

Notas de aula: Probabilidade I

Augusto Teixeira

27 de agosto de 2024

Licença

Esse trabalho é licenciado nos termos da licença [Creative Commons Atribuição-CompartilhaIgual 3.0 Não Adaptada \(CC BY-SA 3.0\)](#). Assim, qualquer um pode usar, distribuir e modificar o conteúdo em obras derivadas livremente, incluindo para fins comerciais, desde com a devida citação da fonte. Qualquer violação dos termos da licença citada será considerado uso ilegal.

Contribuições

Somos gratos especialmente a Hubert Lacoïn, pela revisão do texto, assim como pelas colaborações autorais.

Também gostaríamos de agradecer

Roberto Imbuzeiro de Oliveira

Milton Jara

Cláudio Landim

Conrado Costa

Rangel Baldasso

por diversas discussões, sugestões e correções.

Sumário

Prefácio	iv
1 Fundamentos	1
1.1 Espaços mensuráveis	2
1.2 Espaços de probabilidade	3
1.3 Sistemas λ - π	6
1.3.1 Igualdade de probabilidades	8
1.4 Elementos aleatórios	8
1.4.1 Distribuição de elementos aleatórios	9
Tópico: O paradoxo de Bertrand	11
2 Construção de espaços de probabilidade	13
2.1 Caso enumerável	13
Tópico: Método Probabilístico	15
2.2 Caso absolutamente contínuo	17
2.3 Funções acumuladas de distribuição	17
2.4 Espaços produto finito	19
2.5 Independência	20
2.5.1 Coleções de eventos	21
2.5.2 Independência de σ -álgebras	21
Tópico: Lei dos pequenos números	24
2.6 Espaços produto infinito	29
2.6.1 Recordar é viver...	29
2.6.2 Teorema da Extensão de Kolmogorov	30
Tópico: Percolação	34
2.7 Distribuições conjuntas	37
2.8 Probabilidades condicionais	37
2.8.1 Regra de Bayes	40

SUMÁRIO

2.9	Núcleos de transição	41
2.10	Espaços canônicos	46
2.10.1	Espaços poloneses	48
	Tópico: Cadeias de Markov	52
	Tópico: Urna de Pólya	56
3	Somas de variáveis independentes	61
3.1	Esperança	61
3.1.1	Desigualdade de Markov	64
3.1.2	Esperança e independência	65
3.2	Variância	65
3.3	Lei fraca dos grandes números	68
	Tópico: Contando triângulos	70
3.4	Lei forte dos grandes números	72
3.5	Lei $\{0, 1\}$ de Kolmogorov	74
3.6	Momentos exponenciais	77
3.7	Princípio de Grandes Desvios	79
	Tópico: Funções características	85
3.8	O Teorema Central do Limite	86
3.8.1	A distribuição normal	86
3.8.2	Convergência fraca	90
3.8.3	Convergência fraca em \mathbb{R}	92
3.8.4	O TCL para uma sequência i.i.d.	93
	Tópico: O Teorema de Portmanteau	96
4	Esperança condicional	99
4.1	Esperança condicional	99
4.2	Propriedades básicas da esperança condicional	102
4.3	Probabilidade Condicional Regular	108
4.4	Princípio da substituição	110
	Tópico: Processos de Poisson em \mathbb{R}	114
5	Soluções de exercícios	119
	Referências Bibliográficas	121
	Index	121
	Índice Remissivo	122

SUMÁRIO

Capítulo 1

Fundamentos

A probabilidade moderna se baseia fortemente na Teoria da Medida e supomos durante esse curso que o leitor esteja bem familiarizado com conceitos tais como: Medida de Lebesgue, extensões de medida e teoremas de convergência. Iremos agora justificar brevemente a escolha da Teoria da Medida para o estudo de probabilidade.

No início da Teoria da Probabilidade, a maioria dos fenômenos estudados apresentava apenas um número finito de resultados possíveis, como por exemplo ao se jogar um dado de seis lados ou sortear uma carta em um baralho. Em tais casos é desnecessário o uso de ferramentas sofisticadas pra modelar tais situações. Por exemplo, podemos simplesmente dizer que a probabilidade de se obter cada um dos lados do dado é igual a $1/6$.

Mas digamos por exemplo que queremos um modelo para estudar o volume de chuva em uma cidade durante um ano. Obviamente, esse volume poderia ser qualquer número real positivo e não podemos simplesmente atribuir valores positivos de probabilidade a cada número real (lembramos que somas não enumeráveis de termos positivos são sempre infinitas). Mas como podemos continuar nossa modelagem se nem ao menos podemos dizer qual é a probabilidade de chover um determinado volume esse ano, por exemplo $(\pi/19)mm$?

A solução para tal dilema, se baseia no fato de que na verdade nunca estamos interessados no exato resultado do nosso experimento. Gostaríamos sim de responder perguntas do tipo: qual é a probabilidade de que chova entre zero e $37mm$? Estamos portanto interessados em atribuir probabilidades não a valores exatos do experimento, mas a certos conjuntos de possíveis valores. Chamamos tais conjuntos de *eventos*.

Voltando ao caso do dado de seis lados, poderíamos nos interessar por exemplo pela probabilidade dos seguintes eventos: o lado sorteado foi ímpar

$(P(\{1, 3, 5\}) = 1/2)$ ou o lado sertado foi dois ($P(\{2\}) = 1/6$). E percebemos rapidamente que para eventos disjuntos a probabilidade de sua união é a soma de suas probabilidades (no caso acima, $P(\{1, 2, 3, 5\}) = 1/2 + 1/6 = 2/3$). Esse caráter aditivo da probabilidade certamente nos remete aos conceitos básicos de Teoria da Medida. Vamos agora formalizar a discussão acima com mais calma, sob a ótica dessa teoria.

1.1 Espaços mensuráveis

Denotaremos sempre por Ω o nosso *espaço amostral* (a princípio qualquer conjunto). Um ponto nesse espaço corresponde por exemplo a um possível resultado do nosso experimento aleatório.

Exemplo 1.1.1. *Possíveis exemplos de espaço amostral*

- a) $\Omega_1 = \{1, 2, \dots, 6\}$,
- b) $\Omega_2 = \mathbb{R}_+$,
- c) $\Omega_3 = \{f : [0, 1] \rightarrow \mathbb{R}; f \text{ é contínua}\}$.

Os exemplos acima poderiam ser usados em modelar por exemplo: o resultado de um dado, o volume anual de chuva em uma cidade e o comportamento ao longo do dia do preço de uma ação na bolsa de valores.

Consideraremos sempre Ω 's equipados com uma σ -álgebra denotada por \mathcal{F} . Mais precisamente

Definição 1.1.1. *Dizemos que $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ é uma σ -álgebra se*

- a) $\Omega \in \mathcal{F}$,
- b) $A \in \mathcal{F}$ implica que $A^c \in \mathcal{F}$ e
- c) se $A_1, A_2, \dots \in \mathcal{F}$, então $\cup_i A_i \in \mathcal{F}$.

Nesse caso, dizemos que (Ω, \mathcal{F}) é um *espaço mensurável* e os elementos $A \in \mathcal{F}$ são chamados de *eventos*.

Se $\mathcal{G} \subseteq \mathcal{P}(\Omega)$ (que chamamos de uma classe ou família), denotamos por $\sigma(\mathcal{G})$ a σ -álgebra gerada por \mathcal{G} , que é a menor σ -álgebra contendo \mathcal{G} (ou em outras palavras, a interseção de todas σ -álgebras que contém \mathcal{G}). Um exemplo importante é dado pela σ -álgebra de Borel, gerada pelos abertos de uma topologia em Ω .

Exemplo 1.1.2. *Típicos exemplos de σ -álgebra correspondentes aos espaços amostrais do Exemplo 1.1.1*

- a) $\mathcal{F}_1 = \mathcal{P}(\Omega_1)$,
- b) $\mathcal{F}_2 = \mathcal{B}([0, 1])$ e

c) $\mathcal{F}_3 = \mathcal{B}(C[0, 1])$.

Exemplo 1.1.3. Alguns eventos de $\mathcal{F}_1, \mathcal{F}_2$ e \mathcal{F}_3 acima

- a) $\{x \text{ é ímpar}\}, \{1\} \subset \Omega_1$,
- b) $[0, 1/2], \{0\}, (Q \cap [0, 1]) \subset \Omega_2$ e
- c) $\{f : [0, 1] \rightarrow \mathbb{R}; f(1) > 0\} \subset \Omega_3$.

Exercício 1.1.4. Mostre que $\{f : [0, 1] \rightarrow \mathbb{R}; f(t) \geq 0 \text{ para todo } t \in [0, 1]\} \subset \Omega_3$ é um evento (ou seja, pertence a \mathcal{F}_3).

Notação 1.1.2. Se Q for uma condição qualquer sobre candidatos $\omega \in \Omega$, escreveremos $[\omega \text{ satisfaz } Q]$ para denotar $\{\omega \in \Omega; \omega \text{ satisfaz } Q\}$.

Por exemplo, $\{f : [0, 1] \rightarrow \mathbb{R}; f(1) > 0\}$ pode ser escrita simplesmente como $[f(1) > 0]$.

1.2 Espaços de probabilidade

Agora estamos prontos para introduzir o conceito moderno do que é uma probabilidade.

Definição 1.2.1. Dado (Ω, \mathcal{F}) espaço mensurável, dizemos que $P : \mathcal{F} \rightarrow [0, 1]$ é uma probabilidade se

- a) $P(\Omega) = 1$ e
- b) Seja uma seqüência $(A_i)_{i \in I}$ finita ou enumerável de eventos disjuntos ($A_i \cap A_j = \emptyset$ se $i \neq j$), então
$$P(\bigcup_{i \in I} A_i) = \sum_{i \in I} P(A_i). \quad (1.1)$$

Obviamente, isso nada mais é que uma medida que associa massa um ao espaço todo.

Exemplo 1.2.1. Probabilidades nos espaços do Exemplo 1.1.1

- a) $P_1(A) = (\#A)/6$ em $(\Omega_1, \mathcal{F}_1)$. Ou mais geralmente $P'_1(A) = \sum_{i \in A} p_i$, onde $p_i \geq 0$ e $\sum_i p_i = 1$.
- b) P_2 pode ser a medida de Lebesgue em $([0, 1], \mathcal{B}([0, 1]))$. Mais geralmente também podemos ter $P'_2(A) = \int_A \rho(x) dx$, onde $\rho : [0, 1] \rightarrow \mathbb{R}_+$ é uma função mensurável, chamada densidade, tal que $\int_{[0, 1]} \rho(x) dx = 1$.
- c) $P_3 = \delta_0$, que atribui o valor um se o evento contém a função identicamente nula ($f \equiv 0$) e zero caso contrário.

Obviamente o terceiro exemplo é bastante artificial (e inútil). Mas, futuramente, estaremos prontos para introduzir medidas bem interessantes no espaço $(\Omega_3, \mathcal{F}_3)$.

Proposição 1.2.2. *Valem as afirmativas seguintes*

a) Se $A \subseteq B$ então $P(A) \leq P(B)$.

b) A cota da união: para I finito ou enumerável

$$P\left(\bigcup_{i \in I} A_i\right) \leq \sum_{i \in I} P(A_i). \quad (1.2)$$

c) O que chamamos de princípio da inclusão e exclusão

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{k=1}^n (-1)^{k-1} \sum_{1 \leq i_1 < \dots < i_k \leq n} P(A_{i_1} \cap \dots \cap A_{i_k}). \quad (1.3)$$

Demonstração. a) Como $A \cap (B \setminus A) = \emptyset$, então

$$P(B) = P(A \cup (B \setminus A)) = P(A) + P(B \setminus A) \geq P(A). \quad (1.4)$$

b) $P(A \cup B) = P(A \cup (B \setminus A)) = P(A) + P(B \setminus A) \leq P(A) + P(B)$.

Deixamos o caso enumerável como exercício abaixo.

c) Chamamos de A a união dos A_i . Basta mostrar a validade da equação abaixo e depois integrar com respeito a P .

$$\mathbb{1}_A(\omega) = \sum_{k=1}^n (-1)^{k-1} \sum_{\substack{I \subseteq \{1, \dots, n\} \\ |I|=k}} \prod_{i \in I} \mathbb{1}_{A_i}(\omega). \quad (1.5)$$

Para tanto, observe que para todo $\omega \in \Omega$,

$$(\mathbb{1}_A - \mathbb{1}_{A_1}) \cdot \dots \cdot (\mathbb{1}_A - \mathbb{1}_{A_n})(\omega) = 0. \quad (1.6)$$

Logo, expandindo o produto acima obtemos

$$\mathbb{1}_A + \sum_{k=1}^n \sum_{\substack{I \subseteq \{1, \dots, n\} \\ |I|=k}} (-1)^k \prod_{i \in I} \mathbb{1}_{A_i}(\omega) = 0, \quad (1.7)$$

que equivale a (1.5). □

Exercício 1.2.2. *Mostre que $P\left(\bigcup_i A_i\right) \leq \sum_i P(A_i)$ no caso enumerável.*

Exercício 1.2.3. *Mostre que*

$$P\left(\bigcup_{i=1}^m A_i\right) \leq \sum_{k=1}^m (-1)^{k-1} \sum_{1 \leq i_1 < \dots < i_k \leq m} P(A_{i_1} \cap \dots \cap A_{i_k}) \text{ se } m \text{ é ímpar e}$$

$$P\left(\bigcup_{i=1}^m A_i\right) \geq \sum_{k=1}^m (-1)^{k-1} \sum_{1 \leq i_1 < \dots < i_k \leq m} P(A_{i_1} \cap \dots \cap A_{i_k}) \text{ se } m \text{ é par.}$$

Exercício 1.2.4. Seja $n \geq 1$ um número inteiro e considere $\Omega = \{0, 1\}^n$, o hipercubo de dimensão n (cada $\omega \in \Omega$ pode ser visto como uma função $\omega : \{1, \dots, n\} \rightarrow \{0, 1\}$). Para cada $i \in \{1, \dots, n\}$, definimos o evento $A_i = \{\omega \in \Omega; \omega(i) = 1\}$. Dadas duas probabilidades P e P' em $(\Omega, \mathcal{P}(\Omega))$, mostre que se $P(B) = P'(B)$ para todos conjuntos B dados por interseções de A_i 's, então $P = P'$.

Proposição 1.2.3. Toda probabilidade P é contínua, isto é:

- a) Se $A_1 \subseteq A_2 \subseteq \dots \in \mathcal{F}$ for uma sequência crescente de eventos, então $\lim_{n \rightarrow \infty} P(A_n) = P(\bigcup_{n=1}^{\infty} A_n)$.
- b) Também, se $A_1 \supseteq A_2 \supseteq \dots \in \mathcal{F}$, temos $\lim_{n \rightarrow \infty} P(A_n) = P(\bigcap_{n=1}^{\infty} A_n)$.

Demonstração. a) Observe que

$$\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} \left(A_n \setminus \left(\bigcup_{i=1}^{n-1} A_i \right) \right), \quad (1.8)$$

que são disjuntos. Logo

$$\begin{aligned} P\left(\bigcup_{n=1}^{\infty} A_n\right) &= \sum_{n=1}^{\infty} P\left(A_n \setminus \left(\bigcup_{i=1}^{n-1} A_i\right)\right) \\ &= \lim_{n \rightarrow \infty} P\left(\bigcup_{i=1}^n A_i\right) = \lim_{n \rightarrow \infty} P(A_n). \end{aligned} \quad (1.9)$$

b) A prova é análoga à de (a). \square

Lema 1.2.4 (Borel-Cantelli - primeira parte). Sejam $A_1, A_2, \dots \in \mathcal{F}$ satisfazendo $\sum_{i=1}^{\infty} P(A_i) < \infty$. Então

$$P[A_i \text{ para infinitos } i] := P\left(\bigcap_{n=1}^{\infty} \left(\bigcup_{i \geq n} A_i\right)\right) = 0. \quad (1.10)$$

Demonstração. Estimamos

$$P\left(\bigcap_{n=1}^{\infty} \left(\bigcup_{i \geq n} A_i\right)\right) = \lim_{n \rightarrow \infty} P\left(\bigcup_{i \geq n} A_i\right) \leq \lim_{n \rightarrow \infty} \sum_{i \geq n} P(A_i) = 0. \quad (1.11)$$

O que termina a prova do lema. \square

Demonstração. Sejam $A_1, A_2, \dots \in \mathcal{F}$ satisfazendo $\sum_{i=1}^{\infty} P(A_i) < \infty$, e defina, para $n \geq 1$,

$$f_n = \sum_{k=1}^n \mathbf{1}_{A_k}, \quad (1.12)$$

de modo que

$$[A_i \text{ para infinitos } i] = \left[\lim_n f_n = \infty \right]. \quad (1.13)$$

Note agora que $(f_n)_{n \geq 1}$ forma uma sequência monótona de funções com $\lim_n f_n = \sum_{k=1}^{\infty} \mathbf{1}_{A_k}$. Em particular, o Teorema da convergência monótona implica que

$$\int \sum_{k=1}^{\infty} \mathbf{1}_{A_k} = \lim_n \int f_n dP = \lim_n \sum_{k=1}^n P(A_k) = \sum_{i=1}^{\infty} P(A_i) < \infty. \quad (1.14)$$

Isto nos dá que

$$P\left(\lim_n f_n = \infty\right) = 0, \quad (1.15)$$

e completa a demonstração. \square

Imagine que jogamos todos os dias em uma loteria e que nossa probabilidade de ganhar no dia i é p_i . Então se $\sum_i p_i < \infty$, sabemos que certamente não ganharemos infinitas vezes.

1.3 Sistemas λ - π

Uma importante ferramenta para provar fatos teóricos sobre probabilidades é o Teorema de Dynkin que apresentaremos nessa seção. Ele trata de classes de eventos que não são necessariamente σ -álgebras, mas sistemas λ ou π como definidos abaixo.

