Se $f_n : \Omega \rightarrow \mathbb{C}$ são mensuráveis, $f_n \rightarrow f$ quase certamente e existe $g \in L^1(\mu)$ tal que $|f_n| \leq g$ quase certamente para todo n , então f é integrável e

$$\int f_n \, d\mu \rightarrow \int f \, d\mu.$$

- 2 Fubini.** Se $(\Omega_1, \mathcal{F}_1, \mu)$ e $(\Omega_2, \mathcal{F}_2, \nu)$ são σ -finitos e $f : \Omega_1 \times \Omega_2 \rightarrow \mathbb{C}$ é mensurável com

$$\int_{\Omega_1 \times \Omega_2} |f| \, d(\mu \otimes \nu) < \infty,$$

então as seções são integráveis quase em toda parte e as duas integrais iteradas coincidem com a integral de f no produto.

- 3 Produto de esperanças.** Se Z e W são variáveis aleatórias complexas independentes e integráveis, então ZW é integrável e

$$\mathbf{E}[ZW] = \mathbf{E}[Z] \mathbf{E}[W].$$

Demonstração. Para o primeiro item, aplique a Convergência Dominada real às partes real e imaginária, observando que $|\operatorname{Re} f_n|, |\operatorname{Im} f_n| \leq g$. O segundo item segue da mesma decomposição e do Teorema de Fubini real, pois $|\operatorname{Re} f|, |\operatorname{Im} f| \leq |f|$.

Para o terceiro, $|Z|$ e $|W|$ são variáveis reais independentes e integráveis. Pelo Teorema 8.31,

$$\mathbf{E}|ZW| = \mathbf{E}|Z| \mathbf{E}|W| < \infty.$$

Escreva $Z = Z_1 + iZ_2$ e $W = W_1 + iW_2$. Cada Z_j é independente de cada W_k , e o mesmo resultado real aplicado aos quatro produtos fornece

$$\mathbf{E}[(Z_1 + iZ_2)(W_1 + iW_2)] = (\mathbf{E}Z_1 + i\mathbf{E}Z_2)(\mathbf{E}W_1 + i\mathbf{E}W_2).$$

□

12.1.3. Definição e primeiras propriedades

A integração complexa permite agora definir a função característica e estudar suas primeiras propriedades.

Definição 12.3 — Função característica

Dada uma variável aleatória real X , sua **função característica** é a função $\phi_X : \mathbb{R} \rightarrow \mathbb{C}$ definida por

$$\phi_X(t) = \mathbf{E}[e^{itX}] = \mathbf{E}[\cos(tX)] + i\mathbf{E}[\sin(tX)].$$

Equivalentemente, se $\mathbf{P}_X$ denota a distribuição de X ,

$$\phi_X(t) = \int_{\mathbb{R}} e^{itx} \mathbf{P}_X(dx).$$

Se X possui densidade f_X , a última expressão assume a forma

$$\phi_X(t) = \int_{\mathbb{R}} e^{itx} f_X(x) dx.$$

Observe que a definição não exige a existência de uma densidade.

Definição 12.4 — Transformada de Fourier de uma medida

Se μ é uma medida de probabilidade em $\mathbb{R}$, definimos

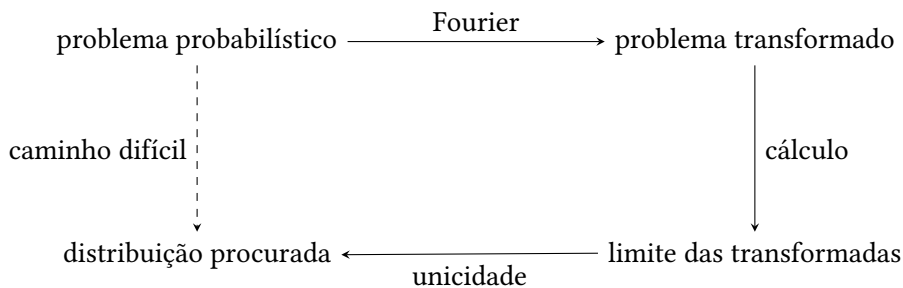
$$\hat{\mu}(t) = \int_{\mathbb{R}} e^{itx} \mu(dx).$$

Assim, $\phi_X = \widehat{\mathbf{P}_X}$.

Como a função característica depende apenas da distribuição,

$$X \stackrel{d}{=} Y \implies \phi_X = \phi_Y.$$

Mais adiante provaremos a recíproca: a função característica determina completamente a distribuição.



Proposição 12.5 — Propriedades elementares

Sejam X e Y variáveis aleatórias reais.

1 $\phi_X(0) = 1$ e $|\phi_X(t)| \leq 1$ para todo $t \in \mathbb{R}$;

2 $\phi_X(-t) = \overline{\phi_X(t)}$;

3 para $a, b \in \mathbb{R}$,

$$\phi_{aX+b}(t) = e^{itb} \phi_X(at);$$

4 se X e Y são independentes, então

$$\phi_{X+Y}(t) = \phi_X(t) \phi_Y(t);$$

5 ϕ_X é uniformemente contínua em $\mathbb{R}$.

Demonstração. Os dois primeiros itens seguem de

$$|e^{itX}| = 1, \quad e^{-itX} = \overline{e^{itX}}.$$

Para a transformação afim,

$$\phi_{aX+b}(t) = \mathbf{E}[e^{it(aX+b)}] = e^{itb} \mathbf{E}[e^{i(at)X}] = e^{itb} \phi_X(at).$$

Se X e Y são independentes, também são independentes e^{itX} e e^{itY} ; portanto,

$$\phi_{X+Y}(t) = \mathbf{E}[e^{itX} e^{itY}] = \mathbf{E}[e^{itX}] \mathbf{E}[e^{itY}] = \phi_X(t) \phi_Y(t).$$

Finalmente, para $s, t \in \mathbb{R}$,

$$|\phi_X(t) - \phi_X(s)| \leq \mathbf{E}[|e^{itX} - e^{isX}|] = \mathbf{E}[|e^{i(t-s)X} - 1|].$$

Quando $t - s \rightarrow 0$, o integrando converge pontualmente para zero e é limitado por 2. O Teorema da Convergência Dominada mostra que o último termo tende a zero. Como ele depende apenas de $t - s$, a continuidade é uniforme. $\square$

Exemplo 12.6 — Distribuição de Bernoulli.

Se $X \sim \text{Bernoulli}(p)$, então

$$\phi_X(t) = e^{it}\mathbf{P}(X = 1) + \mathbf{P}(X = 0) = 1 + p(e^{it} - 1).$$

◁

Exemplo 12.7 — Distribuição binomial.

Se $X_1, \dots, X_n$ são independentes, cada qual com distribuição Bernoulli(p), então $X = X_1 + \dots + X_n \sim \text{Bin}(n, p)$ e

$$\phi_X(t) = \prod_{j=1}^n \phi_{X_j}(t) = [1 + p(e^{it} - 1)]^n.$$

◁

12.1.4. Momentos e derivadas

A exponencial complexa também codifica momentos: quando podemos derivar sob o sinal de esperança, suas derivadas em zero os recuperam.

Proposição 12.8 — Derivação sob o sinal de esperança

Seja $I \subset \mathbb{R}$ um intervalo aberto e seja $f : I \times \Omega \rightarrow \mathbb{C}$. Suponha que:

- 1 para cada $t \in I$, a função $f(t, \cdot)$ é mensurável;
- 2 para quase todo ω , a função $t \mapsto f(t, \omega)$ é continuamente diferenciável;
- 3 $f(t_0, \cdot)$ é integrável para algum $t_0 \in I$;
- 4 existe uma variável aleatória não negativa e integrável Y tal que

$$|\partial_t f(t, \omega)| \leq Y(\omega)$$

para todo $t \in I$ e quase todo ω .

Então $F(t) = \mathbf{E}[f(t, \cdot)]$ é diferenciável em I e

$$F'(t) = \mathbf{E}[\partial_t f(t, \cdot)].$$

Demonstração. Para $t \in I$, o Teorema Fundamental do Cálculo e a hipótese de dominação dão

$$|f(t, \omega) - f(t_0, \omega)| \leq |t - t_0|Y(\omega).$$

Portanto, $f(t, \cdot)$ é integrável para todo $t \in I$.

Fixe $t \in I$. Para h suficientemente pequeno,

$$\frac{f(t+h, \omega) - f(t, \omega)}{h} = \int_0^1 \partial_t f(t+rh, \omega) \, dr.$$

O quociente converge pontualmente para $\partial_t f(t, \omega)$ e seu módulo é limitado por $Y(\omega)$. O resultado segue do Teorema da Convergência Dominada. □

Proposição 12.9 – Momentos e derivadas

Seja X uma variável aleatória tal que $\mathbf{E}[|X|^m] < \infty$. Então ϕ_X possui derivadas contínuas até a ordem m e, para $j = 1, \dots, m$,

$$\phi_X^{(j)}(t) = i^j \mathbf{E}[X^j e^{itX}].$$

Em particular,

$$\phi_X^{(j)}(0) = i^j \mathbf{E}[X^j].$$

Demonstração. Para $j \leq m$, temos $\mathbf{E}[|X|^j] < \infty$. Aplicando repetidamente a Proposição 12.8 a $f(t, \omega) = e^{itX(\omega)}$, obtemos

$$\frac{\partial^j}{\partial t^j} e^{itX} = (iX)^j e^{itX}, \quad |(iX)^j e^{itX}| = |X|^j.$$

Isso fornece a fórmula das derivadas. Sua continuidade segue novamente do Teorema da Convergência Dominada. $\square$

Exemplo 12.10 – A função característica da normal.

Seja $Z \sim N(0,1)$. Pela proposição anterior,

$$\phi_Z'(t) = \frac{i}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{itx} e^{-x^2/2} dx.$$

Como

$$x e^{-x^2/2} = -\frac{d}{dx} e^{-x^2/2},$$

uma integração por partes, com termo de fronteira nulo, fornece

$$\phi_Z'(t) = -t \phi_Z(t).$$

A condição inicial $\phi_Z(0) = 1$ determina a solução:

$$\phi_Z(t) = e^{-t^2/2}.$$

Exemplo 12.11 — Normal com média e variância arbitrárias.

Se $X = \mu + \sigma Z$, com $Z \sim N(0,1)$, então

$$\phi_X(t) = e^{it\mu} \phi_Z(\sigma t) = \exp\left(it\mu - \frac{\sigma^2 t^2}{2}\right).$$

**12.1.5. Unicidade de Fourier**

Provaremos agora que a função característica não perde informação: duas medidas de probabilidade com a mesma transformada de Fourier são iguais. A Gaussiana desempenhará o papel de um filtro de regularização.

Para $\epsilon > 0$, escreva

$$\rho_\epsilon(x) = \frac{1}{\epsilon\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\epsilon^2}\right).$$

Lema 12.12 — Identidade gaussiana de Fourier

Para todo $\epsilon > 0$ e $x \in \mathbb{R}$,

$$\rho_\epsilon(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx} e^{-\epsilon^2 t^2 / 2} dt.$$

Demonstração. Fazendo $u = \epsilon t$ e usando o cálculo da função característica da normal padrão,

$$\begin{aligned} \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx} e^{-\epsilon^2 t^2 / 2} dt &= \frac{1}{2\pi\epsilon} \int_{\mathbb{R}} e^{-iu(x/\epsilon)} e^{-u^2 / 2} du \\ &= \frac{1}{\epsilon\sqrt{2\pi}} e^{-x^2 / (2\epsilon^2)}. \end{aligned}$$



Teorema 12.13 — Teorema da unicidade de Fourier

Se μ e ν são medidas de probabilidade em $\mathbb{R}$ e $\hat{\mu}(t) = \hat{\nu}(t)$ para todo $t \in \mathbb{R}$, então $\mu = \nu$. Equivalentemente,

$$\phi_X = \phi_Y \implies X \stackrel{d}{=} Y.$$

Demonstração. Para $\epsilon > 0$, defina

$$p_{\mu,\epsilon}(x) = \int_{\mathbb{R}} \rho_{\epsilon}(x-y) \mu(dy).$$

Essa é a densidade da convolução de μ com uma normal $N(0, \epsilon^2)$. Pelo lema anterior e pelo Teorema de Fubini,

$$\begin{aligned} p_{\mu,\epsilon}(x) &= \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx} e^{-\epsilon^2 t^2/2} \left(\int_{\mathbb{R}} e^{ity} \mu(dy) \right) dt \\ &= \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx} e^{-\epsilon^2 t^2/2} \hat{\mu}(t) dt. \end{aligned}$$

A aplicação de Fubini é legítima porque $e^{-\epsilon^2 t^2/2}$ é integrável. Se $\hat{\mu} = \hat{\nu}$, concluímos que

$$p_{\mu,\epsilon} = p_{\nu,\epsilon} \quad \text{para todo } \epsilon > 0.$$

Seja $G \sim N(0,1)$. Para toda função contínua limitada f , o Teorema da Convergência Dominada fornece

$$\int_{\mathbb{R}} \int_{\mathbb{R}} f(x + \epsilon z) \mu(dx) \mathbf{P}_G(dz) \longrightarrow \int_{\mathbb{R}} f(x) \mu(dx)$$

quando $\epsilon \downarrow 0$. O mesmo vale para ν . Como as duas convoluções têm a mesma densidade para cada ϵ , segue que

$$\int f d\mu = \int f d\nu$$

para toda função contínua limitada f . Aproximando indicadoras de intervalos nos pontos de continuidade das funções de distribuição, concluímos que $\mu = \nu$. $\square$

12.1.6. Outra rota para a unicidade

A prova anterior regulariza as medidas por meio de Gaussianas. Há uma segunda rota, baseada na densidade dos polinômios trigonométricos. Ela explicita por que conhecer as integrais de todas as funções $x \mapsto e^{itx}$ acaba determinando as integrais de funções muito mais gerais.

Recordemos a forma do Teorema de Stone–Weierstrass complexo que será usada: se K é compacto de Hausdorff e $\mathcal{A} \subseteq C(K; \mathbb{C})$ é uma álgebra que contém as constantes, separa pontos e é fechada por conjugação, então $\mathcal{A}$ é densa em $C(K; \mathbb{C})$ para a norma uniforme. O resultado seguinte é precisamente essa forma geral aplicada ao círculo.

Corolário 12.14 — Stone–Weierstrass para funções periódicas

Fixe $T > 0$. O espaço dos polinômios trigonométricos

$$\mathcal{T}_T = \text{span}_{\mathbb{C}} \{e^{2\pi i k x/T} : k \in \mathbb{Z}\}$$

é denso, para a norma uniforme, no espaço das funções complexas, contínuas e T -periódicas.

Demonstração. Identifique as funções contínuas e T -periódicas em $\mathbb{R}$ com as funções contínuas no círculo compacto $\mathbb{R}/T\mathbb{Z}$. A classe $\mathcal{T}_T$ é uma álgebra complexa que contém as constantes. Ela também é fechada por conjugação, pois

$$\overline{e^{2\pi i k x/T}} = e^{-2\pi i k x/T}.$$

Além disso, separa pontos do círculo: se $x - y \notin T\mathbb{Z}$, então

$$e^{2\pi i x/T} \neq e^{2\pi i y/T}.$$

A forma complexa do Teorema de Stone–Weierstrass enunciada acima implica, portanto, que $\mathcal{T}_T$ é densa na norma uniforme no espaço das funções contínuas sobre $\mathbb{R}/T\mathbb{Z}$, isto é, no espaço das funções contínuas e T -periódicas. $\square$

Com esse teorema, obtemos outra demonstração da unicidade. Suponha que

$\hat{\mu} = \hat{\nu}$ e tome $f \in C_c(\mathbb{R})$. Dado $\delta > 0$, escolha M tão grande que

$$\text{supp}(f) \subset (-M, M) \quad \text{e} \quad \mu([-M, M]^c) + \nu([-M, M]^c) < \delta.$$

A restrição de f a $[-M, M]$ possui uma extensão contínua $2M$ -periódica $\bar{f}$, pois f se anula numa vizinhança dos extremos. Pelo teorema, existem $p_j \in \mathcal{T}_{2M}$ tais que $p_j \rightarrow \bar{f}$ uniformemente.

Cada p_j é combinação linear de funções $e^{i\pi kx/M}$. A igualdade das funções características implica, portanto,

$$\int p_j \, d\mu = \int p_j \, d\nu.$$

Passando ao limite uniforme,

$$\int \bar{f} \, d\mu = \int \bar{f} \, d\nu.$$

Como $f = \bar{f}$ em $[-M, M]$,

$$\left| \int f \, d\mu - \int f \, d\nu \right| \leq \|f\|_\infty (\mu([-M, M]^c) + \nu([-M, M]^c)) < \|f\|_\infty \delta.$$

Fazendo $\delta \downarrow 0$, obtemos, para toda $f \in C_c(\mathbb{R})$,

$$\int f \, d\mu = \int f \, d\nu. \quad (12.1)$$

Novamente, concluímos que $\mu = \nu$.

12.1.7. Convergência de funções características

O último passo antes do Teorema do Limite Central é transformar convergência pontual de funções características em convergência em distribuição. Primeiro precisamos garantir que a massa das distribuições não escape para o infinito.

Proposição 12.15 —

Seja $(X_n)_{n \geq 1}$ uma sequência de variáveis aleatórias, com funções características ϕ_n . Suponha que existem $t_0 > 0$ e uma função $g : (-t_0, t_0) \rightarrow \mathbb{C}$, contínua em 0, tais que

$$\phi_n(t) \rightarrow g(t) \quad \text{para todo } t \in (-t_0, t_0).$$

Então a família $(X_n)_{n \geq 1}$ é tight.

Demonstração. Como $\phi_n(0) = 1$, temos $g(0) = 1$. Fixe $\eta > 0$. Pela continuidade de g em zero, podemos escolher $a \in (0, t_0)$ tal que

$$|1 - g(t)| < \frac{\eta}{4} \quad \text{para } |t| \leq a.$$

Pelo Teorema da Convergência Dominada,

$$\frac{1}{a} \int_{-a}^a (1 - \operatorname{Re} \phi_n(t)) \, dt \rightarrow \frac{1}{a} \int_{-a}^a (1 - \operatorname{Re} g(t)) \, dt.$$

O módulo do limite é menor que $\eta/2$. Logo, para algum n_0 ,

$$\frac{1}{a} \int_{-a}^a (1 - \operatorname{Re} \phi_n(t)) \, dt < \eta \quad \text{se } n \geq n_0.$$

Por Fubini,

$$\begin{aligned} \int_{-a}^a (1 - \operatorname{Re} \phi_n(t)) \, dt &= \mathbf{E} \left[\int_{-a}^a (1 - \cos(tX_n)) \, dt \right] \\ &= 2a \mathbf{E} \left[1 - \frac{\operatorname{sen}(a|X_n|)}{a|X_n|} \right]. \end{aligned}$$

Na expressão acima, o quociente é interpretado como 1 quando $X_n = 0$. No evento $\{|X_n| \geq 2/a\}$,

$$1 - \frac{\operatorname{sen}(a|X_n|)}{a|X_n|} \geq 1 - \frac{1}{a|X_n|} \geq \frac{1}{2}.$$

Consequentemente,

$$\int_{-a}^a (1 - \operatorname{Re} \phi_n(t)) \, dt \geq a \mathbf{P}(|X_n| \geq 2/a).$$

Assim,

$$\mathbf{P}(|X_n| \geq 2/a) < \eta \quad \text{para } n \geq n_0.$$

Restam apenas as variáveis $X_1, \dots, X_{n_0-1}$. Como a família é finita, existe b_0 tal que

$$\mathbf{P}(|X_n| > b_0) < \eta \quad (1 \leq n < n_0).$$

Tomando $b = \max\{2/a, b_0\}$, obtemos

$$\mathbf{P}(|X_n| > b) < \eta \quad \text{para todo } n \geq 1.$$

Portanto, $(X_n)_{n \geq 1}$ é tight. □

Teorema 12.16 — Teorema da continuidade de Lévy

Sejam $X_1, X_2, \dots$ variáveis aleatórias com funções características $\phi_1, \phi_2, \dots$

1 Se $X_n \xrightarrow{d} X$, então

$$\phi_n(t) \longrightarrow \phi_X(t) \quad \text{para todo } t \in \mathbb{R}.$$

2 Reciprocamente, se $\phi_n(t) \rightarrow g(t)$ para todo $t \in \mathbb{R}$ e g é contínua em 0, então existe uma variável aleatória X tal que $g = \phi_X$ e

$$X_n \xrightarrow{d} X.$$

Demonstração. Suponha primeiro que $X_n \xrightarrow{d} X$. Pelo Teorema de Representação de Skorokhod, existem variáveis Y_n e Y , definidas num mesmo espaço, tais que

$$Y_n \stackrel{d}{=} X_n, \quad Y \stackrel{d}{=} X, \quad Y_n \xrightarrow{q.c.} Y.$$

Para cada $t \in \mathbb{R}$, temos $e^{itY_n} \rightarrow e^{itY}$ quase certamente e $|e^{itY_n}| = 1$. Pelo Teorema da Convergência Dominada,

$$\phi_n(t) = \mathbf{E}[e^{itY_n}] \longrightarrow \mathbf{E}[e^{itY}] = \phi_X(t).$$

Para a recíproca, a Proposição 12.15 mostra que (X_n) é tight. Pela Proposição 10.34, alguma subsequência (X_{n_k}) converge em distribuição para uma variável X . Pela primeira parte,

$$\phi_{n_k}(t) \longrightarrow \phi_X(t).$$

Como também $\phi_{n_k}(t) \rightarrow g(t)$, segue que $g = \phi_X$.

Agora tome uma subsequência arbitrária de (X_n) . Ela continua tight e, pela Proposição 10.34, possui uma subsubsequência que converge em distribuição para alguma variável Y . Pela primeira parte, $\phi_Y = g = \phi_X$; o Teorema da Unicidade de Fourier implica $Y \stackrel{d}{=} X$. Assim, toda subsequência possui uma subsubsequência com limite X , e o critério da Proposição 10.36 fornece $X_n \stackrel{d}{\rightarrow} X$. $\square$

12.2. O Teorema do Limite Central

As peças estão agora reunidas: independência transforma somas em produtos de funções características, e a expansão perto da origem identifica o limite.

Teorema 12.17 – Teorema do Limite Central

Sejam $X_1, X_2, \dots$ variáveis aleatórias independentes e identicamente distribuídas, com

$$\mathbf{E}[X_1] = \mu, \quad 0 < \text{Var}(X_1) = \sigma^2 < \infty.$$

Então

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} N(0,1), \quad S_n = X_1 + \dots + X_n.$$

Equivalentemente,

$$\sqrt{n} \left(\frac{S_n}{n} - \mu \right) \xrightarrow{d} N(0, \sigma^2).$$

Demonstração. Defina

$$Y_k = \frac{X_k - \mu}{\sigma}.$$

Então as variáveis Y_k são i.i.d., com média zero e variância um. Seja ϕ sua função característica. Pela Proposição 12.9,

$$\phi(0) = 1, \quad \phi'(0) = i\mathbf{E}[Y_1] = 0, \quad \phi''(0) = -\mathbf{E}[Y_1^2] = -1.$$

Portanto, quando $u \rightarrow 0$,

$$\phi(u) = 1 - \frac{u^2}{2} + o(u^2). \quad (12.2)$$

Fixe $t \in \mathbb{R}$. Pela independência,

$$\begin{aligned} \phi_{\frac{1}{\sqrt{n}} \sum_{k=1}^n Y_k}(t) &= \prod_{k=1}^n \phi\left(\frac{t}{\sqrt{n}}\right) \\ &= \left(1 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right)\right)^n \rightarrow e^{-t^2/2}. \end{aligned}$$

O limite é a função característica de $N(0,1)$. O Teorema da Continuidade de Lévy conclui a demonstração. $\square$

Observação 12.18 —

Se $\sigma^2 = 0$, então $X_1 = \mu$ quase certamente e $S_n - n\mu = 0$ quase certamente. Nesse caso, a versão não normalizada do teorema possui o limite degenerado $N(0,0)$.

12.3. TLC para arranjos triangulares

A hipótese de que todas as parcelas tenham a mesma distribuição não é o verdadeiro motivo para o aparecimento da normal. O ponto essencial é que nenhuma parcela isolada domine a soma. Os arranjos triangulares tornam essa ideia precisa.

Definição 12.19 — Arranjo triangular

Um **arranjo triangular** é uma família

$$\{X_{n,k} : 1 \leq k \leq k_n, n \geq 1\}.$$

A linha n é formada pelas variáveis $X_{n,1}, \dots, X_{n,k_n}$. Exigiremos independência dentro de cada linha, mas não entre linhas diferentes.

Lema 12.20 — Produtos infinitesimais

Para cada n , sejam $z_{n,1}, \dots, z_{n,k_n} \in \mathbb{C}$. Suponha que

$$\max_{1 \leq k \leq k_n} |z_{n,k}| \rightarrow 0, \quad \sum_{k=1}^{k_n} z_{n,k} \rightarrow z, \quad \sum_{k=1}^{k_n} |z_{n,k}|^2 \rightarrow 0.$$

Então

$$\prod_{k=1}^{k_n} (1 + z_{n,k}) \rightarrow e^z.$$

Demonstração. Para n grande, todos os $z_{n,k}$ pertencem a um disco no qual podemos usar o ramo principal do logaritmo. Uniformemente nesse disco,

$$\log(1 + w) = w + O(|w|^2).$$

Logo,

$$\sum_{k=1}^{k_n} \log(1 + z_{n,k}) = \sum_{k=1}^{k_n} z_{n,k} + O\left(\sum_{k=1}^{k_n} |z_{n,k}|^2\right) \rightarrow z.$$

Basta exponenciar. □

Teorema 12.21 — Teorema de Lindeberg–Feller

Para cada n , sejam $X_{n,1}, \dots, X_{n,k_n}$ variáveis aleatórias independentes, centradas e com variâncias finitas. Defina

$$s_n^2 = \sum_{k=1}^{k_n} \text{Var}(X_{n,k}) > 0.$$

Suponha que, para todo $\epsilon > 0$, vale a **condição de Lindeberg**

$$\frac{1}{s_n^2} \sum_{k=1}^{k_n} \mathbf{E} \left[X_{n,k}^2 \mathbb{1}_{\{|X_{n,k}| > \epsilon s_n\}} \right] \rightarrow 0.$$

Então

$$\frac{1}{s_n} \sum_{k=1}^{k_n} X_{n,k} \xrightarrow{d} N(0,1).$$

Demonstração. Ponha

$$Y_{n,k} = \frac{X_{n,k}}{s_n}, \quad v_{n,k} = \mathbf{E}[Y_{n,k}^2].$$

Então

$$\sum_{k=1}^{k_n} v_{n,k} = 1$$

e a condição de Lindeberg se escreve

$$L_n(\epsilon) := \sum_{k=1}^{k_n} \mathbf{E} \left[Y_{n,k}^2 \mathbb{1}_{\{|Y_{n,k}| > \epsilon\}} \right] \rightarrow 0.$$

Primeiro observamos que nenhuma parcela pode conservar uma fração positiva da variância total. De fato, para todo $\epsilon > 0$,

$$\max_{1 \leq k \leq k_n} v_{n,k} \leq \epsilon^2 + L_n(\epsilon).$$

Tomando o limite superior em n e depois fazendo $\epsilon \downarrow 0$, obtemos

$$\max_{1 \leq k \leq k_n} v_{n,k} \rightarrow 0.$$

Fixe $t \in \mathbb{R}$ e escreva

$$R(u) = e^{iu} - 1 - iu + \frac{u^2}{2}.$$

O resto integral de Taylor fornece

$$|R(u)| \leq \frac{|u|^3}{6}.$$

Além disso,

$$|R(u)| \leq 2 + |u| + \frac{u^2}{2}.$$

Como $\mathbf{E}[Y_{n,k}] = 0$, a função característica de $Y_{n,k}$ satisfaz

$$\phi_{n,k}(t) = 1 - \frac{t^2}{2}v_{n,k} + r_{n,k}(t), \quad r_{n,k}(t) = \mathbf{E}[R(tY_{n,k})].$$

Para $\epsilon > 0$, separamos a esperança nos eventos $\{|Y_{n,k}| \leq \epsilon\}$ e $\{|Y_{n,k}| > \epsilon\}$. Com

$$C_{t,\epsilon} = \frac{2}{\epsilon^2} + \frac{|t|}{\epsilon} + \frac{t^2}{2},$$

obtemos

$$\begin{aligned} \sum_{k=1}^{k_n} |r_{n,k}(t)| &\leq \frac{|t|^3 \epsilon}{6} \sum_{k=1}^{k_n} \mathbf{E} \left[Y_{n,k}^2 \mathbb{1}_{\{|Y_{n,k}| \leq \epsilon\}} \right] \\ &\quad + C_{t,\epsilon} \sum_{k=1}^{k_n} \mathbf{E} \left[Y_{n,k}^2 \mathbb{1}_{\{|Y_{n,k}| > \epsilon\}} \right] \\ &\leq \frac{|t|^3 \epsilon}{6} + C_{t,\epsilon} L_n(\epsilon). \end{aligned}$$

Para ϵ fixo, o segundo termo tende a zero. Fazendo depois $\epsilon \downarrow 0$, concluímos que

$$\sum_{k=1}^{k_n} |r_{n,k}(t)| \longrightarrow 0.$$

Defina

$$z_{n,k} = -\frac{t^2}{2}v_{n,k} + r_{n,k}(t).$$

Então

$$\sum_{k=1}^{k_n} z_{n,k} \longrightarrow -\frac{t^2}{2}$$

e

$$\max_{1 \leq k \leq k_n} |z_{n,k}| \rightarrow 0.$$

Aqui usamos $\max_k |r_{n,k}(t)| \leq \sum_k |r_{n,k}(t)|$. Além disso,

$$\sum_{k=1}^{k_n} |z_{n,k}| \leq \frac{t^2}{2} \sum_{k=1}^{k_n} v_{n,k} + \sum_{k=1}^{k_n} |r_{n,k}(t)| = \frac{t^2}{2} + \sum_{k=1}^{k_n} |r_{n,k}(t)|.$$

O lado direito tende a $t^2/2$, de modo que a soma à esquerda permanece limitada. Consequentemente,

$$\sum_{k=1}^{k_n} |z_{n,k}|^2 \leq \left(\max_{1 \leq k \leq k_n} |z_{n,k}| \right) \sum_{k=1}^{k_n} |z_{n,k}| \rightarrow 0.$$

Pelo Lema 12.20 e pela independência dentro de cada linha,

$$\begin{aligned} \phi_{\sum_{k=1}^{k_n} Y_{n,k}}(t) &= \prod_{k=1}^{k_n} \phi_{n,k}(t) \\ &= \prod_{k=1}^{k_n} (1 + z_{n,k}) \rightarrow e^{-t^2/2}. \end{aligned}$$

O Teorema da Continuidade de Lévy encerra a prova. □

A condição de Lindeberg possui uma interpretação concreta: depois da normalização pela variância total, a contribuição quadrática das parcelas excepcionalmente grandes deve desaparecer. Não é necessário que as parcelas tenham a mesma distribuição; é necessário que nenhuma delas carregue sozinha uma parte macroscópica da flutuação.

Corolário 12.22 — Condição de Lyapunov

Nas hipóteses de independência e centralização do teorema anterior, suponha que, para algum $\delta > 0$,

$$\frac{1}{s_n^{2+\delta}} \sum_{k=1}^{k_n} \mathbf{E}[|X_{n,k}|^{2+\delta}] \rightarrow 0.$$

Então

$$\frac{1}{s_n} \sum_{k=1}^{k_n} X_{n,k} \xrightarrow{d} N(0,1).$$

Demonstração. No evento $\{|X_{n,k}| > \epsilon s_n\}$,

$$X_{n,k}^2 \leq \frac{|X_{n,k}|^{2+\delta}}{(\epsilon s_n)^\delta}.$$

Portanto, a expressão de Lindeberg é limitada por

$$\frac{1}{\epsilon^\delta s_n^{2+\delta}} \sum_{k=1}^{k_n} \mathbf{E}[|X_{n,k}|^{2+\delta}],$$

que tende a zero. □

Exemplo 12.23 – Somas de Bernoulli não identicamente distribuídas.

Para cada n , sejam $B_{n,1}, \dots, B_{n,k_n}$ variáveis independentes, com $B_{n,k} \sim \text{Bernoulli}(p_{n,k})$, e suponha que

$$s_n^2 = \sum_{k=1}^{k_n} p_{n,k}(1 - p_{n,k}) \longrightarrow \infty.$$

Como $|B_{n,k} - p_{n,k}| \leq 1$, para cada $\epsilon > 0$ o indicador da condição de Lindeberg é identicamente nulo quando $\epsilon s_n > 1$. Logo,

$$\frac{\sum_{k=1}^{k_n} (B_{n,k} - p_{n,k})}{\sqrt{\sum_{k=1}^{k_n} p_{n,k}(1 - p_{n,k})}} \xrightarrow{d} N(0,1).$$

A soma possui distribuição de Poisson–binomial; não é preciso que os parâmetros $p_{n,k}$ sejam iguais. ◁

O TLC clássico reaparece como caso particular. Basta tomar $X_{n,k} = X_k - \mu$, $1 \leq k \leq n$. Nesse caso $s_n^2 = n\sigma^2$ e, para todo $\epsilon > 0$,

$$\begin{aligned} \frac{1}{s_n^2} \sum_{k=1}^n \mathbf{E} \left[(X_k - \mu)^2 \mathbb{1}_{\{|X_k - \mu| > \epsilon s_n\}} \right] &= \frac{1}{\sigma^2} \mathbf{E} \left[(X_1 - \mu)^2 \mathbb{1}_{\{|X_1 - \mu| > \epsilon \sigma \sqrt{n}\}} \right] \\ &\longrightarrow 0 \end{aligned}$$

pelo Teorema da Convergência Dominada.

EXERCÍCIOS

Funções características

Ex. 12.1 — Demonstre as fórmulas da tabela abaixo.

Distribuição	Função característica
Binomial(n, p)	$[(1 - p) + pe^{it}]^n$
Poisson(λ)	$e^{\lambda(e^{it} - 1)}$
$N(\mu, \sigma^2)$	$e^{i\mu t - \sigma^2 t^2 / 2}$

Ex. 12.2 — Se ξ e η são variáveis aleatórias independentes, prove que

$$\phi_{\xi+\eta}(t) = \phi_{\xi}(t)\phi_{\eta}(t) \quad (t \in \mathbb{R}).$$

Ex. 12.3 — Se $X = aY + b$, com $a, b \in \mathbb{R}$, prove que

$$\phi_X(t) = e^{itb} \phi_Y(at).$$

Ex. 12.4 — Sejam $X_1, \dots, X_n$ variáveis aleatórias independentes, cada uma com distribuição de Poisson de parâmetro λ . Use funções características para determinar a distribuição de $X_1 + \dots + X_n$.

Ex. 12.5 — O objetivo deste exercício é provar a Lei Fraca dos Grandes Números usando funções características. Sejam $\xi_1, \xi_2, \dots$ variáveis aleatórias i.i.d., suponha que $E[|\xi_1|] < \infty$ e escreva

$$S_n = \xi_1 + \dots + \xi_n, \quad \mu = E[\xi_1].$$

1. Mostre que, se ϕ é a função característica de ξ_1 , então

$$\phi_{S_n/n}(t) = \left[\phi\left(\frac{t}{n}\right) \right]^n.$$

2. Mostre que

$$\phi(t) = 1 + i\mu t + r(t), \quad \frac{r(t)}{t} \rightarrow 0$$

quando $t \rightarrow 0$.

3. Conclua que

$$\phi_{S_n/n}(t) = \left[1 + i\mu \frac{t}{n} + r \left(\frac{t}{n} \right) \right]^n \rightarrow e^{i\mu t}.$$

4. Use o Teorema da Continuidade de Lévy e o fato de que convergência em distribuição para uma constante implica convergência em probabilidade para concluir que

$$\frac{S_n}{n} \xrightarrow{P} \mu.$$

Teorema do Limite Central

Ex. 12.6 — Você convidou 64 pessoas para uma festa. O número de sanduíches consumidos por cada convidado assume os valores 0, 1 e 2 com probabilidades $1/4$, $1/2$ e $1/4$, respectivamente, e os consumos são independentes. Usando a aproximação normal e a correção de continuidade, determine o menor número inteiro de sanduíches que deve ser preparado para que a probabilidade aproximada de não haver escassez seja de pelo menos 95%. Registre a média, a variância e o ponto da correção de continuidade usados no cálculo.

Ex. 12.7 — Use o Teorema do Limite Central para mostrar que

$$\lim_{n \rightarrow \infty} \sum_{k=0}^n e^{-n} \frac{n^k}{k!} = \frac{1}{2}.$$

Ex. 12.8 — Suponha que a função geradora de momentos $M_\xi(s) = \mathbf{E}[e^{s\xi}]$ seja finita para s em alguma vizinhança de zero. Prove que, para cada $s \in \mathbb{R}$ fixo,

$$\lim_{n \rightarrow \infty} \left[M_\xi \left(\frac{s}{n} \right) \right]^n = e^{s\mathbf{E}[\xi]}.$$

Sugestão: calcule

$$\lim_{n \rightarrow \infty} n \log M_\xi \left(\frac{s}{n} \right).$$

Ex. 12.9 — Sejam $\xi_1, \xi_2, \dots$ variáveis aleatórias i.i.d., com

$$\mathbf{E}[\xi_1] = \mu, \quad \text{Var}(\xi_1) = \sigma^2 > 0,$$

e suponha que M_{ξ_1} seja finita numa vizinhança de zero. Defina

$$Z_n = \frac{\xi_1 + \cdots + \xi_n - n\mu}{\sigma\sqrt{n}}.$$

Mostre, usando uma expansão de segunda ordem de $\log M_{(\xi_1 - \mu)/\sigma}$ em zero, que, para todo $s \in \mathbb{R}$ fixo,

$$M_{Z_n}(s) \longrightarrow e^{s^2/2}.$$

Conclua o Teorema do Limite Central sob a hipótese adicional de existência da função geradora de momentos numa vizinhança de zero.

Arranjos triangulares

Ex. 12.10 — Para cada n , sejam $X_{n,1}, \dots, X_{n,k_n}$ variáveis aleatórias independentes e centradas. Suponha que

$$|X_{n,k}| \leq M \quad \text{quase certamente, para todos } n, k, \quad s_n^2 = \sum_{k=1}^{k_n} \text{Var}(X_{n,k}) \longrightarrow \infty.$$

Verifique diretamente a condição de Lindeberg e conclua que

$$\frac{1}{s_n} \sum_{k=1}^{k_n} X_{n,k} \xrightarrow{d} N(0,1).$$

Ex. 12.11 — Sejam $X_1, X_2, \dots$ variáveis aleatórias i.i.d., centradas, com variância um. Para cada n , sejam $a_{n,1}, \dots, a_{n,k_n} \in \mathbb{R}$ tais que

$$\sum_{k=1}^{k_n} a_{n,k}^2 = 1, \quad \max_{1 \leq k \leq k_n} |a_{n,k}| \longrightarrow 0.$$

Use o Teorema de Lindeberg–Feller para provar que

$$\sum_{k=1}^{k_n} a_{n,k} X_k \xrightarrow{d} N(0,1).$$

Ex. 12.12 — Para cada n , sejam $B_{n,1}, \dots, B_{n,k_n}$ variáveis independentes, com $B_{n,k} \sim \text{Bernoulli}(p_{n,k})$. Suponha que

$$\sum_{k=1}^{k_n} p_{n,k}(1 - p_{n,k}) \longrightarrow \infty.$$

Defina

$$m_n = \sum_{k=1}^{k_n} p_{n,k}, \quad s_n^2 = \sum_{k=1}^{k_n} p_{n,k}(1 - p_{n,k}).$$

Prove que

$$\frac{\sum_{k=1}^{k_n} B_{n,k} - m_n}{s_n} \xrightarrow{d} N(0,1)$$

verificando diretamente a condição de Lindeberg.

Ex. 12.13 — Oscilações das somas parciais na escala $\sqrt{n}$. Sejam $X_1, X_2, \dots$ variáveis aleatórias i.i.d., centradas, com $0 < \text{Var}(X_1) = \sigma^2 < \infty$, e ponha $S_n = X_1 + \dots + X_n$.