Definição 1.3.1. Dizemos que uma classe $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ é um π -sistema se for fechado por interseções finitas, isto é: para todos $A, B \in \mathcal{A}$ temos $A \cap B \in \mathcal{A}$.

Definição 1.3.2. Dizemos que $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ é um λ -sistema, se

- a) $\Omega \in \mathcal{A}$,
- b) Sempre que $A \in \mathcal{A}$ temos $A^c \in \mathcal{A}$.
- c) Para $A_1, A_2, \dots \in \mathcal{A}$ disjuntos dois a dois, temos $\cup_i A_i \in \mathcal{A}$.

Exercício 1.3.1. Dê um exemplo de λ -sistema que não seja uma σ -álgebra.

Definimos para $\mathcal{A} \subseteq \mathcal{P}(\Omega)$, o menor λ -sistema contendo \mathcal{A} , ou seja,

$$\lambda(\mathcal{A}) = \bigcap_{\substack{\mathcal{B} \text{ } \lambda\text{-sistema} \\ \mathcal{A} \subseteq \mathcal{B}}} \mathcal{B}. \quad (1.16)$$

É fácil ver que $\lambda(\mathcal{A})$ é sempre um λ -sistema.

Teorema 1.3.3 (Dynkin). Se \mathcal{A} é um π -sistema, então $\lambda(\mathcal{A}) = \sigma(\mathcal{A})$.

Note pelo Exercício 1.3.1 que a hipótese de que \mathcal{A} é um π -sistema é necessária em geral.

Demonstração. Obviamente, basta mostrar é que $\lambda(\mathcal{A})$ é fechado por uniões não necessariamente disjuntas. Na verdade, vamos ver que é suficiente provar que

$$\lambda(\mathcal{A}) \text{ é um } \pi\text{-sistema.} \quad (1.17)$$

De fato, caso isso seja provado teremos que $\lambda(\mathcal{A})$ é fechado por diferenças (pois $A \setminus B = A \cap B^c$). Assim, podemos mostrar que $\lambda(\mathcal{A})$ é fechado por uniões enumeráveis, pois se $A_1, A_2, \dots \in \lambda(\mathcal{A})$, definimos $B_n = \bigcup_{i=1}^n A_i = (\bigcap_{i=1}^n A_i^c)^c \in \lambda(\mathcal{A})$ e escrevemos

$$\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} (A_n \setminus B_{n-1}), \quad (1.18)$$

que é uma união disjunta de termos em $\lambda(\mathcal{A})$, logo está em $\lambda(\mathcal{A})$. Isso mostra que $\lambda(\mathcal{A})$ é uma σ -álgebra e que de fato é suficiente demonstrar (1.17).

Vamos primeiramente mostrar que $\lambda(\mathcal{A})$ é fechado por interseções com \mathcal{A} . Para tanto, definimos $\mathcal{B} = \{B \in \lambda(\mathcal{A}); B \cap A \in \lambda(\mathcal{A}) \text{ para todo } A \in \mathcal{A}\}$ e veremos que

$$B = \lambda(\mathcal{A}). \quad (1.19)$$

Obviamente, $\mathcal{A} \subseteq \mathcal{B}$, pois \mathcal{A} é um π -sistema. Então basta mostrar que \mathcal{B} é um λ -sistema.

- a) Ω obviamente pertence a \mathcal{B} .
- b) Se $B \in \mathcal{B}$ e $A \in \mathcal{A}$, então $B^c \cap A = A \setminus (B \cap A) = (A^c \cup (B \cap A))^c$. Mas como $B \in \mathcal{B}$, $(B \cap A) \in \lambda(\mathcal{A})$ e usando o fato que λ -sistemas são fechados por complementos e uniões disjuntas, $B^c \cap A \in \lambda(\mathcal{A})$. Como isso vale para todo $A \in \mathcal{A}$, temos $B^c \in \mathcal{B}$ por definição.
- c) Se $B_1, B_2, \dots \in \mathcal{B}$ são disjuntos e $A \in \mathcal{A}$, então

$$(\bigcup_{n=1}^{\infty} B_n) \cap A = \bigcup_{n=1}^{\infty} (B_n \cap A) \in \lambda(\mathcal{A}), \quad (1.20)$$

pois a união acima é disjunta. Logo $\bigcup_{n=1}^{\infty} B_n \in \mathcal{B}$.

Isso mostra que \mathcal{B} é um λ -sistema com $\mathcal{A} \subseteq \mathcal{B} \subseteq \lambda(\mathcal{A})$, mostrando (1.19).

No próximo passo, definimos $\tilde{\mathcal{B}} = \{A \in \lambda(\mathcal{A}); B \cap A \in \lambda(\mathcal{A}), \forall B \in \lambda(\mathcal{A})\}$ e mostraremos que

$$\tilde{\mathcal{B}} = \lambda(\mathcal{A}), \quad (1.21)$$

que vai na direção de provar (1.17).

Primeiramente, observe que $\mathcal{A} \subseteq \tilde{\mathcal{B}}$ pois $\mathcal{B} = \lambda(\mathcal{A})$ (veja a definição de \mathcal{B}). Mostraremos agora que

$$\tilde{\mathcal{B}} \text{ é um } \lambda\text{-sistema.} \quad (1.22)$$

Para tanto, verificaremos

- a) $\Omega \in \tilde{\mathcal{B}}$, que é claro.

- b) Tomando $A \in \tilde{\mathcal{B}}$ e $B \in \lambda(\mathcal{A})$, $A^c \cap B = B \setminus (A \cap B) = (B^c \cup (A \cap B))^c \in \lambda(\mathcal{A})$, por um argumento análogo ao apresentado para \mathcal{B} . Logo $A^c \in \tilde{\mathcal{B}}$.
- c) Também o caso de uniões disjuntas é bastante análogo ao feito para \mathcal{B} .

Isso mostra que $\tilde{\mathcal{B}}$ é um λ -sistema com $\mathcal{A} \subseteq \tilde{\mathcal{B}} \subseteq \lambda(\mathcal{A})$, estabelecendo (1.22).

Finalmente mostraremos que $\lambda(\mathcal{A})$ é um π -sistema. De fato, dado $A \in \lambda(\mathcal{A})$, segue da igualdade $\tilde{\mathcal{B}} = \lambda(\mathcal{A})$ que $A \cap B \in \lambda(\mathcal{A})$, para todo $B \in \lambda(\mathcal{A})$. Logo estabelecemos (1.17), terminando a prova do teorema. \square

1.3.1 Igualdade de probabilidades

Proposição 1.3.4. Se P_1 e P_2 são probabilidades em (Ω, \mathcal{F}) , tais que $P_1(A) = P_2(A)$ para todo $A \in \mathcal{A}$ e \mathcal{A} é um π -sistema, então $P_1(B) = P_2(B)$ para todo $B \in \sigma(\mathcal{A})$.

Demonstração. Seja $\mathcal{B} = \{A \in \mathcal{F}; P_1(A) = P_2(A)\}$. É fácil ver que \mathcal{B} é um λ -sistema. Logo \mathcal{B} contém $\lambda(\mathcal{A})$ que é igual a $\sigma(\mathcal{A})$ por Dynkin. \square

Corolário 1.3.5. Se P_1 e P_2 são probabilidades em $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2)$, tais que

$$P_1(A_1 \times A_2) = P_2(A_1 \times A_2), \text{ para todos } A_1 \in \mathcal{F}_1, A_2 \in \mathcal{F}_2, \quad (1.23)$$

então $P_1 = P_2$.

Demonstração. Obviamente as caixas do tipo $A_1 \times A_2$ formam um π -sistema que gera $\mathcal{F}_1 \otimes \mathcal{F}_2$ (por definição). \square

Exemplo 1.3.2. Observe portanto que é importante que \mathcal{A} seja um π -sistema na Proposição 1.3.4. Imagine por exemplo que $\Omega = \{0, 1\}^2$ e $P_1 = \frac{1}{4} \sum_{x \in \Omega} \delta_x$ e $P_2 = \frac{1}{2}(\delta_{(0,0)} + \delta_{(1,1)})$. Nesse caso

$$P_1(A) = P_2(A) = 1/2 = P_1(B) = P_2(B), \quad (1.24)$$

com $A = \{(0,0), (0,1)\}$ e $B = \{(0,0), (1,0)\}$. Contudo, $P_1 \neq P_2$, mesmo tendo $\mathcal{P}(\Omega) = \sigma(\{A, B\})$.

1.4 Elementos aleatórios

Muitas vezes não estamos interessados no resultado exato do nosso experimento aleatório, mas sim em uma determinada medição ou função de $\omega \in \Omega$. Por exemplo, no caso do Exemplo 1.1.1 c), talvez não nos interesse toda a função f , mas apenas o seu valor no fim do dia $f(1)$. Essas medições são ditas elementos aleatórios que definimos à seguir.

Seja (E, \mathcal{A}) um espaço mensurável. Nesse caso, se $X : \Omega \rightarrow E$ é uma função $(\mathcal{F}, \mathcal{A})$ -mensurável, dizemos que X é um *elemento aleatório* em (Ω, \mathcal{F}) tomando valores em E , ou um E -elemento aleatório.

Exemplo 1.4.1. Consideramos os casos

- a) $X : \Omega \rightarrow \mathbb{R}$ mensurável é dita *variável aleatória*.
- b) $X : \Omega \rightarrow \mathbb{R}^d$ mensurável é dito *vetor aleatório* (d -dimensional).
- c) $X : \Omega \rightarrow C[0, 1]$ mensurável é dita *função aleatória*.

Seguindo a motivação do Exemplo 1.1.1 c), poderia ser que, por exemplo, estivéssemos interessados apenas na variável aleatória $X : \Omega_3 \rightarrow \mathbb{R}$ dada por $X(f) = f(1)$.

Exercício 1.4.2. Mostre que $X : \Omega_3 \rightarrow \mathbb{R}$ dada por $X(f) = f(1)$ é uma variável aleatória.

Citando Kingman em seu livro Poisson Processes: “a random elephant is a function from Ω into a suitable space of elephants.”

Relembrando a nossa notação: $P[X \in A] = P(\{\omega \in \Omega; X(\omega) \in A\})$.

Proposição 1.4.1. Seja $X : \Omega \rightarrow E$ onde (E, \mathcal{A}) é um espaço mensurável com $\mathcal{A} = \sigma(\mathcal{G})$. Então para verificar que X é um elemento aleatório, basta provar que $X^{-1}(G) \in \mathcal{F}$ para todo $G \in \mathcal{G}$.

Demonstração. Teoria da Medida. □

Exemplo 1.4.3. Se Ω e E são espaços topológicos dotados das correspondentes σ -álgebras de Borel, então toda função contínua é um E -elemento aleatório.

1.4.1 Distribuição de elementos aleatórios

Definição 1.4.2. Se $X : \Omega \rightarrow E$ é um elemento aleatório e Ω é dotado de uma probabilidade P , então denotamos por X_*P , a chamada *distribuição de X* , a medida de probabilidade

$$X_*P(A) := P(\{\omega \in \Omega; X(\omega) \in A\}) = P[X \in A]. \quad (1.25)$$

no espaço mensurável (E, \mathcal{A}) .

Observação 1.4.3. Essa definição corresponde com a de medida imagem vista no curso de integração que tem um papel ainda mais importante em probabilidade.

Fica como exercício verificar que X_*P é de fato uma probabilidade em E .

Exercício 1.4.4. Seja $X : [0, 1] \rightarrow \{0, 1\}$ dada por $X(\omega) = \mathbb{1}_A(\omega)$. Nesse caso, mostre que $X_*P = \text{Ber}(p)$ para algum $p \in [0, 1]$. Calcule o valor de p .

Duas notações importantes nesse contexto são:

- a) Sejam (Ω, \mathcal{F}, P) e $(\Omega', \mathcal{F}', P')$ dois espaços de probabilidade e X e Y dois elementos aleatórios. Dizemos que $X \stackrel{d}{=} Y$, quando $X_*P = Y_*P'$. Note que X e Y nem ao menos precisam pertencer ao mesmo espaço de probabilidade para dizermos que são *igualmente distribuídos*, mas precisam ser elementos aleatórios de mesmo tipo (ou seja, possuir o mesmo contradomínio).

- b) Escrevemos $X \sim_d \mu$, que lê-se *X é distribuída como μ* , onde μ é uma probabilidade em E , caso $X_*P = \mu$.

Exercício 1.4.5. *Sejam X e Y variáveis aleatórias tais que X é nula quase certamente. Mostre que $X + Y$ tem a mesma distribuição de Y .*

O exercício acima é bastante simples, mas o usaremos para fazer uma importante observação sobre como são enunciados tipicamente os resultados de probabilidade.

Raramente encontramos teoremas que explicitam qual é o espaço de probabilidades Ω em questão. Como no exercício acima, o contexto de um teorema frequentemente é dado apenas em termos de elementos aleatórios em Ω e de suas distribuições. Dessa forma, podemos utilizar o resultado em vários contextos diferentes, desde que possamos encontrar elementos aleatórios que satisfaçam as hipóteses. Com o tempo, passamos até mesmo a considerar menos relevante a escolha específica do espaço amostral, focando cada vez mais na distribuição de seus elementos aleatórios.

Tópico: O paradoxo de Bertrand

Vamos estudar um problema que realça a importância do jeito em que escolhemos o espaço amostral. Queremos calcular a probabilidade que uma corda “uniformemente distribuída” em um círculo seja maior do que o lado do triângulo equilátero inscrito nesse círculo (no caso do círculo unitário, o comprimento desse lado vale $\sqrt{3}$). Bertrand propôs dois métodos para realizar esse cálculo.¹

- a) Escolher as duas extremidades da corda uniformemente no círculo.
- b) Escolher o centro da corda uniformemente no interior do disco.

No caso *a*), uma vez que uma extremidade é fixada, o comprimento da corda fica maior do que $\sqrt{3}$ somente se o segundo ponto ficar num setor angular de comprimento $2\pi/3$. Logo, essa probabilidade vale $(2\pi/3)/(2\pi) = 1/3$.

No caso *b*), pra que a corda fique maior do que $\sqrt{3}$, o centro dela deve ficar no círculo inscrito dentro do triângulo equilátero, cujo raio é $1/2$. Então a probabilidade vale a razão dessas áreas, que é $1/4$.

Obtemos então duas respostas diferentes para essa pergunta simples, o que não é nada surpreendente: *a*) e *b*) correspondem a dois experimentos diferentes com espaços amostrais diferentes.

Exercício 1.4.6. *a) Descreva o espaço amostral e as lei de probabilidade associadas aos experimentos a) e b)*

- b) Calcule a lei de probabilidade do comprimento da corda em cada caso.*
- c) Repita os itens anteriores para o seguinte caso: Escolhemos uniformemente um raio do disco. Depois escolhemos o centro da corda uniformemente ao longo desse raio.*

¹Somos gratos a Hubert Lacoïn por sugerir e redigir esse tópico.

Capítulo 2

Construção de espaços de probabilidade

Nessa seção descreveremos diversas maneiras diferentes de construir um espaço de probabilidade, dando diversos exemplos de como elas podem ser usadas na modelagem de diferentes processos reais.

2.1 Caso enumerável

Quando Ω é finito ou enumerável, tipicamente definimos sobre Ω a σ -álgebra das partes, ou seja $\mathcal{F} = \mathcal{P}(\Omega) = \sigma(\{\omega\}_{\omega \in \Omega})$. Além disso podemos definir probabilidades sobre (Ω, \mathcal{F}) de maneira simples tomando $(p_\omega)_{\omega \in \Omega}$ tais que

- a) $p_\omega \geq 0$ para todo $\omega \in \Omega$ e
- b) $\sum_{\omega \in \Omega} p_\omega = 1$.

De fato, nesse caso definimos $P(A) = \sum_{\omega \in A} p_\omega$ que claramente define uma probabilidade.

Exercício 2.1.1. *Mostre que se Ω é finito ou enumerável, toda probabilidade sobre $(\Omega, \mathcal{P}(\Omega))$ é dada como na descrição acima.*

Exemplo 2.1.2.

- a) Dado $p \in [0, 1]$, definimos a medida $\text{Ber}(p)$ (em homenagem a Bernoulli) em $\{0, 1\}$ com $p_1 = p, p_0 = 1 - p$.
- b) Dados $n \geq 1$ e $p \in [0, 1]$, definimos a medida $\text{Bin}(n, p)$ (binomial) em $\Omega = \{0, 1, \dots, n\}$ com

$$p_i = \binom{n}{i} p^i (1-p)^{n-i}, \text{ para } i \in \Omega. \quad (2.1)$$

c) Dado $p \in (0, 1]$, em $\Omega = \{0, 1, \dots\}$ definimos a medida $\text{Geo}(p)$ (geométrica) em Ω induzida pelos pesos

$$p_i = (1 - p)^i p, \text{ para } i \geq 1. \quad (2.2)$$

Exercício 2.1.3. Seja $\Omega = \{0, 1\}^n$ e $p_\omega = \frac{1}{2^n}$ para todo $\omega \in \Omega$ (ou seja a probabilidade uniforme). Considere $X : \Omega \rightarrow \{0, 1, \dots, n\}$ dada por $X(\omega_1, \dots, \omega_n) = \sum_{i=1}^n \omega_i$. Obtenha a distribuição P_X . Dê um exemplo de medida em ω para a qual a distribuição de X seja $\text{Bin}(n, p)$.

Tópico: Método Probabilístico

Uma importante ferramenta em várias áreas da matemática, tais como Teoria dos Números, Combinatória e Teoria da Computação é o que chamamos de Método Probabilístico.

Em várias situações, nós precisamos de mostrar a existência de objetos satisfazendo determinadas propriedades, mas não temos informação suficiente ou capacidade para construí-los explicitamente. Nesse caso, podemos recorrer ao Método Probabilístico, que simplesmente nos sugere tomar um objeto aleatório de uma maneira esperta e mostrar que com probabilidade positiva as propriedades desejadas serão satisfeitas. Esse método, apesar de muito ingênuo, é muito eficiente e em diversos casos provê os melhores exemplos conhecidos de certos objetos (para embaraço da comunidade científica).

Nessa seção daremos um exemplo em Teoria dos Números provido primeiramente por Erdős¹.

Teorema 2.1.1 (Erdős). *Para todo conjunto finito $A \subset \mathbb{N}$, existe um sub-conjunto $B \subseteq A$ satisfazendo*

- a) $\#B \geq \frac{\#A}{3}$ e tal que
- b) não existem x, y e $z \in B$ com $x + y = z$.

A propriedade b) acima é o que chamamos de um conjunto ser livre de somas.

Certamente não temos muita informação sobre A , então vamos usar o método probabilístico para a prova desse teorema.

Demonstração. Fixamos p um número primo maior que três vezes o maior elemento de A e considere o espaço \mathbb{Z}_p dos inteiros módulo p . Seja X um elemento aleatório de \mathbb{Z}_p com distribuição uniforme, isto é $U_{\{0, \dots, p-1\}}$.

Exercício 2.1.4. *Mostre que para todo $a \in A$, a multiplicação por a é uma bijeção em \mathbb{Z}_p , ou seja*

$$\mathbb{Z}_p \cdot a = \mathbb{Z}_p. \quad (2.3)$$

onde o produto $\mathbb{Z}_p \cdot a$ é entendido elemento a elemento. Conclua que

$$P\left[X \cdot a \in \left[\frac{p}{3}, \frac{2p}{3}\right)\right] \geq \frac{1}{3} - \frac{1}{p}. \quad (2.4)$$

Definimos o conjunto aleatório

$$\mathcal{B} = \{a \in A \mid X \cdot a \in \left[\frac{p}{3}, \frac{2p}{3}\right)\}. \quad (2.5)$$

Esse conjunto é livre de soma: se $X = 0$ o conjunto é vazio e, nos outros casos, se $a, b \in \mathcal{B}$ então

$$X(a + b) \in \left[\frac{2p}{3}, \frac{4p}{3}\right) \quad (2.6)$$

¹Somos gratos a Robert Morris por sugerir esse teorema como exemplo do Método Probabilístico.

que é o complemento de $[\frac{p}{3}, \frac{2p}{3})$ em \mathbb{Z}_p .

Basta portanto mostrar que com probabilidade positiva $\#\mathcal{B} \geq \frac{\#A}{3}$, que segue do seguinte argumento. Note inicialmente que

$$\begin{aligned} \int \#\mathcal{B} \, dP &= \int \sum_{a \in A} \mathbb{1}_{[X \cdot a \in [p/3, 2p/3)]} \, dP \\ &= \sum_{a \in A} P\left[X \cdot a \in \left[\frac{p}{3}, \frac{2p}{3}\right)\right] \geq \frac{\#A}{3} - \frac{\#A}{p} > \frac{\#A - 1}{3}, \end{aligned}$$

mas, para qualquer variável aleatória, $P[Y \geq \int Y \, dP] > 0$. Nesse caso, isso implica

$$P[\#\mathcal{B} \geq \frac{\#A}{3}] = P[\#\mathcal{B} > \frac{\#A-1}{3}] \geq P\left[\#\mathcal{B} > \int \#\mathcal{B} \, dP\right] > 0. \quad (2.7)$$

□

2.2 Caso absolutamente contínuo

Uma outra maneira simples de definir um espaço de probabilidade é partindo de um espaço de medida. Seja $(\Omega, \mathcal{F}, \mu)$ um espaço de medida e $\rho : \Omega \rightarrow \mathbb{R}_+$ uma função mensurável com $\int \rho(x) \mu(dx) = 1$. Então podemos definir a probabilidade induzida

$$P(A) = \int_A \rho(x) \mu(dx). \quad (2.8)$$

Nesse caso, chamamos ρ de a *densidade* de P com respeito a μ . Uma outra possível notação para a equação acima é $dP = \rho(x) d\mu$ (lembrando a derivada de Radon-Nikodym).

Observe que o caso discreto pode ser definido em termos de uma densidade, onde $\rho(\omega) = p_\omega$ e μ é a medida da contagem em Ω .

Exemplo 2.2.1. Vários exemplos podem ser obtidos via (2.8) se tomamos $\Omega \subseteq \mathbb{R}$ e μ a medida de Lebesgue restrita a Ω . Nesses casos, escrevemos $P = \rho(x) dx$ em Ω . Alguns exemplos importantes são:

- a) Para $a < b \in \mathbb{R}$, definimos a medida $U[a, b]$ usando $\rho(x) = \frac{1}{b-a} \mathbb{1}_{[a,b]}(x)$.
- b) Para $\lambda > 0$, definimos a medida $\text{Exp}(\lambda)$ (chamada exponencial de parâmetro λ) por meio da densidade $\rho(x) = \lambda \exp\{-\lambda x\}$ em $[0, \infty)$.

Podemos também usar a distribuição de um elemento aleatório para construir outras probabilidades, como mostra o seguinte exemplo.

Exemplo 2.2.2. Considere por exemplo $X : [0, 2\pi] \rightarrow \mathbb{C}$ dada por $X(t) = \exp\{-it\}$. A distribuição imagem $X_* U_{[0, 2\pi]}$ é o que chamamos de distribuição uniforme em S^1 , também denotada por U_{S^1} .

Exercício 2.2.3. Mostre que U_{S^1} não é absolutamente contínua com respeito à medida de Lebesgue em $\mathbb{C} \sim \mathbb{R}^2$.

Exercício 2.2.4. Mostre que U_{S^1} é invariante por rotações rígidas de \mathbb{C} , isto é, se $T : \mathbb{C} \rightarrow \mathbb{C}$ é uma isometria linear, $T_* U_{S^1} = U_{S^1}$.

Exercício 2.2.5. Construa uma probabilidade em S^2 invariante por rotações.

2.3 Funções acumuladas de distribuição

Um caso muito importante de espaço amostral é $\Omega = \mathbb{R}$, principalmente por nos ajudar a entender distribuições de variáveis aleatórias. Para tanto, precisaremos de uma boa ferramenta para descrever probabilidades em \mathbb{R} .

Definição 2.3.1. Dada P em \mathbb{R} , definimos $F_P : \mathbb{R} \rightarrow [0, 1]$ por $F_P(x) = P((-\infty, x])$. Essa função é chamada a função de distribuição acumulada de P .

Notação 2.3.2. Se $X : \Omega \rightarrow \mathbb{R}$ é uma variável aleatória num espaço (Ω, \mathcal{F}, P) , denotamos por F_X a função de distribuição acumulada correspondente à distribuição X_*P .

Lembramos que uma probabilidade em \mathbb{R} é uma função $P : \mathcal{B}(\mathbb{R}) \rightarrow [0, 1]$ e o domínio dessa função é bastante complicado. Por exemplo se quisermos representar uma distribuição de uma variável aleatória no computador através dessa função P , teríamos problemas. Contudo, a função F_P (ou F_X) é muito mais simples de ser compreendida ou representada, por seu domínio ser \mathbb{R} .

Exemplo 2.3.1. Não é difícil verificar que

$$F_{\delta_{x_0}} = \begin{cases} 0 & \text{se } x < x_0, \\ 1 & \text{se } x \geq x_0 \end{cases} \quad (2.9)$$

e que

$$F_{U_{[0,1]}} = \begin{cases} 0 & \text{se } x \leq 0, \\ x & \text{se } x \in [0, 1] \text{ e} \\ 1 & \text{se } x \geq 1. \end{cases} \quad (2.10)$$

Exercício 2.3.2. Calcule $F_{\text{Exp}(\lambda)}$.

Proposição 2.3.3. F_P (e obviamente F_X) satisfazem:

- a) $\lim_{x \rightarrow -\infty} F(x) = 0, \lim_{x \rightarrow \infty} F(x) = 1,$
- b) F é monótona não-decrescente e
- c) F é contínua à direita e possui limite à esquerda (càdlàg, do francês).

Demonstração. a) Se $x_n \rightarrow -\infty$ monotonamente, então $A_n = (-\infty, x_n]$ são encaixados e de interseção vazia. Logo, pela Proposição 1.2.3, temos $P(A_n) \rightarrow 0$. O outro caso é análogo.

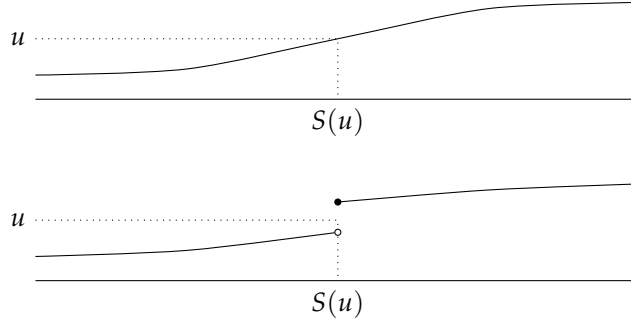
b) Se $x \leq x'$ então $(-\infty, x] \subseteq (-\infty, x']$, donde $F(x) \leq F(x')$.

c) Continuidade à direita (càd) - Se $x_n \downarrow x$ monotonamente, então $A_n = (-\infty, x_n] \downarrow (-\infty, x]$ (eles são encaixados). Logo $F(x_n) \rightarrow F(x)$.

Limite à esquerda (làg) - Segue do fato de F ser monótona e limitada. \square

Teorema 2.3.4. Se F satisfaz as três propriedades listadas na Proposição 2.3.3, então existe uma única P em $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ tal que $F = F_P$.

Poderíamos usar o Teorema da Extensão de Caratheodory para provar tal resultado, de maneira similar ao que foi feito no caso da Medida de Lebesgue. Mas escolhemos abaixo um método mais simples, que parte da existência de $U_{[0,1]}$.


 Figura 2.1: Ilustração da definição de $S(u)$.

Demonstração. A unicidade de tal P segue da Proposição 1.3.4 (consequência do Teorema de Dynkin), pois se P e P' são tais que $F_P = F_{P'}$, então temos que $P((-\infty, x]) = P'((-\infty, x])$. Mas a classe de intervalos semi-infinitos da forma $(-\infty, x]$ forma um π -sistema que gera a σ -álgebra dos borelianos, logo $P = P'$.

Para construir uma P tal que $F_P = F$, definiremos $S : (0, 1) \rightarrow \mathbb{R}$, a inversa generalizada de F , por

$$S(u) = \sup\{x \in \mathbb{R} : F(x) < u\}. \quad (2.11)$$

Seja $P = S_*U_{[0,1]}$, isto é $P(A) = U_{[0,1]}(S^{-1}(A))$ e mostraremos que $F_P = F$. Para tanto, basta ver que

$$\{u \in [0, 1] : S(u) \leq x\} = \{u \in [0, 1] : u \leq F(x)\}, \text{ para todo } x \in \mathbb{R}. \quad (2.12)$$

Pois isso implicaria que $F_P(x) = U_{[0,1]}[S(u) \leq x] = U_{[0,1]}[u \leq F(x)] = F(x)$.

Vamos agora checar (2.12) observando que:

- a) Se $u \leq F(x)$ então todo x' tal que $F(x') < u$ é menor que x . Logo $S(u) \leq x$.
- b) Por outro lado, se $x \geq S(u)$ então todo $x' > x$ satisfaz $F(x') > u$. Pois por continuidade a direita $F(x) \geq u$.

Isso prova (2.12), terminando a prova da proposição. \square

Exercício 2.3.3. Mostre o resultado acima usando o Teorema de Extensão de Caratheodory.

2.4 Espaços produto finito

Dados espaços $\Omega_1, \dots, \Omega_n$ com suas respectivas σ -álgebras $\mathcal{F}_1, \dots, \mathcal{F}_n$, podemos definir o espaço mensurável produto $(\bar{\Omega}, \bar{\mathcal{F}})$ da seguinte forma

$$\bar{\Omega} = \prod_{i=1}^n \Omega_i \quad \text{e} \quad \bar{\mathcal{F}} = \sigma\left(\{A_1 \times \dots \times A_n : \forall i \in \{1, \dots, n\}, A_i \in \mathcal{F}_i\}\right). \quad (2.13)$$

Essa σ -álgebra é chamada de σ -álgebra produto e denotaremos ela por $\bigotimes_{i=1}^n \mathcal{F}_i$, ou $\mathcal{F}_1 \otimes \mathcal{F}_2$ quando $n = 2$.

Proposição 2.4.1. *Se $(\Omega_1, \mathcal{F}_1, P_1), \dots, (\Omega_n, \mathcal{F}_n, P_n)$ são espaços de probabilidade, então existe uma única probabilidade \bar{P} no espaço mensurável $(\bar{\Omega}, \bar{\mathcal{F}})$ tal que*

$$\bar{P}(A_1 \times \dots \times A_n) = \prod_{i=1}^n P_i(A_i), \text{ para todos } A_i \in \mathcal{F}_i, i \leq n. \quad (2.14)$$

Essa probabilidade é chamada probabilidade produto. Usaremos a notação $\bigotimes_{i=1}^n P_i$ o $P_1 \otimes P_2 \otimes \dots \otimes P_n$.

Demonstração. Teoria da Medida. □

Note que a unicidade do produto pode ser concluída por exemplo usando o Corolário 1.3.5.

Exercício 2.4.1. *Mostre que o produto de n cópias de $(\{0, 1\}, \mathcal{P}(\{0, 1\}), \text{Ber}(1/2))$ é a distribuição uniforme em $\{0, 1\}^n$.*

2.5 Independência

Nossa intuição nos diz que quando jogamos duas moedas, o resultado de cada uma delas não deve depender um do outro. Dessa forma, a probabilidade de obtermos um determinado resultado (como por exemplo duas caras) deve ser um quarto, ou seja, meio vezes meio.

Em geral, definimos dois eventos como independentes da seguinte forma.

Definição 2.5.1. *Dizemos que dois eventos $A, B \in \mathcal{F}$ são independentes se*

$$P(A \cap B) = P(A)P(B). \quad (2.15)$$

Exemplo 2.5.1. *Se $\Omega = \{1, \dots, 6\}$ é dotada da σ -álgebra das partes e $P(A) = \#A/6$, então os eventos $A = [\omega \text{ é ímpar}]$ e $B = [\omega \geq 5]$ satisfazem*

$$P(A \cap B) = P(\{5\}) = 1/6 = (1/2)(1/3) = P(A)P(B). \quad (2.16)$$

Logo tais eventos são independentes.

Exercício 2.5.2. *Seja $\Omega = \{0, 1\}^n$ com $P(A) = \#A/2^n$ e $X_i(\omega_1, \dots, \omega_n) = \omega_i$ para $i = 1, \dots, n$. Mostre que*

$$P[X_i = a, X_j = b] = P[X_i = a]P[X_j = b], \quad (2.17)$$

onde $[A, B]$ denota a interseção $A \cap B$.

2.5.1 Coleções de eventos

Definição 2.5.2. Sejam A_1, A_2, \dots, A_k eventos. Dizemos que eles formam uma coleção independente se para todo $I \subseteq \{1, \dots, k\}$ não vazio

$$P\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} P(A_i). \quad (2.18)$$

Vale observar que independência dois a dois não implica independência. Mais precisamente

Exemplo 2.5.3. Seja $\Omega = \{1, 2, 3, 4\}$ com $P(A) = \#A/4$ e sejam os seguintes eventos: $A_1 = \{1, 2\}$, $A_2 = \{2, 3\}$ e $A_3 = \{1, 3\}$. Nesse caso,

- a) $P(A_i) = 1/2$ para $i = 1, 2, 3$,
- b) $P(A_i \cap A_j) = 1/4$ para todo $i \neq j$ mas
- c) $P(A_1 \cap A_2 \cap A_3) = 0 \neq 1/8 = P(A_1)P(A_2)P(A_3)$.

Definição 2.5.3. Dizemos que uma coleção infinita de eventos $(A_n)_{n \geq 1}$ é independente se toda sub-coleção finita de tais eventos forem independentes.

Lema 2.5.4. Se $(A_n)_{n \geq 1}$ forma uma sequência de eventos independentes, então

$$P\left(\bigcap_{i=1}^{\infty} A_i\right) = \prod_{i=1}^{\infty} P(A_i). \quad (2.19)$$

Demonstração. De fato,

$$P\left(\bigcap_{i=1}^{\infty} A_i\right) = \lim_{n \rightarrow \infty} P\left(\bigcap_{i=1}^n A_i\right) = \lim_{n \rightarrow \infty} \prod_{i=1}^n P(A_i) = \prod_{i=1}^{\infty} P(A_i). \quad \square$$

Exercício 2.5.4. Mostre que se $A \in \mathcal{F}$, então $\{B \in \mathcal{F} : B \text{ é independente de } A\}$ é um λ -sistema.

Exercício 2.5.5. Mostre que se B é independente de A para todo $B \in \mathcal{B}$, com \mathcal{B} um π -sistema, então B é independente de A para todo $B \in \sigma(\mathcal{B})$.

2.5.2 Independência de σ -álgebras

Definição 2.5.5. Dado um espaço de probabilidade (Ω, P, \mathcal{F}) Dizemos que as σ -álgebras $\mathcal{F}_1, \dots, \mathcal{F}_n \subset \mathcal{F}$ são independentes se

$$\forall A_1 \in \mathcal{F}_1, \dots, A_n \in \mathcal{F}_n, P\left(\bigcap_{i=1}^n A_i\right) = \prod_{i=1}^n P(A_i). \quad (2.20)$$

Nessa definição podemos tomar uma coleção infinita.

Exercício 2.5.6. Em um espaço produto $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2, P_1 \otimes P_2)$, podemos definir

$$\begin{aligned}\bar{\mathcal{F}}_1 &= \{A \times \Omega_2 : A \in \mathcal{F}_1\}, \\ \bar{\mathcal{F}}_2 &= \{\Omega_1 \times B : B \in \mathcal{F}_2\}.\end{aligned}\tag{2.21}$$

Mostre que essas σ -álgebras são independentes.