Prove que

$$\limsup_{n \rightarrow \infty} \frac{S_n}{\sqrt{n}} = +\infty \quad \text{e} \quad \liminf_{n \rightarrow \infty} \frac{S_n}{\sqrt{n}} = -\infty \quad \text{quase certamente.}$$

Medidas com sinal e o teorema de Radon–Nikodým

Ao reunir conjuntos disjuntos, uma medida positiva soma contribuições não negativas. Diferenças de probabilidades e integrais de funções que mudam de sinal exigem uma noção mais ampla. A aditividade enumerável continua sendo central, mas agora precisamos evitar a expressão indeterminada $\infty - \infty$.

Começaremos pela decomposição de Hahn, que separa o espaço em partes positiva e negativa. Essa separação dará a decomposição de Jordan da própria medida. A comparação com uma medida de referência levará, depois, às noções de continuidade absoluta e singularidade e ao teorema de Lebesgue–Radon–Nikodým. Ao final, veremos como esse resultado prepara a construção da esperança condicional.

13.1. Medidas com sinal

Definição 13.1 — Medida com sinal

Seja $(\Omega, \mathcal{F})$ um espaço mensurável. Uma função

$$\nu : \mathcal{F} \longrightarrow [-\infty, \infty]$$

é uma **medida com sinal** se:

- 1 $\nu(\emptyset) = 0$;
- 2 no máximo um dos valores $+\infty$ e $-\infty$ pertence à imagem de ν ;
- 3 para toda sequência $(E_n)_{n \geq 1}$ de conjuntos mensuráveis dois a dois disjuntos,

$$\nu\left(\bigcup_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty} \nu(E_n).$$

A série do último item é entendida em $[-\infty, \infty]$. A segunda condição impede que nela apareça a forma indeterminada $\infty - \infty$. Se a união tiver medida finita, a série converge absolutamente.

Quando $\nu(E) \in \mathbb{R}$ para todo $E \in \mathcal{F}$, dizemos que ν é uma **medida com sinal finita**. É esse o caso que aparece com mais frequência em probabilidade.

Exemplo 13.2 — Densidade com sinal.

Se $g \in L^1(\Omega, \mathcal{F}, \mu)$, então

$$\nu(E) = \int_E g \, d\mu = \int_E g^+ \, d\mu - \int_E g^- \, d\mu, \quad E \in \mathcal{F},$$

define uma medida com sinal finita. As regiões onde g é positiva e negativa antecipam a decomposição de Hahn.

Exemplo 13.3 — Diferença de medidas.

Se μ_1 e μ_2 são medidas finitas em $(\Omega, \mathcal{F})$, então

$$\nu(E) = \mu_1(E) - \mu_2(E)$$

é uma medida com sinal finita. A hipótese de finitude evita que essa diferença seja $\infty - \infty$.

**13.1.1. Conjuntos positivos, negativos e nulos**

O sinal de $\nu(E)$ não descreve necessariamente o sinal da medida dentro de E : contribuições positivas e negativas podem se cancelar. Para localizar o sinal, precisamos considerar todos os subconjuntos mensuráveis de E .

Definição 13.4 —

Seja ν uma medida com sinal e seja $E \in \mathcal{F}$.

- E é **positivo** para ν se $\nu(A) \geq 0$ para todo $A \in \mathcal{F}$ com $A \subseteq E$.
- E é **negativo** para ν se $\nu(A) \leq 0$ para todo $A \in \mathcal{F}$ com $A \subseteq E$.
- E é **nulo** para ν se $\nu(A) = 0$ para todo $A \in \mathcal{F}$ com $A \subseteq E$.

Um conjunto nulo para ν é ao mesmo tempo positivo e negativo. Essa noção é mais forte que a igualdade $\nu(E) = 0$: um saldo zero pode esconder cancelamentos. Se E é positivo, a restrição $A \mapsto \nu(A)$ aos subconjuntos mensuráveis de E é uma medida positiva.

Lema 13.5 —

Todo subconjunto mensurável de um conjunto positivo é positivo. Além disso, uma união enumerável de conjuntos positivos é positiva. As afirmações análogas valem para conjuntos negativos e nulos.

Demonstração. A primeira afirmação segue diretamente da definição. Para a segunda, sejam $E_1, E_2, \dots$ positivos e disjuntifiquemos a união:

$$F_1 = E_1, \quad F_n = E_n \setminus \bigcup_{j < n} E_j \quad (n \geq 2).$$

Então os F_n são disjuntos, $F_n \subseteq E_n$ e $\bigcup_n F_n = \bigcup_n E_n$. Se $B \subseteq \bigcup_n E_n$ é mensurável,

$$B = \bigcup_{n=1}^{\infty} (B \cap F_n)$$

é uma união disjunta. Como $B \cap F_n \subseteq E_n$,

$$\nu(B) = \sum_{n=1}^{\infty} \nu(B \cap F_n) \geq 0.$$

Logo a união é positiva. Aplicar o mesmo argumento a $-\nu$ prova a afirmação sobre conjuntos negativos; a afirmação sobre nulos segue das duas anteriores. $\square$

O lema seguinte fornece o mecanismo da decomposição de Hahn: enquanto houver um subconjunto de medida negativa, retiramos uma parte desse tipo. O nome descreve a estratégia. Este Pac-Man tem uma dieta seletiva: só retira conjuntos de medida negativa, de modo que a medida do que resta aumenta.

Lema 13.6 — Lema do Pac-Man

Se $E \in \mathcal{F}$ e $0 < \nu(E) < \infty$, então existe um conjunto positivo $P \subseteq E$ tal que $\nu(P) > 0$.

Demonstração. Se E já é positivo, tomamos $P = E$. Caso contrário, ponha $R_0 = E$. Enquanto R_{k-1} não for positivo, existe um subconjunto mensurável de R_{k-1} com medida negativa. Seja n_k o menor inteiro positivo para o qual se pode escolher $N_k \subseteq R_{k-1}$ satisfazendo

$$\nu(N_k) < -\frac{1}{n_k},$$

e defina $R_k = R_{k-1} \setminus N_k$. Se o procedimento terminar, o conjunto residual é positivo e, como apenas retiramos conjuntos de medida negativa, tem medida maior que $\nu(E)$.

Suponha, portanto, que a construção seja infinita. Os conjuntos N_k são dois a dois disjuntos e a sequência (n_k) é não decrescente, pois os resíduos diminuem. A aditividade enumerável dá

$$\nu\left(\bigcup_{k=1}^{\infty} N_k\right) = \sum_{k=1}^{\infty} \nu(N_k).$$

O lado direito não pode ser $-\infty$: de outro modo, a decomposição disjunta de E entre essa união e seu complemento seria incompatível com $0 < \nu(E) < \infty$. Como todos os termos são negativos,

$$\sum_{k=1}^{\infty} |\nu(N_k)| < \infty \quad \text{e, portanto,} \quad \sum_{k=1}^{\infty} \frac{1}{n_k} < \infty.$$

Em particular, $n_k \rightarrow \infty$.

Defina

$$P = E \setminus \bigcup_{k=1}^{\infty} N_k.$$

Temos $\nu(P) \geq \nu(E) > 0$. Resta provar que P é positivo. Se existisse um conjunto mensurável $A \subseteq P$ com $\nu(A) < 0$, poderíamos escolher $m \geq 1$ tal que $\nu(A) < -1/m$. Para todo k , o conjunto A está contido em R_{k-1} ; pela minimalidade de n_k , teríamos $n_k \leq m$. Isso contradiz $n_k \rightarrow \infty$. Logo P é positivo. $\square$

13.1.2. Decomposição de Hahn

O lema extrai um conjunto positivo de um conjunto de medida positiva. Usaremos essa extração para dividir todo o espaço em uma região positiva e outra negativa.

Teorema 13.7 — Decomposição de Hahn

Se ν é uma medida com sinal em $(\Omega, \mathcal{F})$, existem conjuntos mensuráveis disjuntos P e N tais que

$$\Omega = P \cup N,$$

com P positivo e N negativo para ν .

Demonstração. É suficiente tratar o caso em que ν nunca assume o valor $+\infty$. De fato, se esse não for o caso, a medida $-\nu$ nunca assume $+\infty$; uma decomposição para $-\nu$ fornece uma decomposição para ν após trocarmos os papéis das partes positiva e negativa.

Considere

$$\alpha = \sup\{\nu(A) : A \in \mathcal{F} \text{ e } A \text{ é positivo}\}.$$

A classe não é vazia, pois contém $\emptyset$. Além disso, $\alpha < \infty$. Se o supremo fosse infinito, poderíamos escolher conjuntos positivos P_n com $\nu(P_n) > n$. Pelo Lema 13.5, a união $\bigcup_n P_n$ seria positiva e teria medida ao menos n para todo n , obrigando ν a assumir o valor $+\infty$.

Escolha agora conjuntos positivos P_n tais que

$$\nu(P_n) > \alpha - \frac{1}{n}$$

e ponha $P = \bigcup_n P_n$. O conjunto P é positivo. Como $P \setminus P_n$ também é positivo,

$$\nu(P) \geq \nu(P_n) > \alpha - \frac{1}{n}.$$

Pela definição de α , também $\nu(P) \leq \alpha$; logo $\nu(P) = \alpha$.

Seja $N = \Omega \setminus P$. Se N não fosse negativo, existiria $E \subseteq N$ mensurável com $\nu(E) > 0$. Como ν não assume $+\infty$, essa medida seria finita. O Lema 13.6 forneceria um conjunto positivo $A \subseteq E$ com $\nu(A) > 0$. Então $P \cup A$ seria positivo e

$$\nu(P \cup A) = \nu(P) + \nu(A) > \alpha,$$

contradizendo a definição de α . Portanto N é negativo. □

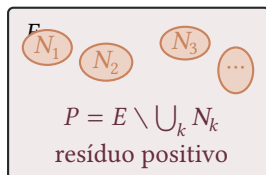
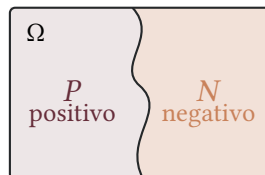
retirada das partes negativas**decomposição de Hahn**

Figura 13.1.: À esquerda, o mecanismo do Lema do Pac-Man: retiram-se partes negativas até restar um conjunto positivo. À direita, $\Omega = P \dot{\cup} N$ é uma decomposição de Hahn. A figura é esquemática: “positivo” significa que todo subconjunto mensurável de P tem medida não negativa, e analogamente para N .

A decomposição de Hahn admite alguma liberdade: conjuntos nulos podem mudar de lado. A proposição seguinte mostra que essa é toda a ambiguidade.

Proposição 13.8 — Unicidade de Hahn

Se (P, N) e (P', N') são duas decomposições de Hahn para ν , então $P \triangle P' = N \triangle N'$ é nulo para ν .

Demonstração. O conjunto $P \cap N'$ é subconjunto de um conjunto positivo e de um conjunto negativo; portanto é nulo. O mesmo vale para $P' \cap N$. Como

$$P \triangle P' = (P \cap N') \cup (P' \cap N),$$

o lema sobre uniões de conjuntos nulos conclui a prova. □

13.2. Decomposição de Jordan e variação total

Hahn separa o espaço; Jordan decompõe a medida. Basta restringir a medida às duas partes de Hahn e trocar o sinal na parte negativa. A singularidade mútua distinguirá essa decomposição de uma diferença arbitrária de medidas positivas.

Definição 13.9 — Medidas mutuamente singulares

Duas medidas positivas ou com sinal μ e ν são **mutuamente singulares**, e escrevemos $\mu \perp \nu$, se existe $E \in \mathcal{F}$ tal que E^c é nulo para μ e E é nulo para ν . Em outras palavras, as duas medidas podem ser concentradas em partes disjuntas de Ω .

Exemplo 13.10 — .

A medida de contagem restrita a $\mathbb{N} \subseteq \mathbb{R}$ e a medida de Lebesgue em $\mathbb{R}$ são mutuamente singulares: a primeira está concentrada em $\mathbb{N}$, enquanto a segunda atribui medida zero a esse conjunto.

◁

Teorema 13.11 — Decomposição de Jordan

Toda medida com sinal ν pode ser escrita de modo único na forma

$$\nu = \nu^+ - \nu^-,$$

em que ν^+ e ν^- são medidas positivas mutuamente singulares.

Demonstração. Seja (P, N) uma decomposição de Hahn e defina, para $E \in \mathcal{F}$,

$$\nu^+(E) = \nu(E \cap P), \quad \nu^-(E) = -\nu(E \cap N).$$

As restrições são medidas positivas, estão concentradas respectivamente em P e N , e a aditividade de ν fornece

$$\nu(E) = \nu^+(E) - \nu^-(E).$$

Para a unicidade, suponha que $\nu = \mu^+ - \mu^-$, com $\mu^+ \perp \mu^-$. Escolha uma partição (P', N') tal que μ^- seja nula em P' e μ^+ seja nula em N' . Para $A \subseteq P'$, $\nu(A) = \mu^+(A) \geq 0$; para $A \subseteq N'$, $\nu(A) = -\mu^-(A) \leq 0$. Logo (P', N') também é

uma decomposição de Hahn. Pela Proposição 13.8, ela difere de (P, N) apenas por conjuntos nulos. Assim, para todo $E \in \mathcal{F}$,

$$\mu^+(E) = \nu(E \cap P') = \nu(E \cap P) = \nu^+(E),$$

e, analogamente, $\mu^-(E) = \nu^-(E)$. □

A variação total soma as duas partes sem permitir que uma cancele a outra.

Definição 13.12 — Variação total

A **variação total** de ν é a medida positiva

$$|\nu| = \nu^+ + \nu^-.$$

Proposição 13.13 — Partes de Jordan

Para todo $E \in \mathcal{F}$,

$$\nu^+(E) = \sup\{\nu(A) : A \in \mathcal{F}, A \subseteq E\},$$

$$\nu^-(E) = -\inf\{\nu(A) : A \in \mathcal{F}, A \subseteq E\},$$

e

$$|\nu|(E) = \sup_{E = \bigcup_{j=1}^m E_j} \sum_{j=1}^m |\nu(E_j)|,$$

onde o supremo percorre as partições mensuráveis finitas de E .

Demonstração. Se (P, N) é uma decomposição de Hahn e $A \subseteq E$, então

$$\nu(A) = \nu(A \cap P) + \nu(A \cap N) \leq \nu(E \cap P) = \nu^+(E).$$

A igualdade é atingida por $A = E \cap P$. A fórmula para ν^- segue do mesmo argumento aplicado a $-\nu$.

Para qualquer partição finita de E ,

$$\sum_j |\nu(E_j)| \leq \sum_j (\nu^+(E_j) + \nu^-(E_j)) = |\nu|(E).$$

A partição $E = (E \cap P) \dot{\cup} (E \cap N)$ realiza a igualdade. □

Podemos agora relacionar essa medida à distância de variação total usada no capítulo de acoplamento. O fator dois decorre de que a diferença entre duas probabilidades tem massa total zero: suas partes positiva e negativa têm a mesma massa.

Corolário 13.14 — Variação total de uma diferença de probabilidades

Se μ e ν são probabilidades em $(\Omega, \mathcal{F})$ e $\rho = \mu - \nu$, então

$$|\mu - \nu|(\Omega) = |\rho|(\Omega) = 2d_{TV}(\mu, \nu).$$

Demonstração. Como $\rho(\Omega) = 0$, temos $\rho(A^c) = -\rho(A)$ para todo $A \in \mathcal{F}$. Portanto

$$\sup_{A \in \mathcal{F}} |\rho(A)| = \sup_{A \in \mathcal{F}} \rho(A) = - \inf_{A \in \mathcal{F}} \rho(A).$$

Pela Proposição 13.13,

$$\rho^+(\Omega) = d_{TV}(\mu, \nu), \quad \rho^-(\Omega) = d_{TV}(\mu, \nu).$$

Somando as duas partes de Jordan obtemos a identidade. □

Uma função mensurável f é integrável em relação a ν quando $f \in L^1(|\nu|)$. Nesse caso definimos

$$\int f \, d\nu = \int f \, d\nu^+ - \int f \, d\nu^-.$$

Não há indeterminação, pois as duas integrais são finitas, e

$$\left| \int f \, d\nu \right| \leq \int |f| \, d|\nu|. \quad (13.1)$$

A integração em relação a uma medida com sinal se reduz, portanto, à integração já construída para as duas partes positivas de Jordan.

13.3. Continuidade absoluta e singularidade

Passemos à comparação com uma medida de referência. A continuidade absoluta exige que seus conjuntos nulos também sejam nulos para a medida comparada. A

singularidade, por sua vez, permite concentrar as duas medidas em partes disjuntas do espaço.

Definição 13.15 — Continuidade absoluta

Se μ é uma medida positiva e ν é uma medida positiva ou com sinal no mesmo espaço mensurável, dizemos que ν é **absolutamente contínua** em relação a μ , e escrevemos

$$\nu \ll \mu,$$

se $\mu(E) = 0$ implica que E é nulo para ν .

Para uma medida positiva, basta exigir $\nu(E) = 0$. Para uma medida com sinal, a formulação com conjuntos nulos explicita que a condição vale também para todos os subconjuntos mensuráveis de E , não apenas para o saldo total.

Observação 13.16 —

Continuidade absoluta e singularidade não são negações uma da outra. Uma medida pode ter uma parte absolutamente contínua e outra singular em relação à mesma medida de referência. O teorema de decomposição a seguir dará uma forma precisa a essa separação.

Lema 13.17 —

Se $\nu \ll \mu$ e $\nu \perp \mu$, então ν é identicamente nula.

Demonstração. Escolha $S \in \mathcal{F}$ tal que $\mu(S) = 0$ e ν esteja concentrada em S . Pela continuidade absoluta, S é nulo para ν ; fora de S , a medida já era nula pela singularidade. Logo $\nu(E) = 0$ para todo $E \in \mathcal{F}$. □

13.4. Decomposição de Lebesgue–Radon–Nikodým

Sob as hipóteses do próximo teorema, a parte absolutamente contínua admite uma densidade em relação à medida de referência. A parte restante é singular. Começaremos com medidas finitas e depois passaremos ao caso σ -finito.

Teorema 13.18 — Lebesgue–Radon–Nikodým: caso finito

Se μ e ν são medidas positivas finitas em $(\Omega, \mathcal{F})$, existem uma função mensurável $f \geq 0$ e uma medida positiva ν_s tais que

$$\nu(E) = \int_E f \, d\mu + \nu_s(E), \quad E \in \mathcal{F},$$

e $\nu_s \perp \mu$. A função f é única μ -quase certamente e a medida ν_s é única.

Demonstração. Considere a classe

$$\mathcal{H} = \left\{ h \geq 0 \text{ mensurável} : \int_E h \, d\mu \leq \nu(E) \text{ para todo } E \in \mathcal{F} \right\}$$

e defina

$$\alpha = \sup_{h \in \mathcal{H}} \int h \, d\mu.$$

A função nula pertence a $\mathcal{H}$ e, tomando $E = \Omega$, vemos que $0 \leq \alpha \leq \nu(\Omega) < \infty$.

A classe $\mathcal{H}$ é fechada por máximos finitos. De fato, para $h, k \in \mathcal{H}$, basta decompor cada E entre $E \cap \{h \geq k\}$ e $E \cap \{h < k\}$ para obter

$$\int_E (h \vee k) \, d\mu \leq \nu(E).$$

Escolha $h_n \in \mathcal{H}$ com $\int h_n \, d\mu \rightarrow \alpha$ e ponha

$$g_n = h_1 \vee \cdots \vee h_n, \quad f = \lim_{n \rightarrow \infty} g_n.$$

Pelo teorema da convergência monótona, $f \in \mathcal{H}$ e

$$\int f \, d\mu = \alpha.$$

Defina a medida positiva

$$\lambda(E) = \nu(E) - \int_E f \, d\mu.$$

Mostraremos que $\lambda \perp \mu$. Para cada $m \geq 1$, aplique o Teorema de Hahn à medida com sinal $\lambda - m^{-1}\mu$ e escolha uma decomposição (P_m, N_m) . Como P_m é positivo,

$$\frac{1}{m}\mu(E \cap P_m) \leq \lambda(E \cap P_m) \quad (E \in \mathcal{F}).$$

Além disso, para todo $E \in \mathcal{F}$,

$$\begin{aligned} \int_E \left(f + \frac{1}{m} \mathbb{1}_{P_m} \right) d\mu &\leq \int_E f \, d\mu + \lambda(E \cap P_m) \\ &\leq \int_E f \, d\mu + \lambda(E) = \nu(E). \end{aligned}$$

Consequentemente, $f + m^{-1}\mathbb{1}_{P_m}$ pertence a $\mathcal{H}$. Pela maximalidade de f ,

$$\alpha + \frac{1}{m}\mu(P_m) \leq \alpha,$$

e portanto $\mu(P_m) = 0$.

Se $S = \bigcup_{m \geq 1} P_m$, então $\mu(S) = 0$. Para $E \subseteq S^c$, temos $E \subseteq N_m$ para todo m , e a negatividade de N_m fornece

$$0 \leq \lambda(E) \leq \frac{1}{m}\mu(E).$$

Fazendo $m \rightarrow \infty$, obtemos $\lambda(E) = 0$. Logo λ está concentrada no conjunto μ -nulo S , isto é, $\lambda \perp \mu$. Tomamos $\nu_s = \lambda$.

Resta a unicidade. Se

$$\nu = f\mu + \nu_s = g\mu + \tilde{\nu}_s$$

são duas decomposições, então a medida com sinal $(f - g)\mu = \tilde{\nu}_s - \nu_s$ é absolutamente contínua em relação a μ . Ela também é singular: se ν_s e $\tilde{\nu}_s$ estão concentradas, respectivamente, nos conjuntos μ -nulos S e $\tilde{S}$, então $\tilde{\nu}_s - \nu_s$ está concentrada em $S \cup \tilde{S}$, que ainda é μ -nulo. Pelo Lema 13.17, essa medida é nula. Logo $f = g$ μ -quase certamente e $\nu_s = \tilde{\nu}_s$. $\square$

A medida

$$\nu_a(E) = \int_E f \, d\mu$$

é a **parte absolutamente contínua** de ν , e ν_s é sua **parte singular**. A função

$$f = \frac{d\nu_a}{d\mu}$$

é a **derivada de Radon–Nikodým**. A ordem importa: f é a densidade de ν_a em relação a μ , não o contrário.

Corolário 13.19 — Radon–Nikodým finito

Se μ e ν são medidas positivas finitas e $\nu \ll \mu$, existe uma função mensurável $f \geq 0$, única μ -quase certamente, tal que

$$\nu(E) = \int_E f \, d\mu, \quad E \in \mathcal{F}. \quad (13.2)$$

Demonstração. Na decomposição do teorema, a parte singular $\nu_s = \nu - \nu_a$ também é absolutamente contínua em relação a μ . Pelo Lema 13.17, $\nu_s = 0$. □

A equação (13.2) tem a forma de um Teorema Fundamental do Cálculo: recuperamos a medida integrando sua densidade. A continuidade absoluta é a hipótese que permite essa recuperação sem uma parcela singular.

Teorema 13.20 — Radon–Nikodým σ -finito

Se μ e ν são medidas positivas σ -finitas em $(\Omega, \mathcal{F})$, então existem, de modo único, medidas positivas ν_a e ν_s tais que

$$\nu = \nu_a + \nu_s, \quad \nu_a \ll \mu, \quad \nu_s \perp \mu.$$

Além disso, existe uma função mensurável $f \geq 0$, única μ -quase certamente, para a qual

$$\nu_a(E) = \int_E f \, d\mu.$$

Se ν é uma medida com sinal cuja variação total $|\nu|$ é σ -finita, a mesma conclusão vale com f real e ν_s uma medida com sinal singular. Em particular, se ν é finita e $\nu \ll \mu$, então $f \in L^1(\mu)$ e

$$\nu(E) = \int_E f \, d\mu.$$

Demonstração. Pela σ -finitude, escolha coberturas mensuráveis

$$\Omega = \bigcup_{m \geq 1} A_m = \bigcup_{n \geq 1} B_n$$

tais que $\mu(A_m) < \infty$ e $\nu(B_n) < \infty$. Os conjuntos $A_m \cap B_n$ formam uma família enumerável que cobre Ω e têm medida finita para ambas as medidas. Enumerando essa família e disjuntificando-a, obtemos uma partição mensurável

$$\Omega = \dot{\bigcup}_{k \geq 1} E_k$$

tal que $\mu(E_k) < \infty$ e $\nu(E_k) < \infty$ para todo k . Aplicamos o caso finito às restrições de μ e ν a cada E_k . As densidades locais podem ser coladas na função

$$f = \sum_{k=1}^{\infty} f_k \mathbb{1}_{E_k},$$

e a união dos suportes singulares locais ainda é μ -nula. Isso produz a decomposição global; a unicidade segue do Lema 13.17.

Para uma medida com sinal, aplicamos o resultado às partes ν^+ e ν^- da decomposição de Jordan e subtraímos as respectivas decomposições. Se ν é finita e absolutamente contínua, a parte singular é nula e $\int |f| \, d\mu = |\nu|(\Omega) < \infty$. $\square$

13.4.1. Exemplos e uma ponte para a esperança condicional

Exemplo 13.21 — Uma medida mista.

Em $\mathbb{R}$, seja

$$\nu(E) = \int_E (1 + x^2) \, dx + 3\delta_0(E),$$

onde δ_0 é a massa de Dirac em 0. Em relação à medida de Lebesgue λ ,

$$\frac{d\nu_a}{d\lambda}(x) = 1 + x^2, \quad \nu_s = 3\delta_0.$$

A parte singular está concentrada num único ponto. Não podemos incorporá-la à densidade mudando apenas o valor desta em 0: a integral de Lebesgue não percebe essa alteração.

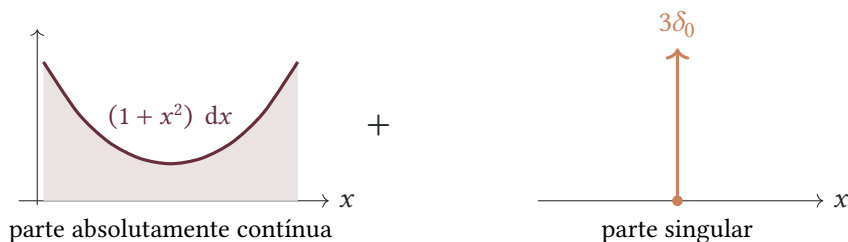


Figura 13.2.: A decomposição da medida mista do exemplo. A área representa a densidade em relação à medida de Lebesgue; a seta vertical representa massa concentrada em $\{0\}$, não o valor de uma densidade.

Exemplo 13.22 — Espaço discreto.

Num espaço enumerável, suponha que $\mu(\{x\}) > 0$ em todos os pontos. Toda medida ν é absolutamente contínua em relação a μ e

$$\frac{d\nu}{d\mu}(x) = \frac{\nu(\{x\})}{\mu(\{x\})}.$$

Se μ e ν são probabilidades, essa razão é o fator que transforma os pesos da primeira lei nos da segunda, ponto a ponto.

◀

A construção da esperança condicional dá uma aplicação direta do teorema. Seja $X \in L^1(\Omega, \mathcal{F}, \mathbf{P})$ e seja $\mathcal{G} \subseteq \mathcal{F}$ uma σ -álgebra. Defina, para $A \in \mathcal{G}$,

$$\nu(A) = \mathbf{E}[X\mathbb{1}_A].$$

Então ν é uma medida com sinal finita em $(\Omega, \mathcal{G})$ e $\nu \ll \mathbf{P}|_{\mathcal{G}}$. Pelo Teorema de Radon–Nikodým, existe uma função integrável e $\mathcal{G}$ -mensurável Y , única $\mathbf{P}$ -quase certamente, tal que

$$\mathbf{E}[X\mathbb{1}_A] = \mathbf{E}[Y\mathbb{1}_A], \quad A \in \mathcal{G}.$$

É essa função que, no capítulo seguinte, chamaremos de esperança condicional de X dada $\mathcal{G}$ e denotaremos por

$$Y = \mathbf{E}[X \mid \mathcal{G}].$$

A existência e a unicidade quase certa da esperança condicional decorrem, portanto, do Teorema de Radon–Nikodým. A restrição a $\mathcal{G}$ incorpora a informação disponível; a densidade Y preserva as integrais que essa informação permite observar.

EXERCÍCIOS

Ex. 13.1 — Seja $(\Omega, \mathcal{F}, \mu)$ um espaço de medida e $f \in L^1(\mu)$. Defina

$$\nu(A) = \int_A f \, d\mu, \quad A \in \mathcal{F}.$$

Prove que ν é uma medida com sinal finita. Encontre uma decomposição de Hahn, sua decomposição de Jordan e a medida de variação total $|\nu|$.

Ex. 13.2 — Seja $\Omega = \{1, \dots, n\}$ e sejam $a_1, \dots, a_n \in \mathbb{R}$. Para

$$\nu(A) = \sum_{i \in A} a_i, \quad A \subseteq \Omega,$$

determine explicitamente:

1. uma decomposição de Hahn de Ω ;
2. as medidas ν^+ e ν^- ;
3. $|\nu|(A)$ para cada $A \subseteq \Omega$.

Ex. 13.3 — Se μ_1 e μ_2 são medidas finitas no mesmo espaço mensurável, prove que $\nu = \mu_1 - \mu_2$ é uma medida com sinal e que

$$|\nu| \leq \mu_1 + \mu_2.$$

Tomando $\rho = \mu_1 + \mu_2$, escreva a decomposição de Jordan de ν em termos das derivadas de Radon–Nikodým de μ_1 e μ_2 com respeito a ρ .

Ex. 13.4 — Se (P, N) e (P', N') são duas decomposições de Hahn da mesma medida com sinal ν , prove que $P \triangle P'$ e $N \triangle N'$ são $|\nu|$ -nulos. Conclua que as partes positiva e negativa da decomposição de Jordan são únicas.

Ex. 13.5 — Seja ν uma medida com sinal finita. Prove que, para todo $A \in \mathcal{F}$,

$$|\nu|(A) = \sup \left\{ \sum_{j=1}^m |\nu(A_j)| : m \geq 1, \quad A = \bigsqcup_{j=1}^m A_j, \quad A_j \in \mathcal{F} \right\}.$$

Mostre também que $|\nu(A)| \leq |\nu|(A)$.

Ex. 13.6 — Sejam λ , ν e μ medidas σ -finitas tais que $\lambda \ll \nu \ll \mu$. Prove a regra da cadeia

$$\frac{d\lambda}{d\mu} = \frac{d\lambda}{d\nu} \frac{d\nu}{d\mu} \quad \mu\text{-quase certamente.}$$

Especifique versões das derivadas para as quais o produto está definido.

Ex. 13.7 — Sejam P e Q probabilidades no mesmo espaço mensurável, com $Q \ll P$, e escreva $Z = dQ/dP$.

1. Prove que $Z \geq 0$ e $\mathbf{E}_P[Z] = 1$.

2. Para toda variável $X \geq 0$, prove a fórmula de mudança de medida

$$\mathbf{E}_Q[X] = \mathbf{E}_P[XZ].$$

3. Estenda a identidade às variáveis Q -integráveis, indicando a condição correspondente em termos de P e Z .

Ex. 13.8 — Em $\mathbb{R}$, defina

$$\mu = \lambda + \delta_0, \quad \nu = 2\lambda + 3\delta_0 + \delta_1,$$

onde λ é a medida de Lebesgue. Determine a decomposição de Lebesgue $\nu = \nu_a + \nu_s$ relativa a μ e calcule $d\nu_a/d\mu$.

Esperança Condicional

Uma sub- σ -álgebra $\mathcal{G} \subseteq \mathcal{F}$ representa informação parcial: determina quais eventos a informação disponível permite distinguir. Se ξ é integrável, queremos substituí-la por uma variável η que use apenas essa informação e preserve as médias sobre os eventos observáveis. Isso significa exigir

$$\int_G \xi \, d\mathbf{P} = \int_G \eta \, d\mathbf{P} \quad \text{para todo } G \in \mathcal{G}.$$

A variável η pode ser vista como uma versão desfocada de ξ : não reproduz todos os seus valores, mas conserva a integral sobre cada evento de $\mathcal{G}$. Em $\mathcal{L}^2$, veremos que ela é também a melhor aproximação de ξ baseada em $\mathcal{G}$, no sentido do erro quadrático médio.

Definição 14.1 —

Seja $(\Omega, \mathcal{F}, \mathbf{P})$ um espaço de probabilidade, seja ξ integrável e seja $\mathcal{G} \subseteq \mathcal{F}$ uma sub- σ -álgebra. Uma **esperança condicional de ξ dada $\mathcal{G}$** , denotada por $\mathbf{E}[\xi \mid \mathcal{G}]$, é uma variável aleatória integrável η tal que:

1 η é $\mathcal{G}$ -mensurável;

2 para todo $G \in \mathcal{G}$,

$$\int_G \xi \, d\mathbf{P} = \int_G \eta \, d\mathbf{P}.$$

Primeiro verificaremos que essas condições determinam a esperança condicional a menos de conjuntos de probabilidade zero.

Teorema 14.2 —

A esperança condicional de uma variável aleatória integrável ξ , se existir, é única quase certamente.

Demonstração. Suponha que η e η' satisfaçam as propriedades da esperança condicional. Pela segunda propriedade da definição,

$$\int_A \eta \, d\mathbf{P} = \int_A \eta' \, d\mathbf{P} \text{ para todo } A \in \mathcal{G}.$$

Logo, $\int_A (\eta - \eta') \, d\mathbf{P} = 0$ para todo $A \in \mathcal{G}$. Para $\varepsilon > 0$, considere $A_\varepsilon = \{\eta - \eta' > \varepsilon\} \in \mathcal{G}$. Então

$$0 = \int_{A_\varepsilon} (\eta - \eta') \, d\mathbf{P} \geq \mathbf{P}(A_\varepsilon) \cdot \varepsilon.$$

Portanto, $\mathbf{P}(A_\varepsilon) = 0$ para todo $\varepsilon > 0$. Como

$$\{\eta - \eta' > 0\} = \bigcup_{n=1}^{\infty} A_{1/n},$$

a subaditividade enumerável implica

$$\mathbf{P}(\eta - \eta' > 0) = 0.$$

Da mesma forma, trocando η e η' , obtemos $\mathbf{P}(\eta' - \eta > 0) = 0$. Logo, $\eta = \eta'$ quase certamente.

□

A existência geral será demonstrada na seção seguinte. Antes disso, vejamos como construir a variável condicional quando a informação disponível é dada por uma partição.

Exemplo 14.3 — .

Seja G um evento tal que $0 < \mathbf{P}(G) < 1$ e seja $\mathcal{G} = \sigma(G) = \{\emptyset, G, G^c, \Omega\}$. Então,

$$\mathbf{E}[\xi|\mathcal{G}](\omega) = \begin{cases} \mathbf{E}[\xi|G], & \text{se } \omega \in G \\ \mathbf{E}[\xi|G^c], & \text{se } \omega \in G^c. \end{cases} = \eta(\omega)$$

onde $\mathbf{E}[\xi | G] = \mathbf{E}[\xi 1_G] / \mathbf{P}(G)$. A variável aleatória definida acima é uma esperança condicional.

◁

Demonstração. A variável η é $\mathcal{G}$ -mensurável, porque $\{\omega : \eta(\omega) \in B\}$ está em $\mathcal{G}$ para todo conjunto de Borel B em $\mathbb{R}$.

Pela definição das médias condicionadas aos eventos,

$$\int_G \xi \, d\mathbf{P} = \int_G \eta \, d\mathbf{P}$$

e

$$\int_{G^c} \xi \, d\mathbf{P} = \int_{G^c} \eta \, d\mathbf{P}.$$

As identidades para $\emptyset$ e Ω seguem dessas duas por aditividade. Logo, a igualdade das integrais vale para todo elemento de $\mathcal{G}$.

□

A mesma construção vale para uma partição enumerável: em cada parte, tomamos a média correspondente.

Exemplo 14.4 — .

Considere $(\Omega, \mathcal{F}, \mathbf{P})$ e uma variável aleatória integrável ξ . Seja $\mathcal{P} = \{A_k : k \geq 1\}$ uma partição enumerável de Ω , com $\mathbf{P}(A_k) > 0$ para todo k , e seja $\sigma(\mathcal{P})$ a σ -álgebra por ela gerada.

Então uma versão da esperança condicional é constante em cada elemento da partição e dada por

$$\mathbf{E}[\xi \mid \sigma(\mathcal{P})] = \sum_{k \geq 1} \mathbf{E}[\xi \mid A_k] \mathbf{1}_{A_k}.$$

A Figura 14.1 ilustra essa construção no caso unidimensional.

◀

Heurística 14.5 —

Suponha que ξ seja o saldo de uma conta no fim do dia, mas que a informação disponível revele apenas a faixa em que ele caiu. A esperança condicional atribui, em cada faixa, a média compatível com a informação observada. Ela não recupera o saldo exato nem altera a média em qualquer união de faixas conhecidas. Faz médias, não adivinhações.

14.1. Existência e unicidade da esperança condicional

14.1.1. Existência em $\mathcal{L}^2$

Em $\mathcal{L}^2$, a existência tem uma interpretação geométrica: condicionar é projetar sobre o subespaço das variáveis $\mathcal{G}$ -mensuráveis.

No espaço de Hilbert $\mathcal{L}^2(\Omega, \mathcal{F}, \mathbf{P})$, a norma é

$$\|\xi\|_2 = (\mathbf{E}[\xi^2])^{1/2}.$$

Ela é induzida pelo produto interno

$$\langle \xi, \eta \rangle = \mathbf{E}[\xi \eta].$$

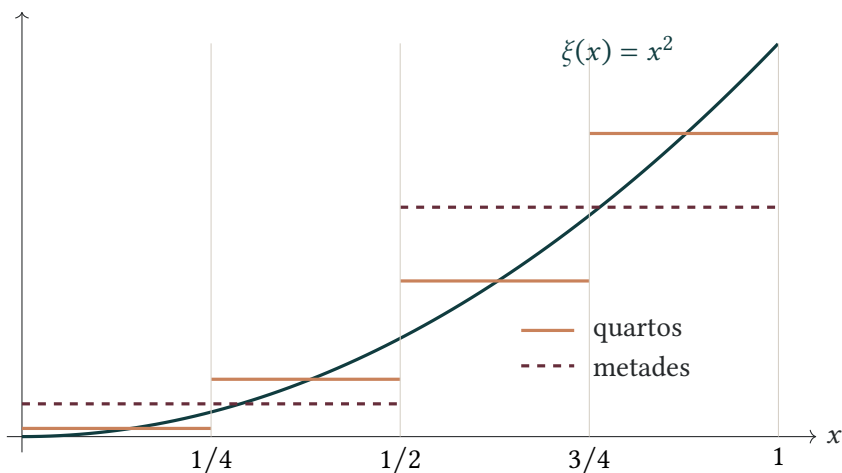


Figura 14.1.: No espaço $[0,1]$ com medida de Lebesgue, as esperanças condicionais de $\xi(x) = x^2$ dadas pelas partições em quartos e em metades são as médias de ξ em cada átomo.

Lema 14.6 —

$\mathcal{L}^2(\Omega, \mathcal{G}, \mathbf{P})$ é um subespaço fechado de $\mathcal{L}^2(\Omega, \mathcal{F}, \mathbf{P})$.