Podemos estender esse conceito a elementos aleatórios, ou seja:

Definição 2.5.6. Dizemos que X_1, \dots, X_k são elementos aleatórios independentes se as respectivas σ -álgebras $\sigma(X_1), \dots, \sigma(X_k)$ o forem.

Quando X_1, \dots, X_k são elementos aleatórios independentes e com a mesma distribuição, escrevemos que X_i são i.i.d. (independentes e identicamente distribuídos).

Exercício 2.5.7. Com a notação do exercício anterior, mostre que as funções $X_i : \Omega_1 \times \Omega_2 \rightarrow \Omega_i$ dadas por

$$X_1(x, y) = x \text{ e } X_2(x, y) = y,\tag{2.22}$$

são elementos aleatórios e são independentes.

Exercício 2.5.8. Mostre que as coordenadas canônicas do exercício anterior no caso $X_i : \mathbb{R}^2 \rightarrow \mathbb{R}$ não são independentes segundo a medida $U_{\mathbb{S}^1}$. Mas o são segundo $U_{[0,1]^2}$ (que é a medida de Lebesgue em \mathbb{R}^2 restrita a $[0, 1]^2$).

Exercício 2.5.9. Seja $\Omega = \{0, 1\}^n$ com $P(A) = \#A/2^n$ e $X_i(\omega_1, \dots, \omega_n) = \omega_i$ para $i = 1, \dots, n$. Mostre que os X_i são independentes.

Exercício 2.5.10. Sejam $(X_i)_{i \geq 1}$ elementos aleatórios independentes tomando valores em espaços $(E_i)_{i \geq 1}$, respectivamente. Mostre que para funções mensuráveis $(f_i)_{i \geq 1}$ temos que $(f_i(X_i))_{i \geq 1}$ são independentes.

Exercício 2.5.11. Mostre que se X, Y são elementos aleatórios e se X é constante quase certamente então X e Y são independentes.

Exercício 2.5.12. Sejam X e Y variáveis aleatórias independentes com distribuição $\text{Exp}(1)$, calcule a distribuição de

- a) $\min\{X, Y\}$ e
- b) $X + Y$.

Exercício 2.5.13. Seja um espaço produto de medidas $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2, \mu_1 \otimes \mu_2)$ e defina a probabilidade P através de

$$dP = \rho(x, y) d(\mu_1 \otimes \mu_2).\tag{2.23}$$

Mostre nesse caso que as coordenadas canônicas X_1 e X_2 são independentes se e somente se existem ρ_1 e ρ_2 em Ω_1 e Ω_2 respectivamente, tais que $\rho(x, y) = \rho_1(x)\rho_2(y)$ quase certamente com respeito a $\mu_1 \otimes \mu_2$.

Exercício 2.5.14. Sejam X, Y variáveis aleatórias tais que

$$P[X \leq x, Y \leq y] = \begin{cases} 0 & \text{if } x < 0, \\ (1 - e^{-x}) \left(\frac{1}{2} + \frac{1}{\pi} \tan^{-1} y \right), & \text{if } x \geq 0. \end{cases} \quad (2.24)$$

a) Mostre que a distribuição conjunta $\mu_{(X,Y)}$ é absolutamente contínua com relação à medida de Lebesgue em \mathbb{R}^2 .

b) Mostre que X e Y são independentes.

Exercício 2.5.15. Mostre que se X, Y são variáveis aleatórias independentes com distribuições $X \sim_d f_X(x) dx$ e $Y \sim_d f_Y(y) dy$, então $X + Y$ tem distribuição absolutamente contínua com respeito a Lebesgue e

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f_Y(z-x) f_X(x) dx. \quad (2.25)$$

Lema 2.5.7 (Borel-Cantelli - segunda parte). Se $A_1, A_2, \dots \in \mathcal{F}$ são independentes e $p_i = P(A_i)$ satisfazem $\sum_i p_i = \infty$, então

$$P[A_i \text{ infinitas vezes}] = 1. \quad (2.26)$$

Demonstração. Queremos mostrar que

$$P\left(\left(\bigcap_n \bigcup_{i=n}^{\infty} A_i\right)^c\right) = 0, \quad (2.27)$$

mas

$$P\left(\left(\bigcap_n \bigcup_{i=n}^{\infty} A_i\right)^c\right) = P\left(\bigcup_n \bigcap_{i=n}^{\infty} A_i^c\right) \leq \sum_n P\left(\bigcap_{i=n}^{\infty} A_i^c\right). \quad (2.28)$$

Logo basta mostrar que a probabilidade à direita é zero para todo n . Mas

$$\begin{aligned} P\left(\bigcap_{i=n}^{\infty} A_i^c\right) &= \prod_{i=n}^{\infty} P(A_i^c) = \prod_{i=n}^{\infty} (1 - p_i) \\ &\leq \prod_{i=n}^{\infty} \exp\{-p_i\} = \exp\left\{-\sum_{i=n}^{\infty} p_i\right\} = 0. \end{aligned} \quad (2.29)$$

Terminando a prova do lema. □

Tópico: Lei dos pequenos números

Nessa seção estudaremos como se comportam limites de algumas variáveis aleatórias bastante importantes, mas primeiramente, uma breve intuição.

Apesar de que descreveremos a nossa motivação a partir desse exemplo do estudo de um material radioativo, podemos encontrar aplicações com justificativas bastante semelhantes para outros problemas, como: chegada de carros em um sinal de trânsito, número de mutações em um gene, número de mortes por ano em uma faixa etária...

Digamos que estamos observando um material radioativo que esporadicamente emite fótons que podemos detectar através de um aparelho. A razão dessas emissões pode ser aproximada pelo seguinte modelo. Na amostra temos um número n grande de átomos instáveis ($n \sim 10^{23}$) e em um determinado tempo de observação, cada um deles tem probabilidade muito baixa de decair emitindo um fóton (digamos $p \sim 10^{-23}$). Nesse caso, supondo que todos decidam emitir de maneira independente, temos para $p \in [0, 1]$,

$$\Omega_n = \{0, 1\}^n, \quad \mathcal{F}_n = \mathcal{P}(\Omega) \quad \text{e} \quad P_p = \otimes_{i=1}^n \text{Ber}(p). \quad (2.30)$$

Dessa forma, o número total de emissões observadas para $\omega = (\omega_1, \dots, \omega_n) \in \Omega$ é

$$X_n(\omega) = \sum_{i=1}^n \omega_i. \quad (2.31)$$

E gostaríamos de entender como se comporta essa distribuição, que nada mais é que $\text{Bin}(n, p)$.

Uma primeira tentativa seria modelar esse processo dizendo que o número de átomos n é tão grande, que somente estamos interessados no comportamento assintótico quando n vai para infinito. Mas para manter o número de emissões sob controle, também gostaríamos que $p = p_n$, que converge a zero. Poderíamos por exemplo escolher

$$p_n = \frac{\lambda}{n}. \quad (2.32)$$

Mas a discussão que se segue é muito mais geral que essa escolha específica.

Como estaremos interessados em um regime assintótico da distribuição de X_p (lembre que apesar do espaço amostral de X_n variar com n , sua distribuição é sempre uma probabilidade em \mathbb{N}), precisamos de definir uma noção de distância entre duas distribuições em \mathbb{N} .

Definição 2.5.8. Dadas duas distribuições μ_1 e μ_2 em (Ω, \mathcal{A}) , definimos

$$\|\mu_1 - \mu_2\|_{\text{VT}} = \sup_{A \in \mathcal{A}} |\mu_1(A) - \mu_2(A)|, \quad (2.33)$$

chamada de distância em variação total entre μ_1 e μ_2 .

No nosso caso, Ω é enumerável. Vamos ver que nesse caso é possível reescrever a definição acima de modo a ver mais facilmente que se trata de uma distância no espaço de probabilidades em Ω .

Lema 2.5.9. Se Ω for finito ou enumerável, então podemos escrever

$$\|\mu_1 - \mu_2\|_{VT} = \frac{1}{2} \sum_{x \in \Omega} |\mu_1(x) - \mu_2(x)|. \quad (2.34)$$

Demonstração. Para mostrar que o lado esquerdo é maior ou igual ao direito, escolhemos $A = \{x \in \Omega : \mu_2(x) \leq \mu_1(x)\}$. Assim

$$\begin{aligned} \sum_{x \in A} \mu_1(x) - \mu_2(x) &= |\mu_1(A) - \mu_2(A)| \\ &= |\mu_1(A^c) - \mu_2(A^c)| = \sum_{x \in A^c} \mu_2(x) - \mu_1(x), \end{aligned} \quad (2.35)$$

donde

$$\|\mu_1 - \mu_2\|_{VT} \geq |\mu_1(A) - \mu_2(A)| = \frac{1}{2} \sum_i |\mu_1(x_i) - \mu_2(x_i)|. \quad (2.36)$$

Na outra direção, observe que para todo $B \subseteq \Omega$,

$$\begin{aligned} \sum_i |\mu_1(x_i) - \mu_2(x_i)| &\geq \sum_{x \in B} \mu_1(x) - \mu_2(x) + \sum_{x \in B^c} \mu_2(x) - \mu_1(x) \\ &= \mu_1(B) - \mu_2(B) + (1 - \mu_2(B)) - (1 - \mu_1(B)) \\ &= 2(\mu_1(B) - \mu_2(B)). \end{aligned} \quad (2.37)$$

O que termina a prova do lema. \square

Fica agora claro que $\|\mu_1 - \mu_2\|_{VT}$ determina uma distância.

Exercício 2.5.16. Mostre um lema análogo ao anterior para (Ω, \mathcal{A}) qualquer, desde que μ_1 e μ_2 sejam absolutamente contínuas com relação à uma medida fixa nesse espaço mensurável. Nesse caso utilizaremos as derivadas de Radon–Nikodym.

Como estaremos interessados em variáveis independentes, precisamos de um resultado que relacione a distância em variação total com produtos de medida. Isso é parte do seguinte

Lema 2.5.10. Sejam μ_1, μ_2 distribuições em Ω e ν_1, ν_2 distribuições em Ω' ambos enumeráveis. Então

$$\|\mu_1 \otimes \nu_1 - \mu_2 \otimes \nu_2\|_{VT} \leq \|\mu_1 - \mu_2\|_{VT} + \|\nu_1 - \nu_2\|_{VT}. \quad (2.38)$$

Demonstração. Basta expandir

$$\begin{aligned} 2\|\mu_1 \otimes \nu_1 - \mu_2 \otimes \nu_2\|_{VT} &= \sum_{x \in \Omega, y \in \Omega'} |\mu_1(x)\nu_1(y) - \mu_2(x)\nu_2(y)| \\ &\leq \sum_{x \in \Omega, y \in \Omega'} |\mu_1(x)\nu_1(y) - \mu_1(x)\nu_2(y)| + |\mu_1(x)\nu_2(y) - \mu_2(x)\nu_2(y)| \\ &\leq 2\|\mu_1 - \mu_2\|_{VT} + 2\|\nu_1 - \nu_2\|_{VT}, \end{aligned} \quad (2.39)$$

onde acima usamos que μ_1 e ν_2 são probabilidades. Isso termina a prova do lema. \square

Finalmente, gostaríamos de entender como a distância de variação total se comporta com respeito à soma de variáveis independentes. Isso estará ligado à convolução de distribuições:

Definição 2.5.11. Dadas, μ e ν distribuições em \mathbb{Z} , definimos a distribuição

$$(\mu \star \nu)(x) := \sum_{y \in \mathbb{Z}} \mu(x - y)\nu(y). \quad (2.40)$$

Essa definição se relaciona com a soma de variáveis independentes graças ao seguinte

Exercício 2.5.17. Se $X \stackrel{d}{\sim} \mu$ e $Y \stackrel{d}{\sim} \nu$ são variáveis aleatórias inteiras e independentes, então $X + Y \stackrel{d}{\sim} \mu \star \nu$. Dica: particione o espaço amostral nos eventos $[X = j]$, para $j \in \mathbb{Z}$, como na prova do Lema 2.5.15 abaixo.

Corolário 2.5.12. Se μ e ν são distribuições em \mathbb{Z} , então $\mu \star \nu = \nu \star \mu$.

Como prometido, obtemos a seguinte relação entre a convolução e a distância de variação total.

Lema 2.5.13. Sejam μ, ν duas medidas em Ω enumerável e $X : (\Omega, \mathcal{P}(\Omega)) \rightarrow (E, \mathcal{A})$ um elemento aleatório

$$\|X_*\mu - X_*\nu\|_{VT} \leq \|\mu - \nu\|_{VT}. \quad (2.41)$$

Em particular se $\mu_1, \mu_2, \nu_1, \nu_2$ são distribuições em \mathbb{Z} , então

$$\|\mu_1 \star \nu_1 - \mu_2 \star \nu_2\|_{VT} \leq \|\mu_1 \otimes \nu_1 - \mu_2 \otimes \nu_2\|_{VT} \quad (2.42)$$

Demonstração. O segundo ponto segue do primeiro aplicado ao caso $\Omega = \mathbb{Z}^2$, $E = \mathbb{Z}$ e $X : (x, y) \mapsto (x + y)$. Para o primeiro, observamos

$$\begin{aligned} 2\|X_*\mu - X_*\nu\|_{VT} &= \sum_{x \in X(\Omega)} \left| \mu(X(\omega) = x) - \nu(X(\omega) = x) \right| \\ &= \sum_{x \in X(\Omega)} \left| \sum_{\omega \in \Omega : X(\omega) = x} \mu(\omega) - \nu(\omega) \right| \\ &\leq \sum_{\omega \in \Omega} |\mu(\omega) - \nu(\omega)| \\ &= 2\|\mu - \nu\|_{VT}, \end{aligned} \quad (2.43)$$

provando o lema. □

Para enunciar o resultado principal dessa seção, vamos apresentar uma distribuição em \mathbb{N} bastante importante, que em particular se comporta muito bem com respeito a somas de variáveis independentes, como veremos.

Definição 2.5.14. Uma variável aleatória X é dita ter distribuição de Poisson com parâmetro λ , se

$$P[X = k] = \frac{\lambda^k e^{-\lambda}}{k!}, \text{ para } k \geq 0 \text{ inteiro.} \quad (2.44)$$

Denotamos isso por $X \stackrel{d}{\sim} \text{Poisson}(\lambda)$.

A distribuição de Poisson se comporta bem com respeito a somas independentes, como mostra o seguinte

Lema 2.5.15. Sejam $X \stackrel{d}{\sim} \text{Poisson}(\lambda_1)$ e $Y \stackrel{d}{\sim} \text{Poisson}(\lambda_2)$ independentes, então $X + Y \stackrel{d}{\sim} \text{Poisson}(\lambda_1 + \lambda_2)$.

Demonstração. Basta calcular

$$\begin{aligned} P[X + Y = k] &= \sum_{j=0}^k P[X = j, Y = k - j] = \sum_{j=0}^k \frac{\lambda_1^j e^{-\lambda_1} \lambda_2^{k-j} e^{-\lambda_2}}{j!(k-j)!} \\ &= e^{-(\lambda_1 + \lambda_2)} \frac{1}{k!} \sum_{j=0}^k \frac{k!}{j!(k-j)!} \lambda_1^j \lambda_2^{k-j} = \frac{e^{(\lambda_1 + \lambda_2)} (\lambda_1 + \lambda_2)^k}{k!}, \end{aligned} \quad (2.45)$$

mostrando o resultado. \square

Nossa próxima tarefa é estimar a distância entre uma variável aleatória com distribuição $\text{Ber}(p)$ e uma $\text{Poisson}(p)$, como segue.

Lema 2.5.16. Para $p \in [0, 1]$, seja $\mu_1 = \text{Ber}(p)$ e $\mu_2 = \text{Poisson}(p)$, então,

$$\|\mu_1 - \mu_2\|_{\text{VT}} \leq p^2. \quad (2.46)$$

Demonstração. Sabemos que

$$\begin{aligned} \|\mu_1 - \mu_2\|_{\text{VT}} &= \frac{1}{2} \sum_x |\mu_1(x) - \mu_2(x)| \\ &= \frac{1}{2} \left(|\mu_1(0) - \mu_2(0)| + |\mu_1(1) - \mu_2(1)| + \sum_{x \geq 2} \mu_2(x) \right) \\ &= \frac{1}{2} \left(e^{-p} - (1 - p) + p(1 - e^{-p}) + (1 - e^{-p} - pe^{-p}) \right) \\ &= \frac{2}{2} p(1 - e^{-p}) \leq p^2, \end{aligned} \quad (2.47)$$

terminando a prova. \square

O teorema principal de convergência dessa seção concerne a soma de variáveis Bernoulli.

Teorema 2.5.17 (Lei dos Pequenos Números). Dado, $n \geq 1$ e $p \in [0, 1]$, suponha que Ω_n , \mathcal{F}_n e P_p sejam dados como em (2.30). Então,

$$\|\text{Bin}(n, p) - \text{Poisson}(pn)\|_{\text{VT}} \leq np^2. \quad (2.48)$$

Demonstração. Basta observar que

$$\begin{aligned}
 \|X_n \circ P_p - \text{Poisson}(pn)\|_{\text{VT}} &\stackrel{\text{Lema 2.5.15}}{=} \|\text{Ber}(p)^{*n} - \text{Poisson}(p)^{*n}\|_{\text{VT}} \\
 &\stackrel{\text{Lema 2.5.13}}{\leq} \|\text{Ber}(p)^{\otimes n} - \text{Poisson}(p)^{\otimes n}\|_{\text{VT}} \\
 &\stackrel{\text{Lema 2.5.10}}{\leq} n \|\text{Ber}(p) - \text{Poisson}(p)\|_{\text{VT}} \stackrel{\text{Lema 2.5.16}}{\leq} np^2,
 \end{aligned} \tag{2.49}$$

provando o teorema. \square

Corolário 2.5.18. *No mesmo contexto do teorema acima, se $p = \lambda/n$, então temos*

$$\|\text{Bin}(n, p) - \text{Poisson}(pn)\|_{\text{VT}} \leq \lambda^2/n, \tag{2.50}$$

que converge a zero com n .