Demonstração. Seja (X_n) uma sequência em $\mathcal{L}^2(\Omega, \mathcal{G}, \mathbf{P})$ que converge em $\mathcal{L}^2$ para $X \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbf{P})$. Escolha uma subsequência (X_{n_k}) tal que

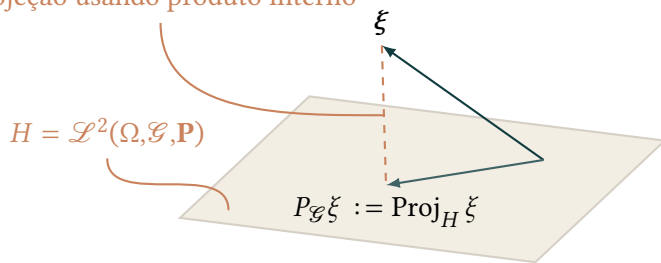
$$\|X_{n_k} - X\|_2^2 \leq 2^{-k}.$$

Para todo $\varepsilon > 0$, a desigualdade de Markov dá

$$\sum_{k=1}^{\infty} \mathbf{P}(|X_{n_k} - X| > \varepsilon) \leq \frac{1}{\varepsilon^2} \sum_{k=1}^{\infty} \|X_{n_k} - X\|_2^2 < \infty.$$

Pelo primeiro lema de Borel–Cantelli, $X_{n_k} \rightarrow X$ quase certamente. Como cada X_{n_k} é $\mathcal{G}$ -mensurável, a variável $\tilde{X} = \limsup_k X_{n_k}$ também é $\mathcal{G}$ -mensurável e satisfaz $\tilde{X} = X$ quase certamente. Assim, a classe de equivalência de X pertence a $\mathcal{L}^2(\Omega, \mathcal{G}, \mathbf{P})$, que é, portanto, fechada. □

Projeção usando produto interno



Teorema 14.7 – Esperança condicional como projeção em $\mathcal{L}^2$

Seja $(\Omega, \mathcal{F}, \mathbf{P})$ um espaço de probabilidade e seja $\mathcal{G} \subseteq \mathcal{F}$ uma sub- σ -álgebra. Seja

$$P_{\mathcal{G}} : \mathcal{L}^2(\Omega, \mathcal{F}, \mathbf{P}) \longrightarrow \mathcal{L}^2(\Omega, \mathcal{G}, \mathbf{P})$$

a projeção ortogonal, cuja existência é garantida pelo Lema 14.6. Então, para qualquer $\xi \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbf{P})$,

$$\mathbf{E}[\xi \mid \mathcal{G}] = P_{\mathcal{G}}\xi$$

quase certamente.

Demonstração. Ponha $\eta = P_{\mathcal{G}}\xi$. Por definição, $\eta \in \mathcal{L}^2(\Omega, \mathcal{G}, \mathbf{P})$ e, portanto, possui uma versão $\mathcal{G}$ -mensurável e integrável.

Para $G \in \mathcal{G}$, a indicadora $\mathbb{1}_G$ pertence a $\mathcal{L}^2(\Omega, \mathcal{G}, \mathbf{P})$. Como $\xi - \eta$ é ortogonal a esse subespaço,

$$0 = \langle \xi - \eta, \mathbb{1}_G \rangle = \int_G \xi \, d\mathbf{P} - \int_G \eta \, d\mathbf{P}.$$

Logo, η satisfaz a definição de $\mathbf{E}[\xi \mid \mathcal{G}]$. A unicidade quase certa segue do Teorema 14.2. □

14.1.2. Existência em $\mathcal{L}^1$

Para variáveis apenas integráveis, não podemos recorrer diretamente à geometria de $\mathcal{L}^2$. O Teorema de Radon–Nikodým substitui a projeção e fornece a construção

em $\mathcal{L}^1$.

Teorema 14.8 —

A esperança condicional de uma variável aleatória integrável ξ existe e é única quase certamente.

Demonstração. Suponha primeiro $\xi \geq 0$. No espaço mensurável $(\Omega, \mathcal{G})$, considere as medidas finitas

$$\mu = \mathbf{P}|_{\mathcal{G}}, \quad \nu(A) = \int_A \xi d\mathbf{P}, \quad A \in \mathcal{G}.$$

A medida ν é finita porque ξ é integrável, e $\nu \ll \mu$: se $\mathbf{P}(A) = 0$, então $\int_A \xi d\mathbf{P} = 0$. Pelo Teorema de Radon–Nikodým, existe uma função $\mathcal{G}$ -mensurável $f \geq 0$ tal que

$$\nu(A) = \int_A f d\mathbf{P} \quad \text{para todo } A \in \mathcal{G}.$$

Além disso, $\int f d\mathbf{P} = \nu(\Omega) = \mathbf{E}\xi < \infty$. Portanto, f é integrável e satisfaz as duas propriedades da esperança condicional.

Para ξ arbitrária, escreva $\xi = \xi^+ - \xi^-$. Pelo caso não negativo, existem

$$\eta_+ = \mathbf{E}[\xi^+ | \mathcal{G}],$$

$$\eta_- = \mathbf{E}[\xi^- | \mathcal{G}].$$

A variável $\eta = \eta_+ - \eta_-$ é $\mathcal{G}$ -mensurável e integrável. Para todo $A \in \mathcal{G}$,

$$\int_A \eta d\mathbf{P} = \int_A \eta_+ d\mathbf{P} - \int_A \eta_- d\mathbf{P} = \int_A \xi^+ d\mathbf{P} - \int_A \xi^- d\mathbf{P} = \int_A \xi d\mathbf{P}.$$

Isso prova a existência. A unicidade é o conteúdo do Teorema 14.2. □

Os casos extremos esclarecem o papel da informação; a partição permite ver os casos intermediários.

Exemplo 14.9 — .

- 1** Para a σ -álgebra trivial $\mathcal{G} = \{\emptyset, \Omega\}$, toda variável $\mathcal{G}$ -mensurável é constante quase certamente. Nesse caso,

$$\mathbf{E}[\xi \mid \mathcal{G}] = \mathbf{E}\xi.$$

Com efeito, para $A = \Omega$,

$$\int_A (\mathbf{E}\xi) d\mathbf{P} = \int_A \xi d\mathbf{P}$$

e ambos os lados são iguais a $\mathbf{E}\xi$; para $A = \emptyset$, ambos são zero.

- 2** Para a informação completa $\mathcal{G} = \mathcal{F}$,

$$\mathbf{E}[\xi \mid \mathcal{G}] = \xi.$$

A própria variável ξ satisfaz as duas condições da definição: com toda a informação, não resta nada a estimar.

- 3** Suponha que $\Omega = \bigsqcup_{k \geq 1} \Omega_k$ seja uma partição enumerável com $\mathbf{P}(\Omega_k) > 0$, e tome $\mathcal{G} = \sigma(\Omega_1, \Omega_2, \dots)$. Então,

$$\mathbf{E}[\xi \mid \mathcal{G}](\omega) = \sum_{k \geq 1} \frac{\mathbf{E}[\xi \mathbf{1}_{\Omega_k}]}{\mathbf{P}(\Omega_k)} \mathbf{1}_{\Omega_k}(\omega).$$

Essa função é $\mathcal{G}$ -mensurável. Se $A = \bigsqcup_{k \in I} \Omega_k \in \mathcal{G}$, a aditividade da integral mostra que sua integral em A é

$$\sum_{k \in I} \mathbf{E}[\xi \mathbf{1}_{\Omega_k}] = \int_A \xi d\mathbf{P},$$

verificando a segunda condição da definição.

14.2. Esperança Condicional em Relação a uma Variável Aleatória

Observar uma variável aleatória significa conhecer os eventos que ela permite distinguir. Por isso, condicionar em uma variável é condicionar na σ -álgebra por ela gerada.

Definição 14.10 —

Dadas duas variáveis aleatórias ξ e η , com ξ integrável, definimos

$$\mathbf{E}[\xi \mid \eta] := \mathbf{E}[\xi \mid \sigma(\eta)],$$

onde

$$\sigma(\eta) = \{\eta^{-1}(B) : B \in \mathcal{B}(\mathbb{R})\}.$$

Exemplo 14.11 — .

Suponha que η tome os valores $y_1, \dots, y_n$, todos com probabilidade positiva. Então $\sigma(\eta) = \sigma(\Omega_1, \dots, \Omega_n)$, onde $\Omega_k = \{\eta = y_k\}$. Logo,

$$\mathbf{E}[\xi \mid \eta](\omega) = \mathbf{E}[\xi \mid \eta = y_k] \quad \text{para } \omega \in \Omega_k.$$



14.3. Propriedades da Esperança Condicional

As propriedades seguintes decorrem das duas condições da definição e da unicidade quase certa. Em cada caso, a estratégia será verificar a mensurabilidade e a igualdade das integrais nos eventos de $\mathcal{G}$.

Proposição 14.12 – Linearidade

Se ξ e η são integráveis e $a, b \in \mathbb{R}$, então

$$\mathbf{E}[a\xi + b\eta \mid \mathcal{G}] = a\mathbf{E}[\xi \mid \mathcal{G}] + b\mathbf{E}[\eta \mid \mathcal{G}]$$

quase certamente.

Demonstração. A variável $a\mathbf{E}[\xi \mid \mathcal{G}] + b\mathbf{E}[\eta \mid \mathcal{G}]$ é $\mathcal{G}$ -mensurável e integrável. Além disso, para todo $A \in \mathcal{G}$,

$$\begin{aligned} \int_A (a\mathbf{E}[\xi \mid \mathcal{G}] + b\mathbf{E}[\eta \mid \mathcal{G}]) d\mathbf{P} &= a \int_A \mathbf{E}[\xi \mid \mathcal{G}], d\mathbf{P} + b \int_A \mathbf{E}[\eta \mid \mathcal{G}], d\mathbf{P} \\ &= a \int_A \xi, d\mathbf{P} + b \int_A \eta, d\mathbf{P} \\ &= \int_A (a\xi + b\eta), d\mathbf{P}. \end{aligned}$$

Portanto, ela satisfaz a definição de $\mathbf{E}[a\xi + b\eta \mid \mathcal{G}]$; a unicidade quase certa conclui a prova. □

Corolário 14.13 –

A aplicação $X \mapsto \mathbf{E}[X \mid \mathcal{G}]$ é um operador linear de $\mathcal{L}^1(\Omega, \mathcal{F}, \mathbf{P})$ em $\mathcal{L}^1(\Omega, \mathcal{G}, \mathbf{P})$.

Demonstração. A Proposição anterior fornece a linearidade. A existência, a integrabilidade e a $\mathcal{G}$ -mensurabilidade da imagem seguem do Teorema 14.8 e da definição. □

A esperança associa um número a cada variável integrável: é um funcional linear. A esperança condicional associa outra variável aleatória, mensurável em relação à informação disponível: é um operador linear.

Proposição 14.14 – Monotonicidade

Se ξ e η são integráveis e $\xi \leq \eta$ quase certamente, então

$$\mathbf{E}[\xi \mid \mathcal{G}] \leq \mathbf{E}[\eta \mid \mathcal{G}]$$

quase certamente.

Demonstração. Para todo $A \in \mathcal{G}$,

$$\begin{aligned} \int_A \mathbf{E}[\xi \mid \mathcal{G}] d\mathbf{P} &= \int_A \xi d\mathbf{P} \\ &\leq \int_A \eta d\mathbf{P} \\ &= \int_A \mathbf{E}[\eta \mid \mathcal{G}] d\mathbf{P} \end{aligned}$$

Ponha $Z = \mathbf{E}[\eta \mid \mathcal{G}] - \mathbf{E}[\xi \mid \mathcal{G}]$. A variável Z é $\mathcal{G}$ -mensurável e $\int_A Z d\mathbf{P} \geq 0$ para todo $A \in \mathcal{G}$. Se $\{Z < 0\}$ tivesse probabilidade positiva, algum conjunto $\{Z \leq -1/m\}$ também teria probabilidade positiva, e a integral de Z nesse conjunto seria estritamente negativa, uma contradição. Logo, $Z \geq 0$ quase certamente. $\square$

Corolário 14.15 –

Se ξ é integrável, então

$$|\mathbf{E}[\xi \mid \mathcal{G}]| \leq \mathbf{E}[|\xi| \mid \mathcal{G}]$$

quase certamente.

Demonstração. Como $\xi \leq |\xi|$ e $-\xi \leq |\xi|$, a monotonicidade e a linearidade dão

$$\mathbf{E}[\xi \mid \mathcal{G}] \leq \mathbf{E}[|\xi| \mid \mathcal{G}], \quad -\mathbf{E}[\xi \mid \mathcal{G}] \leq \mathbf{E}[|\xi| \mid \mathcal{G}].$$

As duas desigualdades equivalem à conclusão. □

Teorema 14.16 — Jensen condicional

Se $Y \in L^1$ e $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ é convexa, com $\varphi(Y) \in L^1$, então

$$\varphi(\mathbf{E}[Y \mid \mathcal{G}]) \leq \mathbf{E}[\varphi(Y) \mid \mathcal{G}]$$

quase certamente.

Demonstração. Uma função convexa real é o supremo de uma família enumerável de minorantes afins. Para ver isso, escolha uma reta de suporte em cada ponto racional. Se $q_k \in \mathbb{Q}$ converge para x , a convexidade mantém limitadas as inclinações dessas retas em qualquer intervalo compacto que contenha x no interior; pela continuidade de φ , os valores das retas em x convergem para $\varphi(x)$. Logo o supremo enumerável dessas retas coincide com φ em todo ponto.

Escreva os minorantes como $\ell_j(x) = a_j x + b_j$. Pela linearidade e pela monotonicidade da esperança condicional,

$$\ell_j(\mathbf{E}[Y \mid \mathcal{G}]) = \mathbf{E}[\ell_j(Y) \mid \mathcal{G}] \leq \mathbf{E}[\varphi(Y) \mid \mathcal{G}]$$

para todo j . Tomando o supremo sobre j , obtemos a desigualdade. □

A convergência dominada também admite uma versão condicional.

Teorema 14.17 — Teorema da Convergência Dominada

Suponha que $\xi_n \rightarrow \xi$ quase certamente e que $|\xi_n| \leq \eta$ quase certamente para todo n , onde η é integrável. Então,

$$\mathbf{E}[\xi_n \mid \mathcal{G}] \longrightarrow \mathbf{E}[\xi \mid \mathcal{G}] \quad \text{quase certamente.}$$

Demonstração. Como $|\xi| \leq \eta$ quase certamente, ξ é integrável. Defina

$$Z_n = \sup_{k \geq n} |\xi_k - \xi|.$$

Então $0 \leq Z_n \leq 2\eta$ e $Z_n \downarrow 0$ quase certamente. Pela linearidade, pelo corolário da monotonicidade e pela própria monotonicidade,

$$\begin{aligned} |\mathbf{E}[\xi_n | \mathcal{G}] - \mathbf{E}[\xi | \mathcal{G}]| &= |\mathbf{E}[\xi_n - \xi | \mathcal{G}]| \\ &\leq \mathbf{E}[|\xi_n - \xi| | \mathcal{G}] \\ &\leq \mathbf{E}[Z_n | \mathcal{G}]. \end{aligned}$$

Pela monotonicidade, e modificando todas as versões numa única união enumerável de conjuntos nulos, podemos escolher $W_n = \mathbf{E}[Z_n | \mathcal{G}]$ de modo que $W_n \downarrow W$ quase certamente, com $W \geq 0$. Para todo n ,

$$0 \leq \mathbf{E}W \leq \mathbf{E}W_n = \mathbf{E}Z_n.$$

Pelo Teorema da Convergência Dominada, $\mathbf{E}Z_n \rightarrow 0$. Logo, $\mathbf{E}W = 0$ e $W = 0$ quase certamente. A estimativa anterior conclui a prova. $\square$

Um fator já conhecido a partir de $\mathcal{G}$ pode ser retirado da esperança condicional, sob as condições de integrabilidade do enunciado seguinte.

Proposição 14.18 — Retirada do que é conhecido

Se η é integrável, ξ é $\mathcal{G}$ -mensurável e $\xi\eta$ é integrável, então

$$\mathbf{E}[\xi\eta | \mathcal{G}] = \xi\mathbf{E}[\eta | \mathcal{G}]$$

quase certamente.

Demonstração. Primeiro suponha que $\xi = 1_B$ para algum $B \in \mathcal{G}$. Para todo $A \in \mathcal{G}$,

$$\int_A \xi \mathbf{E}[\eta \mid \mathcal{G}] d\mathbf{P} = \int_{A \cap B} \mathbf{E}[\eta \mid \mathcal{G}] d\mathbf{P} = \int_{A \cap B} \eta d\mathbf{P} = \int_A \xi \eta d\mathbf{P}.$$

Por linearidade, a identidade vale quando ξ é $\mathcal{G}$ -mensurável simples.

Suponha agora $\xi \geq 0$ e $\eta \geq 0$. Tome funções simples $\mathcal{G}$ -mensuráveis $\xi_n \uparrow \xi$. A identidade já provada para ξ_n e o Teorema da Convergência Monótona, aplicado aos dois lados, dão, para todo $A \in \mathcal{G}$,

$$\int_A \xi \mathbf{E}[\eta \mid \mathcal{G}] d\mathbf{P} = \int_A \xi \eta d\mathbf{P}.$$

Aplicando esse caso a $|\xi|$ e $|\eta|$, com $A = \Omega$, obtemos

$$\mathbf{E}[|\xi| \mathbf{E}[|\eta| \mid \mathcal{G}]] = \mathbf{E}[|\xi| |\eta|] < \infty.$$

Como $|\mathbf{E}[\eta \mid \mathcal{G}]| \leq \mathbf{E}[|\eta| \mid \mathcal{G}]$, o produto $\xi \mathbf{E}[\eta \mid \mathcal{G}]$ é integrável. Finalmente, decompondo ξ e η em partes positiva e negativa e usando a linearidade, a identidade integral vale para toda $A \in \mathcal{G}$. Como $\xi \mathbf{E}[\eta \mid \mathcal{G}]$ é $\mathcal{G}$ -mensurável, esse produto é uma versão de $\mathbf{E}[\xi \eta \mid \mathcal{G}]$. □

Usaremos essa propriedade no estudo de martingalas, ao separar o que já é conhecido da informação que ainda será observada.

14.4. Esperança Condicional como Projeção

As identidades seguintes descrevem o condicionamento sucessivo em duas informações encaixadas. Em qualquer ordem, a informação menos detalhada determina o resultado final.

Proposição 14.19 – Suavização/Elevação

Se ξ é integrável e $\mathcal{G}_1 \subseteq \mathcal{G}_2 \subseteq \mathcal{F}$, então

$$1 \quad \mathbf{E}[\mathbf{E}[\xi \mid \mathcal{G}_1] \mid \mathcal{G}_2] = \mathbf{E}[\xi \mid \mathcal{G}_1];$$

$$2 \quad \mathbf{E}[\mathbf{E}[\xi \mid \mathcal{G}_2] \mid \mathcal{G}_1] = \mathbf{E}[\xi \mid \mathcal{G}_1],$$

quase certamente.

Demonstração. Seja $\mathcal{G}_1 \subseteq \mathcal{G}_2$.

- 1 A variável $\eta = \mathbf{E}[\xi \mid \mathcal{G}_1]$ é $\mathcal{G}_1$ -mensurável e, portanto, também é $\mathcal{G}_2$ -mensurável. Como uma variável integrável mensurável em relação à álgebra condicionante é sua própria esperança condicional,

$$\mathbf{E}[\eta \mid \mathcal{G}_2] = \eta$$

- 2 Mostraremos que $\mathbf{E}[\xi \mid \mathcal{G}_1]$ é a esperança condicional de $\mathbf{E}[\xi \mid \mathcal{G}_2]$ dada $\mathcal{G}_1$.

a) Por definição, $\mathbf{E}[\xi \mid \mathcal{G}_1]$ é $\mathcal{G}_1$ -mensurável.

b) Para $A \in \mathcal{G}_1$, queremos mostrar que

$$\int_A \mathbf{E}[\xi \mid \mathcal{G}_1] d\mathbf{P} = \int_A \mathbf{E}[\xi \mid \mathcal{G}_2] d\mathbf{P}. \quad (14.1)$$

Como $A \in \mathcal{G}_1 \subseteq \mathcal{G}_2$, a definição de esperança condicional mostra que ambos os lados de (14.1) são iguais a

$$\int_A \xi d\mathbf{P}.$$

Isso prova o segundo item. □

Corolário 14.20 —

Se ξ é integrável, então

$$\mathbf{E}[\mathbf{E}[\xi \mid \mathcal{G}] \mid \mathcal{G}] = \mathbf{E}[\xi \mid \mathcal{G}]$$

quase certamente.

Demonstração. Aplique o primeiro item da proposição com $\mathcal{G}_1 = \mathcal{G}_2 = \mathcal{G}$. $\square$

Portanto, a esperança condicional $P : \xi \mapsto \mathbf{E}[\xi \mid \mathcal{G}]$ é uma projeção linear: $P^2 = P$. Sua imagem é $\mathcal{L}^1(\Omega, \mathcal{G}, \mathbf{P})$, o subespaço das classes que possuem representante $\mathcal{G}$ -mensurável.

Em $\mathcal{L}^2$, essa projeção é ortogonal: o erro não tem componente em nenhuma direção representada por uma variável baseada na informação de $\mathcal{G}$.

Proposição 14.21 — Ortogonalidade

Se $\xi \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbf{P})$, então

$$\xi - \mathbf{E}[\xi \mid \mathcal{G}] \perp \mathcal{L}^2(\Omega, \mathcal{G}, \mathbf{P}).$$

Isto é, para todo $\zeta \in \mathcal{L}^2(\Omega, \mathcal{G}, \mathbf{P})$,

$$\mathbf{E}[\zeta(\xi - \mathbf{E}[\xi \mid \mathcal{G}])] = 0.$$

Demonstração. Como $\zeta \in \mathcal{L}^2$ e $\xi \in \mathcal{L}^2$, os produtos envolvidos são integráveis. Pela Proposição 14.18,

$$\mathbf{E}[\zeta \xi \mid \mathcal{G}] = \zeta \mathbf{E}[\xi \mid \mathcal{G}].$$

Tomando esperanças e usando a definição de esperança condicional no evento Ω , obtemos

$$\mathbf{E}[\zeta \xi] = \mathbf{E}[\zeta \mathbf{E}[\xi \mid \mathcal{G}]],$$

que é a identidade desejada. $\square$

Teorema 14.22 — A esperança condicional é o melhor estimador

Se $\xi \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbf{P})$, então, para toda $\eta \in \mathcal{L}^2(\Omega, \mathcal{G}, \mathbf{P})$,

$$\mathbf{E}\left[(\xi - \mathbf{E}[\xi \mid \mathcal{G}])^2\right] \leq \mathbf{E}[(\xi - \eta)^2]. \quad (14.2)$$

Demonstração. Ponha $Z = \mathbf{E}[\xi \mid \mathcal{G}]$ e $\zeta = Z - \eta \in \mathcal{L}^2(\Omega, \mathcal{G}, \mathbf{P})$. Então

$$\xi - \eta = (\xi - Z) + \zeta.$$

Pela Proposição 14.21, o produto cruzado tem esperança zero. Portanto, a identidade de Pitágoras dá

$$\begin{aligned} \mathbf{E}[(\xi - \eta)^2] &= \mathbf{E}[(\xi - Z)^2] + \mathbf{E}[\zeta^2] + 2\mathbf{E}[\zeta(\xi - Z)] \\ &= \mathbf{E}[(\xi - Z)^2] + \mathbf{E}[\zeta^2] \\ &\geq \mathbf{E}[(\xi - Z)^2], \end{aligned}$$

que é (14.2). □

O termo *melhor* tem aqui um sentido preciso: entre os estimadores quadrado-integráveis que usam apenas a informação de $\mathcal{G}$, a esperança condicional minimiza o erro quadrático médio.

Corolário 14.23 —

Em $\mathcal{L}^2(\Omega, \mathcal{F}, \mathbf{P})$, a aplicação $\xi \mapsto \mathbf{E}[\xi \mid \mathcal{G}]$ é a projeção ortogonal sobre o subespaço $\mathcal{L}^2(\Omega, \mathcal{G}, \mathbf{P})$.

Demonstração. A imagem é $\mathcal{G}$ -mensurável, e a Proposição 14.21 mostra que o erro é ortogonal ao subespaço. Essas duas propriedades caracterizam unicamente a projeção ortogonal, em acordo com o Teorema 14.7. □

Nos cálculos condicionais com variáveis normais, a seguinte independência será útil.

Proposição 14.24 – Soma e diferença de normais independentes

Se X e Y são independentes, ambos com distribuição $N(0,1)$, então

$$S = X + Y \quad \text{e} \quad D = X - Y$$

são independentes e ambos têm distribuição $N(0,2)$.

Demonstração. A densidade conjunta de (X,Y) é

$$\frac{1}{2\pi} \exp \left\{ -\frac{x^2 + y^2}{2} \right\}.$$

Faça a mudança linear

$$s = x + y, \quad d = x - y,$$

cujas inversas são $x = (s + d)/2$, $y = (s - d)/2$ e cujo jacobiano inverso tem módulo $1/2$. Como

$$x^2 + y^2 = \frac{s^2 + d^2}{2},$$

a densidade conjunta de (S,D) é

$$\frac{1}{4\pi} \exp \left\{ -\frac{s^2 + d^2}{4} \right\} = \left(\frac{1}{\sqrt{4\pi}} e^{-s^2/4} \right) \left(\frac{1}{\sqrt{4\pi}} e^{-d^2/4} \right).$$

Ela fatora no produto de duas densidades $N(0,2)$, o que prova simultaneamente as distribuições marginais e a independência. □

EXERCÍCIOS

Ex. 14.1 — Seja $\mathcal{G} = \sigma(A_1, \dots, A_m)$, onde $A_1, \dots, A_m$ formam uma partição mensurável de Ω e $\mathbf{P}(A_j) > 0$ para todo j . Se $X \in L^1$, prove que

$$\mathbf{E}[X \mid \mathcal{G}] = \sum_{j=1}^m \frac{\mathbf{E}[X \mathbb{1}_{A_j}]}{\mathbf{P}(A_j)} \mathbb{1}_{A_j} \quad \text{quase certamente.}$$

Ex. 14.2 — Se Y assume os valores distintos $y_1, y_2, \dots$ e $\mathbf{P}(Y = y_j) > 0$, prove que, para $X \in L^1$,

$$\mathbf{E}[X \mid Y] = g(Y), \quad g(y_j) = \mathbf{E}[X \mid Y = y_j].$$

Explique como a fórmula deve ser interpretada nos valores y que têm probabilidade zero.

Ex. 14.3 — Seja $X \in L^1$ e seja $\mathcal{G} \subseteq \mathcal{F}$ uma σ -álgebra independente de $\sigma(X)$. Prove que

$$\mathbf{E}[X \mid \mathcal{G}] = \mathbf{E}[X] \quad \text{quase certamente.}$$

Ex. 14.4 — Sejam $\mathcal{G}_1 \subseteq \mathcal{G}_2 \subseteq \mathcal{F}$ e $X \in L^1$. Prove a propriedade da torre

$$\mathbf{E}[\mathbf{E}[X \mid \mathcal{G}_2] \mid \mathcal{G}_1] = \mathbf{E}[X \mid \mathcal{G}_1].$$

Dê um exemplo mostrando que, em geral, $\mathbf{E}[\mathbf{E}[X \mid \mathcal{G}_1] \mid \mathcal{G}_2]$ não recupera X .

Ex. 14.5 — Sejam $X_1, \dots, X_n$ variáveis de Bernoulli i.i.d. de parâmetro p e $S_n = X_1 + \dots + X_n$. Use a simetria entre as coordenadas para provar que

$$\mathbf{E}[X_k \mid S_n] = \frac{S_n}{n} \quad \text{quase certamente, } k = 1, \dots, n.$$

Ex. 14.6 — Seja $X \in L^2$ e $\mathcal{G} \subseteq \mathcal{F}$. Prove a decomposição da variância

$$\text{Var}(X) = \mathbf{E}[\text{Var}(X \mid \mathcal{G})] + \text{Var}(\mathbf{E}[X \mid \mathcal{G}]),$$

onde

$$\text{Var}(X \mid \mathcal{G}) = \mathbf{E}[(X - \mathbf{E}[X \mid \mathcal{G}])^2 \mid \mathcal{G}].$$

Ex. 14.7 — Prove que a esperança condicional é uma contração:

1. se $X \in L^1$, então $\|\mathbf{E}[X \mid \mathcal{G}]\|_1 \leq \|X\|_1$;
2. se $X \in L^2$, então $\|\mathbf{E}[X \mid \mathcal{G}]\|_2 \leq \|X\|_2$. Indique em cada item se você está usando Jensen condicional ou a interpretação como projeção.

Ex. 14.8 — Sejam X e Y independentes, ambos com distribuição $N(0,1)$, e ponha $S = X + Y$.

1. Prove que

$$\mathbf{E}[X \mid S] = \frac{S}{2}.$$

Dica. Mostre que $X - S/2 = (X - Y)/2$ é independente de S .

2. Calcule o erro quadrático médio de $S/2$ como estimador de X e compare-o com o erro do estimador constante 0.

Ex. 14.9 — Seja $A \in \mathcal{F}$, com $0 < \mathbf{P}(A) < 1$, e $\mathcal{G} = \sigma(A)$. Para $X \in L^1$, prove que

$$\mathbf{E}[X \mid \mathcal{G}] = \mathbf{E}[X \mid A]\mathbb{1}_A + \mathbf{E}[X \mid A^c]\mathbb{1}_{A^c}.$$

Use essa fórmula para calcular uma esperança condicional concreta de sua escolha num espaço amostral finito.

Martingalas

Uma martingala não adivinha o futuro. Sua propriedade é outra: dada a informação disponível em cada instante, a esperança condicional do próximo valor coincide com o valor presente. Essa é a formulação matemática de um jogo justo. Para estudá-la, precisaremos descrever a informação que se acumula, as decisões de interromper a observação e o comportamento das trajetórias no limite. Filtrações, tempos de parada e teoremas de convergência darão forma a essas ideias.

Neste capítulo, todos os processos são indexados por $\mathbb{N}_0 = \{0, 1, 2, \dots\}$ e definidos em um espaço de probabilidade $(\Omega, \mathcal{F}, \mathbf{P})$.

15.1. Informação ao longo do tempo

Definição 15.1 – Processo estocástico

Se (S, Σ) é um espaço mensurável, um **processo estocástico** com espaço de estados S é uma família

$$X = (X_n)_{n \geq 0}$$

de variáveis aleatórias $X_n : \Omega \rightarrow S$. Para $\omega \in \Omega$ fixo, a aplicação $n \mapsto X_n(\omega)$ é uma **trajetória** ou **caminho** do processo.

Na maior parte do capítulo, $S = \mathbb{R}$. O índice n representa o tempo: X_n é o valor observado no instante n .

Definição 15.2 – Filtração

Uma **filtração** é uma sequência crescente de sub- σ -álgebras de $\mathcal{F}$,

$$\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \dots \subseteq \mathcal{F}.$$

O par $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n \geq 0}, \mathbf{P})$ é chamado de espaço de probabilidade filtrado.

Interpretamos $\mathcal{F}_n$ como a informação disponível até o instante n . A inclusão entre essas famílias expressa que a informação já adquirida continua disponível nos instantes seguintes.

Definição 15.3 – Processo adaptado

Um processo $X = (X_n)_{n \geq 0}$ é **adaptado** à filtração $(\mathcal{F}_n)$ se X_n é $\mathcal{F}_n$ -mensurável para todo n .

A menor filtração à qual X é adaptado é sua **filtração natural**,

$$\mathcal{F}_n^X = \sigma(X_0, \dots, X_n).$$

Podemos trabalhar com uma filtração maior, que contenha, além da história de X , outras informações reveladas até o tempo n . A filtração escolhida faz parte do modelo: as esperanças condicionais serão calculadas em relação a ela.

15.2. Martingalas, submartingalas e supermartingalas

A linguagem das filtrações permite distinguir um jogo justo daqueles que apresentam, condicionalmente à informação disponível, uma tendência favorável ou desfavorável.

Definição 15.4 — Martingala

Um processo real $X = (X_n)_{n \geq 0}$, adaptado a $(\mathcal{F}_n)$, é uma **martingala** se $X_n \in L^1$ para todo n e

$$\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] = X_n \quad \text{q.s.}$$

Ele é uma **submartingala** se

$$\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] \geq X_n \quad \text{q.s.,}$$

e uma **supermartingala** se a desigualdade é invertida.

A tendência de crescimento ou decrescimento se refere à esperança condicional, não a cada trajetória. O processo X é uma martingala se, e somente se, é simultaneamente uma submartingala e uma supermartingala. Além disso, X é uma supermartingala se, e somente se, $-X$ é uma submartingala. O prefixo *super*, portanto, não promete ganhos maiores.

Proposição 15.5 — Propriedade de vários passos

Se X é uma martingala e $0 \leq m \leq n$, então

$$\mathbb{E}[X_n \mid \mathcal{F}_m] = X_m \quad \text{q.s.}$$

Para submartingalas, o lado esquerdo é maior ou igual a X_m ; para supermartingalas, é menor ou igual a X_m .

Demonstração. No caso de uma martingala, a propriedade da torre fornece

$$\mathbf{E}[X_n \mid \mathcal{F}_m] = \mathbf{E}[\mathbf{E}[X_n \mid \mathcal{F}_{n-1}] \mid \mathcal{F}_m] = \mathbf{E}[X_{n-1} \mid \mathcal{F}_m].$$

A repetição desse argumento termina em X_m . No caso de uma submartingala, a monotonicidade da esperança condicional substitui cada igualdade por uma desigualdade no sentido correto. O caso das supermartingalas segue aplicando o resultado a $-X$. □

Tomando esperanças, obtemos

$$\mathbf{E}X_n = \mathbf{E}X_0$$

para martingalas. Para submartingalas, $(\mathbf{E}X_n)$ é crescente; para supermartingalas, decrescente. A recíproca não é verdadeira: a igualdade das esperanças não substitui a condição relativa à informação de cada instante.

Teorema 15.6 — Jensen para martingalas

Se X é uma martingala, $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ é convexa e $\varphi(X_n) \in L^1$ para todo n , então $(\varphi(X_n))_{n \geq 0}$ é uma submartingala.

Mais geralmente, se X é uma submartingala e φ é convexa e não decrescente, então $(\varphi(X_n))$ é uma submartingala sempre que os termos forem integráveis.

Demonstração. Pelo Teorema de Jensen condicional 14.16,

$$\mathbf{E}[\varphi(X_{n+1}) \mid \mathcal{F}_n] \geq \varphi(\mathbf{E}[X_{n+1} \mid \mathcal{F}_n]).$$

Se X é uma martingala, o último membro é $\varphi(X_n)$. Se X é uma submartingala, usamos também que φ é não decrescente. □

Em particular, se X é uma martingala e os termos envolvidos são integráveis, então $(|X_n|)$ e $(|X_n|^p)$, para $p \geq 1$, são submartingalas.

15.3. Exemplos fundamentais

Somas centradas, produtos normalizados e esperanças condicionais fornecem três maneiras recorrentes de construir martingalas. O passeio aleatório permite vê-las em ação.

Exemplo 15.7 — Somas de variáveis independentes.

Se $Y_1, Y_2, \dots$ são independentes, integráveis e $\mathbf{E}Y_k = 0$, então

$$S_0 = 0, \quad S_n = \sum_{k=1}^n Y_k$$

é uma martingala para $\mathcal{F}_n = \sigma(Y_1, \dots, Y_n)$. De fato,

$$\mathbf{E}[S_{n+1} \mid \mathcal{F}_n] = S_n + \mathbf{E}[Y_{n+1} \mid \mathcal{F}_n] = S_n.$$

◀

Exemplo 15.8 — Produtos.

Se $Y_1, Y_2, \dots$ são independentes, integráveis e $\mathbf{E}Y_k = 1$, e se todo produto abaixo é integrável, então

$$M_0 = 1, \quad M_n = \prod_{k=1}^n Y_k$$

é uma martingala para a filtração natural das variáveis Y_k , pois

$$\mathbf{E}[M_{n+1} \mid \mathcal{F}_n] = M_n \mathbf{E}[Y_{n+1}] = M_n.$$

◀

Exemplo 15.9 — Esperanças condicionais.

Se $Z \in L^1$, então

$$M_n = \mathbf{E}[Z \mid \mathcal{F}_n]$$

é uma martingala. A propriedade da torre é exatamente a propriedade de martingala:

$$\mathbf{E}[M_{n+1} \mid \mathcal{F}_n] = \mathbf{E}[Z \mid \mathcal{F}_n] = M_n.$$

<

Exemplo 15.10 — Passeio aleatório simples.

Se $Y_1, Y_2, \dots$ são independentes e $\mathbf{P}(Y_k = 1) = \mathbf{P}(Y_k = -1) = 1/2$, o passeio $S_n = Y_1 + \dots + Y_n$ é uma martingala. Também são martingalas

$$S_n^2 - n \quad \text{e} \quad \frac{\exp(\theta S_n)}{(\cosh \theta)^n}, \quad \theta \in \mathbb{R}.$$

Para a primeira, basta usar $\mathbf{E}[Y_{n+1} \mid \mathcal{F}_n] = 0$ e $Y_{n+1}^2 = 1$. Para a segunda, $\mathbf{E}[e^{\theta Y_{n+1}}] = \cosh \theta$.

<

15.4. Estratégias previsíveis e processos parados

Uma estratégia só pode usar a informação disponível antes de cada aposta. A previsibilidade registra essa restrição: o que se conhece de antemão é a aposta, não o ganho.

Definição 15.11 — Processo previsível

Um processo $H = (H_n)_{n \geq 1}$ é **previsível** se H_n é $\mathcal{F}_{n-1}$ -mensurável para todo $n \geq 1$. Sua transformada discreta de X é

$$(H \cdot X)_n = \sum_{k=1}^n H_k (X_k - X_{k-1}), \quad (H \cdot X)_0 = 0.$$

Nessa interpretação, $X_k - X_{k-1}$ é o ganho unitário no passo k , e H_k , a aposta escolhida antes de esse ganho ser revelado. A transformada soma os ganhos produzidos pela estratégia.

Proposição 15.12 — Transformada de martingala

Se X é uma martingala e H é previsível e limitado, então $(H \cdot X)_n$ é uma martingala. Se X é uma submartingala e H é previsível, limitado e não negativo, então a transformada é uma submartingala.

Demonstração. Escrevendo $\Delta X_k = X_k - X_{k-1}$, temos

$$\mathbf{E}[H_k \Delta X_k \mid \mathcal{F}_{k-1}] = H_k \mathbf{E}[\Delta X_k \mid \mathcal{F}_{k-1}].$$

O último membro é zero no caso de uma martingala e é não negativo no caso de uma submartingala com $H_k \geq 0$. A limitação de H garante a integrabilidade. $\square$

Definição 15.13 — Tempo de parada

Uma variável aleatória

$$\tau : \Omega \longrightarrow \mathbb{N}_0 \cup \{\infty\}$$

é um **tempo de parada** em relação a $(\mathcal{F}_n)$ se

$$\{\tau \leq n\} \in \mathcal{F}_n \quad \text{para todo } n \geq 0.$$

A decisão de parar no instante n deve depender somente da informação disponível até n . Como $\{\tau = n\} = \{\tau \leq n\} \setminus \{\tau \leq n-1\}$, a definição é equivalente a exigir $\{\tau = n\} \in \mathcal{F}_n$ para todo n .

Exemplo 15.14 – .

- 1 Todo tempo determinístico $\tau \equiv m$ é um tempo de parada.
- 2 Se X é adaptado e $B \subseteq \mathbb{R}$ é boreliano, então

$$\tau_B = \inf\{n \geq 0 : X_n \in B\}$$

é um tempo de parada, pois $\{\tau_B \leq n\} = \bigcup_{k=0}^n \{X_k \in B\} \in \mathcal{F}_n$.

- 3 Em geral, o último instante em que X visita B não é um tempo de parada: saber que uma visita é a última exige olhar o futuro.

◁

Para um tempo de parada τ , definimos o **processo parado**

$$X_n^\tau = X_{n \wedge \tau}, \quad n \geq 0.$$

Se $\tau = \infty$, convencionamos $n \wedge \tau = n$. A identidade fundamental é

$$X_{n \wedge \tau} = X_0 + \sum_{k=1}^n \mathbb{1}_{\{\tau \geq k\}} (X_k - X_{k-1}). \quad (15.1)$$

Como $\{\tau \geq k\} = \{\tau \leq k-1\}^c \in \mathcal{F}_{k-1}$, os indicadores formam uma estratégia previsível: mantemos os incrementos até a parada e descartamos os seguintes.

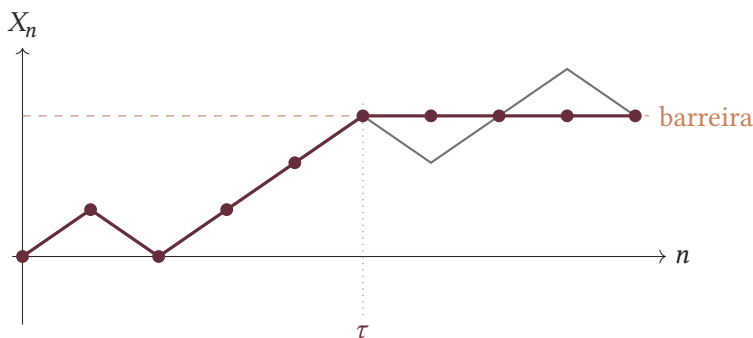


Figura 15.1.: A trajetória original aparece em cinza. Depois de τ , o processo parado X^τ , em vinho, permanece em X_τ .

Teorema 15.15 – Processo parado

Se X é uma martingala e τ é um tempo de parada, então $X^\tau = (X_{n \wedge \tau})_{n \geq 0}$ é uma martingala. O mesmo vale para submartingalas e supermartingalas.

Demonstração. A identidade (15.1) escreve $X^\tau - X_0$ como a transformada de X pela estratégia previsível $H_k = \mathbb{1}_{\{\tau \geq k\}}$, que toma valores em $\{0,1\}$. A Proposição 15.12 conclui a demonstração nos casos de martingalas e submartingalas. O caso das supermartingalas segue por simetria. $\square$

15.5. O teorema da parada de Doob

O processo parado conserva a propriedade de martingala em cada tempo determinístico. Para comparar diretamente os valores em dois tempos de parada, precisamos descrever a informação disponível nesses instantes aleatórios.

Definição 15.16 – Informação em um tempo de parada

Para um tempo de parada τ , definimos

$$\mathcal{F}_\tau = \{A \in \mathcal{F} : A \cap \{\tau \leq n\} \in \mathcal{F}_n \text{ para todo } n \geq 0\}.$$

Essa coleção é uma σ -álgebra. De fato, $\Omega \in \mathcal{F}_\tau$, e as identidades

$$A^c \cap \{\tau \leq n\} = \{\tau \leq n\} \setminus (A \cap \{\tau \leq n\})$$

e

$$\left(\bigcup_{j \geq 1} A_j \right) \cap \{\tau \leq n\} = \bigcup_{j \geq 1} (A_j \cap \{\tau \leq n\})$$

mostram que $\mathcal{F}_\tau$ é fechada por complementos e uniões enumeráveis.

Se $\sigma \leq \tau$ e $A \in \mathcal{F}_\sigma$, então

$$A \cap \{\tau \leq n\} = (A \cap \{\sigma \leq n\}) \cap \{\tau \leq n\} \in \mathcal{F}_n.$$

logo $\mathcal{F}_\sigma \subseteq \mathcal{F}_\tau$. Finalmente, se X é adaptado e $C \subseteq \mathbb{R}$ é boreliano, então

$$\{X_\tau \in C, \tau < \infty\} \cap \{\tau \leq n\} = \bigcup_{k=0}^n (\{\tau = k\} \cap \{X_k \in C\}) \in \mathcal{F}_n.$$

Assim, X_τ é $\mathcal{F}_\tau$ -mensurável no evento $\{\tau < \infty\}$.

Teorema 15.17 — Amostragem opcional limitada

Se X é uma submartingala e $\sigma \leq \tau \leq N$ são tempos de parada, então

$$X_\sigma \leq \mathbf{E}[X_\tau \mid \mathcal{F}_\sigma] \quad \text{q.s.}$$

Se X é uma martingala, há igualdade; se é uma supermartingala, a desigualdade é invertida.

Demonstração. Como σ e τ são limitados, as variáveis X_σ e X_τ são integráveis. Além disso,

$$X_\tau - X_\sigma = \sum_{k=1}^N \mathbf{1}_{\{\sigma < k \leq \tau\}} (X_k - X_{k-1}).$$

Se $A \in \mathcal{F}_\sigma$, então $A \cap \{\sigma < k \leq \tau\} \in \mathcal{F}_{k-1}$. Logo, para uma submartingala,

$$\begin{aligned} \mathbf{E}[\mathbf{1}_A(X_\tau - X_\sigma)] &= \sum_{k=1}^N \mathbf{E}[\mathbf{1}_{A \cap \{\sigma < k \leq \tau\}} \mathbf{E}[X_k - X_{k-1} \mid \mathcal{F}_{k-1}]] \\ &\geq 0. \end{aligned}$$

Essa desigualdade, válida para todo $A \in \mathcal{F}_\sigma$, caracteriza $X_\sigma \leq \mathbf{E}[X_\tau \mid \mathcal{F}_\sigma]$. No caso de uma martingala, todos os termos são zero. O caso das supermartingalas segue de $-X$. □

Corolário 15.18 — Parada de Doob limitada

Se X é uma martingala e $\tau \leq N$ é um tempo de parada, então

$$\mathbf{E}X_\tau = \mathbf{E}X_0.$$

Para uma submartingala, $\mathbf{E}X_\tau \geq \mathbf{E}X_0$; para uma supermartingala, $\mathbf{E}X_\tau \leq$

$\mathbf{E}X_0$.

Demonstração. Aplique o Teorema 15.17 aos tempos de parada $\sigma = 0$ e τ , e depois tome esperanças. Para uma martingala,

$$X_0 = \mathbf{E}[X_\tau \mid \mathcal{F}_0] \implies \mathbf{E}X_0 = \mathbf{E}X_\tau.$$

Nos casos de submartingala e supermartingala, a mesma operação preserva o sentido da desigualdade correspondente. $\square$

Com tempos limitados, não foi necessária uma passagem ao limite. Para estender o resultado a tempos de parada não limitados, precisamos justificar a passagem de $X_{\tau \wedge n}$ a X_τ também nas esperanças. A integrabilidade uniforme fornece esse controle.

Definição 15.19 — Integrabilidade uniforme

Uma família $\{Z_i : i \in I\} \subset L^1$ é **uniformemente integrável** se

$$\lim_{K \rightarrow \infty} \sup_{i \in I} \mathbf{E}[|Z_i| \mathbb{1}_{\{|Z_i| > K\}}] = 0.$$

Lema 15.20 — Critérios úteis

Uma família dominada por uma variável integrável é uniformemente integrável. Além disso, se para algum $p > 1$

$$\sup_{i \in I} \mathbf{E}|Z_i|^p < \infty,$$

então $\{Z_i : i \in I\}$ é uniformemente integrável.

Demonstração. Se $|Z_i| \leq Y \in L^1$, então

$$|Z_i| \mathbb{1}_{\{|Z_i| > K\}} \leq Y \mathbb{1}_{\{Y > K\}},$$

e o lado direito tem esperança tendendo a zero. No segundo caso,

$$|Z_i| \mathbb{1}_{\{|Z_i| > K\}} \leq K^{1-p} |Z_i|^p.$$

□

Lema 15.21 — Lema de Vitali

Se $Z_n \rightarrow Z$ q.s. e a família $\{Z_n : n \geq 1\}$ é uniformemente integrável, então $Z \in L^1$ e $Z_n \rightarrow Z$ em L^1 .

Demonstração. A integrabilidade uniforme implica $\sup_n \mathbf{E}|Z_n| < \infty$. Com efeito, escolhendo K tão grande que

$$\sup_n \mathbf{E} \left[|Z_n| \mathbb{1}_{\{|Z_n| > K\}} \right] \leq 1,$$

temos, para todo n ,

$$\mathbf{E}|Z_n| \leq K + \mathbf{E} \left[|Z_n| \mathbb{1}_{\{|Z_n| > K\}} \right] \leq K + 1.$$

Pelo lema de Fatou, $Z \in L^1$. Seja $\varphi_K(x) = (-K) \vee (x \wedge K)$. Por convergência dominada,

$$\mathbf{E}|\varphi_K(Z_n) - \varphi_K(Z)| \rightarrow 0 \quad (n \rightarrow \infty).$$

Além disso,

$$|Z_n - \varphi_K(Z_n)| \leq |Z_n| \mathbb{1}_{\{|Z_n| > K\}},$$

e vale a desigualdade análoga para Z . Fazendo primeiro $n \rightarrow \infty$ e depois $K \rightarrow \infty$, obtemos $\mathbf{E}|Z_n - Z| \rightarrow 0$. □

Teorema 15.22 – Parada de Doob

Seja X uma martingala e seja $\tau < \infty$ q.s. Se a família

$$\{X_{\tau \wedge n} : n \geq 0\}$$

é uniformemente integrável, então

$$\mathbf{E}X_\tau = \mathbf{E}X_0.$$

Em particular, a conclusão vale em cada uma das situações seguintes:

- 1 existe $Y \in L^1$ tal que $|X_{\tau \wedge n}| \leq Y$ para todo n ;
- 2 os incrementos satisfazem $|X_n - X_{n-1}| \leq C$ q.s. e $\mathbf{E}\tau < \infty$.

Para submartingalas e supermartingalas, sob a mesma integrabilidade uniforme, valem respectivamente as desigualdades $\mathbf{E}X_\tau \geq \mathbf{E}X_0$ e $\mathbf{E}X_\tau \leq \mathbf{E}X_0$.

Demonstração. Pelo Corolário 15.18, $\mathbf{E}X_{\tau \wedge n} = \mathbf{E}X_0$ para todo n . Como $\tau < \infty$ q.s., temos $X_{\tau \wedge n} \rightarrow X_\tau$ q.s. O Lema 15.21 transforma essa convergência em convergência em L^1 , e portanto podemos passar ao limite nas esperanças. Para submartingalas e supermartingalas, fazemos a mesma passagem ao limite nas desigualdades da versão limitada.

A primeira condição particular implica integrabilidade uniforme pelo Lema 15.20. Sob a segunda,

$$|X_{\tau \wedge n}| \leq |X_0| + \sum_{k=1}^{\tau \wedge n} |X_k - X_{k-1}| \leq |X_0| + C\tau,$$

e o último membro é integrável. □

As hipóteses não são ornamento. Parar em tempo finito quase certamente não basta para preservar a esperança, como mostra o exemplo seguinte.

Exemplo 15.23 — Uma hipótese indispensável.

Seja S um passeio aleatório simples, com $S_0 = 0$, e ponha

$$\tau = \inf\{n \geq 0 : S_n = 1\}.$$

Então $\tau < \infty$ q.s., mas $\mathbf{E}\tau = \infty$. Como $S_\tau = 1$,

$$\mathbf{E}S_\tau = 1 \neq 0 = \mathbf{E}S_0;$$

logo, a conclusão do teorema da parada falha.

Para justificar as duas afirmações sobre τ , introduzimos uma segunda barreira. Para $a \geq 1$, seja

$$\tau_a = \inf\{n \geq 0 : S_n \in \{-a, 1\}\}.$$

Em blocos de $a + 1$ passos existe uma probabilidade positiva, uniforme em relação ao ponto interior inicial, de atingir uma das duas barreiras. Assim, τ_a tem cauda geométrica e $\mathbf{E}\tau_a < \infty$. Aplicando o teorema da parada a S_n ,

$$0 = \mathbf{E}S_{\tau_a} = \mathbf{P}(S_{\tau_a} = 1) - a\mathbf{P}(S_{\tau_a} = -a),$$

de onde

$$\mathbf{P}(S_{\tau_a} = 1) = \frac{a}{a+1}.$$

Esses eventos crescem para $\{\tau < \infty\}$; portanto $\mathbf{P}(\tau < \infty) = 1$. Além disso,

$$|S_{n \wedge \tau_a}^2 - (n \wedge \tau_a)| \leq \max\{a^2, 1\} + \tau_a.$$

Logo, a família parada de $S_n^2 - n$ é dominada por uma variável integrável, e o teorema da parada fornece

$$\mathbf{E}\tau_a = \mathbf{E}S_{\tau_a}^2 = a.$$

Como $\tau_a \leq \tau$, segue que $\mathbf{E}\tau \geq a$ para todo a , e portanto $\mathbf{E}\tau = \infty$.

◁

Com duas barreiras fixas, as condições de integrabilidade voltam a valer. A mesma estratégia de prova determina tanto a probabilidade de saída por cada lado quanto o tempo médio até a saída.

Exemplo 15.24 — Ruína do jogador.

Para inteiros $a, b \geq 1$, seja

$$\tau = \inf\{n \geq 0 : S_n \in \{-a, b\}\}.$$

Como no exemplo anterior, τ tem cauda geométrica e esperança finita. Se $p = \mathbf{P}(S_\tau = b)$, a parada de S_n fornece

$$0 = bp - a(1 - p), \quad \text{logo} \quad p = \frac{a}{a + b}.$$

Como a família parada de $S_n^2 - n$ é dominada por $\max\{a^2, b^2\} + \tau$, uma segunda aplicação do teorema fornece

$$\mathbf{E}\tau = \mathbf{E}S_\tau^2 = b^2 \frac{a}{a + b} + a^2 \frac{b}{a + b} = ab.$$

◁

15.6. Desigualdades e convergência de Doob

A desigualdade maximal controla a probabilidade de uma trajetória atingir um nível até certo instante. Para estudar a convergência, controlaremos também suas oscilações, contando os cruzamentos de um intervalo.

Teorema 15.25 — Desigualdade maximal de Doob

Se $X = (X_k)_{0 \leq k \leq n}$ é uma submartingala não negativa e $\lambda > 0$, então

$$\lambda \mathbf{P}\left(\max_{0 \leq k \leq n} X_k \geq \lambda\right) \leq \mathbf{E}\left[X_n \mathbb{1}_{\{\max_{0 \leq k \leq n} X_k \geq \lambda\}}\right] \leq \mathbf{E}X_n.$$

Em particular, se M é uma martingala,

$$\mathbf{P}\left(\max_{0 \leq k \leq n} |M_k| \geq \lambda\right) \leq \frac{\mathbf{E}|M_n|}{\lambda}.$$

Demonstração. Seja $A = \{\max_{0 \leq k \leq n} X_k \geq \lambda\}$ e seja T o primeiro instante em $\{0, \dots, n\}$ no qual $X_T \geq \lambda$, com a convenção $T = n$ em A^c . O evento A pertence a $\mathcal{F}_T$. Pela amostragem opcional entre T e n ,

$$\lambda \mathbf{P}(A) \leq \mathbf{E}[X_T \mathbf{1}_A] \leq \mathbf{E}[X_n \mathbf{1}_A] \leq \mathbf{E}X_n.$$

Para uma martingala M , aplicamos a primeira parte à submartingala $(|M_k|)$. □

Para formular a desigualdade de cruzamentos, escrevemos $x^- = \max\{-x, 0\}$. Fixados $a < b$, um cruzamento ascendente completo de $[a, b]$ ocorre quando a trajetória visita primeiro $(-\infty, a]$ e, depois, $[b, \infty)$. Seja $U_n[a, b]$ o número máximo de tais cruzamentos disjuntos realizados até o instante n .

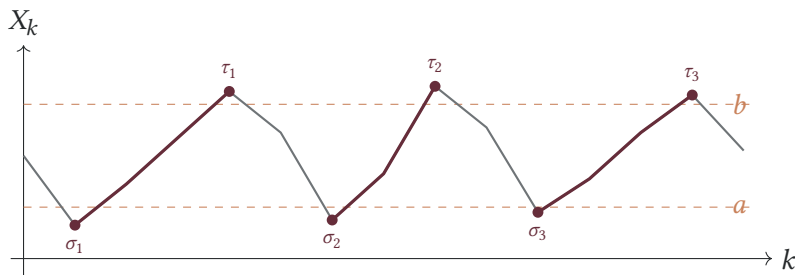


Figura 15.2.: Três cruzamentos ascendentes completos de $[a, b]$, destacados em vinho. Uma nova contagem começa quando a trajetória retorna a $(-\infty, a]$ e termina ao alcançar novamente $[b, \infty)$.

Teorema 15.26 – Desigualdade de cruzamentos de Doob

Se $(X_k)_{0 \leq k \leq n}$ é uma supermartingala, então

$$(b - a) \mathbf{E}U_n[a, b] \leq \mathbf{E}[(X_n - a)^-].$$

Demonstração. Considere a estratégia que compra uma unidade imediatamente depois de a trajetória estar em $(-\infty, a]$ e a conserva até a primeira visita a $[b, \infty)$; então repete a operação. A posição $H_k \in \{0, 1\}$ mantida no passo k é previsível. Cada operação completa produz ao menos $b - a$. Se ao final resta uma posição aberta, sua perda é no máximo $(X_n - a)^-$. Portanto, trajetória por trajetória,

$$(H \cdot X)_n \geq (b - a)U_n[a, b] - (X_n - a)^-.$$

Como X é uma supermartingala e $H \geq 0$, a transformada é uma supermartingala iniciada em zero; logo sua esperança é menor ou igual a zero. Tomando esperanças na desigualdade anterior, obtemos o resultado. $\square$

Se o limite inferior de uma trajetória é menor que seu limite superior, ela atravessa repetidamente algum intervalo entre os dois. A desigualdade anterior permitirá excluir essa oscilação. O controle uniforme das esperanças absolutas também impedirá que o limite seja infinito.

Teorema 15.27 – Convergência de Doob

Se $X = (X_n)_{n \geq 0}$ é uma supermartingala e

$$\sup_{n \geq 0} \mathbf{E}|X_n| < \infty,$$

então existe uma variável aleatória finita X_∞ tal que

$$X_n \longrightarrow X_\infty \quad \text{q.s.}$$

O mesmo resultado vale para submartingalas.

Demonstração. Para racionais $a < b$, a desigualdade de cruzamentos implica

$$\mathbf{E}U_n[a, b] \leq \frac{\mathbf{E}[(X_n - a)^-]}{b - a} \leq \frac{\sup_k \mathbf{E}|X_k| + |a|}{b - a}.$$

Como $U_n[a, b]$ cresce com n , o número total de cruzamentos de $[a, b]$ é finito q.s. Se $\liminf X_n < \limsup X_n$, existem racionais $a < b$ estritamente entre esses dois

limites, e então ocorrem infinitos cruzamentos de $[a, b]$, uma contradição. Logo X_n possui limite no espaço estendido q.s.

Pelo lema de Fatou,

$$\mathbf{E}\left[\liminf_{n \rightarrow \infty} |X_n|\right] \leq \liminf_{n \rightarrow \infty} \mathbf{E}|X_n| < \infty,$$

de modo que o limite é finito q.s. Para submartingalas, aplicamos o resultado a $-X$. □

Corolário 15.28 — Supermartingalas não negativas

Toda supermartingala não negativa converge q.s. para uma variável aleatória finita.

Demonstração. Como $X_n \geq 0$ e $(\mathbf{E}X_n)$ é decrescente, $\sup_n \mathbf{E}|X_n| = \mathbf{E}X_0 < \infty$. □

A convergência quase certa não garante a convergência em L^1 . Para controlar também as esperanças, voltamos à integrabilidade uniforme. O lema seguinte será usado para verificar essa condição em famílias de esperanças condicionais.

Lema 15.29 — Continuidade absoluta da integral

Se $Z \in L^1(\mathbf{P})$, então, para todo $\varepsilon > 0$, existe $\delta > 0$ tal que

$$\mathbf{P}(A) < \delta \implies \mathbf{E}[|Z|1_A] < \varepsilon$$

para todo $A \in \mathcal{F}$.

Demonstração. Escolha $K > 0$ tal que

$$\mathbf{E}[|Z|1_{\{|Z|>K\}}] < \frac{\varepsilon}{2},$$

o que é possível pela Convergência Dominada. Ponha $\delta = \varepsilon/(2K)$. Se $\mathbf{P}(A) < \delta$, então

$$\mathbf{E}[|Z|1_A] \leq K\mathbf{P}(A) + \mathbf{E}[|Z|1_{\{|Z|>K\}}] < \varepsilon.$$

□

Teorema 15.30 – Convergência em L^1

Para uma martingala $M = (M_n)_{n \geq 0}$, são equivalentes:

- 1** a família $\{M_n : n \geq 0\}$ é uniformemente integrável;
- 2** existe $M_\infty \in L^1$ tal que $M_n \rightarrow M_\infty$ q.s. e em L^1 ;
- 3** existe $M_\infty \in L^1$ tal que

$$M_n = \mathbf{E}[M_\infty \mid \mathcal{F}_n] \quad \text{q.s. para todo } n.$$

Demonstração. Suponha primeiro que a família é uniformemente integrável. Em particular, ela é limitada em L^1 . Pelo Teorema 15.27, $M_n \rightarrow M_\infty$ q.s.; o Lema 15.21 fornece a convergência em L^1 .

Se vale a segunda afirmação, então, para $m \geq n$, $M_n = \mathbf{E}[M_m \mid \mathcal{F}_n]$. A contração da esperança condicional em L^1 permite fazer $m \rightarrow \infty$, dando $M_n = \mathbf{E}[M_\infty \mid \mathcal{F}_n]$.

Finalmente, a família de todas as esperanças condicionais de uma variável fixa $Z \in L^1$ é uniformemente integrável. De fato, para $Y = \mathbf{E}[Z \mid \mathcal{G}]$ e $A = \{|Y| > K\}$,

$$\mathbf{P}(A) \leq \frac{\mathbf{E}[|Z|]}{K}.$$

Além disso, por Jensen condicional, $|Y| \leq \mathbf{E}[|Z| \mid \mathcal{G}]$. Como $A \in \mathcal{G}$,

$$\mathbf{E}[|Y| \mathbf{1}_A] \leq \mathbf{E}[\mathbf{E}[|Z| \mid \mathcal{G}] \mathbf{1}_A] = \mathbf{E}[|Z| \mathbf{1}_A].$$

Como a primeira desigualdade torna $\mathbf{P}(A)$ uniformemente pequena quando $K \rightarrow \infty$, o Lema 15.29 conclui a demonstração. □

No exemplo das esperanças condicionais, o crescimento da informação permite identificar o limite: ele é a esperança de Z condicionada a toda a informação reunida pela filtração.

Corolário 15.31 – Teorema ascendente de Lévy

Se $Z \in L^1$, $\mathcal{F}_n \uparrow \mathcal{F}_\infty$ e $\mathcal{F}_\infty = \sigma(\bigcup_n \mathcal{F}_n)$, então

$$\mathbf{E}[Z \mid \mathcal{F}_n] \longrightarrow \mathbf{E}[Z \mid \mathcal{F}_\infty]$$

q.s. e em L^1 .

Demonstração. A martingala $M_n = \mathbf{E}[Z \mid \mathcal{F}_n]$ é uniformemente integrável pelo teorema anterior. Seu limite M_∞ satisfaz $\mathbf{E}[M_\infty \mathbb{1}_A] = \mathbf{E}[Z \mathbb{1}_A]$ para todo $A \in \bigcup_n \mathcal{F}_n$. Um argumento de classe monótona estende a identidade a $A \in \mathcal{F}_\infty$, identificando $M_\infty = \mathbf{E}[Z \mid \mathcal{F}_\infty]$. $\square$

15.7. Decomposição de Doob: o caso discreto de Doob–Meyer

Podemos separar os incrementos de um processo em sua média condicional e no desvio em relação a ela. Somando essas duas parcelas, obtemos uma parte previsível e uma martingala. Para uma submartingala, a parte previsível será crescente.

Definição 15.32 – Previsibilidade em zero

Um processo adaptado $A = (A_n)_{n \geq 0}$ é previsível se A_0 é $\mathcal{F}_0$ -mensurável e A_n é $\mathcal{F}_{n-1}$ -mensurável para $n \geq 1$.

Teorema 15.33 – Decomposição de Doob

Se $X = (X_n)_{n \geq 0}$ é um processo adaptado e integrável, então existe uma única decomposição

$$X_n = M_n + A_n,$$

em que M é uma martingala e A é previsível, integrável e $A_0 = 0$. Explicitamente,

$$A_n = \sum_{k=1}^n \mathbf{E}[X_k - X_{k-1} \mid \mathcal{F}_{k-1}],$$

$$M_n = X_0 + \sum_{k=1}^n (X_k - \mathbf{E}[X_k \mid \mathcal{F}_{k-1}]).$$

Além disso, X é uma submartingala se, e somente se, A é crescente; e X é uma supermartingala se, e somente se, A é decrescente.

Demonstração. A fórmula mostra que $A_0 = 0$ e que $A_n - A_{n-1}$ é $\mathcal{F}_{n-1}$ -mensurável. Como

$$\mathbf{E}[M_n - M_{n-1} \mid \mathcal{F}_{n-1}] = \mathbf{E}[X_n - X_{n-1} \mid \mathcal{F}_{n-1}] - (A_n - A_{n-1}) = 0,$$

M é uma martingala. Se $X = M + A = \tilde{M} + \tilde{A}$ são duas decomposições normalizadas, então $M - \tilde{M} = \tilde{A} - A$ é ao mesmo tempo uma martingala e um processo previsível iniciado em zero. Seus incrementos são $\mathcal{F}_{n-1}$ -mensuráveis e têm esperança condicional zero; portanto são nulos q.s. Isso prova a unicidade.

Por fim,

$$A_n - A_{n-1} = \mathbf{E}[X_n - X_{n-1} \mid \mathcal{F}_{n-1}],$$

que é não negativo exatamente quando X é uma submartingala, e não positivo exatamente quando X é uma supermartingala. $\square$

A decomposição distingue a flutuação de média condicional nula da tendência previsível acumulada. Para uma submartingala, essa tendência é crescente. Ao longo do capítulo, a mesma atenção à informação disponível permitiu definir estratégias, escolher tempos de parada e estudar a convergência. Em cada caso, as hipóteses dizem até onde a ideia de jogo justo pode ser usada.

EXERCÍCIOS

Ex. 15.1 — Seja $(\mathcal{F}_n)_{n \geq 0}$ uma filtração e seja $Z \in L^1$. Prove diretamente, usando a

propriedade da torre, que

$$M_n = \mathbf{E}[Z \mid \mathcal{F}_n], \quad n \geq 0,$$

é uma martingala. Prove também que ela é uniformemente integrável e identifique seu limite, tanto quase certamente quanto em L^1 , tomando $\mathcal{F}_\infty = \sigma(\bigcup_{n \geq 0} \mathcal{F}_n)$.

Ex. 15.2 — Sejam $\xi_1, \xi_2, \dots$ variáveis i.i.d. com $\mathbf{E}[\xi_1] = 0$ e $\text{Var}(\xi_1) = \sigma^2 < \infty$. Para $S_n = \xi_1 + \dots + \xi_n$, adote $S_0 = 0$ e tome $\mathcal{F}_n = \sigma(\xi_1, \dots, \xi_n)$, com $\mathcal{F}_0 = \{\emptyset, \Omega\}$. Prove que

$$(S_n)_{n \geq 0} \quad \text{e} \quad (S_n^2 - n\sigma^2)_{n \geq 0}$$

são martingalas.

Ex. 15.3 — Sejam $\xi_1, \xi_2, \dots$ variáveis i.i.d., defina $S_0 = 0$, $S_n = \xi_1 + \dots + \xi_n$ e tome a filtração natural. Suponha que $m(\theta) = \mathbf{E}[e^{\theta \xi_1}] < \infty$ para algum $\theta \in \mathbb{R}$. Prove que

$$M_n = \frac{e^{\theta S_n}}{m(\theta)^n}, \quad n \geq 0,$$

é uma martingala. Explique o que ocorre quando $m(\theta) = 0$.

Ex. 15.4 — Seja $M = (M_n)$ uma martingala em relação a uma filtração $(\mathcal{F}_n)$ e seja $H = (H_n)_{n \geq 1}$ um processo previsível e limitado em relação à mesma filtração. Defina

$$(H \cdot M)_n = \sum_{k=1}^n H_k(M_k - M_{k-1}).$$

Prove diretamente que $(H \cdot M)_n$ é uma martingala. Interprete o resultado como a impossibilidade de criar ganho médio positivo por uma estratégia de apostas que usa apenas a informação passada.

Ex. 15.5 — Seja (S_n) um passeio aleatório simples iniciado em $S_0 = 0$, munido de sua filtração natural $\mathcal{F}_n = \sigma(S_0, \dots, S_n)$.

1. Prove que para cada inteiro $a \geq 1$, $\tau_a = \inf\{n \geq 0 : S_n \geq a\}$ é um tempo de parada.
2. Fixe N e considere

$$\rho = \sup\{0 \leq n \leq N : S_n = 0\},$$

com $\rho = 0$ se o conjunto estiver vazio. Mostre que, em geral, ρ não é um tempo de parada.

Ex. 15.6 — Ruína do jogador por amostragem opcional. Seja (S_n) um passeio aleatório simples iniciado em 0, e sejam $a, b \geq 1$. Para

$$\tau = \inf\{n \geq 0 : S_n \in \{-a, b\}\},$$

use a amostragem opcional, aplicada primeiro a $\tau \wedge N$, para calcular $\mathbf{P}(S_\tau = b)$. Em seguida, aplique o mesmo procedimento à martingala $S_n^2 - n$ para mostrar que

$$\mathbf{E}[\tau] = ab.$$

Justifique cuidadosamente as hipóteses que permitem cada passagem ao limite. Esse argumento também é frequentemente chamado de *parada opcional*.

Ex. 15.7 — Considere lançamentos independentes de uma moeda justa. Seja $\mathcal{F}_n$ a informação gerada pelos primeiros n lançamentos e defina

$$M_n = 2^n \mathbb{1}_{\{\text{os } n \text{ primeiros lançamentos são caras}\}}.$$

Prove que (M_n) é uma martingala não negativa com $\mathbf{E}[M_n] = 1$ para todo n , mas $M_n \rightarrow 0$ quase certamente. Conclua que a família $\{M_n : n \geq 0\}$ não é uniformemente integrável e que a convergência não ocorre em L^1 .

Ex. 15.8 — Sejam $\xi_1, \xi_2, \dots$ variáveis i.i.d. centradas, com $\text{Var}(\xi_1) = \sigma^2 < \infty$, e defina $S_n = \xi_1 + \dots + \xi_n$ e $\mathcal{F}_n = \sigma(\xi_1, \dots, \xi_n)$. Calcule a decomposição de Doob da submartingala $X_n = S_n^2$. Identifique explicitamente a parte martingala e o compensador previsível crescente.

Ex. 15.9 — Seja $S_n = \xi_1 + \dots + \xi_n$, onde os incrementos são independentes, centrados e têm variância comum $\sigma^2 < \infty$. Use a desigualdade maximal de Doob aplicada à submartingala S_k^2 para provar que, para todo $r > 0$,

$$\mathbf{P}\left(\max_{1 \leq k \leq n} |S_k| \geq r\right) \leq \frac{n\sigma^2}{r^2}.$$

Compare o resultado com a desigualdade maximal de Kolmogorov apresentada no Capítulo 11.

Definição e primeiras propriedades do movimento browniano

Uma trajetória browniana reúne duas propriedades que, à primeira vista, parecem mal acomodadas na mesma função. Ela é contínua: não há saltos. Ao mesmo tempo, oscila em todas as escalas e não admite tangente em ponto algum. A quantidade que concilia esses dois fatos é a escala $\sqrt{h}$: num intervalo de duração h , um incremento típico tem ordem $\sqrt{h}$, muito maior do que a ordem linear h exigida pela diferenciabilidade.

Neste capítulo construímos o processo a partir de variáveis gaussianas independentes e tornamos essa intuição precisa. A construção e as demonstrações centrais seguem a estratégia de Paul Lévy, na forma apresentada por Mörters e Peres: primeiro se define o processo nos pontos diádicos, depois se completa a trajetória por interpolação e convergência uniforme. A mesma decomposição em escalas será

usada para medir a regularidade das trajetórias.

Ao longo do capítulo, $(\Omega, \mathcal{F}, \mathbf{P})$ denota um espaço de probabilidade. Escrevemos $\mathcal{N}(\mu, \sigma^2)$ para a distribuição normal de média μ e variância σ^2 . A Figura 16.1 antecipa o objeto que construiremos: três realizações da mesma lei, contínuas e visivelmente irregulares.

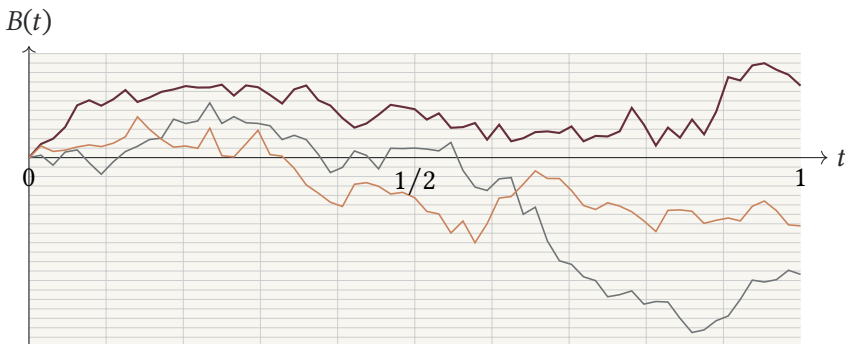


Figura 16.1.: Três aproximações de trajetórias brownianas numa malha diádica. A lei é a mesma; a geometria de cada realização é irrepetível.

16.1. A construção de Paul Lévy

16.1.1. Definição e primeiras consequências

Definição 16.1 — Movimento browniano

Um processo estocástico real $B = (B(t))_{t \geq 0}$ é um **movimento browniano iniciado em $x \in \mathbb{R}$** se:

- 1 $B(0) = x$ quase certamente;
- 2 seus incrementos são independentes: sempre que $0 \leq t_0 < t_1 < \dots < t_n$, as variáveis

$$B(t_1) - B(t_0), \dots, B(t_n) - B(t_{n-1})$$

são independentes;

3 para $0 \leq s < t$,

$$B(t) - B(s) \sim \mathcal{N}(0, t - s);$$

4 quase certamente, a aplicação $t \mapsto B(t)$ é contínua.

Quando $x = 0$, dizemos que B é um **movimento browniano padrão**.

As três primeiras condições descrevem o processo em qualquer coleção finita de tempos. A quarta fala da trajetória inteira e não pode ser descartada. O exemplo seguinte isola essa diferença.

Exemplo 16.2 — As distribuições finito-dimensionais não veem tudo.

Seja B um movimento browniano e seja U , independente de B , uniforme em $[0, 1]$. Defina

$$\tilde{B}(t) = \begin{cases} B(t) + 1, & t = U, \\ B(t), & t \neq U. \end{cases}$$

Para tempos determinísticos $t_1, \dots, t_n$, a probabilidade de U coincidir com algum t_i é zero. Assim, $(\tilde{B}(t_1), \dots, \tilde{B}(t_n))$ tem a mesma distribuição que $(B(t_1), \dots, B(t_n))$. Entretanto, quase toda trajetória de $\tilde{B}$ é descontínua em U . Portanto, conhecer todas as distribuições finito-dimensionais não determina, por si só, a regularidade das trajetórias.

◀

Recordemos dois fatos sobre vetores gaussianos. Um vetor aleatório é gaussiano quando toda combinação linear de suas coordenadas tem distribuição normal. A lei de um vetor gaussiano é determinada por seu vetor de médias e por sua matriz de covariâncias; além disso, coordenadas gaussianas não correlacionadas são independentes. Esses fatos serão usados repetidamente.

Proposição 16.3 — Caracterização gaussiana

Um processo gaussiano centrado $X = (X(t))_{t \geq 0}$ é um movimento browniano padrão, desde que tenha trajetórias contínuas e

$$\text{Cov}(X(s), X(t)) = s \wedge t, \quad s, t \geq 0.$$

Demonstração. Para $0 \leq t_0 < t_1 < \dots < t_n$, o vetor dos incrementos é gaussiano e centrado. Se $i < j$, então

$$\begin{aligned} & \text{Cov}(X(t_i) - X(t_{i-1}), X(t_j) - X(t_{j-1})) \\ &= (t_i \wedge t_j) - (t_i \wedge t_{j-1}) - (t_{i-1} \wedge t_j) + (t_{i-1} \wedge t_{j-1}) = 0. \end{aligned}$$

Logo os incrementos são independentes. A mesma identidade dá

$$\text{Var}(X(t_i) - X(t_{i-1})) = t_i - t_{i-1}.$$

Por fim, $\text{Var}(X(0)) = 0$, de modo que $X(0) = 0$ quase certamente. A continuidade completa as quatro condições da definição. $\square$

Reciprocamente, todo movimento browniano padrão é um processo gaussiano centrado com essa covariância. De fato, qualquer vetor de valores $(B(t_1), \dots, B(t_n))$ é uma transformação linear do vetor de incrementos gaussianos independentes. Além disso, para $s \leq t$,

$$\text{Cov}(B(s), B(t)) = \text{Cov}(B(s), B(s)) = s,$$

pois $B(t) - B(s)$ é independente de $B(s)$. Assim, entre processos contínuos, a função de covariância $s \wedge t$ caracteriza completamente a lei browniana.

16.1.2. Construção nos pontos diádicos

Não é imediato que as quatro exigências da definição sejam compatíveis. A continuidade impõe simultaneamente condições a uma família não enumerável de variáveis aleatórias. A construção de Lévy resolve o problema de maneira concreta: acrescenta-se uma escala diádica por vez e controla-se o erro que ela produz.

Para $n \geq 0$, seja

$$D_n = \left\{ \frac{k}{2^n} : 0 \leq k \leq 2^n \right\}, \quad D = \bigcup_{n=0}^{\infty} D_n.$$

Considere uma família $(Z_d)_{d \in D \setminus \{0\}}$ de variáveis aleatórias independentes, todas com distribuição $\mathcal{N}(0,1)$. Começamos com $B(0) = 0$ e $B(1) = Z_1$. Suponha que os valores de B já tenham sido definidos em D_{n-1} . Para cada ponto novo $d \in D_n \setminus D_{n-1}$, defina

$$B(d) = \frac{B(d - 2^{-n}) + B(d + 2^{-n})}{2} + 2^{-(n+1)/2} Z_d. \quad (16.1)$$

A primeira parcela é o valor dado pela interpolação linear dos dois extremos. A segunda é uma perturbação gaussiana independente, exatamente na escala necessária.

Lema 16.4 — Incrementos na malha diádica

Para cada $n \geq 0$, os incrementos

$$B(k2^{-n}) - B((k-1)2^{-n}), \quad k = 1, \dots, 2^n,$$

são independentes e têm distribuição $\mathcal{N}(0, 2^{-n})$.

Demonstração. A afirmação vale em D_0 . Suponha que valha em D_{n-1} e considere um intervalo $[a, b]$ dessa malha, de comprimento $2^{-(n-1)}$, com ponto médio d . Ponha

$$Y = B(b) - B(a), \quad Z = Z_d, \quad \sigma = 2^{-(n+1)/2}.$$

Pela hipótese de indução, $Y \sim \mathcal{N}(0, 2^{-(n-1)})$; além disso, Y e Z são independentes. A definição (16.1) fornece

$$B(d) - B(a) = \frac{Y}{2} + \sigma Z, \quad B(b) - B(d) = \frac{Y}{2} - \sigma Z.$$

Cada uma dessas variáveis é centrada e tem variância

$$\frac{1}{4} 2^{-(n-1)} + \sigma^2 = 2^{-n}.$$

Sua covariância é

$$\frac{1}{4} 2^{-(n-1)} - \sigma^2 = 0.$$

Como o par é gaussiano, os dois novos incrementos são independentes. Em intervalos-pais distintos, usamos incrementos antigos independentes e variáveis Z_d distintas. Portanto, todos os incrementos da nova malha são independentes, concluindo a indução. $\square$

A figura representa uma única etapa da construção. O novo valor é a interpolação dos extremos acrescida de uma flutuação independente.

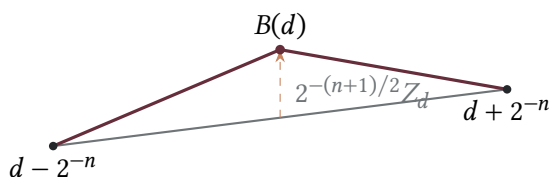


Figura 16.2.: Correção do ponto médio na construção de Lévy.

16.1.3. Da malha diádica à trajetória contínua

Escrevamos a construção como uma série de funções contínuas. Seja $F_0(t) = tZ_1$.

Para $n \geq 1$, defina F_n nos pontos de D_n por

$$F_n(d) = \begin{cases} 2^{-(n+1)/2} Z_d, & d \in D_n \setminus D_{n-1}, \\ 0, & d \in D_{n-1}, \end{cases}$$

e estenda F_n linearmente entre pontos consecutivos de D_n . Essas são funções-tenda aleatórias. Por indução em n , a própria definição nos dá

$$B(d) = \sum_{j=0}^n F_j(d), \quad d \in D_n. \quad (16.2)$$

Teorema 16.5 — Existência de Wiener

Existe um movimento browniano padrão.