Veremos mais tarde que existem outros tipos de convergência.

Exercício 2.5.18. *Fixado $\lambda > 0$, seja N uma variável aleatória com distribuição $\text{Poisson}(\lambda)$, isto é*

$$P[N = k] = \frac{\lambda^k e^{-\lambda}}{k!} \text{ para } k = 0, 1, \dots \tag{2.51}$$

Considere no mesmo espaço de probabilidade uma sequência de variáveis aleatórias X_1, X_2, \dots que sejam i.i.d., com distribuição $\text{Ber}(1/2)$ e independentes de N .

- a) *Calcule a distribuição de $Z = \sum_{i=1}^N X_i$.*
- b) *Mostre que Z e $N - Z$ são independentes.*

2.6 Espaços produto infinito

Nessa seção estudaremos Ω que são dados por produtos enumeráveis de outros espaços de probabilidade. Mas antes iremos recordar o Teorema da Extensão de Caratheodory.

2.6.1 Recordar é viver...

Vamos lembrar o enunciado do Teorema da Extensão de Caratheodory. Antes, vamos relembrar uma definição importante. Uma família $\mathcal{G} \subseteq \mathcal{P}(\Omega)$ é dita uma álgebra de conjuntos se valem:

- a) $\Omega \in \mathcal{G}$.
- b) Se $A \in \mathcal{G}$, então $A^c \in \mathcal{G}$.
- c) Para todo $n \geq 1$, se $A_1, \dots, A_n \in \mathcal{G}$, então $\bigcup_{i=1}^n A_i \in \mathcal{G}$.

Teorema 2.6.1 (Teorema da Extensão de Caratheodory). *Seja $\mathcal{G} \subseteq \mathcal{P}(\Omega)$ uma álgebra de conjuntos em Ω e suponha que $\mu : \mathcal{G} \rightarrow \mathbb{R}_+$ satisfaça a seguinte propriedade:*

$$\begin{aligned} &\text{Se } (A_i)_{i \in I} \text{ e uma família finita ou enumerável de elementos disjuntos de} \\ &\quad \mathcal{G} \text{ tal que } \bigcup_{i \in I} A_i \in \mathcal{G}, \\ &\text{temos } \mu(\bigcup_{i \in I} A_i) = \sum_{i \in I} \mu(A_i). \end{aligned} \quad (2.52)$$

Então existe uma medida $\bar{\mu} : \sigma(\mathcal{G}) \rightarrow \mathbb{R}_+$ tal que $\bar{\mu}(A) = \mu(A)$ para todo $A \in \mathcal{G}$.

Mostraremos agora uma consequência simples do teorema acima, que é muito utilizada em probabilidade.

Lema 2.6.2 (Extensão por continuidade no vazio). *Seja $\mathcal{G} \subseteq \mathcal{P}(\Omega)$ uma álgebra de conjuntos em Ω e suponha que $P : \mathcal{G} \rightarrow \mathbb{R}_+$ satisfaça as seguintes propriedades:*

- a) $P(\Omega) = 1$,
- b) P é finitamente aditiva e
- c) sempre que $B_1 \supseteq B_2 \supseteq \dots \in \mathcal{G}$ forem tais que $\bigcap_i B_i = \emptyset$ (denotamos isso por $B_i \downarrow \emptyset$), temos que $\lim_i \mu(B_i) = 0$.

Então existe uma única medida $\bar{P} : \sigma(\mathcal{G}) \rightarrow \mathbb{R}_+$ tal que $\bar{P}(A) = P(A)$ para $A \in \mathcal{G}$.

Observe que $P(\Omega) = 1$ somente é necessário para provar a unicidade de \bar{P} , então poderíamos tentar mostrar uma versão mais geral desse lema. Mas no contexto de medidas infinitas, não é de se esperar que $B_i \downarrow \emptyset$ implique $\lim_i \mu(B_i) = 0$, como foi assumido acima (veja também a Proposição 1.2.3). Portanto resolvemos escrever o enunciado com probabilidades.

Exercício 2.6.1. *Dê um exemplo de medida que não satisfaz a terceira hipótese do Lema 2.6.2.*

Demonstração. Primeiro observe que a unicidade segue da Proposição 1.3.4, já que \mathcal{G} é um π -sistema. Iremos agora mostrar que a propriedade (2.52) é válida para P , logo tome $A_1, A_2, \dots \in \mathcal{G}$ disjuntos e tais que $A = \cup_{i \in \mathbb{N}} A_i \in \mathcal{G}$. Definimos o “resto da união” por

$$B_n = A \setminus \bigcup_{i=1}^n A_i. \quad (2.53)$$

Claramente

- a) $B_n \downarrow \emptyset$ e
- b) $B_n \in \mathcal{G}$, pois \mathcal{G} é uma álgebra.

Logo podemos escrever A como a união disjunta $A = \bigcup_{i=1}^n A_i \cup B_n$ e já que P é finitamente aditiva,

$$P(A) = \sum_{i=1}^n P(A_i) + P(B_n), \quad (2.54)$$

mas como $\lim_{n \rightarrow \infty} P(B_n) = 0$, temos

$$P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i), \quad (2.55)$$

mostrando a propriedade (2.52) e concluindo o teorema. \square

2.6.2 Teorema da Extensão de Kolmogorov

O objetivo desta seção é provar um resultado que nos permitirá construir probabilidades em espaços produtos infinitos. Antes precisaremos de introduzir algumas notações. Dada uma coleção de espaços $(E_i)_{i \in \mathbb{N}}$, definimos o espaço produto

$$\Omega = \prod_{i=1}^{\infty} E_i = \{(\omega_i)_{i \in \mathbb{N}} : \omega_i \in E_i \text{ para todo } i \geq 1\} \quad (2.56)$$

e os mapas $X_i : \Omega \rightarrow E_i$, definidos para $i = 1, 2, \dots$ por

$$X_i(\omega_1, \omega_2, \dots) = \omega_i, \quad (2.57)$$

que chamamos de *coordenadas canônicas* associadas ao produto Ω .

Se cada E_i é dotado de uma σ -álgebra \mathcal{A}_i , então definimos

$$\mathcal{F} = \sigma((X_i)_{i \geq 1}), \quad (2.58)$$

que é claramente uma σ -álgebra em Ω . Chamamos \mathcal{F} de σ -álgebra canônica.

Exercício 2.6.2. Mostre que em $(\mathbb{R}^{\mathbb{N}}, \mathcal{F})$ temos que os conjuntos

$$a) A = \{\liminf_{n \rightarrow \infty} X_n \notin \{\infty, -\infty\}\},$$

$$b) B = \{\lim_{n \rightarrow \infty} X_n = 4\} \text{ e}$$

$$c) C = \{\lim_{n \rightarrow \infty} \frac{1}{n} X_n \text{ existe}\}$$

são todos mensuráveis (eventos) com respeito a \mathcal{F} . Além disso $Y = \mathbb{1}_A \liminf_{n \rightarrow \infty} X_n$ é uma variável aleatória em (Ω, \mathcal{F}) .

Exercício 2.6.3. Verifique as seguintes afirmações

a) $\mathcal{F} = \sigma(A_1 \times \cdots \times A_k \times E_{k+1} \times E_{k+2} \times \cdots : k \geq 1, A_i \in \mathcal{A}_i, i \leq k)$, os chamados eventos retangulares.

b) $\mathcal{F} = \sigma(A \times E_{k+1} \times E_{k+2} \times \cdots : k \geq 1, A \in \mathcal{A}_i \otimes \cdots \otimes \mathcal{A}_k)$, conhecidos como eventos cilíndricos.

Definição 2.6.3. Seja $\Omega = \prod_{i \in I} E_i$ um espaço produto (infinito ou finito) dotado de uma probabilidade P . Se X_i é uma coordenada canônica, então chamamos a probabilidade $(X_i)_* P$ de distribuição marginal de P na coordenada i .

Teorema 2.6.4 (Extensão de Kolmogorov). Seja para cada $n \geq 1$ uma medida de probabilidade P_n em \mathbb{R}^n tal que seja satisfeita a seguinte condição de compatibilidade

$$P_{n+1}(A \times \mathbb{R}) = P_n(A), \text{ para todo } A \in \mathcal{B}(\mathbb{R}^n). \quad (2.59)$$

Então existe uma única probabilidade P no espaço produto infinito (Ω, \mathcal{F}) tal que $P(A \times \mathbb{R} \times \cdots) = P_n(A)$ para todo n e todo boreliano A de \mathbb{R}^n .

Demonstração. Considere a classe de conjuntos

$$\mathcal{S}_l = \left\{ \bigcup_{j=1}^k [a_{1,j}, b_{1,j}) \times \cdots \times [a_{l,j}, b_{l,j}) \subseteq \mathbb{R}^l : a_{i,j} \in \mathbb{R} \cup \{-\infty\}, b_{i,j} \in \mathbb{R} \cup \{\infty\} \right\}.$$

Que é obviamente uma álgebra em \mathbb{R}^l e seja também

$$\mathcal{S} = \{A \times \mathbb{R} \times \cdots : \text{onde } l \geq 1 \text{ e } A \in \mathcal{S}_l\}. \quad (2.60)$$

Claramente, \mathcal{S} também é uma álgebra.

Se $B = A \times \mathbb{R} \times \cdots \in \mathcal{S}$ com $A \in \mathcal{S}_l$ como acima, definimos

$$P(B) = P_l(A). \quad (2.61)$$

Note que por (2.59) essa definição independe da escolha de l que usamos na definição de B .

Gostaríamos agora de utilizar o Lemma 2.6.2. Para tanto, tome uma sequência encaixada $B_1 \supseteq B_2 \supseteq \cdots \in \mathcal{S}$ e, supondo que $P(B_n) \geq \delta > 0$ para todo $n \geq 1$, temos de mostrar que sua interseção não pode ser vazia.

Como $B_n \in \mathcal{S}$, podemos escrever

$$B_n = A_n \times \mathbb{R} \times \cdots, \text{ onde } A_n \in \mathcal{S}_{l_n} \text{ e } n \geq 1. \quad (2.62)$$

Podemos obviamente supor que

$$I_n \text{ são estritamente crescentes.} \quad (2.63)$$

A fim de obter um ponto na interseção de B_n , gostaríamos de aproximá-lo usando conjuntos compactos encaixados. Para tanto definimos os conjuntos

$$C_n = C_n^* \times \mathbb{R} \times \dots, \text{ com } C_n^* \in \mathcal{S}_{I_n} \quad (2.64)$$

de forma que C_n^* seja compacto, $C_n^* \subseteq A_n$ e

$$P(B_n \setminus C_n) \leq \frac{\delta}{2^{I_n+1}}, \quad (2.65)$$

o que pode ser feito graças à continuidade de P_{I_n} , que é uma probabilidade.

Temos ainda um problema, pois os conjuntos C_n não são encaixados, e isso nos impede de utilizar resultados sobre interseções de compactos. Introduzimos pois $D_n = \bigcap_{i=1}^n C_i$, que obviamente pertence à álgebra \mathcal{S} , e estimamos

$$P(B_n \setminus D_n) = P\left(\bigcup_{i=1}^n (B_n \setminus C_i)\right) \leq \sum_{i=1}^n P(B_n \setminus C_i) \leq \frac{\delta}{2}, \quad (2.66)$$

donde $P(D_n) = P(B_n) - P(B_n \setminus D_n) \geq \delta/2$. De forma que os D_n são encaixados e não vazios.

Nosso próximo obstáculo vem do fato de que os conjuntos D_n estão definidos em $\mathbb{R}^{\mathbb{N}}$, e gostaríamos de ter conjuntos em espaços de dimensão finita. Isso pode ser feito observando que podemos escrever $D_n = D_n^* \times \mathbb{R} \times \mathbb{R} \times \dots$, onde $D_n^* \in \mathcal{S}_{I_n}$ e

$$D_n^* = \underbrace{C_n^*}_{\text{compacto}} \cap \underbrace{\left(\bigcap_{i=1}^{n-1} C_i^* \times \mathbb{R}^{I_n - I_i}\right)}_{\text{fechado}}, \quad (2.67)$$

de forma que os $D_n^* \subseteq \mathbb{R}^{I_n}$ são compactos e não vazios.

Para cada $n \geq 1$ considere um $\omega^n \in D_n$. Usando um argumento de diagonal de Cantor, podemos obter um $\omega \in \Omega$ e uma sub-sequência de ω^{n_j} que convirja para $\omega \in \Omega$ coordenada a coordenada (observe que $\omega^{n_j} \in \mathbb{R}^{I_{n_j}}$). Para concluir a prova mostramos que $\omega \in \bigcap_{n \geq 1} B_n$. Para isso é suficiente mostrar (lembramos que por definição $C_n \subseteq B_n$) que para todo $n \in \mathbb{N}$

$$\omega = (\omega_1, \omega_2, \dots) \in C_n.$$

O que é equivalente a $(\omega_1, \omega_2, \dots, \omega_n) \in C_n^*$, que vale por compacidade. \square

Observe que usamos muito poucos atributos de \mathbb{R} na prova. Poderíamos na verdade substituir \mathbb{R} por um espaço métrico que satisfaça certas propriedades, como por exemplo a existência de uma álgebra cujos conjuntos possam ser aproximados por compactos. Contudo, decidimos não apresentar essa versão mais geral aqui porque muito em breve obteremos uma versão bem mais geral do Teorema de Kolmogorov usando apenas o resultado para \mathbb{R} .

Exercício 2.6.4. Mostre que a hipótese (2.59) pode ser substituída por

$$P_{n+1}(I_1 \times \dots \times I_n \times \mathbb{R}) = P_n(I_1 \times \dots \times I_n), \quad (2.68)$$

para todo $n \geq 1$ e $I_i = (-\infty, b_i]$, onde $b_i \in \mathbb{R}$, $i \leq n$.

Um importante exemplo do uso deste teorema é o seguinte.

Exemplo 2.6.5. Se P_i são probabilidades em $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, podemos definir $\mathbb{P}_n = \bigotimes_{i=1}^n P_i$ (relembrando, \mathbb{P}_n é a única distribuição em \mathbb{R}^n tal que $\mathbb{P}_n(A_1 \times \dots \times A_n) = \prod_{i=1}^n P_i(A_i)$). Não é difícil verificar que essa lei satisfaz as equações de consistência (2.59). Desta forma, podemos construir uma única \mathbb{P} em $\mathbb{R}^{\mathbb{N}}$ para os quais as coordenadas canônicas X_i são independentes e possuem distribuições marginais P_i . Denotamos nesse caso $\mathbb{P} = \bigotimes_{i \geq 1} P_i$.

Mais adiante no texto daremos outros exemplos bastante interessantes do uso do Teorema 2.6.4.

Exercício 2.6.6. Mostre que se $p > 0$ e $\mathbb{P} = \bigotimes_{i \geq 1} \text{Ber}(p)$ em $\mathbb{R}^{\mathbb{N}}$, então

$$\limsup_{n \rightarrow \infty} X_n = 1 \text{ quase certamente.} \quad (2.69)$$

Exercício 2.6.7. Mostre que se $\mathbb{P} = \bigotimes_{i \geq 1} U_{[0,1]}$ em $\mathbb{R}^{\mathbb{N}}$, então

$$\limsup_{n \rightarrow \infty} X_n = 1 \text{ quase certamente.} \quad (2.70)$$

Exercício 2.6.8. Mostre que se $\mathbb{P} = \bigotimes_{i \geq 1} \text{Exp}(i)$ em $\mathbb{R}^{\mathbb{N}}$, então

$$\limsup_{n \rightarrow \infty} X_n < \infty \text{ quase certamente.} \quad (2.71)$$

Tópico: Percolação

Imagine que gostaríamos de modelar o movimento de um líquido em um meio poroso, como uma rocha ou uma esponja. A primeira tarefa nesse estudo seria modelar esse meio poroso de maneira matematicamente rigorosa, que é o que faremos a seguir.

Fixamos uma dimensão $d \geq 1$ e consideramos o seguinte grafo (\mathbb{Z}^d, E) , onde a rede quadrada \mathbb{Z}^d é o conjunto de vértices e o conjunto de elos é dado por

$$E = \{\{x, y\} \subset \mathbb{Z}^d : |x - y| = 1\},$$

onde $|\cdot|$ representa a distância euclidiana em \mathbb{R}^d .

No nosso modelo, esse grafo pode ser entendido como um cristal periódico onde cada vértice representa uma cavidade do material poroso e os elos são potenciais conexões entre poros vizinhos.

Até agora nosso grafo é apenas uma rede periódica, mas as coisas começam a ficar interessantes a partir de agora. Imaginamos que nosso material poroso está sujeito a variações durante sua formação. Isso se reflete no fato que alguns elos de E podem estar abertos ou não aleatoriamente.

Para o nosso modelo, o espaço amostral vai ser $\Omega := \{0, 1\}^E$ considerado com a σ -álgebra produto. Fixamos um $p \in [0, 1]$ e definimos uma coleção de variáveis aleatórias ω_e , para $e \in E$, que sejam i.i.d. e com distribuição $\text{Ber}(p)$. Chamamos P_p a probabilidade correspondente. Essas variáveis aleatórias induzem um grafo aleatório $G(\omega) = (\mathbb{Z}^d, \mathcal{E}(\omega))$, subgrafo do grafo original, que corresponde a incluir apenas os elos e com $\omega_e = 1$. Mais precisamente

$$\mathcal{E}(\omega) = \{e \in E : \omega_e = 1\}. \quad (2.72)$$

Podemos ver na Figura 2.2 algumas simulações desse grafo aleatório.