Demonstração. Construiremos primeiro o processo em $[0,1]$. Fixe $c > \sqrt{2 \log 2}$. Para uma normal padrão Z e $u > 0$, usaremos a estimativa

$$\mathbf{P}(|Z| \geq u) \leq 2e^{-u^2/2}.$$

Como há no máximo $2^n + 1$ pontos em D_n , segue que

$$\begin{aligned} \sum_{n=1}^{\infty} \mathbf{P} \left(\max_{d \in D_n \setminus D_{n-1}} |Z_d| \geq c\sqrt{n} \right) \\ \leq \sum_{n=1}^{\infty} 2(2^n + 1)e^{-c^2 n/2} < \infty. \end{aligned}$$

Pelo lema de Borel–Cantelli, quase certamente existe um índice aleatório N tal que, para $n \geq N$ e $d \in D_n \setminus D_{n-1}$, $|Z_d| < c\sqrt{n}$. Consequentemente,

$$\|F_n\|_{\infty} \leq c\sqrt{n} 2^{-(n+1)/2}, \quad n \geq N. \quad (16.3)$$

A série numérica no lado direito é convergente. O teste de Weierstrass mostra que, quase certamente, a série

$$B(t) = \sum_{n=0}^{\infty} F_n(t)$$

converge uniformemente em $[0,1]$. Seu limite é contínuo e, por (16.2), coincide com os valores já construídos em D .

Resta verificar as distribuições dos incrementos. Dados $0 \leq t_0 < \dots < t_r \leq 1$, escolha, para cada m , pontos diádicos $t_{0,m} \leq \dots \leq t_{r,m}$ que convergem para os tempos correspondentes. A continuidade implica

$$B(t_i) - B(t_{i-1}) = \lim_{m \rightarrow \infty} (B(t_{i,m}) - B(t_{i-1,m})).$$

Pelo Lema 16.4, para cada m o vetor à direita é gaussiano centrado, com matriz de covariâncias diagonal. As variâncias convergem para $t_i - t_{i-1}$. Pela convergência das

funções características, o vetor limite é gaussiano com coordenadas independentes e

$$B(t_i) - B(t_{i-1}) \sim \mathcal{N}(0, t_i - t_{i-1}).$$

Temos, portanto, um movimento browniano em $[0, 1]$.

Para estender a construção, tome cópias independentes $B^{(0)}, B^{(1)}, \dots$ desse processo. Se $k = \lfloor t \rfloor$, defina

$$W(t) = \sum_{j=0}^{k-1} B^{(j)}(1) + B^{(k)}(t - k),$$

com a soma vazia interpretada como zero. As peças se encontram nos extremos, por isso W é contínuo. Os incrementos que atravessam números inteiros se decompõem em somas de incrementos gaussianos independentes; logo têm a lei e a independência requeridas. Assim, W é um movimento browniano padrão em $[0, \infty)$. □

Observação 16.6 — O que a construção realmente produz

A prova não apenas mostra que o movimento browniano existe. Ela fornece a representação de Lévy–Ciesielski: uma trajetória é uma soma uniforme de funções-tenda independentes, organizadas por escala. Os níveis baixos descrevem o contorno largo; os níveis altos acrescentam oscilações cada vez menores e cada vez mais numerosas.

16.1.4. Invariâncias elementares

A escala $\sqrt{t}$ já está inscrita na variância dos incrementos. A proposição seguinte diz que ampliar o espaço por a exige ampliar o tempo por a^2 .

Proposição 16.7 — Invariância de escala

Se B é um movimento browniano padrão e $a > 0$, então

$$X(t) = \frac{1}{a}B(a^2t), \quad t \geq 0,$$

também é um movimento browniano padrão.

Demonstração. A continuidade e a independência dos incrementos são preservadas pela mudança de escala. Para $0 \leq s < t$,

$$X(t) - X(s) = \frac{1}{a}(B(a^2t) - B(a^2s)) \sim \mathcal{N}(0, t - s).$$

□

Outra simetria troca o comportamento próximo de zero pelo comportamento em tempos grandes.

Teorema 16.8 — Inversão temporal

Se B é um movimento browniano padrão, então o processo

$$X(t) = \begin{cases} 0, & t = 0, \\ tB(1/t), & t > 0, \end{cases}$$

é um movimento browniano padrão.

Demonstração. O processo X é gaussiano e centrado. Para $0 < s \leq t$,

$$\text{Cov}(X(s), X(t)) = st \text{Cov}(B(1/s), B(1/t)) = st \left(\frac{1}{s} \wedge \frac{1}{t} \right) = s.$$

Assim, suas distribuições finito-dimensionais coincidem com as do movimento browniano. Suas trajetórias são contínuas em $(0, \infty)$. Falta apenas o ponto zero.

Restrito aos racionais positivos, X tem a mesma lei que um movimento browniano restrito aos racionais. Como esse último converge a zero ao longo dos ra-

cionais quando $t \downarrow 0$, o mesmo ocorre com X , quase certamente. A continuidade de X em $(0, \infty)$ transforma esse limite racional no limite completo. Portanto, $X(t) \rightarrow 0 = X(0)$ quando $t \downarrow 0$ e a caracterização gaussiana conclui a prova. $\square$

Corolário 16.9 — Lei dos grandes números para o movimento browniano

Quase certamente,

$$\lim_{t \rightarrow \infty} \frac{B(t)}{t} = 0.$$

Demonstração. Pela inversão temporal, $X(u) = uB(1/u)$ é um movimento browniano contínuo em zero. Fazendo $u = 1/t$,

$$\frac{B(t)}{t} = X(1/t) \rightarrow X(0) = 0.$$

$\square$

16.2. Propriedades de continuidade

A definição garante continuidade, mas ainda não diz quão rapidamente uma trajetória pode oscilar. Em $[0, 1]$, considere

$$\omega_B(h) = \sup_{0 \leq t \leq 1-h} |B(t+h) - B(t)|, \quad 0 < h < 1.$$

Uma função $\varphi(h) \downarrow 0$ que domina $\omega_B(h)$ para h pequeno é chamada de módulo de continuidade. A construção por escalas fornece um módulo determinístico, válido para quase toda trajetória.

Teorema 16.10 – Estimativa superior do módulo

Existe uma constante determinística $C > 0$ tal que, quase certamente, para todo $h > 0$ suficientemente pequeno,

$$\omega_B(h) \leq C\sqrt{h \log(1/h)}.$$

Demonstração. Retomemos a série $B = \sum_{n \geq 0} F_n$ da construção de Lévy. Como F_n é linear em cada intervalo de comprimento 2^{-n} e se anula nos extremos de cada tenda,

$$\|F'_n\|_\infty \leq 2^{n+1}\|F_n\|_\infty$$

nos pontos em que a derivada existe. Pela estimativa (16.3), quase certamente existe N tal que, para $n \geq N$,

$$\|F_n\|_\infty \leq c\sqrt{n} 2^{-n/2}, \quad \|F'_n\|_\infty \leq 2c\sqrt{n} 2^{n/2},$$

após aumentar c por uma constante absoluta.

Fixe $t, t+h \in [0,1]$ e escolha ℓ de modo que $2^{-(\ell+1)} < h \leq 2^{-\ell}$. Separando as escalas mais largas e mais finas do que h , obtemos

$$\begin{aligned} |B(t+h) - B(t)| &\leq \sum_{n=0}^{\ell} |F_n(t+h) - F_n(t)| + \sum_{n=\ell+1}^{\infty} |F_n(t+h) - F_n(t)| \\ &\leq h \sum_{n=0}^{\ell} \|F'_n\|_\infty + 2 \sum_{n=\ell+1}^{\infty} \|F_n\|_\infty. \end{aligned}$$

A contribuição dos níveis $n < N$ é $O_\omega(h)$. Nas duas somas restantes, o maior termo está na escala ℓ ; comparações com séries geométricas dão

$$h \sum_{n=N}^{\ell} \sqrt{n} 2^{n/2} + \sum_{n=\ell+1}^{\infty} \sqrt{n} 2^{-n/2} \leq C_0 \sqrt{\ell} 2^{-\ell/2}.$$

Como $2^{-\ell}$ é comparável a h e ℓ é comparável a $\log(1/h)$, o último membro é no máximo $C_1 \sqrt{h \log(1/h)}$. Por fim, $O_\omega(h) = o(\sqrt{h \log(1/h)})$. A constante final pode ser escolhida deterministicamente; apenas o limiar a partir do qual a estimativa vale depende da trajetória. □

A ordem obtida é correta. O próximo resultado mostra que não se pode remover o fator logarítmico nem substituir sua constante por uma arbitrariamente pequena.

Teorema 16.11 – Estimativa inferior do módulo

Se $c < \sqrt{2}$, então, quase certamente, para todo $\varepsilon > 0$ existem $0 < h < \varepsilon$ e $t \in [0, 1 - h]$ tais que

$$|B(t + h) - B(t)| \geq c\sqrt{h \log(1/h)}.$$

Demonstração. Podemos supor $c > 0$. Para inteiros $n \geq 1$ e $0 \leq k < \lfloor e^n \rfloor$, considere

$$A_{k,n} = \{B((k+1)e^{-n}) - B(ke^{-n}) > c\sqrt{n}e^{-n/2}\}.$$

Para n fixo, esses eventos são independentes. Se Z é normal padrão,

$$p_n := \mathbf{P}(A_{k,n}) = \mathbf{P}(Z > c\sqrt{n}).$$

Uma estimativa inferior usual para a cauda gaussiana fornece

$$p_n \geq C \frac{e^{-c^2 n/2}}{1 + \sqrt{n}}$$

com uma constante $C > 0$. Como $c^2/2 < 1$, $e^n p_n \rightarrow \infty$. Logo

$$\mathbf{P}\left(\bigcap_{k=0}^{\lfloor e^n \rfloor - 1} A_{k,n}^c\right) = (1 - p_n)^{\lfloor e^n \rfloor} \leq \exp\{-\lfloor e^n \rfloor p_n\} \rightarrow 0.$$

Fixe $\varepsilon > 0$. O evento em que nenhum incremento com $0 < h < \varepsilon$ ultrapassa a barreira do enunciado está contido, para todo n suficientemente grande, no evento exibido acima. Sua probabilidade é, portanto, zero. Intersectando os eventos de probabilidade um correspondentes a $\varepsilon = 1/m$, $m \geq 1$, obtemos incrementos desse tamanho em escalas arbitrariamente pequenas. $\square$

O valor exato da constante é um dos resultados clássicos de Lévy. Incluímos a prova porque ela também ensina uma técnica recorrente: controlar primeiro uma família esparsa de intervalos e depois aproximar todos os demais.

Teorema 16.12 – Módulo de continuidade de Lévy

Quase certamente,

$$\limsup_{h \downarrow 0} \frac{\omega_B(h)}{\sqrt{2h \log(1/h)}} = 1.$$

Para a estimativa superior, precisaremos de duas observações. Fixe $m \geq 1$ e ponha

$$A_m = \left\{0, \frac{1}{m}, \dots, \frac{m-1}{m}\right\}.$$

Denote por $\Lambda_n(m)$ a família dos intervalos contidos em $[0,1]$ da forma

$$[(k-1+b)2^{-n+a}, (k+b)2^{-n+a}], \quad a, b \in A_m, \quad k \geq 1.$$

Lema 16.13 – Controle na família esparsa

Fixados $m \geq 1$ e $c > \sqrt{2}$, quase certamente existe n_0 tal que, para $n \geq n_0$ e $[s, t] \in \Lambda_n(m)$,

$$|B(t) - B(s)| \leq c \sqrt{(t-s) \log \frac{1}{t-s}}.$$

Demonstração. A família $\Lambda_n(m)$ tem no máximo $C_m 2^n$ intervalos, e o comprimento q de cada um deles pertence a $[2^{-n}, 2^{-n+1})$. Pela estimativa superior da cauda normal,

$$\begin{aligned} \mathbf{P}\left(|B(t) - B(s)| > c \sqrt{q \log(1/q)}\right) \\ = \mathbf{P}\left(|Z| > c \sqrt{\log(1/q)}\right) \leq 2q^{c^2/2} \leq C 2^{-nc^2/2}. \end{aligned}$$

A probabilidade de alguma violação em $\Lambda_n(m)$ é, portanto, no máximo

$$C'_m 2^{n(1-c^2/2)}.$$

Como $c > \sqrt{2}$, essa expressão é somável em n . Borel–Cantelli conclui a prova. $\square$

Lema 16.14 – Aproximação de intervalos

Para todo $\eta > 0$ existe $m \geq 1$ tal que, para qualquer intervalo $[s, t] \subset [0, 1]$, existe $[s_0, t_0] \in \bigcup_n \Lambda_n(m)$ com

$$|s - s_0| \leq \eta(t - s), \quad |t - t_0| \leq \eta(t - s).$$

Demonstração. Escolha m tão grande que $m^{-1} + 2^{1/m} - 1 < \eta$. Ponha $\ell = t - s$ e escolha n com $2^{-n} \leq \ell < 2^{-n+1}$. Existe $a \in A_m$ tal que

$$q := 2^{-n+a} \leq \ell < 2^{1/m} q.$$

Os números $(k - 1 + b)q$, com $k \in \mathbb{N}$ e $b \in A_m$, formam uma malha de passo q/m . Tome s_0 como o maior desses números que não excede s e ponha $t_0 = s_0 + q$. Então $[s_0, t_0]$ pertence à família indicada, está contido em $[0, 1]$ e satisfaz $t_0 \leq t$. Além disso,

$$0 \leq s - s_0 < \frac{q}{m} \leq \frac{\ell}{m}$$

e

$$0 \leq t - t_0 = (s - s_0) + (\ell - q) \leq \left(\frac{1}{m} + 2^{1/m} - 1 \right) \ell < \eta \ell.$$

□

Demonstração do Teorema 16.12. O Teorema 16.11, com $c \uparrow \sqrt{2}$, dá

$$\limsup_{h \downarrow 0} \frac{\omega_B(h)}{\sqrt{2h \log(1/h)}} \geq 1.$$

Provemos a desigualdade oposta. Fixe $c > \sqrt{2}$ e escolha $c_0 \in (\sqrt{2}, c)$. Tome $\eta > 0$ pequeno e depois m como no Lema 16.14. Para um intervalo pequeno $[s, t] \subset [0, 1]$, escolha a aproximação $[s_0, t_0]$. Pela desigualdade triangular,

$$|B(t) - B(s)| \leq |B(t) - B(t_0)| + |B(t_0) - B(s_0)| + |B(s_0) - B(s)|.$$

O termo central é controlado pelo Lema 16.13, com constante c_0 . Os dois termos das extremidades são controlados pelo Teorema 16.10. Se $\ell = t - s$ e $\delta \leq \eta \ell$, então

$$\frac{\sqrt{\delta \log(1/\delta)}}{\sqrt{\ell \log(1/\ell)}} \leq \sqrt{\eta \left(1 + \frac{\log(1/\eta)}{\log(1/\ell)} \right)}.$$

Além disso, $(t_0 - s_0)/\ell \rightarrow 1$ uniformemente quando $\eta \downarrow 0$. Primeiro escolhendo η suficientemente pequeno e depois ℓ suficientemente pequeno, obtemos

$$|B(t) - B(s)| \leq c\sqrt{\ell \log(1/\ell)}.$$

Como $c > \sqrt{2}$ era arbitrário, o limsup normalizado por

$$\sqrt{2h \log(1/h)}$$

é no máximo um. □

Definição 16.15 – Continuidade de Hölder

Uma função $f : [0, \infty) \rightarrow \mathbb{R}$ é **localmente Hölder de expoente $\alpha > 0$ em t** se existem $C, \delta > 0$ tais que

$$|f(s) - f(t)| \leq C|s - t|^\alpha$$

sempre que $s \geq 0$ e $|s - t| < \delta$.

Corolário 16.16 – Regularidade abaixo de $1/2$

Para todo $0 < \alpha < 1/2$, quase certamente a trajetória browniana é localmente Hölder de expoente α em todos os tempos.

Demonstração. Para h pequeno,

$$\sqrt{h \log(1/h)} = o(h^\alpha), \quad \alpha < \frac{1}{2}.$$

O Teorema 16.10 fornece, portanto, o controle Hölder uniforme em $[0, 1]$. Aplicando-o aos processos transladados $B(k+t) - B(k)$, $0 \leq t \leq 1$, obtemos o mesmo controle em cada bloco $[k, k+1]$. Se dois tempos próximos estiverem em blocos consecutivos, inserimos o inteiro comum entre eles e usamos a desigualdade triangular; a constante é no máximo a soma das constantes dos dois blocos. Assim, o controle vale em todo compacto. A interseção sobre $k \in \mathbb{N}$ ainda tem probabilidade um. □

O expoente crítico é $1/2$. A trajetória é Hölder para todo expoente menor do que ele e, como veremos a seguir, não pode ser localmente Hölder com expoente maior do que $1/2$ em ponto algum. No próprio expoente crítico há tempos excepcionais, mas eles são raros; sua análise exige ferramentas mais finas.

16.3. Não diferenciabilidade

A continuidade browniana não deve ser confundida com suavidade. Um primeiro sinal da irregularidade é simples: nenhuma trajetória típica permanece monótona durante um intervalo, por menor que ele seja. A Figura 16.3 sugere o fenômeno que as provas tornarão preciso: ampliar a curva revela novas oscilações, não uma tangente.



Figura 16.3.: Ampliações sucessivas de uma mesma aproximação browniana, com os eixos reescalados. A trajetória muda de escala, mas não se torna lisa.

Proposição 16.17 — Ausência de intervalos monótonos

Quase certamente, não existe intervalo não degenerado no qual a trajetória browniana seja monótona.

Demonstração. Fixe $a < b$ e divida $[a, b]$ em n subintervalos de mesmo comprimento. Se a trajetória for crescente em $[a, b]$, os n incrementos da partição serão não negativos; se for decrescente, serão não positivos. Como esses incrementos são independentes, simétricos e não se anulam,

$$\mathbf{P}(B \text{ é monótono em } [a, b]) \leq 2^{1-n}.$$

Fazendo $n \rightarrow \infty$, essa probabilidade é zero. Agora tome a união sobre os intervalos com extremos racionais. Todo intervalo não degenerado contém um subintervalo racional não degenerado, e monotonicidade passa a subintervalos. $\square$

Para estudar derivadas, usaremos a inversão temporal. Precisamos antes de uma forma elementar de oscilação em tempos grandes.

Lema 16.18 — Lei zero-um de Hewitt–Savage

Se $(X_n)_{n \geq 1}$ é uma sequência de variáveis aleatórias independentes e identicamente distribuídas, todo evento $A \in \sigma(X_1, X_2, \dots)$ invariante por permutações finitas das coordenadas satisfaz $\mathbf{P}(A) \in \{0, 1\}$.

Demonstração. É exatamente a Lei zero-um de Hewitt–Savage demonstrada no Teorema 8.40, aplicada à sequência (X_n) . $\square$

Proposição 16.19 — Oscilações em tempos inteiros

Quase certamente,

$$\limsup_{n \rightarrow \infty} \frac{B(n)}{\sqrt{n}} = +\infty, \quad \liminf_{n \rightarrow \infty} \frac{B(n)}{\sqrt{n}} = -\infty.$$

Demonstração. Fixe $c > 0$ e ponha

$$A_c = \{B(n) > c\sqrt{n} \text{ para infinitos } n\}.$$

Pela desigualdade de Fatou e pela invariância de escala,

$$\mathbf{P}(A_c) \geq \limsup_{n \rightarrow \infty} \mathbf{P}(B(n) > c\sqrt{n}) = \mathbf{P}(B(1) > c) > 0.$$

Escreva $X_j = B(j) - B(j-1)$. A sequência (X_j) é i.i.d., e uma permutação finita altera as somas parciais apenas em um número finito de índices. Logo A_c é permutável. Pelo Lema 16.18, $\mathbf{P}(A_c) = 1$. Intersectando esses eventos para $c \in \mathbb{N}$, obtemos a primeira igualdade. A segunda segue aplicando o mesmo argumento a $-B$. $\square$

Para uma função f , definimos as derivadas superiores e inferiores à direita no sentido de Dini por

$$D^+f(t) = \limsup_{h \downarrow 0} \frac{f(t+h) - f(t)}{h}, \quad D_+f(t) = \liminf_{h \downarrow 0} \frac{f(t+h) - f(t)}{h}.$$

Teorema 16.20 – Um tempo fixo

Fixe $t \geq 0$. Quase certamente,

$$D^+B(t) = +\infty, \quad D_+B(t) = -\infty.$$

Em particular, B não é diferenciável em t .

Demonstração. Começemos em $t = 0$. Pelo Teorema 16.8, $X(h) = hB(1/h)$ é um movimento browniano. Ao longo de $h = 1/n$,

$$\frac{X(1/n) - X(0)}{1/n} = B(n).$$

A Proposição 16.19 afirma que $B(n)/\sqrt{n}$ é ilimitado tanto acima quanto abaixo; em particular, o limsup dos quocientes exibidos é $+\infty$ e o liminf é $-\infty$. O resultado vale, portanto, no zero para o movimento browniano X . Como todo movimento browniano padrão tem a mesma lei no espaço de trajetórias contínuas, a mesma afirmação vale no zero para B .

Para $t > 0$, o processo

$$X(s) = B(t + s) - B(t), \quad s \geq 0,$$

é novamente um movimento browniano padrão. Os quocientes diferenciais de X em zero são exatamente os quocientes à direita de B em t . $\square$

Observação 16.21 — A ordem dos quantificadores

O teorema anterior afirma que, para cada t fixado, a probabilidade de diferenciabilidade em t é zero. Isso ainda não prova que, com probabilidade um, não há nenhum tempo de diferenciabilidade: a união não enumerável dos conjuntos excepcionais pode ter probabilidade positiva.

Ainda assim, Fubini já fornece uma conclusão intermediária. Em qualquer intervalo $[0, T]$, se $D(B)$ é o conjunto dos tempos em que a trajetória é diferenciável, então

$$\mathbf{E}[\text{Leb}(D(B))] = \int_0^T \mathbf{P}(B \text{ é diferenciável em } t) dt = 0.$$

Logo $D(B)$ tem medida de Lebesgue zero quase certamente. O próximo teorema elimina também os possíveis tempos excepcionais.

Teorema 16.22 — Paley–Wiener–Zygmund

Quase certamente, a trajetória do movimento browniano não é diferenciável em ponto algum. Mais precisamente, quase certamente, para todo $t \geq 0$,

$$D^+B(t) = +\infty \quad \text{ou} \quad D_+B(t) = -\infty,$$

possivelmente ambas as alternativas.

Demonstração. Basta trabalhar em $[0,1]$ e depois tomar uma interseção enumerável de eventos de probabilidade um. Suponha que exista $t_0 \in [0,1]$ para o qual

$$-\infty < D_+ B(t_0) \leq D^+ B(t_0) < +\infty.$$

Pela continuidade e pela limitação da trajetória em $[0,2]$, existe um inteiro $M \geq 1$ tal que

$$|B(t_0 + h) - B(t_0)| \leq Mh, \quad 0 \leq h \leq 1. \quad (16.4)$$

Fixe n e seja k o índice da célula diádica de ordem n que contém t_0 :

$$t_0 \in [(k-1)2^{-n}, k2^{-n}].$$

Para $j = 1, 2, 3$, a desigualdade triangular e (16.4) dão

$$|B((k+j)2^{-n}) - B((k+j-1)2^{-n})| \leq M(2j+1)2^{-n} \leq 7M2^{-n}.$$

Denote por $\Omega_{n,k}$ o evento em que essas três desigualdades ocorrem. Os três incrementos são independentes e cada um tem a distribuição de $2^{-n/2}Z$, com Z normal padrão. Como a densidade normal é limitada,

$$\mathbf{P}(\Omega_{n,k}) \leq C(7M)^3 2^{-3n/2}.$$

Há no máximo $2^n + 1$ escolhas relevantes de k . Logo

$$\mathbf{P}\left(\bigcup_k \Omega_{n,k}\right) \leq C_M 2^{-n/2},$$

e essa série é convergente. Pelo lema de Borel–Cantelli, quase certamente apenas um número finito desses eventos ocorre. Isso contradiz a existência de t_0 satisfazendo (16.4), pois sua célula produziria um evento $\Omega_{n,k}$ para todo n suficientemente grande.

Tomando a união enumerável sobre M , concluímos que não existe t_0 com as duas derivadas de Dini finitas. Se $D^+ B(t)$ não for $+\infty$ e $D_+ B(t)$ não for $-\infty$, ambas serão finitas, caso já excluído. Isso prova a afirmação mais forte e, em particular, a não diferenciabilidade em todos os pontos. $\square$

Corolário 16.23 — O limiar Hölder é exato

Se $\alpha > 1/2$, então, quase certamente, a trajetória browniana não é localmente Hölder de expoente α em ponto algum.

Demonstração. Na prova do Teorema 16.22, substitua a hipótese linear por

$$|B(t_0 + h) - B(t_0)| \leq Mh^\alpha.$$

Escolha um inteiro r tal que $r(\alpha - 1/2) > 1$ e examine os r incrementos diádicos imediatamente à direita da célula que contém t_0 . Cada um fica limitado por $C_r M 2^{-\alpha n}$. Depois de o normalizarmos por seu desvio padrão $2^{-n/2}$, a estimativa de pequena bola para a normal dá

$$\mathbf{P}(\text{os } r \text{ incrementos são pequenos}) \leq C_{M,r} 2^{-nr(\alpha-1/2)}.$$

A união sobre as $O(2^n)$ células é limitada por $C_{M,r} 2^{-n(r(\alpha-1/2)-1)}$, uma série convergente. Borel–Cantelli e a união enumerável sobre M concluem a prova. $\square$

16.3.1. Variação quadrática e variação total

A ausência de derivada tem uma assinatura ainda mais concreta. A variação total da trajetória é infinita, mas a soma dos quadrados de seus incrementos converge para o tempo transcorrido.

Definição 16.24 — Variação total

Para uma função $f : [0, T] \rightarrow \mathbb{R}$, sua variação total em $[0, T]$ é

$$V_f(T) = \sup_{\pi} \sum_{j=1}^k |f(t_j) - f(t_{j-1})|,$$

onde o supremo percorre todas as partições $\pi : 0 = t_0 < t_1 < \dots < t_k = T$.

Lema 16.25 — Uma simetria quadrática

Se X e Z são independentes, simétricas e pertencem a $\mathcal{L}^2$, então

$$\mathbf{E}[(X + Z)^2 \mid X^2 + Z^2] = X^2 + Z^2.$$

Demonstração. A transformação $(X, Z) \mapsto (X, -Z)$ preserva a lei conjunta e mantém $X^2 + Z^2$ inalterado. Assim, para toda função mensurável limitada g ,

$$\mathbf{E}[XZg(X^2 + Z^2)] = -\mathbf{E}[XZg(X^2 + Z^2)],$$

de onde $\mathbf{E}[XZ \mid X^2 + Z^2] = 0$. Basta agora expandir $(X + Z)^2$. □

O lema anterior condiciona apenas na soma quadrática produzida por uma divisão. Para a variação quadrática, precisaremos saber que a mesma simetria continua válida depois de observar todas as subdivisões posteriores.

Lema 16.26 — Simetria preservada pelos refinamentos

Suponha que cada partição π_j seja obtida de π_{j-1} acrescentando um único ponto e defina

$$Q_j = \sum_{[a,b] \in \pi_j} (B(b) - B(a))^2, \quad \mathcal{G}_j = \sigma(Q_j, Q_{j+1}, \dots).$$

Então

$$\mathbf{E}(Q_{j-1} \mid \mathcal{G}_j) = Q_j.$$

Demonstração. Se o novo ponto s divide $[u, v] \in \pi_{j-1}$, escreva

$$X = B(s) - B(u), \quad Z = B(v) - B(s).$$

Os demais termos não mudam, e portanto

$$Q_{j-1} = Q_j + 2XZ.$$

Resta provar que $\mathbf{E}(XZ \mid \mathcal{G}_j) = 0$.

Para $N \geq j$, ponha $\mathcal{H}_{j,N} = \sigma(Q_j, \dots, Q_N)$. Considere o vetor dos incrementos brownianos determinados pela partição π_N . Trocar o sinal de todos os incrementos contidos em $[u, s]$ preserva a lei conjunta desse vetor, pois suas coordenadas são gaussianas centradas e independentes. A transformação troca X por $-X$, mantém Z e preserva cada soma quadrática $Q_j, \dots, Q_N$: como todas essas partições contêm s , nenhum de seus intervalos cruza o ponto onde os sinais foram trocados. Mais precisamente, cada incremento de uma dessas partições contido em $[u, s]$ muda apenas de sinal, enquanto os demais permanecem inalterados; portanto, todos os quadrados que formam $Q_j, \dots, Q_N$ são preservados. Assim, para toda função mensurável limitada h ,

$$\mathbf{E}[XZ h(Q_j, \dots, Q_N)] = -\mathbf{E}[XZ h(Q_j, \dots, Q_N)],$$

e $\mathbf{E}(XZ \mid \mathcal{H}_{j,N}) = 0$.

As σ -álgebras $\mathcal{H}_{j,N}$ são crescentes e $\sigma(\bigcup_{N \geq j} \mathcal{H}_{j,N}) = \mathcal{G}_j$. A convergência das esperanças condicionais em L^1 fornece, portanto, $\mathbf{E}(XZ \mid \mathcal{G}_j) = 0$. A identidade desejada segue da expansão de Q_{j-1} . $\square$

Teorema 16.27 — Variação quadrática

Se (π_n) é uma sequência encaixada de partições determinísticas de $[0, T]$, com malha $|\pi_n| \rightarrow 0$, então, quase certamente,

$$\sum_{[u,v] \in \pi_n} (B(v) - B(u))^2 \rightarrow T.$$

Consequentemente, a trajetória browniana tem variação total infinita em todo intervalo não degenerado.

Demonstração. Acrescentar partições intermediárias não altera o problema. Podemos, pois, supor que cada passo insere um único ponto. Defina

$$Q_n = \sum_{[u,v] \in \pi_n} (B(v) - B(u))^2, \quad \mathcal{G}_n = \sigma(Q_n, Q_{n+1}, \dots).$$

A sequência $(\mathcal{G}_n)$ é decrescente. Pelo Lema 16.26,

$$\mathbf{E}(Q_{n-1} \mid \mathcal{G}_n) = Q_n.$$

Portanto, $(Q_n, \mathcal{G}_n)$ é um martingal reverso. Recordemos por que o teorema de convergência se aplica: para cada N , a sequência finita $Q_N, Q_{N-1}, \dots, Q_0$, com as σ -álgebras na ordem inversa, é uma martingala usual. Aplicando a desigualdade de cruzamentos a ela e a seu oposto, obtemos, uniformemente em N , cotas para os cruzamentos ascendentes e descendentes de qualquer intervalo racional pela sequência $Q_0, \dots, Q_N$. Essas cotas dependem apenas de $\sup_n \mathbb{E}|Q_n| = T$ e, portanto, não crescem com o horizonte N . Fazendo $N \rightarrow \infty$, o mesmo argumento usado na prova do Teorema de Convergência de Doob mostra que (Q_n) possui limite quase certamente.

Se os comprimentos dos intervalos de π_n são $\Delta t_1, \dots, \Delta t_{k(n)}$, então os incrementos são independentes e

$$\mathbb{E}Q_n = \sum_j \Delta t_j = T, \quad \text{Var}(Q_n) = 2 \sum_j (\Delta t_j)^2 \leq 2T|\pi_n| \rightarrow 0.$$

Logo $Q_n \rightarrow T$ em $\mathcal{L}^2$. O limite quase certo do martingal reverso só pode ser T .

Fixe agora $0 < \alpha < 1/2$. Pelo Corolário 16.16, quase certamente existe uma constante aleatória C_T tal que, para n grande,

$$\max_{[u,v] \in \pi_n} |B(v) - B(u)| \leq C_T |\pi_n|^\alpha.$$

Consequentemente,

$$\begin{aligned} Q_n &\leq \left(\max_{[u,v] \in \pi_n} |B(v) - B(u)| \right) \sum_{[u,v] \in \pi_n} |B(v) - B(u)| \\ &\leq C_T |\pi_n|^\alpha \sum_{[u,v] \in \pi_n} |B(v) - B(u)|. \end{aligned}$$

Como $Q_n \rightarrow T > 0$ e $|\pi_n|^\alpha \rightarrow 0$, as somas de variações ao longo de π_n tendem a infinito. Portanto, $V_B(T) = \infty$.

Finalmente, aplique o argumento, com uma sequência fixa de partições, a cada intervalo $[a, b]$ com extremos racionais. Há apenas uma quantidade enumerável desses intervalos, de modo que, quase certamente, todos eles têm variação infinita simultaneamente. Todo intervalo não degenerado contém um subintervalo racional não degenerado; logo a variação total é infinita em todo intervalo não degenerado. □

Esses resultados dão uma boa descrição local das trajetórias. Em intervalos de comprimento h , a oscilação máxima é da ordem de $\sqrt{h \log(1/h)}$; isso garante continuidade, mas está longe do comportamento linear que a diferenciabilidade exigiria. Ao mesmo tempo, a variação total é infinita e a variação quadrática coincide com o tempo. É esse contraste que estará por trás do cálculo estocástico.

EXERCÍCIOS

Ex. 16.1 — Seja B um movimento browniano padrão.

1. Prove que $\mathbf{E}[B(s)B(t)] = s \wedge t$ para $s, t \geq 0$.
2. Para $0 \leq t_1 < \dots < t_n$, determine a matriz de covariâncias do vetor $(B(t_1), \dots, B(t_n))$.
3. Use essa matriz e a gaussianidade para recuperar a independência dos incrementos em intervalos disjuntos.

Ex. 16.2 — Na construção de Lévy, considere um intervalo diádico $[a, b]$ de comprimento $2^{-(n-1)}$ e seu ponto médio d . Supondo conhecidos $B(a)$ e $B(b)$ e tomando $Z_d \sim N(0, 1)$ independente desses valores, mostre que a escolha

$$B(d) = \frac{B(a) + B(b)}{2} + 2^{-(n+1)/2} Z_d$$

é exatamente a distribuição condicional do valor browniano no ponto médio. Calcule a média e a variância condicionais.

Ex. 16.3 — Com a notação das funções-tenda F_n da construção de Lévy, prove que, para todo $c > \sqrt{2 \log 2}$,

$$\|F_n\|_\infty \leq c\sqrt{n} 2^{-(n+1)/2}$$

para todo n suficientemente grande, quase certamente. Conclua que $\sum_n F_n$ converge uniformemente em $[0, 1]$.

Ex. 16.4 — Seja B um movimento browniano padrão e fixe $a > 0$.

1. Verifique diretamente, pelas distribuições finito-dimensionais, que

$$\tilde{B}(t) = a^{-1} B(a^2 t)$$

é novamente um movimento browniano padrão.

2. Para $t > 0$, defina $\widehat{B}(t) = tB(1/t)$ e $\widehat{B}(0) = 0$. Calcule sua covariância e explique por que a continuidade em 0 é o único ponto adicional a ser verificado na inversão temporal.

Ex. 16.5 — Ponte browniana. Para $0 \leq t \leq 1$, defina

$$\beta(t) = B(t) - tB(1).$$

1. Mostre que β é um processo gaussiano centrado e calcule

$$\text{Cov}(\beta(s), \beta(t)).$$

2. Prove que $\beta(0) = \beta(1) = 0$ e explique por que β não é um movimento browniano.
3. Mostre que β e $B(1)$ são independentes.

Ex. 16.6 — Fixe $0 < \alpha < 1/2$. Use o módulo de continuidade de Lévy para provar que, em todo intervalo compacto, uma trajetória browniana é Hölder-contínua de ordem α quase certamente. Mostre também por que o mesmo argumento falha no expoente crítico $\alpha = 1/2$.

Ex. 16.7 — Fixe $T > 0$. Para a partição uniforme

$$\pi_n = \left\{0, \frac{T}{n}, \dots, T\right\},$$

defina

$$Q_n = \sum_{k=1}^n \left(B\left(\frac{kT}{n}\right) - B\left(\frac{(k-1)T}{n}\right) \right)^2.$$

Calcule $\mathbf{E}[Q_n]$ e $\text{Var}(Q_n)$ e conclua que $Q_n \rightarrow T$ em L^2 . Compare essa conclusão com o teorema quase certo para partições encaixadas.

Ex. 16.8 — Prove que toda função contínua de variação total finita possui variação quadrática nula ao longo de qualquer sequência de partições cuja malha tende a zero. Use esse fato e a variação quadrática browniana para obter novamente que as trajetórias de B têm variação total infinita em todo intervalo não degenerado.

Ex. 16.9 — Para cada tempo determinístico t , o teorema de não diferenciabilidade em tempo fixo fornece

$$\mathbf{P}(B \text{ é diferenciável em } t) = 0.$$

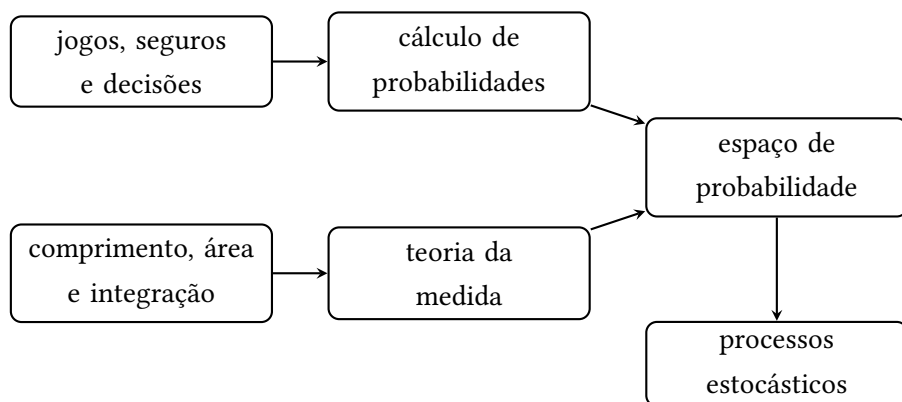
Explique por que não se pode concluir a não diferenciabilidade simultânea em todos os tempos tomando simplesmente uma união não enumerável de eventos nulos. Identifique o argumento adicional do teorema de Paley–Wiener–Zygmund que supera essa dificuldade.

A

História da medida e da probabilidade

A probabilidade nasceu ao menos duas vezes. Na primeira, no século XVII, apareceu como uma aritmética do futuro: como dividir uma aposta interrompida, quanto vale uma chance, que preço é justo pagar por um risco? Na segunda, no início do século XX, reapareceu como uma teoria de medida: antes de calcular a probabilidade de um evento, tornou-se necessário dizer com precisão o que são o espaço de resultados, os eventos e a própria probabilidade.

Entre esses dois nascimentos, a pergunta mudou. De início, queria-se saber *quanto*. Mais tarde, foi preciso perguntar *o que pode ser medido e quais operações preservam essa medida*. A passagem de uma pergunta à outra não apagou o cálculo clássico; deu-lhe uma linguagem capaz de tratar limites, variáveis contínuas e trajetórias inteiras.



Este capítulo não pretende oferecer uma história exaustiva, nem transformar o desenvolvimento da área numa procissão de nomes e datas. Seu fio condutor é a formação da linguagem usada neste livro. A primeira versão deste texto nasceu de uma tradução de um ensaio de Maurice Sion sobre a história da teoria da medida no século XX. A presente reescrita conserva essa dívida e amplia o foco para a história da probabilidade; para panoramas mais extensos, veja [34, 35].

A.1. O primeiro nascimento: calcular o acaso

Jogos de azar, seguros, anuidades e decisões sob incerteza são muito anteriores à teoria matemática da probabilidade. Durante séculos, porém, os cálculos permaneceram locais: uma contagem para este jogo, uma regra prática para aquele contrato. Em meados do século XVI, Girolamo Cardano já comparava casos favoráveis e casos possíveis em problemas com dados. Seu tratado sobre jogos foi publicado apenas postumamente, em 1663, e não chegou a organizar uma teoria geral.