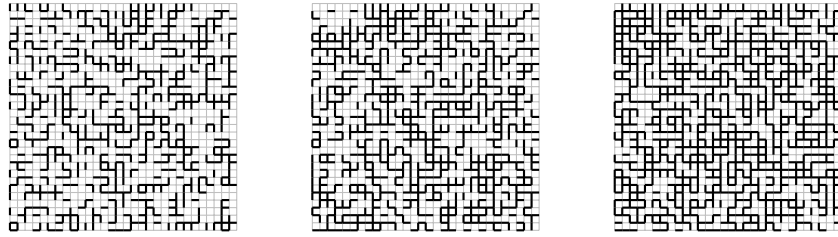


Figura 2.2: Três simulações do grafo aleatório $(\mathbb{Z}^d, \mathcal{E})$, para valores de $p = 0,4$ (esquerda), $p = 0,5$ (centro) e $p = 0,6$ (direita). Tente imaginar como seria caminhar nesse grafo como se ele fosse um labirinto.

Agora que temos um modelo de meio poroso bem definido, precisamos pensar em quais perguntas nos interessam sobre $\mathcal{G} = (\mathbb{Z}^d, \mathcal{E})$. Sendo esse um modelo para passagem de fluidos, as primeiras perguntas que faremos concerne a conectividade de \mathcal{G} .

Exercício 2.6.9. *Mostre que quase certamente $G(\omega)$ é desconexo. Mais precisamente, mostre que existem quase certamente infinitos vértices isolados em $G(\omega)$.*

Como não podemos esperar que $G(\omega)$ seja conexo, podemos nos perguntar algo mais fraco, como por exemplo se a componente conexa da origem $0 \in \mathbb{Z}^d$ em $G(\omega)$ é infinita.

Voltando à Figura 2.2 vemos que, dependendo do valor de $p \in [0, 1]$, pode ser bem difícil ou bem fácil encontrar um caminho longo à partir da origem. Isso é o que estudaremos em mais detalhes no que segue.

Mais precisamente estamos interessados em:

$$A = \{\omega \in \Omega : \text{a componente conexa de } 0 \in \mathbb{Z}^d \text{ em } G(\omega) \text{ é infinita}\}. \quad (2.73)$$

Para estudar A , vamos fazer uma aproximação de A por eventos mais simples

$$A_n = \{\omega \in \Omega : \text{a componente conexa de } 0 \text{ sai da caixa } [-n, n]^d\}, \quad (2.74)$$

para $n \geq 1$.

Exercício 2.6.10. *Mostre que $A = \bigcap_{n=1}^{\infty} A_n$ e consequentemente que A é de fato mensurável e $P(A) = \lim_{n \rightarrow \infty} P(A_n)$.*

Definimos portanto a função $\theta : [0, 1] \rightarrow [0, 1]$ por

$$\theta(p) = P_p(A), \quad (2.75)$$

onde P_p denota a probabilidade correspondente ao valor escolhido de $p \in [0, 1]$.

Exercício 2.6.11. *Mostre que $\theta(p) \leq 1 - (1 - p)^{2d}$.*

Nosso objetivo é entender algumas das propriedades de θ . A nossa intuição diz que quanto maior o valor de p , maiselos serão abertos em \mathcal{G} e portanto maior será o valor de θ , ou em outras palavras, θ deve ser monótona não decrescente.

Exercício 2.6.12. *Construiremos nosso modelo de uma maneira alternativa num espaço de probabilidade maior. Definimos $\Omega_0 := [0, 1]^E$ (com a σ -álgebra produto correspondente), e $(U_e)_{e \in E}$ uma coleção de variáveis aleatórias i.i.d. com distribuição $U[0, 1]$, e \mathbb{P} a probabilidade correspondente. Definimos para cada $p \in [0, 1]$, $X^p : \Omega_0 \rightarrow \Omega$ do jeito seguinte*

$$X_e^p = \mathbb{1}_{[\omega_e \leq p]}. \quad (2.76)$$

Mostre que para todo $p \in [0, 1]$ $(X^p)_ \mathbb{P} = P_p$. Use isso para concluir que θ é monótona não decrescente.*

Iremos agora mostrar a existência de um regime para o qual a componente conexa da origem não é infinita.

Teorema 2.6.5. *Para $p < 1/(2d)$, temos que $\theta(p) = 0$.*

Antes da prova, alguns exercícios.

Exercício 2.6.13. Definimos um caminho como sendo uma sequência x_1, \dots, x_k ($k \in \mathbb{N}$), tal que $\{x_i, x_{i+1}\} \in E$ para todo $i = 1, \dots, k-1$. Tal caminho é dito aberto se $\omega_{\{x_i, x_{i+1}\}} = 1$ para todo $i \leq k-1$. E dizemos que ele é auto-evitante se $x_i \neq x_j$ para todo $1 \leq i < j < k$. Mostre que

$$A_n = \left\{ \omega \in \Omega : \text{existe um caminho aberto } (x_i)_{i=1}^k \text{ com } x_1 = 0 \text{ e } x_k \notin [-n, n]^d \right\}$$

$$A_n = \left\{ \omega \in \Omega : \text{existe um caminho auto-evitante como acima} \right\}.$$

Demonstração. Dado $p < 1/(2d)$ e $n \in \mathbb{N}$, lembramos que

$$\begin{aligned} \theta(p) &\leq P_p(A_n) = P_p \left[\text{existe } k \in \mathbb{N} \text{ e um caminho auto-evitante } (x_i)_{i=1}^k \right. \\ &\quad \left. \text{aberto e com } x_1 = 0 \text{ e } x_k \notin [-n, n]^d \right] \\ &\leq \sum_{k \geq n} \sum_{(x_i)_{i=1}^k \text{ auto-evit.}} P_p[(x_i)_{i=1}^k \text{ aberto}] = \sum_{k \geq n} \sum_{(x_i)_{i=1}^k \text{ auto-evit.}} p^k \\ &\leq \sum_{k \geq n} \sum_{(x_i)_{i=1}^k \text{ caminho}} P_p[(x_i)_{i=1}^k \text{ aberto}] = \sum_{k \geq n} (2d)^k p^k. \end{aligned}$$

Como $p < 1/(2d)$, a soma acima é finita e converge a zero quando n diverge, provando o teorema. \square

Notas - O teorema acima ajuda a compreender o comportamento que observamos no lado esquerdo da Figura 2.2. Mais precisamente, ele nos diz que para valores de p baixos (na verdade 0,4 não é baixo o suficiente para podermos aplicar esse teorema) é difícil encontrar um caminho aberto do centro à borda da caixa.

Na verdade, é possível mostrar que para $d = 2$,

$$\begin{aligned} \theta(p) &= 0 \text{ para todo } p \leq 1/2 \text{ e} \\ \theta(p) &> 0 \text{ para todo } p > 1/2, \end{aligned} \tag{2.77}$$

como foi mostrado por Harris e Kesten, veja por exemplo [Gri99] e [BR06]. De fato, algo bastante interessante está acontecendo nesse modelo para $p = 1/2$, como nos mostrou o trabalho de grandes matemáticos, como: Oded Schramm, Wendelin Werner, Stanislav Smirnov, entre outros.

2.7 Distribuições conjuntas

Um caso bastante importante de distribuição de um elemento aleatório é o caso de vetores. Digamos por exemplo que temos dois elementos aleatórios $X : \Omega \rightarrow E$ e $Y : \Omega \rightarrow E'$. Já sabemos a definição de X_*P e Y_*P (vamos também usar a notação P_X e P_Y) que nada mais são que as distribuições de X e Y , respectivamente.

Mas podemos considerar o vetor (X, Y) que será um elemento aleatório tomando valores em $E \times E'$ e possui também sua própria distribuição dada por $(X, Y)_*P$ (também denotada por $P_{(X,Y)}$). A essa probabilidade em $E \times E'$ damos o nome de distribuição conjunta deste par. .

Vejam as relações que existem entre P_X , P_Y e $P_{(X,Y)}$. Primeiramente, é fácil ver que a distribuição conjunta nos fornece as demais, pois para todo $A \subseteq E$ mensurável

$$P_{(X,Y)}(A \times E') = P[(X, Y) \in A \times E'] = P[X \in A] = P_X(A) \quad (2.78)$$

e analogamente para P_Y . De acordo com a Definição 2.6.3, as distribuições P_X e P_Y nada mais são do que as marginais da distribuição conjunta.

Apesar de podermos extrair as marginais P_X e P_Y de $P_{(X,Y)}$, o contrário não é sempre possível como mostra o seguinte exemplo.

Exemplo 2.7.1. *Sejam X, Y i.i.d. com distribuição $\text{Ber}(1/2)$. Então (X, Y) não tem a mesma distribuição de (X, X) , apesar de que esses vetores possuem as mesmas marginais.*

Exercício 2.7.2. *Mostre que se X e Y são independentes, então $P_{(X,Y)} = P_X \otimes P_Y$.*

Exercício 2.7.3. *Sejam X, Y i.i.d. com distribuição $U_{[0,1]}$ e calcule $P_{(X,X+Y)}$.*

Note que a discussão acima se estende naturalmente para coleções maiores de elementos aleatórios. Mais precisamente, considere um conjunto I qualquer (finito, enumerável ou não enumerável) de índices e seja $(X_i)_{i \in I}$ uma coleção de elementos aleatórios tomando valores em $(E_i)_{i \in I}$. Então a distribuição conjunta destes elementos aleatórios é $P_{(X_i)_{i \in I}}$.

Exercício 2.7.4. *Mostre que no caso acima, se $P_{(X_i)_{i \in J}} = P_{(X'_i)_{i \in J}}$ para todo $J \subseteq I$ finito, então $P_{(X_i)_{i \in I}} = P_{(X'_i)_{i \in I}}$.*

2.8 Probabilidades condicionais

Uma outra maneira de se construir espaços de probabilidade é através de condicionamento, como mostra a seguinte definição.

Definição 2.8.1. *Se (Ω, \mathcal{F}, P) é espaço de probabilidade e $B \in \mathcal{F}$ é tal que $P(B) > 0$, então definimos a probabilidade $P(\cdot|B) : \mathcal{F} \rightarrow [0, 1]$ por*

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \quad (2.79)$$

chamada probabilidade condicional dado o evento B .

Obviamente $P(\cdot|B)$ é uma probabilidade em (Ω, \mathcal{F}) e podemos entendê-la de duas formas: como uma normalização ou como uma tentativa de sucesso. Explicaremos abaixo cada uma dessas interpretações.

Quando restringimos o espaço amostral Ω ao conjunto B (e associamos a $A \in \mathcal{F}$ o valor $P(A \cap B)$), temos uma sub-probabilidade, isto é, possivelmente $P(\Omega \cap B) < 1$. Logo podemos entender o denominador de (2.79) como uma normalização para obtermos novamente uma probabilidade.

Mas a interpretação mais natural de (2.79) é dada pela seguinte proposição. Para enunciá-la, considere (Ω, \mathcal{F}, P) um espaço de probabilidade e defina o produto infinito

$$\bar{\Omega} = \Omega^{\mathbb{N}}, \quad \bar{\mathcal{F}} = \mathcal{F}^{\otimes \mathbb{N}} \quad \text{e} \quad \bar{P} = P^{\otimes \mathbb{N}}. \quad (2.80)$$

Na verdade somente definimos esse produto para $\Omega = \mathbb{R}$, mas como mencionamos abaixo do Teorema da Extensão de Kolmogorov, isso pode ser facilmente generalizado e o faremos posteriormente.

Proposição 2.8.2. *Na situação acima, seja $B \in \mathcal{F}$ com $P(B) > 0$ e defina $T : \bar{\Omega} \rightarrow \mathbb{N}$ por $T(\omega) = \inf\{n \geq 1 : X_n(\omega) \in B\}$, onde os X_n são as coordenadas canônicas. Então $T < \infty$ quase certamente e*

$$X_{T(\omega)}(\omega) \text{ é um elemento aleatório em } \Omega \text{ com distribuição } P(\cdot|B). \quad (2.81)$$

A intuição desta proposição é que se repetimos o experimento (Ω, \mathcal{F}, P) independentemente até obter uma amostra em B , essa terá a distribuição condicional.

Demonstração. Sejam os eventos $A_n = [X_n \in B]$, $n \geq 1$ que são claramente independentes segundo \bar{P} . Logo, como $\sum_n \bar{P}(A_n) = \sum_n P(B) = \infty$, temos pelo Lema de Borel-Cantelli (segunda parte) que $\bar{P}(A_n \text{ infinitas vezes}) = 1$, logo $T < \infty$ quase certamente.

Para ver que $X_{T(\omega)}(\omega)$ é um elemento aleatório, basta escrever

$$[X_T \in A] = \bigcup_{t=1}^{\infty} [X_t \in A, T = t], \quad (2.82)$$

e observar que tanto $[X_t \in A]$ quanto $[T = t] = [X_1 \notin B, \dots, X_{t-1} \notin B, X_t \in B]$ são mensuráveis.

Finalmente podemos usar a decomposição (disjunta) acima para calcular

$$\begin{aligned} \bar{P}[X_T \in A] &= \sum_{t=1}^{\infty} \bar{P}[X_t \in A, T = t] \\ &= \sum_{t=1}^{\infty} \bar{P}[X_t \in A, X_t \in B, X_s \notin B \text{ for } s < t] \\ &= \sum_{t=1}^{\infty} P(A \cap B) P(B^c)^{t-1} = \frac{P(A \cap B)}{1 - P(B^c)} = P(A|B), \end{aligned} \quad (2.83)$$

terminando a prova da proposição. \square

Exercício 2.8.1. Sejam $\lambda > 0$ e $X \sim_d \text{Exp}(\lambda)$ (lembrando a definição da distribuição exponencial: $\text{Exp}(\lambda)(dx) = \lambda \exp\{-\lambda x\} dx$). Mostre que as variáveis com distribuição exponencial não possuem memória, ou seja:

$$P[X > t + s \mid X > t] = P[X > s], \text{ para todo } s, t > 0. \quad (2.84)$$

Ou em outras palavras, sabendo que X é maior que t , a distribuição condicional de $X - t$ ainda é $\text{Exp}(\lambda)$.

Definimos a distribuição geométrica de parâmetro $p \in (0, 1]$ por

$$\text{Geo}(p) = \sum_{i=1}^{\infty} (1-p)^{i-1} p \delta_i. \quad (2.85)$$

Exercício 2.8.2. Inspirado no exercício anterior, mostre que a distribuição geométrica $\text{Geo}(p)$ também satisfaz (2.84) para todos $t, s \in \mathbb{N}$. Mostre que essas são as únicas distribuições com suporte em \mathbb{N} satisfazendo tal propriedade

Exercício 2.8.3. Sejam Y_i , para $i \geq 1$, i.i.d. com distribuição $\text{Ber}(p)$ e defina

$$T = \inf\{i : Y_i = 1\}. \quad (2.86)$$

Mostre que $T \stackrel{d}{\sim} \text{Geo}(p)$.

Exercício 2.8.4. Barry James: Cap. 2-5, Ex: 5, 10, 21, 22 (a) e (b).

Exercício 2.8.5 (Porta dos desesperados). Nas tardes da década de 80, as crianças tinham poucas opções de entretenimento além de assistir Sérgio Malandro, que todos os dias apresentava o seguinte jogo. O participante era apresentado a três portas ($\Omega = \{1, 2, 3\}$) e apenas uma delas (chamada de X) continha um prêmio $X \sim_d U_{\Omega}$ e o jogo seguia três fases:

- a) O participante escolhia uma porta arbitrariamente (digamos $y \in \Omega$),
- b) o Sérgio Malandro abria uma porta X' que não fosse a escolhida nem a premiada ($X' \sim_d U_{\Omega \setminus \{y, X\}}$)
- c) ao participante era dada a oportunidade de trocar sua porta X pela porta restante em $\Omega \setminus \{X, X'\}$.

Mostre que o participante sempre aumenta suas chances ao trocar sua escolha. Tente interpretar esse aparente paradoxo tomando o número de portas para infinito.

Exercício 2.8.6. Emílio e Cristina tiveram dois filhos cujos sexos X, X' são i.i.d. e distribuídos como $U_{\{\sigma^*, \varnothing\}}$. Enunciando hipóteses adequadas se for necessário, calcule

- a) $P[X, X' = \sigma^* \mid \text{pelo menos um é } \sigma^*]$ e
- b) $P[X, X' = \sigma^* \mid \text{pelo menos um é } \sigma^* \text{ e nasceu em uma segunda-feira}]$.

Interprete esses resultados trocando “segunda-feira” por “primeiro de abril”.²

Exercício 2.8.7. Supondo que $P(A \cap B) > 0$, mostre que “ $P(\cdot|A|B) = P(\cdot|B|A)$ ”. Mais precisamente, podemos condicionar P em B e depois a probabilidade resultante em A ou vice-versa.

Exercício 2.8.8. Sejam X, Y variáveis aleatórias em um espaço (Ω, \mathcal{F}, P) , independentes e com distribuição $U_{[0,1]}$.

a) Calcule P_{X+Y} .

b) Considere $P'(\cdot) = P(\cdot | X + Y \leq 1)$ e calcule $X_* P'$.

2.8.1 Regra de Bayes

Frequentemente definimos um espaço de probabilidade através de probabilidades condicionais. Consideramos por exemplo um exame médico para detectar uma doença, caso em que temos

$$\Omega = \{(\text{doente}, +), (\text{doente}, -), (\text{saudável}, +), (\text{saudável}, -)\}, \quad (2.87)$$

com obviamente a σ -álgebra das partes.

Contudo, ao contrário do que fizemos anteriormente, não daremos probabilidades $p_\omega \in [0, 1]$ para cada $\omega \in \Omega$. Poderíamos por exemplo fornecer

$$P(\text{doente}) = 0.005, \quad P(+|\text{saudável}) = 0.01, \quad P(-|\text{doente}) = 0.05. \quad (2.88)$$

Obviamente podemos obter as probabilidades dos complementos dos eventos acima. As probabilidades acima podem ser facilmente estimadas num laboratório e as duas últimas são chamadas respectivamente de probabilidades de *falso positivo* e *falso negativo*. Outra vantagem da representação em (2.88) é que as probabilidades descritas são mais “compartimentadas” no seguinte sentido. Note que $P(\text{doente})$ somente depende da população em questão, enquanto as outras duas dependem apenas do exame e não da população. Isso não pode ser dito das probabilidades de pontos individuais em Ω .

Agora fica fácil construir nosso espaço de probabilidade escrevendo, para $r \in \{+, -\}$ e $e \in \{\text{saudável}, \text{doente}\}$,

$$P(r \cap e) = P(r|e)P(e). \quad (2.89)$$

E as probabilidades do lado direito da equação acima estão todas determinadas em (2.88) (possivelmente tomando complementos).

Contudo, o que estamos interessado muitas vezes é em como interpretar resultados de um exame. Por exemplo, quanto vale $P(\text{doente}|+)$? Isso nos é fornecido em geral pela regra de Bayes enunciada na seguinte proposição.

²Gratos ao Ricardo Misturini por sugerir esse problema

Proposição 2.8.3. Se $(A_j)_{j \in I}$ formam uma partição (finita o enumeável) de Ω e $B \in \mathcal{F}$ tem probabilidade positiva, então

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{\sum_{j \in I} P(A_j)P(B|A_j)}. \quad (2.90)$$

Demonstração. Basta notar que

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{P(B)} = \frac{P(A_i)P(B|A_i)}{\sum_{j \in I} P(B \cap A_j)} = \frac{P(A_i)P(B|A_i)}{\sum_{j \in I} P(A_j)P(B|A_j)}. \quad (2.91)$$

□

Exercício 2.8.9. Utilize a fórmula acima para calcular $P(\text{doente}|+)$ com os dados em (2.88). Comente o resultado.

Exercício 2.8.10. Barry James: Cap. 1, Ex: 18 e 19.

2.9 Núcleos de transição

Já focamos bastante energia em variáveis aleatórias independentes. Por exemplo, estudamos em detalhes o que acontece com a soma de tais variáveis. Agora passaremos a estudar elementos aleatórios dependentes e o primeiro passo para isso é obter um método geral de construí-los.

Definiremos agora um núcleo de transição. Intuitivamente, ele nos dá uma maneira de usar um elemento aleatório em um espaço para induzir uma probabilidade em outro espaço. Um exemplo em que poderíamos utilizar essa construção seria o seguinte.

Digamos que estamos preocupados com a possibilidade de um deslizamento de terra em uma determinada região. A ocorrência desse deslizamento é algo aleatório, mas que certamente depende da quantidade de chuva no período, que também podemos modelar como sendo aleatória.

Após estudarmos alguns trabalhos anteriores, descobrimos uma função $F : \mathbb{R}_+ \rightarrow [0, 1]$ que nos dá a probabilidade de um deslizamento ocorrer, como função da quantidade de chuva em milímetros.

Lendo o histórico pluvial da região, podemos estimar a distribuição Q em \mathbb{R}_+ correspondente à quantidade de chuva naquele período. A lei F_*Q (também chamada de Q_F) é uma lei em $[0, 1]$ que nos dá a distribuição da probabilidade de deslizamento, mas como seguimos em frente para obter a probabilidade de deslizamento (um número entre zero e um)? Sabemos como fazer isso ao terminar essa seção.

Sejam (E_1, \mathcal{A}_1) e (E_2, \mathcal{A}_2) espaços mensuráveis.

Definição 2.9.1. Um núcleo de transição entre E_1 e E_2 é uma função

$$K : E_1 \times \mathcal{A}_2 \rightarrow [0, 1], \quad (2.92)$$

tal que

- a) para todo $y \in E_1$, $K(y, \cdot)$ é uma probabilidade em (E_2, \mathcal{A}_2) e
 b) para todo $A \in \mathcal{A}_2$, a função $K(\cdot, A) : E_1 \rightarrow [0, 1]$ é \mathcal{A}_1 -mensurável.

Exemplo 2.9.1. Daremos agora o exemplo da probabilidade de deslizamento como função de F (que será possivelmente uma variável aleatória). Nesse caso, seja $E_1 = [0, 1]$ e $E_2 = \{0, 1\}$ com as σ -álgebras naturais e defina

$$K(p, A) = ((1 - p)\delta_0 + p\delta_1)(A). \quad (2.93)$$

Vamos verificar que K definido acima é um núcleo de transição. De fato,

- i) $K(p, \cdot)$ é a distribuição Bernoulli com parâmetro p , que obviamente é uma probabilidade,
 ii) além disso, $K(\cdot, \Omega) = 1$, $K(\cdot, \emptyset) = 0$ e $K(\cdot, \{0\}) = 1 - p = 1 - K(\cdot, \{1\})$, que obviamente são mensuráveis.

Isso prova que esse K específico é um núcleo de transição

Exemplo 2.9.2 (Caso discreto). Sejam E_1 e E_2 espaços finitos ou enumeráveis. Se $p : E_1 \times E_2 \rightarrow [0, 1]$ é tal que para todo $y \in E_1$ temos $\sum_{z \in E_2} p(y, z) = 1$, então

$$K(y, A) := \sum_{z \in A} p(y, z) \text{ é um núcleo de transição entre } E_1 \text{ e } E_2. \quad (2.94)$$

Nesse caso $p(y, z)$ representa a probabilidade que a segunda coordenada seja z , se a primeira é y .

Exercício 2.9.3. Mostre que se E_1 e E_2 são enumeráveis então todo núcleo entre E_1 e E_2 pode ser escrito na forma do exemplo acima.

Exemplo 2.9.4 (Caso absolutamente contínuo). Digamos que E_1 e E_2 sejam dotados de medidas μ_1 e μ_2 σ -finitas. Seja $\rho : E_1 \times E_2 \rightarrow \mathbb{R}_+$ mensurável e tal que para μ_1 -quase todo $y \in E_1$, tenhamos $\int_{E_2} \rho(y, z) \mu_2(dz) = 1$. Então

$$K(y, A) := \int_A \rho(y, z) \mu_2(dz) \text{ é um núcleo de transição entre } E_1 \text{ e } E_2. \quad (2.95)$$

Note que $K(\cdot, A)$ está bem definido para μ_2 -quase todo ponto pelo Teorema de Fubini.

Exercício 2.9.5. Prove que os dois exemplos acima de fato definem um núcleo.

Tipicamente, definimos os núcleos de transição introduzindo $K(y, \cdot)$ como sendo uma medida que depende de y . Nesse caso, uma das condições para que K seja um núcleo está automaticamente satisfeita, restando apenas mostrar que $K(\cdot, A)$ é mensurável para quaisquer $A \in \mathcal{A}_2$. Mas obviamente o conjunto \mathcal{A}_2 pode ser muito complexo, então gostaríamos de apenas verificar que $K(\cdot, A)$ é mensurável para os conjuntos A em uma classe rica o suficiente.

Proposição 2.9.2. Seja $K : E_1 \times \mathcal{A}_2 \rightarrow [0, 1]$, tal que $K(y, \cdot)$ é uma medida para todo $y \in E_1$. Se $K(\cdot, A)$ é mensurável para todo $A \in \mathcal{G}$, onde \mathcal{G} é um π -sistema que gera \mathcal{A}_2 , então K é um núcleo de transição.

Demonstração. Como de costume, vamos definir

$$\mathcal{B} = \{B \in \mathcal{A}_2 : K(\cdot, B) \text{ é } \mathcal{A}_1\text{-mensurável}\}. \quad (2.96)$$

Obviamente, como $K(y, \cdot)$ é uma probabilidade, vale que

- a) $\Omega \in \mathcal{B}$, pois a função constante igual a um é mensurável.
- b) Se $B \in \mathcal{B}$, então $B^c \in \mathcal{B}$, pois $1 - f$ é mensurável se f o é.
- c) E se $B_1, B_2, \dots, B_n \in \mathcal{B}$ são disjuntos, então $\bigcup_{i=1}^n B_i \in \mathcal{B}$, pois a soma de funções mensuráveis também é mensurável.

A discussão acima mostra que \mathcal{B} é um λ -sistema que contém o π -sistema \mathcal{G} . Daí, vemos pelo Teorema 1.3.3 que $\mathcal{A}_2 = \sigma(\mathcal{G}) \subseteq \mathcal{B}$, provando a proposição. \square

Exercício 2.9.6. Seja $K : \mathbb{R} \times \mathcal{B}(\mathbb{R}) \rightarrow [0, 1]$ dada por $K(y, \cdot) = U_{[y-1, y+1]}$. Mostre que K define um núcleo de transição.

Apesar de interessante, a definição acima ainda não nos permitiu definir espaços de probabilidade novos. Isso será possibilitado pelo próximo resultado, que pode ser visto como uma generalização do Teorema de Fubini.

Teorema 2.9.5 (Fubini para Núcleos de Transição). *Dado um núcleo de transição K de (E_1, \mathcal{A}_1) para (E_2, \mathcal{A}_2) e uma probabilidade P_1 em E_1 , existe uma única probabilidade P em $(E_1 \times E_2, \mathcal{A}_1 \otimes \mathcal{A}_2)$ tal que*

$$\int_{E_1 \times E_2} f dP = \int_{E_1} \int_{E_2} f(y, z) K(y, dz) P_1(dy), \quad (2.99)$$

para toda $f : E_1 \times E_2 \rightarrow \mathbb{R}_+$. Em particular, $P(A_1 \times A_2) = \int_{A_1} K(y, A_2) P_1(dy)$. Nesse caso escrevemos $P = P_1 \star K$.

Antes de iniciar a prova do teorema, vamos ver que as integrais do lado direito de (2.99) estão bem definidas. Para isso, definimos para $y \in E_1$ a função fatiadora $\phi_y : E_2 \rightarrow E_1 \times E_2$ dada por $\phi_y(z) = (y, z)$. Obviamente essa função é mensurável, pois

$$\phi_y^{-1}(A_1 \times A_2) = \begin{cases} \emptyset, & \text{se } y \notin A_1 \text{ e} \\ A_2, & \text{se } y \in A_1. \end{cases} \quad (2.100)$$

Dessa forma, para definirmos $\int f(y, z) K(y, dz)$, introduzimos $f_y : A_2 \rightarrow \mathbb{R}_+$ dada por $f_y(z) = f(y, z)$, que é mensurável pois $f_y = f \circ \phi_y$.

Assim, gostaríamos de integrar a função $y \mapsto \int f_y(z) K(y, dz)$, que está obviamente bem definida. Porém resta a pergunta, será que essa expressão define uma função mensurável de y ?

Lema 2.9.6. *Se K é um núcleo de transição, então para toda $f : E_1 \times E_2 \rightarrow \mathbb{R}_+$ que seja $\mathcal{A}_1 \otimes \mathcal{A}_2$ mensurável, temos que $g^f : A_1 \rightarrow \mathbb{R}_+$ dada por*

$$g^f(y) = \int f_y(z) K(y, dz) \quad (2.101)$$

é \mathcal{A}_1 -mensurável.

Demonstração. Se $f = \mathbb{1}_{A_1 \times A_2}$ para $A_i \in \mathcal{A}_i, i = 1, 2$, então temos que $g^f(y) = K(y, A_2)\mathbb{1}_{A_1}$, que obviamente é mensurável pois K é um núcleo.

Definimos $\mathcal{D} = \{B \in \mathcal{A}_1 \otimes \mathcal{A}_2 : g^{\mathbb{1}_B} \text{ é } \mathcal{A}_1\text{-mensurável}\}$. É fácil ver que \mathcal{D} é um λ -sistema que contém o π -sistema dos retângulos, logo $\mathcal{D} = \mathcal{A}_1 \otimes \mathcal{A}_2$.

Acabamos de ver que g^f é mensurável para toda f indicadora, donde o mesmo vale para f simples por linearidade e para toda f positiva pelo Teorema da Convergência Monótona (lembre que limite de funções mensuráveis é mensurável). \square

Estamos prontos agora para fornecer a

Demonstração do Teorema 2.9.5. Já sabemos que a integral do lado direito de (2.99) está bem definida (assumindo possivelmente o valor infinito). A unicidade vale obviamente pois (2.99) aplicado a funções indicadoras temos necessariamente para todos B

$$P(B) = \int_{E_1} \int_{E_2} \mathbb{1}_B K(y, dz) P_1(dy). \quad (2.102)$$

Só temos que verificar a fórmula acima nos define uma probabilidade em $(E_1 \times E_2, \mathcal{A}_1 \otimes \mathcal{A}_2)$.

De fato,

- a) obviamente $P(\Omega) = \int_{E_1} \int_{E_2} K(y, dz) P_1(dy) = 1$ e
- b) se $(B_i)_{i \in I}$ é uma família finita ou enumerável de eventos disjuntos (em $\mathcal{A}_1 \otimes \mathcal{A}_2$) então $\mathbb{1}_{\bigcup_{i \in I} B_i} = \sum_{i \in I} \mathbb{1}_{B_i}$ a σ -aditividade de P segue das propriedades básicas (linearidade e Teorema de convergência monótona) da integração.

Isto demonstra o teorema. \square

Exercício 2.9.7. Considere duas probabilidades P_i em (E_i, \mathcal{A}_i) para $i = 1, 2$ e $K : E_1 \times E_2 \rightarrow [0, 1]$ dado por $K(y, A) = P_2(A)$. Mostre que K é núcleo e que $P_1 \star K = P_1 \otimes P_2$. Relacione esse resultado ao Teorema de Fubini clássico para produtos de medidas.

Exercício 2.9.8. Considere o núcleo do Exemplo 2.9.1 e calcule:

- a) $U_{[0,1]} \star K[X_2 = 1]$,
- b) $P_1 \star K[X_2 = 1]$, onde $dP_1 = 2x \, dx$ e
- c) encontre a distribuição de $(X_1)_* (U_{[0,1]} \star K[\cdot | X_2 = 1])$. Interprete o resultado.

Exercício 2.9.9. Seja $P = P_1 \star K$ como acima e $Q(\cdot) = P[\cdot | X_2 = 1]$. Calcule

$$\int_{[0,1] \times \{0,1\}} X_1 \, dQ \quad (2.103)$$

Exercício 2.9.10. Para $0 \leq a < b \leq 1$, definimos a probabilidade $U_{[a,b]}$ em $([0, 1], \mathcal{B}([0, 1]))$ através da seguinte fórmula $U_{[a,b]}(B) = \mathcal{L}(B \cap [a, b]) / (b - a)$. Consideramos também a função $K : [0, 1] \times \mathcal{B}([0, 1]) \rightarrow [0, 1]$ dada por $K(x, \cdot) = U_{[0,x]}(\cdot)$, se $x > 0$ e $K(0, \cdot) = \delta_0(\cdot)$.

- a) Mostre que K é um núcleo de transição.
- b) Calcule $U_{[0,1]} \star K[X_1 < 1/2]$ e $U_{[0,1]} \star K[X_2 < 1/2]$, onde X_1 e X_2 são as projeções canônicas em $[0, 1]^2$.
- c) Mostre que $U_{[0,1]} \star K$ é absolutamente contínua com respeito à medida de Lebesgue em $[0, 1]^2$ e calcule sua densidade.

Exercício 2.9.11. Considere $K : E_1 \times \mathcal{A}_2 \rightarrow [0, 1]$ dada por $K(p, \cdot) = \text{Exp}(p)$. Mostre que K é núcleo de transição e calcule $U_{[0,1]} \star K[X_2 > 1]$.

Exercício 2.9.12. Se K é um núcleo de transição entre E_1 e E_2 e $\{y\} \in \mathcal{A}_1$ satisfaz $P_1(\{y\}) > 0$, mostre que

$$P_1 \star K[X_2 \in \cdot | X_1 = y] = K(y, \cdot). \quad (2.104)$$

Ou em outras palavras, K nos dá a distribuição condicional de X_2 dado $X_1 = y$.

Posteriormente estenderemos o resultado acima para o caso $P_1(\{y\}) = 0$, mas isso demandará algum esforço.

Vamos introduzir uma última notação com respeito a núcleos de transição. Muitas vezes, não estamos interessados na distribuição conjunta de $P_1 \star K$ em $E_1 \times E_2$, mas apenas na distribuição marginal da segunda coordenada. No nosso problema da chuva por exemplo, talvez poderíamos estar interessados apenas na probabilidade final de ocorrer um deslizamento. Nesse caso, é conveniente escrever

$$P_1 K := (X_2)_*(P_1 \star K) = (P_1 \star K)_{X_2}. \quad (2.105)$$

Exercício 2.9.13. Seja $K : \mathbb{R}_+ \times \mathcal{B}(\mathbb{R}_+) \rightarrow [0, 1]$ dada pela equação $K(x, A) = \int_A x \exp\{-xt\} dt$.

- a) Prove que K é um núcleo de transição.
- b) Seja P dada por $P = \text{Exp}(1) \star K$. Obtenha $P[X_2 > x_2]$ para todo $x_2 \geq 0$ (lembrando que X_2 denota a segunda coordenada no espaço produto onde está definida P). Compare a probabilidade acima com $K(1, [x_2, \infty))$.
- c) Mostre que $P[X_1 + X_2 \geq z] = \int_0^z \exp\{-x(z-x+1)\} dx + \exp\{-z\}$.

2.10 Espaços canônicos

Em várias áreas da matemática, existe um importante conceito de equivalência entre duas estruturas, como por exemplo: homeomorfismos, isometrias e isomorfismos. Nessa seção estudaremos o caso análogo para espaços mensuráveis, que nos trará uma grande surpresa.

Definição 2.10.1. *Uma função $\phi : E \rightarrow E'$ entre dois espaços mensuráveis é dita bi-mensurável quando ϕ é uma bijeção mensurável, com inversa mensurável.*

Vamos agora tentar classificar os espaços a menos de bi-mensurabilidade. Descobriremos que na verdade os borelianos da reta incluem praticamente tudo em que podemos estar interessados. Começamos com a seguinte definição.