Em 1654, a correspondência entre Blaise Pascal e Pierre de Fermat inaugurou a probabilidade matemática. Um dos problemas discutidos era o **problema dos pontos**: se uma partida é interrompida antes de haver vencedor, como dividir honestamente o prêmio? A pontuação já obtida não basta. É preciso enumerar os futuros possíveis e pesar o valor presente de cada jogador pelas chances de vitória que ainda possui.

Essa mudança de perspectiva é pequena apenas em aparência. Pascal e Fermat

transformaram uma disputa sobre dinheiro numa pergunta acerca da estrutura do futuro. Seus métodos eram diferentes — Pascal explorava uma recorrência, Fermat enumerava continuações possíveis — mas conduziam à mesma divisão. Em 1657, Christiaan Huygens publicou o primeiro tratado impresso dedicado ao cálculo de probabilidades e colocou a esperança matemática no centro do assunto.

Ao mesmo tempo, outra tradição tomava forma fora das mesas de jogo. Em 1662, John Graunt analisou registros de mortalidade de Londres, procurando regularidades em dados coletivos. De um lado havia a combinatória dos casos possíveis; de outro, frequências observadas em populações reais. Grande parte da história posterior da probabilidade pode ser lida como a tentativa de explicar por que essas duas linguagens conversam.

A.2. Frequências, erros e leis-limite

O primeiro resultado geral a aproximar probabilidade teórica e frequência observada apareceu na obra de Jakob Bernoulli. Publicado postumamente em 1713, o *Ars Conjectandi* contém uma forma da Lei dos Grandes Números. Na notação atual, se $X_1, X_2, \dots$ são ensaios de Bernoulli independentes, com $\mathbf{P}(X_k = 1) = p$, e $S_n = X_1 + \dots + X_n$, então, para todo $\epsilon > 0$,

$$\mathbf{P}\left(\left|\frac{S_n}{n} - p\right| > \epsilon\right) \longrightarrow 0.$$

A afirmação não define p como uma frequência observada. Ela diz algo mais delicado: partindo de um modelo com parâmetro p , as frequências tendem a se concentrar perto desse parâmetro. A teoria passava a produzir regularidade a partir da repetição do acaso.

No século XVIII, Abraham de Moivre descobriu que a distribuição binomial, depois de centralizada e reescalada, podia ser aproximada pela curva normal. Laplace ampliou essa ideia e reuniu, em sua *Théorie analytique des probabilités*, de 1812, técnicas combinatórias, funções geradoras, inferência e aproximações assintóticas. A normal deixou de ser apenas uma fórmula ligada a um problema específico e começou a revelar uma forma universal das flutuações.

Durante o século XIX, Poisson, Chebyshev e Markov, entre outros, deram às leis dos grandes números hipóteses e demonstrações cada vez mais gerais. No início do século XX, Lyapunov formulou uma versão ampla do Teorema do Limite Central. Essa linhagem chega diretamente aos capítulos de leis-limite deste livro. Média, variância, independência e normalização já formavam uma teoria poderosa antes da axiomatização moderna.

Mas o sucesso dos cálculos escondia uma fragilidade. O que significava, em geral, atribuir probabilidade a um subconjunto infinito de resultados? Como integrar uma variável aleatória sem supor que ela possuísse uma densidade bem comportada? E o que justificava trocar limite e esperança nas demonstrações? As respostas exigiriam uma teoria que, à primeira vista, não havia sido criada para falar de acaso.

A.3. Medir o que não é um intervalo

A noção de medida nasceu de problemas de comprimento, área, volume e integração. No cálculo de Newton e Leibniz, a integral ligava a área sob um gráfico à operação inversa da diferenciação. No século XIX, as séries de Fourier e o estudo da equação do calor forçaram os analistas a considerar funções e convergências mais irregulares do que as admitidas pelo cálculo elementar.

A integral de Riemann organizou de modo rigoroso uma parte importante dessa teoria. Ainda assim, sequências de funções integráveis podiam convergir para funções problemáticas, limites não podiam ser trocados livremente com integrais, e conjuntos definidos analiticamente escapavam da geometria intuitiva de intervalos e retângulos. Não faltavam apenas técnicas de cálculo; faltava uma noção mais flexível de tamanho.

Em 1898, Émile Borel estendeu o comprimento dos intervalos a uma classe muito maior de subconjuntos da reta. A novidade decisiva era a aditividade enumerável: se $A_1, A_2, \dots$ são disjuntos, então

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n).$$

A família obtida a partir dos abertos por complementos e uniões enumeráveis é hoje chamada de σ -álgebra de Borel. É importante separar os dois nomes: uma σ -álgebra é qualquer família de conjuntos fechada por essas operações; ela é de Borel quando é gerada pelos abertos de um espaço topológico.

Henri Lebesgue deu o passo seguinte. Em sua tese de 1902, construiu uma noção de medida mais ampla e uma integral adaptada a ela. Em vez de controlar uma função apenas por partições do eixo horizontal, sua teoria permitia organizar o domínio segundo os valores assumidos pela função. Conjuntos de medida zero passaram a registrar exceções que podiam ser ignoradas pela integral sem serem apagadas da matemática.

A nova linguagem produziu teoremas de convergência com hipóteses simples e fortes conclusões. O Teorema da Convergência Dominada, em sua forma moderna, afirma que, se $f_n \rightarrow f$ quase sempre e existe $g \in L^1$ tal que $|f_n| \leq g$ para todo n , então

$$\int f_n \, d\mu \rightarrow \int f \, d\mu.$$

A função dominante não é um detalhe técnico dispensável: é justamente o que impede que massa relevante fuja para regiões cada vez menores ou mais distantes. A teoria tornou precisas as condições sob as quais as operações que os analistas desejavam fazer eram de fato legítimas.

Nos anos seguintes, Carathéodory abstraiu a construção por medida exterior. Radon relacionou medidas e funcionais lineares, enquanto Daniell mostrou que a integral também podia ser tomada como ponto de partida. O teorema de Radon–Nikodým permitiu representar uma medida em relação a outra por meio de uma densidade. O objeto já não precisava viver apenas na reta: podia ser definido em espaços abstratos, exatamente onde a probabilidade logo precisaria dele.

A.4. O segundo nascimento: probabilidade como medida

Na virada do século XX, problemas probabilísticos já exigiam espaços muito maiores do que conjuntos finitos de resultados. Em 1900, Louis Bachelier usou um modelo contínuo para descrever flutuações de preços. Poucos anos depois, Einstein

e Smoluchowski explicaram estatisticamente o movimento errático de partículas suspensas num fluido, observado por Robert Brown no século XIX. Markov, por sua vez, iniciou o estudo de sequências de variáveis dependentes que hoje levam seu nome.

Esses exemplos mudavam a natureza do resultado aleatório. O resultado de um lançamento de dado é um número; o resultado de um movimento ao longo do tempo é uma trajetória inteira. Dizer “a partícula permanece nesta região até o instante t ” significa atribuir probabilidade a um conjunto de funções. A geometria elementar dos casos favoráveis já não alcançava esse espaço.

Em 1923, Norbert Wiener construiu rigorosamente uma medida no espaço das funções contínuas para descrever o movimento browniano. A medida de Wiener mostrou que a teoria de Lebesgue podia sustentar probabilidades em dimensão infinita. Não se tratava apenas de calcular a distribuição da posição em um instante: a própria curva aleatória tornava-se o objeto observado.

Essa construção não surgiu isolada. Borel já havia usado ideias de medida em probabilidade; os trabalhos de Fréchet e Daniell haviam ampliado o cenário abstrato; e o sexto problema de Hilbert pedira uma axiomatização matemática das teorias físicas, mencionando explicitamente o cálculo de probabilidades. Ao início da década de 1930, muitas peças estavam disponíveis. Faltava reuni-las num sistema pequeno o bastante para servir de fundamento e amplo o bastante para acomodar os exemplos.

A.5. A síntese de Kolmogorov

Em 1933, Andrei Kolmogorov publicou os *Grundbegriffe der Wahrscheinlichkeitsrechnung*. O livro não inventou a teoria da medida nem apagou as tradições anteriores da probabilidade. Sua força foi reconhecer que elas se encaixavam. Na formulação moderna, um modelo probabilístico é uma trinca

$$(\Omega, \mathcal{F}, \mathbf{P}),$$

em que Ω é o espaço de resultados, $\mathcal{F}$ é uma σ -álgebra de eventos e $\mathbf{P}$ é uma medida de probabilidade, isto é,

$$\mathbf{P}(\Omega) = 1 \qquad \text{e} \qquad \mathbf{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathbf{P}(A_n)$$

para toda sequência de eventos dois a dois disjuntos.

O vocabulário da probabilidade passou então a ser uma leitura particular do vocabulário da medida:

teoria da medida	probabilidade
espaço mensurável $(\Omega, \mathcal{F})$	experimento e seus eventos
medida finita com massa total 1	probabilidade $\mathbf{P}$
função mensurável	variável aleatória
imagem de uma medida	distribuição de uma variável
integral $\int X \, d\mathbf{P}$	esperança $\mathbf{E}[X]$
medida produto	experimentos independentes
derivada de Radon–Nikodým	densidade e esperança condicional

A aditividade enumerável permitia passar de eventos finitos para eventos definidos por limites. A integração de Lebesgue permitia tratar variáveis sem densidade, esperanças infinitas e convergências quase certas. Os teoremas de extensão permitiam construir uma lei em espaços de sequências ou trajetórias a partir de distribuições finito-dimensionais compatíveis.

Os axiomas, entretanto, não dizem por que um modelo descreve bem um fenômeno, nem resolvem sozinhos o significado filosófico da palavra “probabilidade”. Eles separam duas tarefas que antes apareciam misturadas: construir uma teoria matemática coerente e justificar sua relação com a experiência. Essa separação foi uma virtude, não uma fuga. Depois dela, modelos frequentistas, bayesianos, físicos ou combinatórios puderam compartilhar a mesma gramática matemática.

A.6. Processos, condicionamento e martingalas

Com os fundamentos disponíveis, a probabilidade deixou de ser apenas uma teoria de distribuições isoladas e passou a estudar sistemas que evoluem. Um **processo**

estocástico é uma família $(X_t)_{t \in T}$ de variáveis aleatórias definida num mesmo espaço de probabilidade. Dependendo da pergunta, podemos observá-lo por suas distribuições finito-dimensionais, por suas trajetórias ou pela informação acumulada ao longo do tempo.

Paul Lévy desenvolveu profundamente as leis estáveis, o movimento browniano e os processos com incrementos independentes. Jean Ville e Joseph Doob deram à ideia de jogo justo sua forma moderna por meio das martingalas. Uma filtração $(\mathcal{F}_t)$ registra a informação disponível até o instante t , e a condição

$$\mathbf{E}[X_t \mid \mathcal{F}_s] = X_s, \quad s \leq t,$$

diz que, conhecendo o passado, não existe ganho médio previsível. O antigo vocabulário das apostas retornou, agora dentro da teoria abstrata que ele ajudara a iniciar.

Também a esperança condicional ganhou uma definição que não depende de densidades ou partições finitas: ela pode ser construída como uma derivada de Radon–Nikodým. Tempos de parada formalizaram a decisão de observar um processo até que algo aconteça. O cálculo de Itô, criado na década de 1940, tornou possível integrar em relação ao movimento browniano e abriu um novo ramo da análise.

O problema dos pontos e os tempos de parada tratam, em épocas e linguagens muito diferentes, de uma questão aparentada: o que se pode dizer quando um jogo ou processo é interrompido antes de terminar? Não há uma linha direta entre as duas teorias, mas a comparação ajuda a reconhecer a persistência de certas perguntas probabilísticas.

A.7. A linguagem que ficou

A ordem adotada neste livro não é a ordem histórica. Aqui começamos por σ -álgebras e medidas, passamos pela integral e pelas variáveis aleatórias e só depois chegamos aos teoremas-limite, ao condicionamento e às martingalas. Na história, muitas dessas ideias apareceram na direção contrária: cálculos e problemas concretos vieram antes da linguagem que hoje usamos para justificá-los.

A teoria da medida não foi apenas uma mudança de nomes. Ela tornou possíveis perguntas que antes sequer cabiam no vocabulário: probabilidades de conjuntos de trajetórias, convergência de leis, condicionamento por informação e comportamento assintótico de processos inteiros. Também não eliminou a intuição dos jogos, das frequências ou dos fenômenos físicos; deu-lhe uma casa comum.

O acaso não ficou menos acaso. Tornou-se um objeto sobre o qual se podia integrar.

Respostas dos Exercícios

1.1 Como Ω pertence às duas álgebras, pertence à interseção. Se $A \in \mathcal{A}_1 \cap \mathcal{A}_2$, então A^c pertence a ambas; se A, B pertencem à interseção, o mesmo vale para $A \cup B$. Portanto a interseção satisfaz os três axiomas de álgebra.

1.2 Temos $\Omega = (0,1] \in \mathcal{A}$. Para uma coleção finita de intervalos, ordene todos os seus extremos. As células determinadas por esses extremos são intervalos semiabertos disjuntos; tanto uma união quanto o complemento em $(0,1]$ são uniões de algumas dessas células. Isso prova o fechamento por união finita e complemento.

1.3 $\emptyset$ é finito. O complemento troca conjuntos finitos por cofinitos. Para uniões, finito $\cup$ finito é finito; se ao menos um dos conjuntos é cofinito, o complemento da união está contido num conjunto finito e, portanto, é finito.

1.4 Escolha pontos distintos $x_1, x_2, \dots$ em Ω e ponha $A_n = \{x_2, x_4, \dots, x_{2n}\}$. Cada A_n é finito. A união $A = \{x_{2k} : k \geq 1\}$ é infinita e seu complemento contém todos os x_{2k-1} , logo também é infinito. Assim $A \notin \mathcal{A}$.

1.5 $\emptyset$ é enumerável e o complemento troca as duas alternativas da definição. Se $A = \bigcup_n A_n$ e algum A_n é coenumerável, então $A^c \subseteq A_n^c$ é enumerável. Caso contrário, todos os A_n são enumeráveis e sua união enumerável também o é.

1.6 Se $A_n \uparrow A$, todo ponto de A pertence a algum A_N e, pela monotonicidade, a todos os A_n com $n \geq N$; pontos fora de A não pertencem a nenhum termo. Logo os limites inferior e superior são A . Para $A_n \downarrow A$, um ponto de A pertence a todos os termos, enquanto um ponto fora de A fica fora a partir de algum índice. A mesma conclusão segue.

1.7 Se C_n é a união dos 2^n intervalos que restam na etapa n , então $C = \bigcap_n C_n$; cada C_n é fechado, logo C é fechado e boreliano. Para $(\varepsilon_n) \in \{0,1\}^{\mathbb{N}}$, defina $\Phi(\varepsilon) = \sum_{n \geq 1} 2\varepsilon_n 3^{-n}$. A expansão usa apenas os algarismos ternários 0 e 2, portanto $\Phi(\varepsilon) \in C$. No primeiro índice em que duas sequências diferem, o termo principal domina a cauda; assim Φ é injetiva. Como $\{0,1\}^{\mathbb{N}}$ é não enumerável, C também é.

1.8 $\mathcal{G}$ é uma σ -álgebra que contém $\mathcal{C}$. Por definição, $\sigma\langle\mathcal{C}\rangle$ é a interseção de todas as σ -álgebras que contêm $\mathcal{C}$; em particular, está contida em $\mathcal{G}$. É exatamente nessa última inclusão que se usa a minimalidade.

1.9 Em $[0,1]^2$, declare $x \sim y$ quando $x - y \in \mathbb{Q}^2$ e escolha um representante de cada classe, formando V . Os conjuntos $V+q$, $q \in \mathbb{Q}^2$, são dois a dois disjuntos. Para $q \in \mathbb{Q}^2 \cap [-1,1]^2$, sua união contém $[0,1]^2$ e está contida em $[-1,2]^2$. Se V fosse mensurável, invariância daria a todos os transladados medida $\mu(V)$. Se $\mu(V) = 0$, a união teria medida zero; se $\mu(V) > 0$, teria medida infinita. Ambos contradizem $0 < \mu([0,1]^2) \leq \mu([-1,2]^2) < \infty$, pois o quadrado maior é coberto por finitos transladados do quadrado unitário.

1.10 Como $\sigma\langle\mathcal{F}_0\rangle$ é uma classe monótona contendo $\mathcal{F}_0$, a minimalidade dá $\mathcal{M} \subseteq \sigma\langle\mathcal{F}_0\rangle$. A classe $\{A : A^c \in \mathcal{M}\}$ é monótona e contém $\mathcal{F}_0$, logo contém $\mathcal{M}$. Fixe então A e aplique duas vezes o mesmo artifício à classe dos B para os quais $A \cup B \in \mathcal{M}$: primeiro para $A \in \mathcal{F}_0$, depois para $A \in \mathcal{M}$. Obtemos fechamento por uniões finitas. Pelo Teorema 1.22, $\mathcal{M}$ é uma σ -álgebra e contém $\sigma\langle\mathcal{F}_0\rangle$; as duas inclusões dão a igualdade.

1.11 De (λ_1) e (λ'_2) segue $A^c = \Omega \setminus A \in \mathcal{L}$; junto com (λ_3) , isso é a definição de λ -sistema. Reciprocamente, se $A \subseteq B$, então $B \setminus A = (B^c \sqcup A)^c$, usando união disjunta e complementos.

Num λ -sistema, (λ_4) é o caso finito de (λ_3) . Para $A_n \uparrow A$, escreva $A = A_1 \sqcup \bigsqcup_{n \geq 2} (A_n \setminus A_{n-1})$; complementos dão o caso decrescente. Se valem $(\lambda_1), (\lambda'_2), (\lambda_6)$, obtemos complementos e uniões finitas disjuntas; por complementação, também limites crescentes, e as uniões parciais fornecem (λ_3) . Por fim, se $\mathcal{L} \neq \emptyset$ satisfaz (λ_2) e (λ_3) , escolha $A \in \mathcal{L}$: de $A \sqcup A^c = \Omega$ segue (λ_1) .

1.12 Cada elemento de $\mathcal{A}$ pertence a $\mathcal{C} \subseteq \mathcal{F} \subseteq \mathcal{U}$. Fixe $A \in \mathcal{A}$: então $A \cap A^c = \emptyset \in \mathcal{F}$ e $A \cup A^c = \Omega \in \mathcal{U}$. Distribuindo interseções sobre uniões, $\mathcal{U}$ é fechada por interseções finitas (o fecho por uniões finitas é imediato da definição); por De Morgan, o complemento de uma união de interseções de literais é novamente uma união finita de interseções de literais. Logo $\mathcal{U}$ é uma álgebra contendo $\mathcal{A}$, de modo

que $f\langle\mathcal{A}\rangle \subseteq \mathcal{U}$. A inclusão oposta vale porque toda álgebra que contém $\mathcal{A}$ contém os complementos e todas as interseções e uniões finitas usadas na definição de $\mathcal{U}$.

1.13 Uma combinação booleana finita envolve apenas finitos extremos. Ordenando-os, eles particionam $(0,1]$ em células semiabertas da forma $(c_{j-1}, c_j]$. Em cada célula, todos os intervalos geradores têm função indicadora constante; logo o conjunto é a união das células que satisfazem a combinação. Elas são disjuntas, e células adjacentes podem ainda ser fundidas.

1.14 Um espaço adequado é a união disjunta

$$\Omega = \bigsqcup_{n \geq 1} (\{1, \dots, 5\}^{n-1} \times \{6\}) \sqcup \{1, \dots, 5\}^{\mathbb{N}}.$$

A primeira parcela reúne as trajetórias finitas: antes do último símbolo só podem aparecer as faces $1, \dots, 5$, e o último símbolo é o primeiro 6. A segunda parcela reúne precisamente as trajetórias infinitas sem nenhuma ocorrência de 6.

1.15 No modelo de Maxwell–Boltzmann, uma configuração é uma função de $\{1, \dots, k\}$ em $\{1, \dots, n\}$; há n^k configurações. No modelo de Bose–Einstein, ela é um vetor $(x_1, \dots, x_n) \in \mathbb{N}_0^n$ com $\sum_i x_i = k$; estrelas e barras dá $\binom{n+k-1}{k}$. No modelo de Fermi–Dirac, exige-se ainda $x_i \in \{0,1\}$; há $\binom{n}{k}$ configurações se $k \leq n$ e nenhuma se $k > n$.

1.16 Se B é o conjunto das bolas, os resultados ordenados são as injeções de $\{1, \dots, k\}$ em B , em número $N!/(N-k)!$. Sem a ordem, são os subconjuntos de B com k elementos, em número $\binom{N}{k}$. Registrando apenas as cores, o espaço é

$$\left\{ (x_1, \dots, x_r) \in \mathbb{N}_0^r : \sum_i x_i = k, 0 \leq x_i \leq n_i \right\}.$$

A fibra do vetor x contém $\prod_i \binom{n_i}{x_i}$ amostras não ordenadas; portanto

$$\mathbf{P}(X_1 = x_1, \dots, X_r = x_r) = \frac{\prod_i \binom{n_i}{x_i}}{\binom{N}{k}}.$$

2.1 Há 5^{n-1} palavras de comprimento n em W , logo

$$\mathbf{P}(\Omega) = \sum_{n \geq 1} 5^{n-1} 6^{-n} = 1.$$

A aditividade enumerável segue da reordenação de uma série de termos não negativos. Além disso,

$$\mathbf{P}(T = n) = 5^{n-1}6^{-n} = \left(\frac{5}{6}\right)^{n-1} \frac{1}{6}, \quad \mathbf{P}(T > n) = \left(\frac{5}{6}\right)^n.$$

O conjunto $\{T = \infty\} = \Omega \setminus W$ não contém palavras finitas e, portanto, tem probabilidade zero.

2.2 Se E é mensurável, aplique o critério de Carathéodory a X e use $\mu^*(X) = \mu_0(X)$. Para a volta, introduza a medida interior $\mu_*(E) = \mu_0(X) - \mu^*(E^c)$. A igualdade dada diz $\mu_*(E) = \mu^*(E)$. Pela construção da medida exterior a partir da álgebra, escolha conjuntos mensuráveis $G_n \supseteq E$ com $\mu^*(G_n) \downarrow \mu^*(E)$ e ponha $G = \bigcap_n G_n$. Então $\mu^*(G \setminus E) = \mu^*(G) - \mu_*(E) = 0$. Logo E difere de um conjunto Carathéodory-mensurável por um conjunto de medida exterior nula e também é mensurável.

2.3 Se P é pré-medida, monotonicidade dá que F é crescente; continuidade da medida para $(-\infty, x + 1/n] \downarrow (-\infty, x]$ dá continuidade à direita, e os limites seguem das exaustões de $\mathbb{R}$ pelas semirretas. Reciprocamente, em intervalos ponha $P((a, b]) = F(b) - F(a)$, com as fórmulas análogas para intervalos ilimitados. As condições sobre F garantem não negatividade, massa total um e continuidade no vazio; após disjuntar uma união finita de intervalos, isso prova a aditividade enumerável sempre que a união ainda pertence à álgebra. Assim P é pré-medida. Carathéodory fornece a extensão, única porque P é finita e $\mathcal{A}_0$ gera $\mathcal{B}(\mathbb{R})$.

2.4 Disjunte a união, pondo $B_1 = A_1$ e $B_n = A_n \setminus \bigcup_{k < n} A_k$. Então $B_n \subseteq A_n$ e $\bigcup_n B_n = \bigcup_n A_n$. Pela aditividade enumerável e monotonicidade, $\mathbf{P}(\bigcup_n A_n) = \sum_n \mathbf{P}(B_n) \leq \sum_n \mathbf{P}(A_n)$.

2.5 No caso crescente, disjunte como no exercício anterior e use aditividade enumerável. No caso decrescente, aplique o primeiro resultado aos complementos e use $\mathbf{P}(A_n) = 1 - \mathbf{P}(A_n^c)$. Para o item (c), ponha $B_n = \bigcap_{k \geq n} A_k$ e $C_n = \bigcup_{k \geq n} A_k$, de modo que $B_n \uparrow \liminf A_n$ e $C_n \downarrow \limsup A_n$. De $B_n \subseteq A_n \subseteq C_n$ e dos dois itens

anteriores segue

$$\begin{aligned}\mathbf{P}(\liminf A_n) &= \lim_n \mathbf{P}(B_n) \leq \liminf_n \mathbf{P}(A_n), \\ \limsup_n \mathbf{P}(A_n) &\leq \lim_n \mathbf{P}(C_n) = \mathbf{P}(\limsup A_n).\end{aligned}$$

Se os dois limites de conjuntos coincidem com A , as três quantidades intermediárias ficam espremidas entre $\mathbf{P}(A)$ e $\mathbf{P}(A)$.

2.6 Provaremos a contrapositiva. Suponha que $\sum_n \mathbf{P}(A_n) < \infty$ e ponha

$$B_N = \bigcup_{n \geq N} A_n.$$

Pela subaditividade,

$$\mathbf{P}(B_N) \leq \sum_{n \geq N} \mathbf{P}(A_n) \longrightarrow 0.$$

Como $B_N \downarrow \limsup_n A_n = \{A_n \text{ i.v.}\}$, a continuidade por cima fornece

$$\mathbf{P}(A_n \text{ i.v.}) = \mathbf{P}(\limsup A_n) = \lim_{N \rightarrow \infty} \mathbf{P}(B_N) = 0.$$

Logo, se $\mathbf{P}(A_n \text{ i.v.}) = 1$, a série das probabilidades não pode convergir.

2.7 Os conjuntos $[0, n]$ crescem para $[0, \infty)$, logo a primeira igualdade é a continuidade por baixo. Já $[n, \infty)$ decresce para o vazio; como a probabilidade é finita, a continuidade por cima dá limite zero.

2.8 As interseções e diferenças de $\{1, 2\}$ e $\{1, 3\}$ isolam os quatro pontos, logo a σ -álgebra gerada é $\mathcal{P}(\Omega)$. Tome P_1 com massas $(1/4, 1/4, 1/4, 1/4)$ e P_2 com massas $(1/10, 2/5, 2/5, 1/10)$. Ambas atribuem $1/2$ a cada conjunto de $\mathcal{A}$, mas, por exemplo, $P_1(\{1\}) = 1/4 \neq 1/10 = P_2(\{1\})$.

2.9 Para $J \subseteq \{1, \dots, n\}$, seja E_J o evento em que ocorrem exatamente os A_i com $i \in J$. Os E_J formam uma partição finita e

$$S_j = \sum_{J \subseteq \{1, \dots, n\}} \binom{|J|}{j} \mathbf{P}(E_J).$$

Para $N \geq 1$,

$$\sum_{j=1}^k (-1)^{j-1} \binom{N}{j} = 1 - (-1)^k \binom{N-1}{k},$$

interpretando o coeficiente binomial como zero quando $N \leq k$. Portanto, a soma à esquerda é maior ou igual a 1 para k ímpar e menor ou igual a 1 para k par. Somando sobre a partição e observando que $\bigcup_i A_i = \bigsqcup_{J \neq \emptyset} E_J$, obtemos as duas desigualdades.

2.10 Se A_i é o evento de o pai i receber seu filho, então a interseção de quaisquer k desses eventos tem probabilidade $(n-k)!/n!$. Inclusão-exclusão dá

$$\mathbf{P}\left(\bigcup_i A_i\right) = \sum_{k=1}^n (-1)^{k-1} \binom{n}{k} \frac{(n-k)!}{n!} = \sum_{k=1}^n \frac{(-1)^{k-1}}{k!}.$$

A série exponencial mostra que o limite é $1 - e^{-1}$.

2.11 Se r é o inraio e $\theta \in [-\pi/3, \pi/3]$ é o ângulo em torno da normal à face mais próxima, a condição radial é $0 \leq \rho < r/(1 + \cos \theta)$. Por simetria, a área favorável é

$$\frac{3}{2} r^2 \int_{-\pi/3}^{\pi/3} \frac{d\theta}{(1 + \cos \theta)^2}.$$

Como a integral vale $10/(9\sqrt{3})$ e a área do triângulo vale $3\sqrt{3}r^2$, a probabilidade é $5/27$.

2.12 Por simetria, tome como espaço amostral $\Omega = (0, d/2) \times (0, \pi/2)$, com a medida uniforme de densidade $4/(\pi d)$. As coordenadas X e Θ representam, respectivamente, a distância do centro à reta mais próxima e o ângulo agudo com as retas. Há interseção exatamente quando $X \leq (d/2) \sin \Theta$. Logo

$$\mathbf{P}(\text{interseção}) = \frac{4}{\pi d} \int_0^{\pi/2} \frac{d}{2} \sin \theta \, d\theta = \frac{2}{\pi}.$$

O complemento tem probabilidade $1 - 2/\pi$.

2.13 Para uma palavra w de comprimento k , seja $N_w(x, n)$ o número de suas ocorrências entre as primeiras n posições. Cada evento que fixa $N_w(x, n)$ depende de

finitos dígitos. O conjunto dos normais na base b é

$$\bigcap_{k \geq 1} \bigcap_{w \in \{0, \dots, b-1\}^k} \bigcap_{m \geq 1} \bigcup_{N \geq 1} \bigcap_{n \geq N} \left\{ x : \left| \frac{N_w(x, n)}{n} - b^{-k} \right| < \frac{1}{m} \right\}.$$

Logo é boreliano, assim como seu complemento. Os absolutamente normais formam a interseção, sobre $b \geq 2$, desses conjuntos borelianos.

2.14 Singletons são fechados; portanto todo conjunto enumerável é uma união enumerável de fechados, e todo coenumerável é o complemento de tal união. Em $\mathbb{R}$, um conjunto perfeito é fechado por definição, logo boreliano. Finalmente, $(0,1] \setminus \mathbb{Q}$ é coenumerável em $(0,1]$.

2.15 Toda bola aberta é aberta, logo a σ -álgebra que elas geram está em $\mathcal{B}(\Omega)$. Se D é enumerável e denso, cada aberto U é a união das bolas $B(q,r)$ com $q \in D$, $r \in \mathbb{Q}_{>0}$ e $B(q,r) \subseteq U$; a união é enumerável. Assim todo aberto pertence à σ -álgebra das bolas. Em $\mathbb{R}$, entre $a < b$ existe um racional: escolha n com $n(b-a) > 1$ e um inteiro entre na e nb . Logo $\mathbb{Q}$ é denso e a conclusão se aplica.

2.16 No caso finito, ponha $\rho = \mu_1 + \mu_2$. Todo conjunto da σ -álgebra gerada pode ser aproximado em diferença simétrica, para ρ , por elementos de $\mathcal{F}_0$ (a classe dos conjuntos aproximáveis é uma classe monótona que contém a álgebra). Se $F_n \in \mathcal{F}_0$ e $\rho(F_n \triangle A) \rightarrow 0$, então $\mu_i(F_n) \rightarrow \mu_i(A)$; passe ao limite na desigualdade. No caso σ -finito, faça isso nas partes de uma exaustão comum por conjuntos de $\mathcal{F}_0$ de medida finita e depois use convergência monótona.

Sem σ -finitude, em $\mathbb{R}$ tome a álgebra das uniões finitas de intervalos semiabertos $(a,b]$, admitindo intervalos ilimitados, e as medidas de contagem em $\mathbb{R}$ e em $\mathbb{R} \setminus \mathbb{Q}$. Elas valem 0 no vazio e ∞ em todo elemento não vazio da álgebra, mas em $\{0\}$ a primeira vale 1 e a segunda, 0.

Para o segundo caso, em $\Omega = \{1,2,3,4\}$ use o π -sistema $\{\Omega, \{1\}, \{1,2\}, \{1,3\}\}$, que gera $\mathcal{P}(\Omega)$. Se μ_1 e μ_2 têm, respectivamente, as massas

$$(1/4, 1/4, 1/4, 1/4) \quad \text{e} \quad (7/20, 1/5, 1/5, 1/4),$$

então $\mu_1 \leq \mu_2$ no π -sistema, mas $\mu_1(\{2\}) > \mu_2(\{2\})$.

2.17 Escolha uma exaustão por conjuntos da álgebra nos quais μ é finita e disjunta. Em cada peça, a caracterização de Carathéodory fornece uma cobertura de B por elementos de $\mathcal{F}_0$ com excesso menor que $\varepsilon 2^{-k}$; disjuntando todos os termos obtemos A_n e, por aditividade, excesso total no máximo ε .

Se $\mu(B) < \infty$, escolha N tal que $\mu(B \setminus \bigcup_{n \leq N} A_n) < \varepsilon/2$ e faça a primeira construção com excesso $\varepsilon/2$. Então $A = \bigcup_{n \leq N} A_n \in \mathcal{F}_0$ satisfaz $\mu(A \triangle B) \leq \varepsilon$.

Para mostrar a necessidade da finitude, use Lebesgue em $\mathbb{R}$ e a álgebra das uniões finitas de intervalos semiabertos, admitindo intervalos ilimitados. Para $B = \bigcup_{k \in \mathbb{Z}} [2k, 2k+1)$, tanto B quanto B^c ocupam comprimento infinito nas caudas, enquanto todo elemento da álgebra é eventualmente constante em cada cauda. Logo $\lambda(A \triangle B) = \infty$ para todo $A \in \mathcal{F}_0$.

2.18 Para toda escolha finita de índices em I , agrupe-os pelas células I_k ; isso dá “ $\leq$ ”. Para “ $\geq$ ”, escolha finitas células e, dentro de cada uma, uma soma finita arbitrariamente próxima do respectivo supremo. Se $I = \{i_1, i_2, \dots\}$, as somas parciais são crescentes e majoram toda soma finita a partir de algum índice, provando a segunda fórmula. Finalmente, se $A = \bigsqcup_n A_n$, aplique a primeira identidade à partição de A pelos A_n : $\mu(A) = \sum_n \sum_{a \in A_n} f(a) = \sum_n \mu(A_n)$. Também $\mu(\emptyset) = 0$, logo μ é medida.

2.19 Seja $\mathcal{D}$ a classe dos conjuntos $\bigcap_n U_n$, com cada U_n uma união finita de elementos de $\mathcal{K}$. A distributividade mostra que $\mathcal{D}$ contém $\mathcal{K}$ e é fechada pelas duas operações; logo $\widehat{\mathcal{K}} = \mathcal{D}$.

Dada uma sequência em $\mathcal{D}$ com a propriedade da interseção finita, enumere todas as uniões finitas de uma representação de seus termos. As escolhas de um elemento de $\mathcal{K}$ em cada união que mantêm a propriedade da interseção finita formam uma árvore finitamente ramificada com nós em todos os níveis. Para justificar o passo usado do lema de König, parta da raiz e escolha sucessivamente um filho que tenha descendentes em níveis arbitrariamente altos; tal filho existe porque há apenas finitos filhos. As escolhas formam um ramo infinito. A compacidade de $\mathcal{K}$ dá um ponto comum aos conjuntos desse ramo e, portanto, a todos os conjuntos da sequência original.

2.20 Se $\pi_n : \Omega \rightarrow \{0,1\}^n$ é a projeção, então

$$\mathcal{F}_n = \{\pi_n^{-1}(B) : B \subseteq \{0,1\}^n\}.$$

Imagens inversas preservam o espaço total, complementos e uniões enumeráveis. Como $\mathcal{P}(\{0,1\}^n)$ é uma σ -álgebra, $\mathcal{F}_n$ também é.

2.21 As famílias são crescentes: $\mathcal{F}_n \subseteq \mathcal{F}_{n+1}$. O espaço total pertence a todas elas e $A \in \mathcal{F}_m$ implica $A^c \in \mathcal{F}_m$. Se $A \in \mathcal{F}_m$ e $B \in \mathcal{F}_n$, ambos pertencem a $\mathcal{F}_{\max\{m,n\}}$, que também contém $A \cup B$. Isso dá uma álgebra. Numa união enumerável, porém, os níveis necessários podem ser ilimitados, e não há em geral um único $\mathcal{F}_n$ que contenha todos os termos.

2.22 A não enumerabilidade de Ω torna as duas alternativas disjuntas. Complementos apenas as trocam. Numa união enumerável, se todos os termos são enumeráveis, a união é enumerável; se algum é coenumerável, o complemento da união é enumerável. Logo $\mathcal{F}$ é uma σ -álgebra.

Para conjuntos disjuntos (A_j) , no máximo um pode ser coenumerável. Se nenhum o é, tanto $\mathbf{P}(\bigcup_j A_j)$ quanto $\sum_j \mathbf{P}(A_j)$ valem zero. Se A_{j_0} é coenumerável, todos os outros estão contidos em seu complemento e são enumeráveis; ambos os lados valem um. Assim $\mathbf{P}$ é enumeravelmente aditiva e $\mathbf{P}(\Omega) = 1$.

3.1 Cada evento tem probabilidade $1/3$. Além disso, $A \cap B = \{1\}$, $A \cap C = \{2\}$ e $B \cap C = \{4\}$, todos de probabilidade $1/9 = (1/3)^2$. Porém $A \cap B \cap C = \emptyset$, cuja probabilidade é diferente de $(1/3)^3$.

3.2 No espaço equiprovável $\Omega = \{1, \dots, 8\}$, tome $A = B = \{1, 2, 3, 4\}$ e $C = \{1, 5, 6, 7\}$. Os três eventos têm probabilidade $1/2$ e $A \cap B \cap C = \{1\}$, logo $\mathbf{P}(A \cap B \cap C) = 1/8 = \mathbf{P}(A)\mathbf{P}(B)\mathbf{P}(C)$. Entretanto, $\mathbf{P}(A \cap B) = 1/2 \neq 1/4 = \mathbf{P}(A)\mathbf{P}(B)$.

3.3 Para subconjuntos disjuntos de índices I, J , inclusão-exclusão dá

$$\mathbf{P}\left(\bigcap_{i \in I} A_i \cap \bigcap_{j \in J} A_j^c\right) = \sum_{K \subseteq J} (-1)^{|K|} \mathbf{P}\left(\bigcap_{i \in I \cup K} A_i\right) = \prod_{i \in I} \mathbf{P}(A_i) \prod_{j \in J} (1 - \mathbf{P}(A_j)).$$

Essa é exatamente a condição de independência para qualquer escolha dos B_i .

3.4 Tome $\Omega = \{x \in \{0,1\}^k : x_1 + \dots + x_k = 0 \pmod{2}\}$ com lei uniforme e $A_i = \{x : x_i = 1\}$. Quaisquer $k-1$ coordenadas são uniformes em $\{0,1\}^{k-1}$, portanto as subfamílias próprias são independentes e cada A_i tem probabilidade $1/2$; a paridade torna a família completa dependente. Há 2^{k-1} pontos. A minimalidade segue porque as 2^{k-1} configurações de quaisquer $k-1$ eventos têm probabilidade $2^{-(k-1)} > 0$ e, portanto, exigem ao menos um ponto cada.

3.5 Em A_n , todas as primeiras n retiradas foram brancas. Na etapa $j+1$ há $m+j$ bolas brancas e uma preta, logo

$$\mathbf{P}(A_n) = \prod_{j=0}^{n-1} \frac{m+j}{m+j+1} = \frac{m}{m+n}.$$

A série diverge. Como $A_n \downarrow \{\text{a preta nunca é retirada}\}$, $\mathbf{P}(A_n \text{ i.v.}) = \lim_n m/(m+n) = 0$. Portanto divergência da soma, sem independência, não implica ocorrência infinita quase certa.

3.6 Ponha $B_N = \bigcup_{n \geq N} A_n$. Então $B_N \downarrow \limsup A_n$ e $\lambda(B_N) \geq \lambda(A_N) \geq \eta$. Como $\lambda([0,1]) < \infty$, continuidade por cima dá $\lambda(\limsup A_n) = \lim_N \lambda(B_N) \geq \eta$. Em particular o limsup não é vazio, e qualquer de seus pontos pertence a infinitos A_n .

3.7 A necessidade é imediata. Para a suficiência, todo $E_k \in \sigma\langle \mathcal{A}_k \rangle$ é união de células da partição. As interseções correspondentes às diferentes escolhas de células são disjuntas; pela aditividade enumerável e pela distributividade do produto sobre séries de termos não negativos,

$$\mathbf{P}\left(\bigcap_k E_k\right) = \sum_{A_k \subseteq E_k} \prod_k \mathbf{P}(A_k) = \prod_k \sum_{A_k \subseteq E_k} \mathbf{P}(A_k) = \prod_k \mathbf{P}(E_k).$$

Logo as σ -álgebras são independentes.

3.8 Para a primeira parte use $\Omega = \{1, \dots, 8\}$ equiprovável, $A_1 = A_2 = \{1, 2, 3, 4\}$ e $A_3 = \{1, 5, 6, 7\}$, e ponha $\mathcal{A}_k = \{A_k\}$. A identidade tripla vale, mas as duas primeiras σ -álgebras não são independentes.

Sob a hipótese de cobertura, fixe as demais coordenadas. Inclusão-exclusão estende a identidade a uniões finitas disjuntas de elementos de um π -sistema; a cobertura enumerável e continuidade por baixo permitem inserir Ω . A classe dos conjuntos da primeira coordenada para os quais a identidade vale é então um λ -sistema contendo $\mathcal{A}_1$. O Teorema π - λ a amplia para $\sigma\langle\mathcal{A}_1\rangle$. Repita coordenada por coordenada para obter a independência das n σ -álgebras.

3.9 $\{N < \infty\} = (\limsup A_n)^c$. Se esse evento tem probabilidade positiva, como $\{N < \infty\} = \bigcup_{r \geq 0} \{N \leq r\}$, algum $B = \{N \leq r\}$ tem probabilidade positiva. A recíproca é imediata, pois $B \subseteq \{N < \infty\}$.

Para a segunda equivalência, lembre que $\mathbf{P}(\cdot | B)$ é ela mesma uma probabilidade. Se $\sum_n \mathbf{P}(A_n | B) < \infty$ para algum B com $\mathbf{P}(B) > 0$, o primeiro lema de Borel-Cantelli aplicado a $\mathbf{P}(\cdot | B)$ dá $\mathbf{P}(\limsup A_n | B) = 0$ e portanto $\mathbf{P}(\limsup A_n) \leq 1 - \mathbf{P}(B) < 1$. Reciprocamente, se $\mathbf{P}(\limsup A_n) < 1$, tome $B = \{N \leq r\}$ como acima e particione B pelos conjuntos B_S em que ocorrem exatamente os índices de um conjunto finito S , necessariamente com $|S| \leq r$. Como $\mathbf{P}(A_n \cap B_S) = 0$ para $n \notin S$, a aditividade dá $\sum_n \mathbf{P}(A_n \cap B) = \sum_S |S| \mathbf{P}(B_S) \leq r \mathbf{P}(B)$, isto é, $\sum_n \mathbf{P}(A_n | B) \leq r < \infty$.

3.10 Na álgebra $\mathcal{A}$ das uniões finitas disjuntas de intervalos semiabertos, defina ℓ pela soma dos comprimentos; refinamentos por extremos mostram que ela é bem definida e finitamente aditiva. Restrinja primeiro a um intervalo compacto $[-m, m]$. A classe $\mathcal{K}$ dos compactos desse intervalo é compacta: a propriedade da interseção finita, junto com compacidade, implica interseção total não vazia. Encolhendo ligeiramente cada componente de $A \in \mathcal{A}$, obtemos a aproximação interior exigida pelo Teorema de Extensão Compacta; logo ℓ é pré-medida em cada restrição. Fazendo $m \uparrow \infty$ conclui-se a pré-medida em $\mathbb{R}$. Carathéodory a estende a $\sigma(\mathcal{A}) = \mathcal{B}(\mathbb{R})$; a extensão é a medida de Lebesgue e é única pela σ -finitude.

3.11 Refinar um cilindro de nível n para o nível $n + 1$ multiplica seus pesos por p e $1 - p$, cuja soma é um; por indução, a prescrição independe do nível. Refinando simultaneamente finitos cilindros, define-se de modo único uma probabilidade finitamente aditiva na álgebra.

O espaço produto é compacto (ou use diagonalização com a métrica produto). Cilindros são clopen; suas uniões finitas são compactas. A própria álgebra é, portanto, uma classe compacta e satisfaz a aproximação interior. O Teorema de Extensão Compacta dá a pré-medida e Carathéodory, a única extensão. Finalmente, para índices distintos e valores ε_i ,

$$\mathbf{P}_p(X_{i_1} = \varepsilon_1, \dots, X_{i_r} = \varepsilon_r) = \prod_{j=1}^r p^{\varepsilon_j} (1-p)^{1-\varepsilon_j},$$

o que prova que as coordenadas são i.i.d. Bernoulli(p).

3.12 Dados finitos cilindros, refine todos para o cilindro determinado pela união de suas bases. Somar sobre as configurações das novas arestas usa $p + (1-p) = 1$, provando consistência e aditividade finita.

Com a topologia produto em $\{0,1\}^E$, os cilindros são clopen. A base de cada cilindro é um produto finito, portanto compacta; Tychonoff mostra que o espaço produto e esses cilindros são compactos. A álgebra dos cilindros é assim uma classe compacta e satisfaz a aproximação interior. O Teorema de Extensão Compacta torna a função uma pré-medida. Como sua massa total é um, Carathéodory fornece uma extensão única à σ -álgebra gerada.

3.13 Os eventos B_{100k+1} , $k \geq 0$, dependem de blocos disjuntos, são independentes e têm probabilidade 2^{-100} . A soma dessas probabilidades diverge. Pelo segundo lema de Borel–Cantelli, B_{100k+1} ocorre infinitas vezes quase certamente; logo também B_n ocorre infinitas vezes e $\mathbf{P}(B_n \text{ i.v.}) = 1$.

3.14 Cada $|Z_n|^{1/n}$ é mensurável, e supremos e ínfimos enumeráveis mostram que seu limite superior também o é. A fórmula de Cauchy–Hadamard identifica R com o raio de convergência.

Alterar finitos coeficientes não muda o limite superior; portanto R é mensurável em relação à σ -álgebra caudal dos Z_n . Para cada racional q , o evento $\{R \leq q\}$ é caudal e, pela lei 0–1 de Kolmogorov, tem probabilidade zero ou um. Esses eventos crescentes determinam um único limiar r ; usando racionais que se aproximam de r por ambos os lados, conclui-se que $R = r$ quase certamente.

4.1 Para $A \in \mathcal{F}$, $(g \circ f)^{-1}(A) = f^{-1}(g^{-1}(A))$. Como $g^{-1}(A) \in \mathcal{S}$ e f é mensurável, essa imagem inversa pertence a $\mathcal{F}$.

4.2 O mapa $(X, Y) : \Omega \rightarrow \mathbb{R}^2$ é mensurável, pois as pré-imagens de retângulos são interseções de eventos mensuráveis. Componha-o com as funções contínuas $(x, y) \mapsto x + y$, xy , $\max\{x, y\}$ e $\min\{x, y\}$. Para o primeiro item, componha X com a função contínua $x \mapsto |x|$.

4.3 Como $\{X > 0\} = \bigcup_{n \geq 1} \{X > 1/n\}$, se todos os eventos da união tivessem probabilidade zero, sua união também teria. Logo algum $\delta = 1/n$ satisfaz a conclusão.

4.4 Tome $g(u) = -\lambda^{-1} \log u$. Para $x \geq 0$,

$$\mathbf{P}(g(U) \leq x) = \mathbf{P}(U \geq e^{-\lambda x}) = 1 - e^{-\lambda x},$$

e a probabilidade é zero para $x < 0$. Essa é a função de distribuição exponencial de taxa λ .

4.5 No primeiro item ponha $t = u - s$; a volta é a mesma substituição. Para uma exponencial, a razão $\mathbf{P}(X > s + t)/\mathbf{P}(X > s) = e^{-\lambda t}$ prova a propriedade.

Na recíproca, a não degenerescência, junto com o caso $s = 0$, implica $\mathbf{P}(X > 0) = 1$. Para $G(t) = \mathbf{P}(X > t)$, a hipótese dá $G(s + t) = G(s)G(t)$. Como X é finita, $G(t) \downarrow 0$. Se $G(1) = 1$, então $G(n) = 1$ para todo n ; se $G(1) = 0$, então $G(1/n) = 0$, contradizendo $G(1/n) \uparrow \mathbf{P}(X > 0) = 1$. Portanto $0 < G(1) < 1$. Pondo $\lambda = -\log G(1) > 0$, a equação dá $G(q) = e^{-\lambda q}$ para todo racional $q \geq 0$. A monotonicidade e aproximações racionais por ambos os lados estendem a fórmula a todo real $t \geq 0$.

4.6 Para todo boreliano B , o evento $\{X \in B\}$ é independente de si mesmo; logo sua probabilidade p satisfaz $p = p^2$ e pertence a $\{0, 1\}$. Em particular a função de distribuição de X só assume 0 e 1. Se $c = \inf\{x : \mathbf{P}(X \leq x) = 1\}$, continuidade à direita e limites da função de distribuição dão $\mathbf{P}(X = c) = 1$.

4.7 Os eventos $A_i = \{X_i > \delta\}$ são independentes e $\sum_i \mathbf{P}(A_i) = \infty$. Pelo segundo lema de Borel–Cantelli, infinitos A_i ocorrem quase certamente. Em cada um deles $X_i > \delta$; como os termos são não negativos, as somas parciais tendem a infinito.

4.8 As partes positiva e negativa são transformações borelianas: $X^+ = \max(X, 0)$ e $X^- = \max(-X, 0)$. Transformações mensuráveis separadas preservam independência; aplique esse fato a cada um dos quatro pares.

4.9 A família de todas as uniões de células contém os E_n e é uma σ -álgebra; pela minimalidade, ela é $\mathcal{F}$. Se uma função mensurável assumisse dois valores distintos numa mesma célula, um racional entre eles produziria um conjunto $\{X \leq q\} \in \mathcal{F}$ que cortaria essa célula, o que é impossível. Logo X é constante em cada E_n . Reciprocamente, escolhidos $c_n \in \mathbb{R}$, a função

$$X = \sum_{n \geq 1} c_n \mathbb{1}_{E_n}$$

é mensurável, pois a imagem inversa de cada boreliano é uma união das células correspondentes.

5.1 O conjunto de Cantor tem medida de Lebesgue zero. Como x^2 é limitada em $[0, 1]$, sua integral sobre um conjunto nulo é nula: $\int_C x^2 \, d\lambda = 0$.

5.2 Se $t_n \rightarrow t$, então $u(x) \cos(xt_n) \rightarrow u(x) \cos(xt)$ pontualmente e $|u(x) \cos(xt_n)| \leq |u(x)| \in L^1$. Pela Convergência Dominada, $U(t_n) \rightarrow U(t)$.

5.3 Ponha $B_N = \bigcap_{n \geq N} A_n$. Então $B_N \uparrow \liminf A_n$ e, para todo $n \geq N$, $\mu(B_N) \leq \mu(A_n)$; logo $\mu(B_N) \leq \inf_{n \geq N} \mu(A_n)$. Pela continuidade por baixo da medida (equivalentemente, Convergência Monótona aplicada a $\mathbb{1}_{B_N}$),

$$\mu(\liminf A_n) = \lim_N \mu(B_N) \leq \lim_N \inf_{n \geq N} \mu(A_n) = \liminf_n \mu(A_n).$$

5.4 f_n assume apenas os valores $k2^{-n}$, $0 \leq k < n2^n$, e n ; suas fibras são pré-imagens de intervalos por f , logo é simples e mensurável. A desigualdade $2^{-(n+1)} \lfloor 2^{n+1}x \rfloor \geq 2^{-n} \lfloor 2^n x \rfloor$, junto com o aumento do truncamento, dá $f_n \leq f_{n+1} \leq f$. Se $f(x) < \infty$, o erro de arredondamento é menor que 2^{-n} quando $n > f(x)$; se $f(x) = \infty$, $f_n(x) = n$. Portanto $f_n \uparrow f$. Assim $\int f = \lim_n \int f_n$ pela Convergência Monótona, que é a aproximação canônica por simples.

6.1 Os eventos $\{X = k, Y = n - k\}$ são disjuntos e cobrem $\{X + Y = n\}$; a independência dá

$$\mathbf{P}(X + Y = n) = \sum_{k \in \mathbb{Z}} \mathbf{P}(X = k) \mathbf{P}(Y = n - k).$$

Substituindo as massas de Poisson e usando o binômio de Newton, obtemos

$$\mathbf{P}(X + Y = r) = e^{-(\lambda + \mu)} \frac{(\lambda + \mu)^r}{r!},$$

logo $X + Y \sim \text{Poisson}(\lambda + \mu)$. Para as binomiais, a identidade de Vandermonde (ou a interpretação como soma de ensaios de Bernoulli) dá $X + Y \sim \text{Binomial}(n + m, p)$.

6.2 Como X é mensurável, $P_X(A)$ está definida para todo boreliano. Além disso, $P_X(\emptyset) = 0$ e $P_X(\mathbb{R}) = 1$. Se A_j são disjuntos, também o são $X^{-1}(A_j)$ e

$$P_X\left(\bigsqcup_j A_j\right) = \mathbf{P}\left(\bigsqcup_j X^{-1}(A_j)\right) = \sum_j P_X(A_j).$$

Logo P_X é uma probabilidade.

6.3 No espaço $((0,1), \mathcal{B}, \lambda)$, tome $U(u) = u$ e

$$X = F^{-1}(U), \quad F^{-1}(v) = \inf\{x : F(x) \geq v\}.$$

Essa inversa é crescente, logo mensurável, e $\{F^{-1}(U) \leq x\} = \{U \leq F(x)\}$ (a continuidade à direita resolve os pontos de salto). Portanto $\mathbf{P}(X \leq x) = \mathbf{P}(U \leq F(x)) = F(x)$.

6.4 A série converge absolutamente e fica entre 0 e $\sum_{k \geq 1} 2^{-k} = 1$. Cada uma das 2^n configurações iniciais tem probabilidade 2^{-n} , de modo que U_n assume os valores $i2^{-n}$, $i = 0, \dots, 2^n - 1$, todos com massa 2^{-n} , e

$$\mathbf{P}(U_n \leq i2^{-n}) = (i + 1)2^{-n}, \quad i = 0, \dots, 2^n - 1.$$

Como $0 \leq U - U_n \leq 2^{-n}$ e o evento $U - U_n = 2^{-n}$ (todos os dígitos seguintes iguais a 1) tem probabilidade zero, segue que, para $j = 0, \dots, 2^n$,

$$\mathbf{P}(U < j2^{-n}) = \mathbf{P}(U_n \leq (j - 1)2^{-n}) = j2^{-n}.$$

Passando ao limite e aproximando qualquer $x \in [0,1]$ por diádicos, $\mathbf{P}(U \leq x) = x$. As duas expansões de um diádico correspondem a uma cauda eventualmente toda 0 ou toda 1; cada evento tem probabilidade zero, por isso a ambiguidade não altera a lei.

6.5 As funções $g_+(x) = x^+ = \max\{x, 0\}$ e $g_-(x) = x^- = \max\{-x, 0\}$ são contínuas. Para todo boreliano B ,

$$\mathbf{P}(g_{\pm}(X) \in B) = \mathbf{P}(X \in g_{\pm}^{-1}(B)) = \mathbf{P}(Y \in g_{\pm}^{-1}(B)) = \mathbf{P}(g_{\pm}(Y) \in B).$$

6.6 A série converge absolutamente e sua expansão ternária usa somente 0 e 2, logo $Y \in C$. Para todo y , o evento $\{Y = y\}$ fixa uma sequência infinita de dígitos (no máximo duas nas ambiguidades) e tem probabilidade zero; portanto F_Y não tem saltos e é contínua. Em cada intervalo removido na construção de C não há massa, de modo que F_Y é constante; como $[0,1] \setminus C$ tem medida de Lebesgue um, $F_Y' = 0$ quase em toda parte. Por fim, escolha $A = C$: $\lambda(C) = 0$ e $\mu_Y(C^c) = \mathbf{P}(Y \notin C) = 0$, provando singularidade mútua.

6.7 Se existe o acoplamento ordenado, então $\{Y^* \leq x\} \subseteq \{X^* \leq x\}$ e, portanto, $F_X(x) \geq F_Y(x)$. Reciprocamente, tome um único $U \sim U(0,1)$ e ponha $X^* = F_X^{-1}(U)$, $Y^* = F_Y^{-1}(U)$. A desigualdade entre as funções de distribuição implica $F_X^{-1}(u) \leq F_Y^{-1}(u)$ para todo u , e a construção por quantis preserva as leis marginais.

6.8 Para um conjunto não vazio $S_t = \{y : F(y) \geq t\}$, monotonicidade e continuidade à direita mostram que sua extremidade esquerda é simultaneamente $\inf S_t$ e $\sup\{y : F(y) < t\}$; logo $F_1^{-1} = F_2^{-1}$. Aplicando o mesmo argumento a $\{F > t\}$ obtém-se $F_3^{-1} = F_4^{-1}$.

Se X é Bernoulli(1/2) em $\{0,1\}$ e $t = 1/2$, então $F_1^{-1}(t) = F_2^{-1}(t) = 0$, enquanto $F_3^{-1}(t) = F_4^{-1}(t) = 1$.

7.1 Escreva $X^+ = \sum_k x_k^+ \mathbf{1}_{\{X=x_k\}}$ e $X^- = \sum_k x_k^- \mathbf{1}_{\{X=x_k\}}$. Por Convergência Monótona, $\mathbf{E}[X^{\pm}] = \sum_k x_k^{\pm} \mathbf{P}(X = x_k)$. A hipótese torna ambas as somas finitas; subtraindo-as resulta a fórmula pedida.

7.2 Em $[0,1]$, ponha $(X,Y) = (1,0)$ em $[0,3/4]$ e $(X,Y) = (0,4)$ em $(3/4,1]$. Então $\mathbf{P}(X > Y) = 3/4$, enquanto $\mathbf{E}[X] = 3/4 < 1 = \mathbf{E}[Y]$.

7.3 Para todo boreliano B ,

$$\mathbf{P}(g(\xi) \in B) = \mathbf{P}(\xi \in g^{-1}(B)) = \mathbf{P}(\eta \in g^{-1}(B)) = \mathbf{P}(g(\eta) \in B),$$

pois $g^{-1}(B)$ é boreliano. Isso prova a afirmação para qualquer transformação boreliana e contém, em particular, os exemplos do Capítulo 6.

7.4 Eventos que dependem de finitas coordenadas ímpares são independentes de eventos que dependem de finitas coordenadas pares. Essas duas famílias de cilindros são π -sistemas e geram as σ -álgebras exibidas. Duas aplicações do Teorema π - λ estendem a identidade de produto às σ -álgebras geradas.

7.5 Decomponha pela partição $\{N = n\}$, lembrando que N é independente da sequência. Para cada n fixo, $\mathbf{E}[X_1 + \cdots + X_n] = n\mu$ e $\mathbf{E}[(X_1 + \cdots + X_n)^2] = n\sigma^2 + n^2\mu^2$. Somando,

$$\mathbf{E}[S] = \sum_{n \geq 0} \mathbf{P}(N = n) n\mu = m\mu, \quad \mathbf{E}[S^2] = \sum_{n \geq 0} \mathbf{P}(N = n) (n\sigma^2 + n^2\mu^2) = \sigma^2 \mathbf{E}[N] + \mu^2 \mathbf{E}[N^2].$$

Portanto

$$\text{Var}(S) = \sigma^2 m + \mu^2 (\mathbf{E}[N^2] - m^2) = m\sigma^2 + v\mu^2.$$

7.6 Para todo $t \in \mathbb{R}$, $0 \leq \mathbf{E}[(X - tY)^2] = \mathbf{E}[X^2] - 2t\mathbf{E}[XY] + t^2\mathbf{E}[Y^2]$. Se $\mathbf{E}[Y^2] > 0$, o discriminante não pode ser positivo, o que dá a desigualdade; o caso $Y = 0$ é imediato. Há igualdade exatamente quando o mínimo quadrático é zero, isto é, quando $X = (\mathbf{E}[XY]/\mathbf{E}[Y^2])Y$ quase certamente. Incluindo o caso nulo, isso equivale à dependência linear em L^2 .

7.7 $\text{Cov}(X,Y) = a$, $\text{Var}(X) = 1$ e $\text{Var}(Y) = a^2 + b^2$, logo

$$\text{Corr}(X,Y) = \frac{a}{\sqrt{a^2 + b^2}}.$$

Seu módulo é no máximo um porque $a^2 \leq a^2 + b^2$. A correlação é 1 se, e somente se, $b = 0$ e $a > 0$; é -1 se, e somente se, $b = 0$ e $a < 0$.

7.8 Escreva $X = \sum_{i=1}^n I_i$, onde I_i indica um ponto fixo. Então $\mathbf{E}I_i = 1/n$ e, para $i \neq j$, $\mathbf{E}[I_i I_j] = 1/[n(n-1)]$. Assim

$$\mathbf{E}X = 1, \quad \text{Var } X = n \frac{1}{n} \left(1 - \frac{1}{n}\right) + n(n-1) \left(\frac{1}{n(n-1)} - \frac{1}{n^2}\right) = 1.$$

7.9 Ponha $Z = \lfloor X \rfloor$. Pontualmente, o número de inteiros $n \geq 1$ com $n < Z$ é $\lfloor Z \rfloor - 1 \leq Z$ e o número de inteiros $n \geq 0$ com $n < Z$ é $\lfloor Z \rfloor \geq Z$, isto é,

$$\sum_{n \geq 1} \mathbb{1}_{\{Z > n\}} \leq Z \leq \sum_{n \geq 0} \mathbb{1}_{\{Z > n\}}.$$

Tomando esperanças e trocando soma com esperança pela Convergência Monótona (os termos são não negativos),

$$\sum_{n=1}^{\infty} \mathbf{P}(Z > n) \leq \mathbf{E}Z \leq \sum_{n=0}^{\infty} \mathbf{P}(Z > n).$$

Se $\mathbf{E}|X| = \infty$, a desigualdade da direita força a divergência da série que começa em $n = 0$.

7.10 Como $\{|X| > M\} \downarrow \emptyset$, escolha M com $\mathbf{P}(|X| > M) < \varepsilon$ e defina $X_\varepsilon = (-M) \vee (X \wedge M)$. Então $|X_\varepsilon| \leq M$ e $X_\varepsilon = X$ em $\{|X| \leq M\}$, de modo que $\mathbf{P}(X \neq X_\varepsilon) < \varepsilon$.

7.11 Depois de obter k tipos, a probabilidade de o pacote seguinte trazer um tipo novo é $(n-k)/n$. O tempo de espera correspondente é geométrico de média $n/(n-k)$. Somando para $k = 0, \dots, n-1$,

$$\mathbf{E}T = \sum_{k=0}^{n-1} \frac{n}{n-k} = n \sum_{j=1}^n \frac{1}{j} = nH_n.$$

7.12 Para $0 < s < 1$, derive termo a termo. Quando $s \uparrow 1$, os termos são não negativos e a Convergência Monótona dá $G'_X(1-) = \mathbf{E}X$ e $G''_X(1-) = \mathbf{E}[X(X-1)]$; ambos são finitos pela hipótese de segundo momento. Como $\mathbf{E}[X^2] = \mathbf{E}[X(X-1)] + \mathbf{E}X$, subtraímos $(\mathbf{E}X)^2$ e obtemos a fórmula.

7.13 $\mathbf{E}(S_n/n) = 1/2$ e $\text{Var}(S_n/n) = 1/(4n)$. Portanto, por Chebyshev,

$$\mathbf{P}(|S_n/n - 1/2| \geq \varepsilon) \leq \frac{1}{4n\varepsilon^2} \rightarrow 0.$$

7.14 Nos 2^n intervalos diádicos, os vetores $(d_1, \dots, d_n)$ percorrem $\{0, 1\}^n$ com pesos 2^{-n} (os extremos não importam). Assim

$$M_n(t) = 2^{-n} \sum_{z_1, \dots, z_n \in \{-1, 1\}} e^{t(z_1 + \dots + z_n)} = \left(\frac{e^t + e^{-t}}{2} \right)^n.$$

Como a soma é finita, podemos derivar: $M'_n(0) = 0$ e $M''_n(0) = n$, que são, respectivamente, as duas integrais pedidas.

7.15 Para $t > 0$, Markov e (7.20) dão

$$\mathbf{P}(s_n \geq n\varepsilon) \leq e^{-tn\varepsilon} M_n(t) \leq \exp\{n(t^2/2 - t\varepsilon)\}.$$

O mínimo ocorre em $t = \varepsilon$ e vale $e^{-n\varepsilon^2/2}$. Aplique o mesmo argumento a $-s_n$ e use a subaditividade.

7.16 Se $S_n \sim \text{Binomial}(n, x)$, o polinômio é $B_n f(x) = \mathbf{E}[f(S_n/n)]$. Dado $\varepsilon > 0$, a continuidade uniforme fornece $\delta > 0$ tal que $|u - x| < \delta$ implica $|f(u) - f(x)| < \varepsilon$. Se $M = \|f\|_\infty$, então, uniformemente em x ,

$$|B_n f(x) - f(x)| \leq \varepsilon + 2M\mathbf{P}(|S_n/n - x| \geq \delta) \leq \varepsilon + \frac{M}{2n\delta^2},$$

pois $\text{Var}(S_n/n) = x(1-x)/n \leq 1/(4n)$. Tome $n \rightarrow \infty$ e depois $\varepsilon \downarrow 0$.

7.17 Defina $X = \mu + t\sigma$ e $X = \mu - t\sigma$, cada um com probabilidade $1/(2t^2)$, e $X = \mu$ com probabilidade $1 - 1/t^2$. Por simetria, $\mathbf{E}X = \mu$; além disso, $\text{Var } X = t^2\sigma^2(1/t^2) = \sigma^2$, e o evento indicado ocorre exatamente nos dois valores extremos.

7.18 Um exemplo é $\mathbf{P}(X = k) = \zeta(3)^{-1}k^{-3}$, $k \geq 1$: o primeiro momento é finito e o segundo diverge. Para $p > 2$, $(X \wedge n)^2 \leq nX$, logo

$$\sum_n n^{-p} \mathbf{E}[X_n^2] \leq \mathbf{E}X \sum_n n^{1-p} < \infty.$$

Para $p = 2$, troque soma e esperança pela Convergência Monótona (os termos são não negativos) e use a separação sugerida: com uma constante universal C ,

$$\sum_{n \geq 1} \frac{(x \wedge n)^2}{n^2} \leq [x] + x^2 \sum_{n > x} n^{-2} \leq C(1 + x).$$

Integrando, a soma continua finita porque $X \in L^1$.

7.19 Ponha $p = \max_{k \leq n} \mathbf{P}(|S_k| \geq r)$. Se $p \geq 1/3$, a cota é trivial. Suponha $p < 1/3$. Pela desigualdade triangular,

$$\mathbf{P}(|S_n - S_j| > 2r) \leq \mathbf{P}(|S_n| > r) + \mathbf{P}(|S_j| > r) \leq 2p.$$

Como A_j depende de $X_1, \dots, X_j$ e é independente de $S_n - S_j$,

$$(1 - 2p) \sum_j \mathbf{P}(A_j) \leq \sum_j \mathbf{P}(A_j, |S_n - S_j| \leq 2r) \leq \mathbf{P}(|S_n| \geq r) \leq p.$$

Assim $\mathbf{P}(\max_{k \leq n} |S_k| \geq 3r) = \sum_j \mathbf{P}(A_j) \leq 3p$. Tome $r = t/3$.

7.20 Num espaço equiprovável de três pontos, tome $X = -1, 0, 1$ e $Y = \mathbb{1}_{\{X=0\}}$. Então $\mathbf{E}X = 0$ e $XY = 0$ sempre, logo $\text{Cov}(X, Y) = 0$. As variáveis são dependentes, pois $\mathbf{P}(Y = 1) = 1/3$, mas $\mathbf{P}(Y = 1 \mid X = 0) = 1$. Ambas são limitadas.

8.1 Cada identidade segue testando um ponto y : por exemplo, $y \in (E \cap F)^x$ se, e somente se, $(x, y) \in E$ e $(x, y) \in F$, isto é, $y \in E^x \cap F^x$. O mesmo argumento, trocando “e” por complemento ou por “existe n ”, prova as outras duas.

8.2 Escolha $A_n \uparrow \Omega_1$ e $B_m \uparrow \Omega_2$ com $\mu_1(A_n), \mu_2(B_m) < \infty$. Os retângulos $A_n \times B_m$ cobrem o produto e têm medida $\mu_1(A_n)\mu_2(B_m) < \infty$. Enumerando os pares (n, m) obtemos uma cobertura enumerável por conjuntos de medida finita.

8.3 As projeções coordenadas são contínuas; portanto seus cilindros borelianos são borelianos e $\bigotimes_i \mathcal{B}(\mathbb{R}) \subseteq \mathcal{B}(\mathbb{R}^d)$. Reciprocamente, retângulos abertos com extremos racionais pertencem ao produto de σ -álgebras e formam uma base enumerável de $\mathbb{R}^d$. Todo aberto é união de uma subfamília dessa base, logo pertence ao produto; portanto ele contém a σ -álgebra boreliana.

8.4 Refinamentos por retângulos mostram primeiro que μ_0 é bem definida e finitamente aditiva. Se $E = \bigsqcup_{j=1}^r A_j \times B_j$, a regularidade interior permite escolher compactos $K_j \subseteq A_j$, $L_j \subseteq B_j$ com perda total arbitrariamente pequena; $K = \bigcup_j K_j \times L_j$ é compacto, está em E e aproxima $\mu_0(E)$ por dentro. Os compactos de $\mathbb{R}^{m+n}$ formam uma classe compacta. O Teorema de Extensão Compacta dá a pré-medida, e Carathéodory a estende à $\mathcal{B}(\mathbb{R}^{m+n})$, preservando $\mu(A \times B) = \mu_1(A)\mu_2(B)$.

8.5 Em $((0,1)^{\mathbb{N}}, \lambda^{\otimes \mathbb{N}})$, sejam U_j as projeções e defina $\xi_j = F^{-1}(U_j)$, onde $F^{-1}(u) = \inf\{x : F(x) \geq u\}$. As projeções são i.i.d. uniformes; transformações coordenada a coordenada preservam a independência, e a construção por quantis dá função de distribuição F a cada ξ_j . Assim obtemos diretamente uma construção alternativa ao argumento de existência do texto.

8.6 Pela convolução,

$$f_Z(z) = \int_0^1 \alpha e^{-\alpha(z-y)} \mathbb{1}_{\{y \leq z\}} dy.$$

Logo

$$f_Z(z) = \begin{cases} 0, & z < 0, \\ 1 - e^{-\alpha z}, & 0 \leq z < 1, \\ e^{-\alpha(z-1)} - e^{-\alpha z}, & z \geq 1. \end{cases}$$

8.7 Como ξ é uniforme, $T(\xi)$ é uniforme em $\text{Im } T$. Suas coordenadas são independentes e uniformes em $\mathbb{F}_q$ exatamente quando $\text{Im } T = \mathbb{F}_q^n$, isto é, quando T é sobrejetora. Pela não degenerescência da forma, uma relação entre os funcionais $x \mapsto \langle x, v_i \rangle$ equivale à mesma relação linear entre os v_i . Assim $\text{rank } T = n$ se, e somente se, $v_1, \dots, v_n$ são linearmente independentes.

8.8 Condicionando em $\eta = y$ e usando a independência,

$$\mathbf{P}(\xi \eta \leq z) = \int \mathbf{P}(\xi \leq z/y) dG(y) = \int F(z/y) dG(y).$$

No caso uniforme, para $0 < z < 1$,

$$H(z) = \int_0^z 1 dy + \int_z^1 \frac{z}{y} dy = z(1 - \log z).$$

Além disso, $H(z) = 0$ para $z \leq 0$ e $H(z) = 1$ para $z \geq 1$; derivando no intervalo $(0,1)$, obtemos a densidade $h(z) = -\log z$ ali e zero fora dele.

8.9 Para cada instante a partir do qual todos os dígitos valem 1, há apenas finitos prefixos; logo E é enumerável. Cada sequência tem probabilidade zero, portanto $\mathbf{P}(E) = 0$.

As únicas ambiguidades da expansão binária são as expansões que terminam em zeros e as correspondentes que terminam em uns. Retirar E escolhe a primeira convenção e elimina também a expansão de 1, produzindo uma bijeção sobre $[0,1)$. A fórmula dos d_k fornece seus dígitos e mostra que a inversa é mensurável; a mensurabilidade de Φ segue das somas parciais.

Fixados os primeiros n dígitos, o cilindro tem probabilidade 2^{-n} e, fora de E , corresponde ao intervalo diádico semiaberto de comprimento 2^{-n} . A igualdade das medidas nesses intervalos estende-se à σ -álgebra de Borel pelo teorema da classe monótona. Assim Φ preserva medida e é um isomorfismo bimensurável módulo os conjuntos nulos E e $\{1\}$.

8.10 Por Tonelli,

$$\int_{\mathbb{R}} \mathbf{P}(x < X < x + a) \, dx = \mathbf{E} \left[\int_{\mathbb{R}} \mathbb{1}_{\{x < X < x+a\}} \, dx \right].$$

Para $X(\omega) = y$, a condição equivale a $y - a < x < y$; esse intervalo tem comprimento a . A variável dentro da esperança é, portanto, identicamente igual a a .

9.1 A tabela conjunta é

$$\mathbf{P}(1,1) = a, \quad \mathbf{P}(1,0) = q - a, \quad \mathbf{P}(0,1) = r - a, \quad \mathbf{P}(0,0) = 1 - q - r + a.$$

Logo $\max\{0, q + r - 1\} \leq a \leq q$. O acoplamento independente tem $a = qr$; o monótono, $a = q$. Como $\mathbf{P}(X \neq Y) = q + r - 2a$, o mínimo é atingido pelo monótono e vale $r - q$.

9.2 Cada uma das quatro variáveis é Bernoulli(1/2), logo as marginais em cada tempo coincidem. Porém (X_0, X_1) é uniforme nos quatro pares de $\{0,1\}^2$, enquanto (Y_0, Y_1) vale (0,0) ou (1,1), cada um com probabilidade 1/2. Um acoplamento de processos deve preservar a lei do vetor inteiro (todas as distribuições finito-dimensionais), não apenas cada marginal temporal.

9.3 As marginais de $(f(X), f(Y))$ são as imagens de μ e ν , portanto ele é o acoplamento afirmado. Em espaço enumerável, escolha um acoplamento maximal de

(X, Y) ; então

$$d_{TV}(f_{\#}\mu, f_{\#}\nu) \leq \mathbf{P}(f(X) \neq f(Y)) \leq \mathbf{P}(X \neq Y) = d_{TV}(\mu, \nu).$$

Diretamente, para todo mensurável $B \subseteq S$, $|(\mu_{\#}f)(B) - (\nu_{\#}f)(B)| = |\mu(f^{-1}B) - \nu(f^{-1}B)| \leq d_{TV}(\mu, \nu)$; tome o supremo.

9.4 Se os vetores diferem, alguma coordenada difere; a cota da união dá a primeira desigualdade. Escolha acoplamentos maximais independentes, para os quais $\mathbf{P}(X_i \neq Y_i) = d_{TV}(\mu_i, \nu_i)$. A desigualdade de acoplamento aplicada aos vetores, seguida da primeira parte, fornece a segunda fórmula.

9.5 Suponha, por simetria, $\lambda_n \geq \lambda$. Tome $X \sim \text{Po}(\lambda)$ e, independentemente, $Z_n \sim \text{Po}(\lambda_n - \lambda)$; então $Y_n = X + Z_n \sim \text{Po}(\lambda_n)$ e

$$d_{TV}(\mathcal{L}(Y_n), \mathcal{L}(X)) \leq \mathbf{P}(Z_n > 0) = 1 - e^{-(\lambda_n - \lambda)}.$$

Com $|\lambda_n - \lambda|$ a mesma cota vale nos dois casos e tende a zero.

9.6 Abrir arestas adicionais preserva qualquer caminho aberto, logo $\{x \leftrightarrow y\}$ e cada A_n são crescentes. Se C_x é infinito, ele alcança distância arbitrariamente grande. Reciprocamente, num grafo localmente finito cada bola de raio finito tem finitos vértices; portanto alcançar todas as distâncias força C_x a ser infinito. Assim $\{|C_x| = \infty\} = \bigcap_n A_n$. No acoplamento por uniformes, a configuração em p está contida na de q quando $p \leq q$; a probabilidade do evento crescente não pode diminuir, logo $\theta_x(p) \leq \theta_x(q)$.

10.1 Para $x \geq 0$,

$$\mathbf{P}(\eta_n > x) = \left(1 - \frac{\lambda}{n}\right)^{\lfloor nx \rfloor + 1} \longrightarrow e^{-\lambda x};$$

para $x < 0$ a cauda vale um. Portanto as funções de distribuição convergem em todos os pontos para a da exponencial de taxa λ .

10.2 Pela desigualdade triangular e pela união,

$$\mathbf{P}(|(\xi_n + \eta_n) - (\xi + \eta)| > \varepsilon) \leq \mathbf{P}(|\xi_n - \xi| > \varepsilon/2) + \mathbf{P}(|\eta_n - \eta| > \varepsilon/2) \longrightarrow 0.$$

10.3 Para $0 < \varepsilon < n$, $\mathbf{P}(|\xi_n| > \varepsilon) = n^{-2} \rightarrow 0$. Além disso, $\mathbf{E}|\xi_n|^r = n^{r-2}$, que tende a zero exatamente para $r < 2$ (vale um para $r = 2$ e diverge para $r > 2$). Por fim, $\sum_n \mathbf{P}(\xi_n \neq 0) = \sum_n n^{-2} < \infty$; Borel–Cantelli implica que apenas finitos ξ_n são não nulos, quase certamente.

10.4 A desigualdade triangular reversa em L^r fornece

$$|\|\xi_n\|_r - \|\xi\|_r| \leq \|\xi_n - \xi\|_r \rightarrow 0.$$

Elevando à potência r , obtemos a convergência dos momentos.

10.5 A convergência em distribuição torna (ξ_n) tight. Para $M > 0$,

$$\mathbf{P}(|(\eta_n - c)\xi_n| > \varepsilon) \leq \mathbf{P}(|\xi_n| > M) + \mathbf{P}(|\eta_n - c| > \varepsilon/M).$$

Escolha primeiro M grande e depois n grande: o termo de erro $D_n = (\eta_n - c)\xi_n$ converge em probabilidade para zero. Por mudança de escala nas funções de distribuição, $c\xi_n \xrightarrow{d} c\xi$. Para concluir, escreva $\eta_n\xi_n = c\xi_n + D_n$ e tome x ponto de continuidade de $F_{c\xi}$ e $\delta > 0$ com $x \pm \delta$ também pontos de continuidade (eles são densos). Então

$$\mathbf{P}(\eta_n\xi_n \leq x) \leq \mathbf{P}(c\xi_n \leq x+\delta) + \mathbf{P}(|D_n| > \delta), \quad \mathbf{P}(\eta_n\xi_n \leq x) \geq \mathbf{P}(c\xi_n \leq x-\delta) - \mathbf{P}(|D_n| > \delta).$$

Tomando $\limsup$ e $\liminf$ e depois $\delta \downarrow 0$, a continuidade de $F_{c\xi}$ em x dá $\mathbf{P}(\eta_n\xi_n \leq x) \rightarrow F_{c\xi}(x)$.

10.6 Dado $\varepsilon > 0$, escolha K com $\mathbf{P}(|\xi| > K) < \varepsilon$. A continuidade uniforme de g em $[-K-1, K+1]$ fornece $0 < \delta \leq 1$ tal que $|x-y| < \delta$ implica $|g(x) - g(y)| < \varepsilon$ nesse compacto. Logo

$$\mathbf{P}(|g(\xi_n) - g(\xi)| > \varepsilon) \leq \mathbf{P}(|\xi| > K) + \mathbf{P}(|\xi_n - \xi| > \delta),$$

e o limite superior é no máximo ε ; faça $\varepsilon \downarrow 0$.

10.7 Escolha índices disjuntos n_k, m_k . Então $(\xi_{n_k}, \xi_{m_k}) \rightarrow_{\mathbf{P}} (\xi, \xi)$. Por outro lado, cada par tem lei $\mu_{n_k} \otimes \mu_{m_k}$ e as marginais convergem fracamente para a lei μ de ξ ; logo a lei-limite do par é $\mu \otimes \mu$. Portanto duas cópias independentes $X, Y \sim \mu$ satisfazem $X = Y$ quase certamente. Para todo boreliano A , isso implica $0 = \mathbf{P}(X \in A, Y \notin A) = \mu(A)(1 - \mu(A))$; assim μ só assume valores 0 e 1 e é uma massa de Dirac.

10.8 Convergência em L^2 implica convergência das normas L^2 e também em L^1 , pois $\|Y\|_1 \leq \|Y\|_2$. Logo $\mathbf{E}[\xi_n^2] \rightarrow \mathbf{E}[\xi^2]$ e $\mathbf{E}[\xi_n] \rightarrow \mathbf{E}[\xi]$. Passe ao limite em $\text{Var}(\xi_n) = \mathbf{E}[\xi_n^2] - \mathbf{E}[\xi_n]^2$.

10.9 Por Markov,

$$\mathbf{P}(|\xi_n - \xi| > \varepsilon) \leq \varepsilon^{-p} \mathbf{E}|\xi_n - \xi|^p \rightarrow 0.$$

A recíproca falha para $\xi_n = n$ com probabilidade $1/n$ e 0 caso contrário: $\xi_n \rightarrow_{\mathbf{P}} 0$, mas $\mathbf{E}|\xi_n| = 1$, logo não há convergência em L^1 .

10.10 Dado $\varepsilon > 0$, escolha $a < b$ com $F(a) < \varepsilon$ e $F(b) > 1 - \varepsilon$. Pela continuidade uniforme de F em $[a, b]$, escolha uma partição $a = x_0 < \dots < x_m = b$ tal que $F(x_j) - F(x_{j-1}) < \varepsilon$. A convergência pontual nos finitos x_j torna $|F_n(x_j) - F(x_j)| < \varepsilon$. Monotonicidade, espremendo $F_n(x)$ entre os valores nos extremos da célula, dá $|F_n(x) - F(x)| < 2\varepsilon$ em $[a, b]$; as caudas dão a mesma cota fora. Logo a convergência é uniforme.

10.11 Se $\xi_n \rightarrow_d \xi$, os pontos $k \pm 1/2$ são continuidades de F_ξ e

$$\mathbf{P}(\xi_n = k) = F_n(k + 1/2) - F_n(k - 1/2) \rightarrow \mathbf{P}(\xi = k).$$

Na volta, dado $\varepsilon > 0$, escolha K com $\mathbf{P}(|\xi| \leq K) > 1 - \varepsilon$. A convergência das finitas massas em $[-K, K]$ implica $\mathbf{P}(|\xi_n| \leq K) > 1 - 2\varepsilon$ para n grande; logo não há fuga de massa. Somando as massas até $\lfloor x \rfloor$ e controlando as duas caudas, obtemos $F_n(x) \rightarrow F(x)$ em todo ponto de continuidade de F .

10.12 Para $0 < \varepsilon < 1$, $\mathbf{P}(Z_n \leq 1 - \varepsilon) = (1 - \varepsilon)^n \rightarrow 0$, o que prova convergência em probabilidade. Além disso, Z_n é crescente e limitada por 1, logo converge quase certamente para algum $Z \leq 1$. Por continuidade por cima, $\mathbf{P}(Z \leq 1 - \varepsilon) = \lim_n (1 - \varepsilon)^n = 0$. Intersectando sobre $\varepsilon = 1/m$, concluimos $Z = 1$ quase certamente.

10.13 Para $k = 0, \dots, 10$, a contagem de subconjuntos dá

$$\mathbf{P}(X_n = k) = \frac{\binom{n}{k} \binom{n}{10-k}}{\binom{2n}{10}}.$$

Equivalentemente, usando fatoriais decrescentes,

$$\mathbf{P}(X_n = k) = \binom{10}{k} \frac{(n)_k (n)_{10-k}}{(2n)_{10}} \longrightarrow \binom{10}{k} 2^{-10}.$$

Como o suporte finito é comum a todas as variáveis, a convergência das massas prova a convergência em distribuição.

11.1 As duas cotas são imediatas da definição. Para cada valor finito de ξ , $\xi^{(M)} = \xi$ quando $M > |\xi|$, logo há convergência quase certa. Como $|\xi^{(M)}| \leq M$, todos os seus momentos são finitos. Finalmente, $|\xi^{(M)}| \leq |\xi| \in L^1$ e a Convergência Dominada dá $\mathbf{E}[\xi^{(M)}] \rightarrow \mathbf{E}[\xi]$; a função dominante é $|\xi|$.

11.2 Pela Lei Forte, $n^{-1} \sum_{k \leq n} \xi_k^2 \rightarrow \mathbf{E} \xi_1^2 = 1$ e $n^{-1} \sum_{k \leq n} (\xi_k - 1)^2 \rightarrow \mathbf{E}(\xi_1 - 1)^2 = 2$ quase certamente. O denominador é então positivo para n grande, e o quociente converge a $1/2$.

11.3 $\mathbf{E}|\log \xi_1| = \int_0^1 -\log x \, dx = 1$ e $\mathbf{E} \log \xi_1 = -1$. Pela Lei Forte,

$$\frac{1}{n} \log \prod_{k=1}^n \xi_k = \frac{1}{n} \sum_{k=1}^n \log \xi_k \longrightarrow -1$$

quase certamente. Pela continuidade da exponencial, a média geométrica converge a e^{-1} .

11.4 Independência dá $M_{\bar{\xi}_n}(s) = [M_{\xi}(s/n)]^n$. A diferenciabilidade em zero fornece $M_{\xi}(u) = 1 + \mu u + o(u)$; portanto

$$n \log M_{\xi}(s/n) \longrightarrow s\mu \quad \text{e} \quad M_{\bar{\xi}_n}(s) \longrightarrow e^{s\mu}.$$

Essa é a função geradora de momentos da constante μ . Para concluir, use a cota exponencial de Markov diretamente: para $t \in (0, c]$, a independência dá

$$\mathbf{P}(\bar{\xi}_n \geq \mu + \varepsilon) \leq e^{-tn(\mu+\varepsilon)} M_{\xi}(t)^n = [e^{-t(\mu+\varepsilon)} M_{\xi}(t)]^n,$$

e de $M_{\xi}(t) = 1 + \mu t + o(t)$ segue $e^{-t(\mu+\varepsilon)} M_{\xi}(t) = 1 - t\varepsilon + o(t)$, que é < 1 para $t > 0$ pequeno. Logo a probabilidade decai geometricamente em n . Aplicando o mesmo a $-\bar{\xi}_k$ controla-se a cauda inferior, e portanto $\bar{\xi}_n \xrightarrow{\mathbf{P}} \mu$.

11.5 As variáveis ξ_k/k são independentes, centradas, e a soma de suas variâncias é $\sum_k \sigma_k^2/k^2 < \infty$. Pelo Teorema 11.16, a série $\sum_k \xi_k/k$ converge quase certamente. O Lema 11.17 implica então $n^{-1} \sum_{k=1}^n \xi_k = S_n/n \rightarrow 0$ quase certamente.

11.6 A série é dominada por $\sum 2/3^k = 1$ e sua expansão ternária usa somente 0 e 2, logo $Z \in C$. Os dígitos são independentes e equiprováveis, o que produz precisamente a distribuição de Cantor. Por Tonelli,

$$\mathbf{E}Z = \sum_{k \geq 1} \frac{\mathbf{E}\xi_k}{3^k} = \sum_{k \geq 1} 3^{-k} = \frac{1}{2}.$$

Como $\text{Var}(\xi_k) = 1$ e os termos são independentes, $\text{Var } Z = \sum_{k \geq 1} 3^{-2k} = 1/8$; a série converge em L^2 , justificando essa passagem.

11.7 Cada η_i é função apenas do par independente (ξ_i, c_i) , logo as variáveis são i.i.d. Além disso, condicionando em ξ_i ,

$$\mathbf{P}(\eta_i = 1) = \mathbf{E}[\mathbf{P}(c_i < f(\xi_i) \mid \xi_i)] = \mathbf{E}[f(\xi_i)] = \int_0^1 f(x) \, dx.$$

Assim são Bernoulli com esse parâmetro, e a Lei Forte aplicada às η_i fornece a convergência exibida.

12.1 Uma Bernoulli(p) tem função característica $(1 - p) + pe^{it}$; somando n cópias independentes, obtemos a potência indicada. Para Poisson,

$$\sum_{k \geq 0} e^{itk} e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda e^{it}}.$$

Se $Z \sim N(0,1)$, completar o quadrado na integral gaussiana dá $\phi_Z(t) = e^{-t^2/2}$; como $\mu + \sigma Z \sim N(\mu, \sigma^2)$, a fórmula afim produz $e^{i\mu t - \sigma^2 t^2/2}$.

12.2 Pela independência,

$$\phi_{\xi+\eta}(t) = \mathbf{E}[e^{it\xi} e^{it\eta}] = \mathbf{E}[e^{it\xi}] \mathbf{E}[e^{it\eta}] = \phi_\xi(t) \phi_\eta(t).$$

12.3 Diretamente, $\phi_X(t) = \mathbf{E}[e^{it(aY+b)}] = e^{itb} \mathbf{E}[e^{i(at)Y}] = e^{itb} \phi_Y(at)$.

12.4 O produto das funções características é

$$\prod_{j=1}^n \exp\{\lambda(e^{it} - 1)\} = \exp\{n\lambda(e^{it} - 1)\},$$

que caracteriza a distribuição Poisson($n\lambda$).

12.5 Independência e a regra de escala dão $\phi_{S_n/n}(t) = [\phi(t/n)]^n$. Além disso,

$$\frac{e^{it\xi_1} - 1}{t} \rightarrow i\xi_1, \quad \left| \frac{e^{it\xi_1} - 1}{t} \right| \leq |\xi_1|,$$

de modo que a Convergência Dominada fornece $\phi(t) = 1 + i\mu t + o(t)$. Logo $n \log \phi(t/n) \rightarrow i\mu t$ e $\phi_{S_n/n}(t) \rightarrow e^{i\mu t}$. Por Lévy, $S_n/n \rightarrow_d \mu$; como o limite é constante, a convergência é em probabilidade.

12.6 Para um convidado, $EX = 1$ e $\text{Var } X = 1/2$. Assim o total tem média 64, variância 32 e desvio $\sqrt{32}$. Com m sanduíches, a correção de continuidade dá

$$\mathbf{P}(S \leq m) \approx \Phi\left(\frac{m + 0.5 - 64}{\sqrt{32}}\right).$$

Como $z_{0.95} \approx 1.6449$, precisamos de $m + 0.5 \geq 64 + 1.6449\sqrt{32} \approx 73.305$. O menor inteiro é $\boxed{m = 73}$ (para $m = 72$ a aproximação fica abaixo de 95%).

12.7 Se $N_n \sim \text{Poisson}(n)$, a soma é $\mathbf{P}(N_n \leq n)$. Escreva N_n como soma de n Poisson(1) independentes. Pelo TLC, $(N_n - n)/\sqrt{n} \rightarrow_d N(0,1)$; como zero é ponto de continuidade da normal, $\mathbf{P}(N_n \leq n) \rightarrow \Phi(0) = 1/2$.

12.8 A existência da f.g.m. numa vizinhança de zero implica $M_\xi(u) = 1 + u\mathbf{E}\xi + o(u)$. Assim

$$n \log M_\xi(s/n) = n \log \left(1 + \frac{s\mathbf{E}\xi}{n} + o(n^{-1}) \right) \rightarrow s\mathbf{E}\xi.$$

Exponenciando obtemos o limite pedido.

12.9 Ponha $Y = (\xi_1 - \mu)/\sigma$. Então $\mathbf{E}Y = 0$, $\mathbf{E}Y^2 = 1$ e

$$\log M_Y(u) = \frac{u^2}{2} + o(u^2).$$

Por independência,

$$\log M_{Z_n}(s) = n \log M_Y(s/\sqrt{n}) = \frac{s^2}{2} + o(1) \rightarrow \frac{s^2}{2}.$$

Logo $M_{Z_n}(s) \rightarrow e^{s^2/2}$ numa vizinhança de zero. Para passar disso à convergência em distribuição, repita a mesma expansão de segunda ordem com a função característica — ela requer apenas $\mathbf{E}Y^2 < \infty$ e dá $\phi_{Z_n}(s) = [\phi_Y(s/\sqrt{n})]^n \rightarrow e^{-s^2/2}$ — e aplique o Teorema da Continuidade de Lévy 12.16, obtendo $Z_n \rightarrow_d N(0,1)$.

12.10 Para todo $\varepsilon > 0$, como $s_n \rightarrow \infty$, eventualmente $\varepsilon s_n > M$. Então

$$\frac{1}{s_n^2} \sum_k \mathbf{E} \left[X_{n,k}^2 \mathbb{1}_{\{|X_{n,k}| > \varepsilon s_n\}} \right] = 0.$$

A condição de Lindeberg está satisfeita; como a variância da soma normalizada é um, Lindeberg–Feller fornece o limite normal padrão.

12.11 Considere o arranjo $Y_{n,k} = a_{n,k}X_k$. Ele é centrado e a soma das variâncias é $\sum_k a_{n,k}^2 = 1$. Se $a_n^* = \max_k |a_{n,k}|$, então a soma de Lindeberg é majorada por

$$\sum_k a_{n,k}^2 \mathbf{E} \left[X_1^2 \mathbb{1}_{\{|X_1| > \varepsilon/a_n^*\}} \right] = \mathbf{E} \left[X_1^2 \mathbb{1}_{\{|X_1| > \varepsilon/a_n^*\}} \right] \rightarrow 0.$$

Lindeberg–Feller conclui a convergência.

12.12 Ponha $X_{n,k} = B_{n,k} - p_{n,k}$. Então $|X_{n,k}| \leq 1$, $\sum_k \text{Var}(X_{n,k}) = s_n^2 \rightarrow \infty$ e, para cada $\varepsilon > 0$, $|X_{n,k}| \leq \varepsilon s_n$ para todo k e todo n suficientemente grande. A soma de Lindeberg é, portanto, eventualmente zero. Lindeberg–Feller dá $s_n^{-1} \sum_k X_{n,k} \rightarrow_d N(0,1)$, que é a expressão do enunciado.

12.13 Para $M > 0$, o evento $\{\limsup_n S_n/\sqrt{n} \geq M\}$ é caudal: retirar finitos termos altera $S_n/\sqrt{n}$ por uma quantidade que tende a zero. Pelo teorema do limite central,

$$\mathbf{P}(S_n/\sqrt{n} > M) \rightarrow \mathbf{P}(N(0,\sigma^2) > M) > 0.$$

Logo a desigualdade de Fatou para eventos dá probabilidade positiva ao evento caudal acima; pela lei 0–1 de Kolmogorov, sua probabilidade é um. Intersectando sobre $M \in \mathbb{N}$, obtemos o limite superior infinito. Aplicando o mesmo argumento a $-X_n$, obtemos o limite inferior igual a $-\infty$.

13.1 Aditividade enumerável segue de Convergência Dominada aplicada às somas de indicadoras disjuntas, e $|\nu(\Omega)| \leq \int |f| \, d\mu < \infty$. Tome $P = \{f \geq 0\}$ e $N = \{f < 0\}$. Então

$$\nu^+(A) = \int_A f^+ \, d\mu, \quad \nu^-(A) = \int_A f^- \, d\mu, \quad |\nu|(A) = \int_A |f| \, d\mu.$$

13.2 Uma decomposição é $P = \{i : a_i \geq 0\}$, $N = \{i : a_i < 0\}$. Para todo A ,

$$\nu^+(A) = \sum_{i \in A} a_i^+, \quad \nu^-(A) = \sum_{i \in A} a_i^-, \quad |\nu|(A) = \sum_{i \in A} |a_i|.$$

Os índices com $a_i = 0$ podem ser postos em qualquer uma das duas partes.

13.3 A diferença de duas medidas finitas é uma medida com sinal. Para qualquer partição $A = \bigsqcup_j A_j$, $\sum_j |\nu(A_j)| \leq \sum_j (\mu_1(A_j) + \mu_2(A_j)) = \rho(A)$; tomando o supremo, $|\nu| \leq \rho$. Se $h_i = d\mu_i / d\rho$, então $\nu(A) = \int_A (h_1 - h_2) \, d\rho$ e

$$\nu^+ = (h_1 - h_2)^+ \rho, \quad \nu^- = (h_1 - h_2)^- \rho.$$

13.4 $P \cap N'$ é simultaneamente positivo e negativo: toda parte mensurável tem medida zero. Logo é $|\nu|$ -nulo; o mesmo vale para $N \cap P'$. Como $P \triangle P' \subseteq (P \cap N') \cup (N \cap P')$, essa diferença é nula, e analogamente para $N \triangle N'$. As fórmulas $\nu^+(A) = \nu(A \cap P)$ e $\nu^-(A) = -\nu(A \cap N)$ não mudam ao trocar a decomposição, provando a unicidade.

13.5 Para qualquer partição,

$$\sum_j |\nu(A_j)| \leq \sum_j (\nu^+(A_j) + \nu^-(A_j)) = |\nu|(A).$$

Se (P, N) é uma decomposição de Hahn, a partição $A = (A \cap P) \sqcup (A \cap N)$ atinge a igualdade, pois seus termos somam $\nu^+(A) + \nu^-(A)$. A partição trivial $\{A\}$ dá imediatamente $|\nu(A)| \leq |\nu|(A)$.

13.6 Escolha versões mensuráveis não negativas $f = d\lambda / d\nu$ e $g = d\nu / d\mu$, tornando-as finitas fora de conjuntos nulos e definindo $f = 0$ onde sua versão não

estiver determinada. Como esse conjunto é ν -nulo, $g = 0$ nele μ -quase certamente. Para todo A ,

$$\lambda(A) = \int_A f \, d\nu = \int_A fg \, d\mu.$$

Pela unicidade de Radon–Nikodým, $fg = d\lambda/d\mu$ μ -quase certamente.

13.7 Por Radon–Nikodým, $Z \geq 0$ P -q.c. e $\mathbf{E}_P Z = Q(\Omega) = 1$. Para uma indicadora, $\mathbf{E}_Q \mathbb{1}_A = Q(A) = \int_A Z \, dP$; linearidade, Convergência Monótona e decomposição em partes positiva e negativa estendem a fórmula a todo $X \geq 0$. Para X com sinal, ela vale exatamente quando $\mathbf{E}_P(|X|Z) < \infty$, condição equivalente a $X \in L^1(Q)$.

13.8 Como δ_1 está concentrada em $\{1\}$ e $\mu(\{1\}) = 0$, ela é singular em relação a μ . Já $2\lambda + 3\delta_0 \ll \mu$. Portanto

$$\nu_a = 2\lambda + 3\delta_0, \quad \nu_s = \delta_1,$$

e uma versão da densidade é

$$\frac{d\nu_a}{d\mu}(x) = 2\mathbb{1}_{\mathbb{R} \setminus \{0\}}(x) + 3\mathbb{1}_{\{0\}}(x).$$

14.1 O lado direito é $\mathcal{G}$ -mensurável. Se $G \in \mathcal{G}$, então G é união de algumas células A_j , e a integral do lado direito sobre G é

$$\sum_{A_j \subseteq G} \frac{\mathbf{E}[X\mathbb{1}_{A_j}]}{\mathbf{P}(A_j)} \mathbf{P}(A_j) = \mathbf{E}[X\mathbb{1}_G].$$

Isso caracteriza a esperança condicional.

14.2 $g(Y)$ é $\sigma(Y)$ -mensurável. Para cada átomo $\{Y = y_j\}$,

$$\int_{\{Y=y_j\}} g(Y) \, dP = g(y_j) \mathbf{P}(Y = y_j) = \mathbf{E}[X\mathbb{1}_{\{Y=y_j\}}].$$

Somando átomos obtemos a identidade para todo evento de $\sigma(Y)$. Nos valores que Y assume com probabilidade zero, g pode ser definida arbitrariamente, pois isso não altera $g(Y)$ quase certamente.

14.3 A constante $\mathbf{E}X$ é $\mathcal{G}$ -mensurável. Para $G \in \mathcal{G}$, a independência de G e X (primeiro para simples e depois por aproximação) dá $\mathbf{E}[X1_G] = \mathbf{E}X \mathbf{P}(G) = \int_G \mathbf{E}X \, dP$. A caracterização da esperança condicional conclui.

14.4 $Y = \mathbf{E}[X \mid \mathcal{G}_2]$ é integrável. Para $G \in \mathcal{G}_1 \subseteq \mathcal{G}_2$, $\int_G Y \, dP = \int_G X \, dP$; logo $\mathbf{E}[Y \mid \mathcal{G}_1] = \mathbf{E}[X \mid \mathcal{G}_1]$. Para a recíproca, tome $\mathcal{G}_1 = \{\emptyset, \Omega\}$, $\mathcal{G}_2 = \sigma(X)$ e X integrável não constante. Então $\mathbf{E}[\mathbf{E}[X \mid \mathcal{G}_1] \mid \mathcal{G}_2] = \mathbf{E}X \neq X$.

14.5 Condiicionado a $S_n = m$, todas as configurações com m sucessos têm a mesma probabilidade. Por simetria, $\mathbf{P}(X_k = 1 \mid S_n = m) = m/n$; portanto $\mathbf{E}[X_k \mid S_n = m] = m/n$. Equivalentemente, as n esperanças condicionais são iguais e sua soma é $\mathbf{E}[S_n \mid S_n] = S_n$.

14.6 Ponha $M = \mathbf{E}[X \mid \mathcal{G}]$. Como $X - \mathbf{E}X = (X - M) + (M - \mathbf{E}X)$, ao elevar ao quadrado o termo cruzado tem esperança zero: $\mathbf{E}[(X - M)(M - \mathbf{E}X)] = \mathbf{E}[(M - \mathbf{E}X)\mathbf{E}[X - M \mid \mathcal{G}]] = 0$. Logo $\text{Var } X = \mathbf{E}[(X - M)^2] + \mathbf{E}[(M - \mathbf{E}X)^2]$, que são os dois termos exibidos.

14.7 Em L^1 bastam a linearidade e a positividade da esperança condicional: de $\mathbf{E}[X \mid \mathcal{G}] = \mathbf{E}[X^+ \mid \mathcal{G}] - \mathbf{E}[X^- \mid \mathcal{G}]$, com as duas parcelas não negativas, segue

$$|\mathbf{E}[X \mid \mathcal{G}]| \leq \mathbf{E}[X^+ \mid \mathcal{G}] + \mathbf{E}[X^- \mid \mathcal{G}] = \mathbf{E}[|X| \mid \mathcal{G}],$$

e integrando obtemos $\|\mathbf{E}[X \mid \mathcal{G}]\|_1 \leq \|X\|_1$. (É a desigualdade de Jensen condicional aplicada a $|\cdot|$, aqui obtida diretamente.) Em L^2 usa-se a interpretação como projeção: como $\mathbf{E}[\cdot \mid \mathcal{G}]$ é a projeção ortogonal sobre $L^2(\mathcal{G})$, a decomposição $X = \mathbf{E}[X \mid \mathcal{G}] + (X - \mathbf{E}[X \mid \mathcal{G}])$ é ortogonal e o teorema de Pitágoras dá

$$\|X\|_2^2 = \|\mathbf{E}[X \mid \mathcal{G}]\|_2^2 + \|X - \mathbf{E}[X \mid \mathcal{G}]\|_2^2 \geq \|\mathbf{E}[X \mid \mathcal{G}]\|_2^2.$$

14.8 Ponha $U = X - Y$ e $V = X + Y = S$. Como X e Y são independentes e normais padrão, a densidade conjunta de (X, Y) é

$$f_{X,Y}(x,y) = \frac{1}{2\pi} \exp\left\{-\frac{x^2 + y^2}{2}\right\}.$$

A transformação inversa é

$$x = \frac{u+v}{2}, \quad y = \frac{v-u}{2},$$

cujo jacobiano tem módulo $1/2$. Além disso, $x^2 + y^2 = (u^2 + v^2)/2$. Pela fórmula de mudança de variáveis,

$$f_{U,V}(u,v) = \frac{1}{4\pi} \exp\left\{-\frac{u^2 + v^2}{4}\right\} = \left(\frac{1}{\sqrt{4\pi}} e^{-u^2/4}\right) \left(\frac{1}{\sqrt{4\pi}} e^{-v^2/4}\right).$$

Portanto U e V são independentes, ambos com distribuição $N(0,2)$. Em particular, $(X - Y)/2 = U/2$ é independente de $S = V$ e tem média zero. Como

$$X = \frac{S}{2} + \frac{X - Y}{2},$$

segue que

$$\mathbf{E}[X | S] = \frac{S}{2}.$$

Finalmente,

$$\mathbf{E}[(X - S/2)^2] = \text{Var}\left(\frac{X - Y}{2}\right) = \frac{1}{4}(\text{Var } X + \text{Var } Y) = \frac{1}{2},$$

enquanto o estimador constante 0 tem erro quadrático $\mathbf{E}[X^2] = 1$.

14.9 As variáveis $\mathcal{G}$ -mensuráveis são constantes em A e em A^c . Impor a igualdade das integrais sobre esses dois átomos determina as constantes como $\mathbf{E}[X | A]$ e $\mathbf{E}[X | A^c]$. Por exemplo, em $\Omega = \{1,2,3,4\}$ equiprovável, tome $A = \{1,2\}$ e $X(1), \dots, X(4) = (1,3,0,2)$. Então $\mathbf{E}[X | \sigma(A)] = 2\mathbb{1}_A + 1\mathbb{1}_{A^c}$.

15.1 M_n é $\mathcal{F}_n$ -mensurável, integrável e, pela torre, $\mathbf{E}[M_{n+1} | \mathcal{F}_n] = \mathbf{E}[Z | \mathcal{F}_n] = M_n$. A família das esperanças condicionais de uma única variável L^1 é uniformemente integrável: aproxime Z em L^1 por uma variável limitada e use a contração condicional em L^1 . Pelo Teorema de Convergência de Doob e pela propriedade caracterizadora,

$$M_n \longrightarrow \mathbf{E}[Z | \mathcal{F}_\infty]$$

quase certamente e em L^1 .

15.2 Como ξ_{n+1} é independente de $\mathcal{F}_n$ e centrada, $\mathbf{E}[S_{n+1} \mid \mathcal{F}_n] = S_n$. Além disso,

$$\mathbf{E}[S_{n+1}^2 \mid \mathcal{F}_n] = S_n^2 + 2S_n\mathbf{E}[\xi_{n+1} \mid \mathcal{F}_n] + \mathbf{E}[\xi_{n+1}^2] = S_n^2 + \sigma^2,$$

o que fornece a segunda propriedade após subtrair $(n+1)\sigma^2$.

15.3 Independência do novo incremento dá

$$\mathbf{E}[M_{n+1} \mid \mathcal{F}_n] = \frac{e^{\theta S_n}}{m(\theta)^{n+1}} \mathbf{E}[e^{\theta \xi_{n+1}}] = M_n.$$

Integrabilidade e adaptação são imediatas. Como $e^{\theta \xi_1} > 0$ quase certamente, sua esperança, quando finita, é estritamente positiva; portanto o caso $m(\theta) = 0$ não pode ocorrer.

15.4 Cada soma é adaptada e integrável, pois H é previsível e limitado. Como H_k é $\mathcal{F}_{k-1}$ -mensurável,

$$\mathbf{E}[H_k(M_k - M_{k-1}) \mid \mathcal{F}_{k-1}] = H_k \mathbf{E}[M_k - M_{k-1} \mid \mathcal{F}_{k-1}] = 0.$$

Logo as somas parciais formam uma martingala. Assim qualquer aposta baseada apenas no passado multiplica um incremento de média condicional zero e não cria ganho esperado.

15.5 Temos $\{\tau_a \leq n\} = \{\max_{0 \leq k \leq n} S_k \geq a\} \in \mathcal{F}_n$, logo τ_a é tempo de parada. Para $n < N$, decidir $\{\rho \leq n\}$ exige saber se haverá uma visita a zero entre $n+1$ e N . Duas trajetórias com o mesmo prefixo até n podem diferir por um retorno posterior a zero; portanto esse evento não pertence, em geral, a $\mathcal{F}_n$.

15.6 Aplique primeiro o Teorema de Amostragem Opcional Limitada a $\tau \wedge N$. O processo parado $S_{\tau \wedge N}$ fica entre $-a$ e b . Além disso, em cada bloco de $a+b$ passos há uma probabilidade positiva, uniforme no ponto inicial, de atingir uma das barreiras; portanto $\tau < \infty$ quase certamente. Logo a passagem $N \rightarrow \infty$ é justificada por convergência dominada. Se $p = \mathbf{P}(S_\tau = b)$, então $0 = \mathbf{E}S_\tau = bp - a(1-p)$ e $p = a/(a+b)$.

Para $S_n^2 - n$, a amostragem no tempo limitado $\tau \wedge N$ dá $\mathbf{E}[S_{\tau \wedge N}^2] = \mathbf{E}[\tau \wedge N]$. O lado esquerdo é dominado por $\max\{a^2, b^2\}$; convergência dominada à esquerda e monótona à direita mostram que $\mathbf{E}\tau = \mathbf{E}S_\tau^2 < \infty$. Assim

$$\mathbf{E}\tau = b^2 \frac{a}{a+b} + a^2 \frac{b}{a+b} = ab.$$

15.7 Se H_n é o evento das primeiras n caras, então

$$\mathbf{E}[M_{n+1} \mid \mathcal{F}_n] = 2^{n+1} \mathbf{1}_{H_n}(1/2) = M_n, \quad \mathbf{E}M_n = 2^n 2^{-n} = 1.$$

Quase toda sequência apresenta uma coroa em tempo finito; depois disso $M_n = 0$, logo $M_n \rightarrow 0$ quase certamente. Se a família fosse uniformemente integrável, essa convergência implicaria convergência em L^1 e, portanto, $\mathbf{E}M_n \rightarrow 0$, contradizendo $\mathbf{E}M_n = 1$.

15.8 Como $\mathbf{E}[S_n^2 - S_{n-1}^2 \mid \mathcal{F}_{n-1}] = \sigma^2$, o compensador é $A_n = n\sigma^2$ (com $A_0 = 0$). Logo a decomposição é

$$S_n^2 = M_n + A_n, \quad M_n = S_n^2 - n\sigma^2,$$

onde M é martingala e A é previsível, crescente e inicia em zero.

15.9 S_k^2 é uma submartingala não negativa. A desigualdade maximal de Doob fornece

$$\mathbf{P}\left(\max_{k \leq n} S_k^2 \geq r^2\right) \leq \frac{\mathbf{E}[S_n^2]}{r^2}.$$

Independência e centramento dão $\mathbf{E}[S_n^2] = n\sigma^2$, e o evento à esquerda é exatamente $\{\max_{k \leq n} |S_k| \geq r\}$. Assim recuperamos, neste caso, a desigualdade maximal de Kolmogorov por uma rota mais geral.

16.1 Para $s \leq t$, escreva $B(t) = B(s) + [B(t) - B(s)]$; o incremento é independente de $B(s)$ e centrado, logo $\mathbf{E}[B(s)B(t)] = \mathbf{E}[B(s)^2] = s$. Assim a matriz é $\Sigma_{ij} = t_i \wedge t_j$. Incrementos sobre intervalos disjuntos são combinações lineares do vetor gaussiano e têm covariância zero, calculada com essa fórmula. Um vetor gaussiano com covariâncias cruzadas nulas tem blocos independentes, recuperando a independência dos incrementos.

16.2 Ponha $\ell = b - a = 2^{-(n-1)}$. Condicionados os extremos, as duas metades devem somar $B(b) - B(a)$ e, antes do condicionamento, são normais independentes $N(0, \ell/2)$. A primeira metade, condicionada à soma, tem média $[B(b) - B(a)]/2$ e variância $\ell/4$. Portanto

$$\mathbf{E}[B(d) \mid B(a), B(b)] = \frac{B(a) + B(b)}{2}, \quad \text{Var}(B(d) \mid B(a), B(b)) = 2^{-(n+1)},$$

que dá exatamente o coeficiente do Z_d exibido.

16.3 No nível n há no máximo 2^n coeficientes normais padrão. Pela cauda gaussiana e união,

$$\mathbf{P}\left(\max_k |Z_{n,k}| > c\sqrt{n}\right) \leq 2^{n+1} e^{-c^2 n/2}.$$

A série converge quando $c^2/2 > \log 2$; Borel–Cantelli dá a cota máxima eventualmente. Como cada tenda tem altura $2^{-(n+1)/2} |Z_{n,k}|$ e os suportes no mesmo nível são disjuntos, $\|F_n\|_\infty \leq c\sqrt{n} 2^{-(n+1)/2}$. A soma dessas cotas converge, e o teste de Weierstrass dá convergência uniforme.

16.4 $\tilde{B}$ é gaussiano centrado, contínuo, e

$$\text{Cov}(\tilde{B}(s), \tilde{B}(t)) = a^{-2}(a^2 s \wedge a^2 t) = s \wedge t.$$

Logo tem as leis finito-dimensionais brownianas. Para $0 < s \leq t$,

$$\text{Cov}(\hat{B}(s), \hat{B}(t)) = st(1/s \wedge 1/t) = s = s \wedge t.$$

O processo é gaussiano centrado e é contínuo em $(0, \infty)$; resta apenas $tB(1/t) \rightarrow 0$ em $t \downarrow 0$, equivalente a $B(u)/u \rightarrow 0$ quando $u \rightarrow \infty$, que segue da lei forte para os incrementos e do controle entre tempos inteiros.

16.5 Como transformação linear de um processo gaussiano, β é gaussiano e centrado. Para $s, t \in [0, 1]$,

$$\text{Cov}(\beta(s), \beta(t)) = s \wedge t - st.$$

Além disso, $\beta(0) = \beta(1) = 0$; em particular sua variância no tempo um é zero, portanto não é browniano. Finalmente, $\text{Cov}(\beta(t), B(1)) = t - t = 0$ para todo t . Gaussianidade conjunta implica independência entre o processo β e $B(1)$.

16.6 O módulo de Lévy fornece, para h pequeno em um compacto, $|B(t+h) - B(t)| \leq C\sqrt{h \log(1/h)}$ quase certamente. Como

$$\frac{\sqrt{h \log(1/h)}}{h^\alpha} = h^{1/2-\alpha} \sqrt{\log(1/h)} \rightarrow 0$$

quando $\alpha < 1/2$, obtém-se uma cota de Hölder. Para $\alpha = 1/2$, o quociente é $\sqrt{\log(1/h)}$, que diverge; por isso o argumento não fecha no expoente crítico.

16.7 Os incrementos são independentes $N(0, T/n)$. Se $Z \sim N(0, v)$, então $\mathbf{E}Z^2 = v$ e $\text{Var}(Z^2) = 2v^2$. Portanto

$$\mathbf{E}Q_n = T, \quad \text{Var}(Q_n) = n \cdot 2(T/n)^2 = \frac{2T^2}{n}.$$

Logo $\mathbf{E}[(Q_n - T)^2] \rightarrow 0$. Isso dá convergência em L^2 para as partições uniformes; o teorema para partições encaixadas reforça a conclusão para convergência quase certa.

16.8 Para uma função contínua f de variação total $V < \infty$,

$$\sum_{[u,v] \in \pi} (f(v) - f(u))^2 \leq \max_{[u,v] \in \pi} |f(v) - f(u)| V.$$

Uniforme continuidade faz o máximo tender a zero com a malha, logo a variação quadrática é nula. A variação quadrática browniana em $[0, T]$ é $T > 0$ quase certamente; portanto uma trajetória browniana não pode ter variação total finita em nenhum intervalo não degenerado (aplique o fato a intervalos com extremos racionais e use inclusão).

16.9 Uma união não enumerável de eventos nulos pode ter probabilidade positiva ou um; a subaditividade só controla uniões enumeráveis. Paley–Wiener–Zygmund substitui os tempos arbitrários por uma hierarquia enumerável de intervalos diádicos: se houvesse derivada finita em algum ponto, os incrementos em todos os intervalos suficientemente pequenos que o contêm obedeceriam a uma cota linear. Estimativas gaussianas uniformes sobre cada nível tornam as probabilidades desses eventos somáveis; Borel–Cantelli exclui-os simultaneamente. A discretização multiescala enumerável é o passo que falta ao argumento ponto a ponto.

Bibliografia

- [1] Mark (<https://mathoverflow.net/users/7392/mark>). *Demystifying the Carathéodory Approach to Measurability*. MathOverflow. Disponível em <https://mathoverflow.net/q/34029>. Versão de 2010-07-31.
- [2] Charalambos D Aliprantis e Kim Border. *Infinite Dimensional Analysis: A Hitchhiker's Guide*. Springer Science & Business Media, 2006.
- [3] Robert Ash. *Real Analysis and Probability*. Vol. 6. Academic Press, 1972, p. 70.
- [4] Robert G Bartle. *The elements of integration and Lebesgue measure*. John Wiley & Sons, 2014.
- [5] Andrej Bauer. *Why do roots of polynomials tend to have absolute value close to 1?* MathOverflow <https://mathoverflow.net/q/182412>. 2014.
- [6] Heinz Bauer e R... B. Burckel. *Probability theory and elements of measure theory*. Academic Press London, 1981.
- [7] V Bentkus et al. *Limit theorems of probability theory*. Springer Science & Business Media, 2013.
- [8] Patrick Billingsley. *Convergence of probability measures*. John Wiley & Sons, 2013.
- [9] Patrick Billingsley. *Probability and measure*. John Wiley & Sons, 2008.

- [10] Salomon Bochner. *Harmonic analysis and the theory of probability*. Courier Corporation, 2013.
- [11] Vladimir I Bogachev. *Measure theory*. Vol. 1. Springer Science & Business Media, 2007.
- [12] Leo Breiman. *Probability*. 1992.
- [13] Marek Capinski e Peter E Kopp. *Measure, integral and probability*. Springer Science & Business Media, 2013.
- [14] Jair Donadelli. *Processos de Ramificação*. <https://anotacoesdeaula.wordpress.com>. 2016.
- [15] Rick Durrett. *Probability: theory and examples*. Cambridge university press, 2010.
- [16] Terrence L Fine. *Theories of probability: An examination of foundations*. Academic Press, 2014.
- [17] Gerald B Folland. *Real analysis: modern techniques and their applications*. John Wiley & Sons, 2013.
- [18] Geoffrey Grimmett e Dominic Welsh. *Probability: an introduction*. Oxford University Press, 2014.
- [19] Allan Gut. *Probability: a graduate course*. Springer Science & Business Media, 2012.
- [20] Frank den Hollander. “Probability theory: The coupling method”. Em: *Lectures Notes-Mathematical Institute, Leiden University* (2012).
- [21] Alan Karr. *Probability*. Springer Science & Business Media, 1992.
- [22] Achim Klenke. *Probability theory: a comprehensive course*. Springer Science & Business Media, 2013.
- [23] Emmanuel Kowalski. “Bernstein polynomials and Brownian motion”. Em: *American Mathematical Monthly* 113.10 (2006), pp. 865–886.
- [24] Emmanuel Lesigne. *Heads or tails: An introduction to limit theorems in probability*. Vol. 28. American Mathematical Soc., 2005.
- [25] KR Parthasarathy. *CONVERGENCE OF PROBABILITY MEASURES*. 1970.

- [26] Marcus Pivato. "Analysis, Measure, and Probability: A visual introduction". Em: (2003).
- [27] David Pollard. *A user's guide to measure theoretic probability*. Vol. 8. Cambridge University Press, 2002.
- [28] Sidney I Resnick. *A probability path*. Springer Science & Business Media, 2013.
- [29] Jeffrey S Rosenthal. *A first look at rigorous probability theory*. World Scientific Publishing Co Inc, 2006.
- [30] Halsey Lawrence Royden e Patrick Fitzpatrick. *Real analysis*. Vol. 198. 8. Macmillan New York, 1988.
- [31] Luis A Santaló. *Integral geometry and geometric probability*. Cambridge university press, 2004.
- [32] Klaus D Schmidt. "The Cantor set in probability theory". Em: (1991).
- [33] Albert N Shiryaev. *Probability*. Springer-Verlag, New York, 1996.
- [34] Maurice Sion. "HISTORY on MEASURE THEORY IN THE TWENTIETH CENTURY". Em: (1990).
- [35] Jan Von Plato. *Creating modern probability: Its mathematics, physics and philosophy in historical perspective*. Cambridge University Press, 1994.