Definição 2.10.2. *Dizemos que o espaço mensurável (E, \mathcal{A}) é canônico se existe uma função $\phi : E \rightarrow B$ bi-mensurável para algum $B \in \mathcal{B}(\mathbb{R})$.*

Antes de mostrar que essa classe de espaços canônicos inclui muitíssimos exemplos, vamos motivar a definição acima exemplificando como esse conceito pode ser utilizado.

Teorema 2.10.3 (Extensão de Kolmogorov Extendida). *Se $(E_1, \mathcal{F}_1), (E_2, \mathcal{F}_2), \dots$ são espaços mensuráveis canônicos, então o Teorema 2.6.4 (da extensão de Kolmogorov) também é válido no espaço produto $\Omega = E_1 \times E_2 \times \dots$: Se a seguinte condição de consistência for válida*

$$\forall n \geq 0, \forall A \in \bigotimes_{i=1}^n \mathcal{F}_i, \quad P_{n+1}(A \times E_{n+1}) = P_n(A). \quad (2.106)$$

então existe uma probabilidade P em Ω tal que

$$\forall n \geq 0, \forall A \in \bigotimes_{i=1}^n \mathcal{F}_i, \quad P(A \times E_{n+1} \times E_{n+2} \times \dots) = P_n(A). \quad (2.107)$$

Demonstração. Sejam $\phi_i : E_i \rightarrow B_i \in \mathcal{B}(\mathbb{R})$ bijeções bi-mensuráveis e defina $\bar{\phi}_n : E_1 \times \dots \times E_n \rightarrow \mathbb{R}^n$ por $\bar{\phi}_n(\omega_1, \dots, \omega_n) = (\phi_1(\omega_1), \dots, \phi_n(\omega_n))$. Assim podemos introduzir as medidas de probabilidade

$$\bar{P}_n = (\bar{\phi}_n)_* P_n, \text{ em } \mathbb{R}^n. \quad (2.108)$$

É fácil verificar que as \bar{P}_n são consistentes como em (2.59). Logo, existe \bar{P} em $(\mathbb{R}^n, \mathcal{F})$ estendendo \bar{P}_n .

Vamos agora definir uma medida em $\prod_{i=1}^{\infty} E_i$. Para tanto, primeiramente fixamos para cada $i \geq 1$ um elemento arbitrário w_i de E_i e definimos $\psi_i : \mathbb{R} \rightarrow E_i$ por

$$\psi_i(x) = \begin{cases} \phi_i^{-1}(x), & \text{se } x \in B_i, \\ w_i & \text{no caso contrário.} \end{cases}$$

Como $B_i \in \mathcal{B}(\mathbb{R})$, concluímos que ψ_i é mensurável.

Finalmente, consideramos o mapa $\Psi : \mathbb{R}^{\mathbb{N}} \rightarrow \Omega$ dado por

$$\Psi(x_1, x_2, \dots) = (\psi_1(x_1), \psi_2(x_2), \dots). \quad (2.109)$$

Resta mostrar que a medida $P = \Psi_* \bar{P}$ estende as probabilidades P_n . Observe que

$$\begin{aligned} P(A_1 \times \dots \times A_n \times E_{n+1} \times \dots) &= \bar{P}(\Psi^{-1}(A_1 \times \dots \times A_n \times E_{n+1} \times \dots)) \\ &= \bar{P}(\psi_1^{-1}(A_1) \times \dots \times \psi_n^{-1}(A_n) \times \mathbb{R} \times \dots) \\ &= \bar{P}_n(\psi_1^{-1}(A_1) \times \dots \times \psi_n^{-1}(A_n)) \\ &= P_n(\phi_1^{-1}(\psi_1^{-1}(A_1)) \times \dots \times \phi_n^{-1}(\psi_n^{-1}(A_n))) \\ &= P_n(A_1 \times \dots \times A_n), \end{aligned}$$

concluindo a prova do teorema. \square

Uma ferramenta importante para construirmos espaços canônicos é a seguinte.

Lema 2.10.4. *Seja (E, \mathcal{A}) é um espaço canônico e $A \in \mathcal{A}$, então A também é canônico quando dotado da σ -álgebra $\{A \cap C : C \in \mathcal{A}\}$ induzida por \mathcal{A} em A .*

Demonstração. Seja $\phi : E \rightarrow B \in \mathcal{B}(\mathbb{R})$ uma função bi-mensurável que mostra que E é canônico. Consideramos $\phi' : A \rightarrow \mathbb{R}$ dada pela restrição de ϕ a A e precisamos mostrar as seguintes afirmativas:

- a) ϕ' é injetiva.
- b) ϕ' é mensurável.
- c) $\phi(A) \in \mathcal{B}(\mathbb{R})$.
- d) A inversa de ϕ' (chamada ψ') de $\phi'(A)$ em A é mensurável.

Vejamos,

- a) ϕ ser injetiva implica que ϕ' também o é.
- b) Dado $D \in \mathcal{B}(\mathbb{R})$, $(\phi')^{-1}(D) = A \cap \phi^{-1}(D)$ which is of the form $A \cap C$ with $C \in \mathcal{B}(\mathbb{R}^d)$.
- c) Denotando por $\psi : B \rightarrow E$ a inversa de ϕ , temos que $\phi(A) = \psi^{-1}(A) \in \mathcal{B}(B)$ pois ψ é mensurável.
- d) Finalmente, se $D \in \mathcal{B}(A)$, então $(\psi')^{-1}(D) = \psi^{-1}(D) \in \mathcal{B}(B)$, novamente pela mensurabilidade de ψ .

Concluindo portanto a bi-mensurabilidade de ϕ' quando o seu contra-domínio é restrito a sua imagem. \square

A seguir daremos um exemplo de espaço canônico que será importante na seção seguinte.

Lema 2.10.5. *O espaço produto $E = \mathbb{N} \times \mathbb{N} \times \dots$, dotado da σ -álgebra produto é canônico.*

Demonstração. Primeiramente definimos em E a Métrica de Hamming:

$$d_H(x, y) = \sum_{i \geq 1} \frac{1}{2^{i+1}} \mathbb{1}_{\{x_i \neq y_i\}}. \quad (2.110)$$

Fica como exercício mostrar que a σ -álgebra dos borelianos induzida por essa métrica coincide com a σ -álgebra produto em E . Definimos agora o mapa $\phi : E \rightarrow \mathbb{R}$ dado por

$$\phi(n_1, n_2, \dots) = 2^{-n_1} + 2^{-1-n_1-n_2} + \dots + 2^{-k-\sum_{i=1}^k n_i} + \dots \quad (2.111)$$

Também deixamos a cargo do leitor mostrar que ϕ define um homeomorfismo entre (E, d_H) e um boreliano de \mathbb{R} . \square

2.10.1 Espaços poloneses

Nessa seção mostraremos que todos espaços chamados poloneses são canônicos.

Definição 2.10.6. *Um espaço métrico (E, d) é dito polonês se é separável e completo.*

Exemplo 2.10.1.

- a) *Todo espaço enumerável Ω pode ser feito em um espaço métrico polonês de forma que a σ -álgebra de Borel seja $\mathcal{P}(\Omega)$.*
- b) *\mathbb{R}^n e $C([0, 1])$ são notoriamente poloneses.*

Exercício 2.10.2. *Se $(E_i, d_i)_{i=1}^\infty$ é uma sequência de espaços métricos poloneses, mostre que $E = \prod_{i=1}^\infty E_i$ com a métrica*

$$d(x, y) = \sum_{i=1}^\infty \frac{1}{2^{i+1}} \frac{d_i(x_i, y_i)}{1 + d_i(x_i, y_i)} \quad (2.112)$$

também é polonês. Mostre também que a topologia induzida por essa métrica é equivalente à topologia produto em E .

Outros exemplos de espaços poloneses são dados pelo seguinte lema, que também será útil para provar o resultado principal desta seção.

Lema 2.10.7. *Seja (E, d) um espaço polonês e $G, F \subseteq E$ um aberto e um fechado de E respectivamente. Então, existe uma métrica d' em $F \cap G$ tal que*

- a) *d e d' são equivalentes em $F \cap G$ (induzem a mesma noção de convergência),*
- b) *$d(x, y) \leq d'(x, y)$ para todo $x, y \in F \cap G$ e*

c) $(F \cap G, d')$ é polonês.

Demonstração. A primeira observação que faremos é que $F \cap G$ é separável com respeito a d . Isso segue do fato de separabilidade ser equivalente à existência de uma base enumerável.

Vamos definir para x, y em G ,

$$d'(x, y) = d(x, y) + \left| \frac{1}{d(x, G^c)} - \frac{1}{d(y, G^c)} \right|, \quad (2.113)$$

onde $d(x, A) = \inf\{d(x, x') : x' \in A\}$. Não é difícil ver que com a definição acima (e deixamos como exercício) que:

- a) As métricas d e d' são equivalentes em G .
- b) $F \cap G$ é separável quando dotado da métrica d' .
- c) $(F \cap G, d')$ é completo.

Isso termina a prova do lema. \square

Exemplo 2.10.3. Um importante exemplo é dado por espaços produto. Seja $(E_i, d_i)_{i=1}^{\infty}$ uma sequência de espaços poloneses e introduza em $E = \prod_{i=1}^{\infty} E_i$ a métrica d definida em (2.112). Então, se $A_1 \subseteq E_1, \dots, A_k \subseteq E_k$ forem abertos, o retângulo $R = A_1 \times \dots \times A_k \times E_{k+1} \times \dots$ é aberto. Dessa forma vemos que tanto R como R^c podem ser dotados de métricas com as quais se tornam espaços poloneses. Além disso tais métricas podem ser escolhidas satisfazendo as hipóteses do Lema 2.10.7

O próximo lema é o ingrediente chave para provarmos o resultado principal dessa seção. Ele nos dá uma maneira de fatiar um espaço polonês em uma partição de espaços poloneses pequenos.

Lema 2.10.8. Seja (E, d) um espaço polonês e $r > 0$. Então existe uma partição finita ou enumerável $(A_i)_{i \in I}$ de E e métricas $(d_i)_{i \in I}$ nesses respectivos subconjuntos de forma que para todo $i \in I$,

- a) (A_i, d_i) são espaços poloneses disjuntos.
- b) d_i e d são equivalentes em A_i e $d_i \geq d$.
- c) O diâmetro de A_i (com respeito a d) é menor ou igual a r .

Observe que podemos sempre escolher $I = \mathbb{N}$ mas nesse caso os A_i podem ser vazios.

Demonstração. Obtemos através da separabilidade de E , uma coleção de bolas $(B_i)_{i \geq 1}$ com diâmetros limitados por r e cobrindo E . Então definimos

$$A_1 = B_1, \quad \text{e} \quad A_n = B_n \setminus \bigcup_{i=0}^{n-1} B_i \quad \text{para } n \geq 1. \quad (2.114)$$

Agora podemos dotar cada um dos A_i com a métrica d_i obtida através do Lema 2.10.7 (observe para tanto que os A_i são dados por interseções de um aberto com um fechado). As propriedades enunciadas no lema são trivialmente satisfeitas. \square

Terminamos essa seção com esse importante resultado, que confirma nossa afirmação de que quase todos os espaços mensuráveis que podemos nos interessar são canônicos.

Teorema 2.10.9. *Todo sub-conjunto boreliano de espaço polonês (E, d) é canônico.*

Demonstração. Primeiramente, pelo Lema 2.10.4, basta mostrar que todo espaço E polonês é canônico. Pelo Lema 2.10.5 e novamente o Lema 2.10.4,

$$\text{basta construir uma função bi-mensurável } \phi : E \rightarrow B \in \mathcal{B}(\mathbb{N}^{\mathbb{N}}) \quad (2.115)$$

e depois compô-la com uma função bi-mensurável $\phi' : B \rightarrow C \in \mathcal{B}(\mathbb{R})$.

Para começar, construiremos uma partição encaixada de E . Mais precisamente, defina os conjuntos M_n que serão utilizados como índices

$$M_n = \mathbb{N}^n \quad \text{para } n \geq 1 \text{ e } M = \cup_n M_n. \quad (2.116)$$

Vamos definir borelianos A_m de E e métricas d_m em A_m para cada $m \in M$. Faremos isso da seguinte forma:

- a) se $m = i \in M_1$, então definimos A_1, A_2, A_3, \dots e d_1, d_2, d_3, \dots como no Lema 2.10.8 com $r = 1$,
- b) se (A_m, d_m) já foi definido para algum $m \in M_n$, então utilizamos também o Lema 2.10.8 com $r = 1/n$ para particionar o conjunto A_m (com a métrica d_m) em $A_{(m,1)}, A_{(m,2)}, \dots$ com suas respectivas métricas $d_{(m,1)}, d_{(m,2)}, \dots$

Obviamente suporemos que são válidas as propriedades de tais métricas garantidas pelo Lema 2.10.8.

Podemos desde já definir $\phi : E \rightarrow \mathbb{N}^{\mathbb{N}}$ e para tanto, considere $x \in E$. Indutivamente

- a) como $\{A_m\}_{m \in M_1}$ formam uma partição de E , definimos $\phi_1(x)$ como o único índice tal que $x \in A_{\phi_1(x)}$,
- b) se já encontramos $\phi_1(x), \dots, \phi_n(x)$ tal que $x \in A_{(\phi_1(x), \dots, \phi_n(x))}$, então o fato que particionamos o último conjunto na definição de $A_m, m \in M_{n+1}$ nos garante que podemos definir unicamente $\phi_{n+1}(x)$ de forma a continuar a indução.

Da maneira acima já obtivemos $\phi(x) = (\phi_1(x), \phi_2(x), \dots)$. Para terminar, devemos mostrar que ϕ é bi-mensurável quando seu contra-domínio é restrito à sua imagem.

Isso começa com a prova de que ϕ é injetiva. Se $\phi(x) = \phi(y)$, então existe uma sequência $m_n \in M_n$ tal que $x, y \in A_{m_n}$ para todo n . Mas isso não é possível dado que o diâmetro de $A_{m_{n+1}}$ é menor ou igual a $1/n$ na métrica $d_{m_n} \geq d$. Isso mostra que $x = y$.

Veamos agora que ϕ é mensurável. Seja $w \in \mathbb{N}^{\mathbb{N}}$ tal que $\phi(x) = w$ e tome $G \subseteq \mathbb{N}^{\mathbb{N}}$ com $G = \{(w_1, \dots, w_l)\} \times \mathbb{N}^{\mathbb{N}}$ (esses conjuntos geram a σ -álgebra canônica em $\mathbb{N}^{\mathbb{N}}$). Claramente, $\phi^{-1}(G) = A_{(\phi_1(x), \dots, \phi_l(x))}$, de forma que mostramos que ϕ é mensurável.

Para mostrar que sua inversa $\psi : \phi(E) \rightarrow E$ é mensurável, veremos que ela é de fato contínua com respeito à Métrica de Hamming definida em (2.110). Dado $n \geq 1$, tomamos $\delta < 2^{-n}$. Se $w, w' \in \phi(E)$ são tais que $d_H(w, w') < \delta$ em $\mathbb{N}^{\mathbb{N}}$, então $w_i = w'_i$ para todo $i \leq n$, de forma que $\phi^{-1}(w)$ e $\phi^{-1}(w')$ pertencem a $A_{(w_1, \dots, w_n)}$. A continuidade de ϕ^{-1} segue do fato que o diâmetro de $A_{(w_1, \dots, w_n)}$ é no máximo $1/n$ (com respeito a $d_{(w_1, \dots, w_{n-1})}$ e portanto com respeito a d).

Mas atenção, apesar de que parece que provamos o teorema, ainda falta mostrar que $\phi(E)$ é mensurável. Para tanto, afirmamos que

$$\phi(E) = \mathbb{N}^{\mathbb{N}} \setminus \left(\bigcup_{(w_1, \dots, w_k) \in \mathcal{E}} \{w_1\} \times \{w_k\} \times \mathbb{N} \times \dots \right), \quad (2.117)$$

onde

$$\mathcal{E} := \{(w_1, \dots, w_k) \in \bigcup_{n \geq 1} \mathbb{N}^n : A_{w_1, \dots, w_k} = \emptyset\}.$$

A igualdade acima será mostrada no que segue.

Dado $w \in \phi(E)$ existe $x \in E$ tal que $\phi(x) = w$. Como $x \in A_{w_1, \dots, w_n}$ para todo $n \geq 1$, esses conjuntos não são vazios. Logo w não pertence à união em (2.117), mostrando o lado (\subseteq) da igualdade. Finalmente, suponha que $w = (w_1, w_2, \dots)$ é tal que para todo $k \geq 1$, $A_{w_1, \dots, w_k} \neq \emptyset$. Tomamos portanto para todo $k \geq 1$ um ponto $x_k \in A_{w_1, \dots, w_k}$.

Afirmamos que

$$\text{para todo } n, (x_k)_{k \geq n} \text{ é Cauchy em } (A_{w_1, \dots, w_n}, d_{w_1, \dots, w_n}), \quad (2.118)$$

o que segue logo do fato que por $k \geq n+1$, $x_k \in A_{w_1, \dots, w_k}$ cujo d_{w_1, \dots, w_n} -diâmetro é menor que $1/k$.

Consideramos x^n o limite de $(x_k)_{k \geq n}$ em $(A_{w_1, \dots, w_n}, d_{w_1, \dots, w_n})$. É fácil de mostrar que $x^n = x^0 := x$ (o limite da sequência em (E, d)) para todo valor de n . É suficiente ver que $d(x^n, x_k) \leq d_{w_1, \dots, w_n}(x^n, x_k)$, para todo $k \geq n$, o que implica que x^n é o limite em (E, d) .

Como consequência podemos concluir que $x \in A_{w_1, \dots, w_n}$ para todo n e então que $\phi(x) = w$, o que conclui a prova do teorema. \square

Tópico: Cadeias de Markov

Um exemplo de como usar núcleos de transição é a construção de Cadeias de Markov. Esse tipo de processo é bastante útil em diversas aplicações, desde a biologia até a computação.

Considere um espaço mensurável canônico fixo (E, \mathcal{A}) e seja K um núcleo de E nele mesmo. Seria bastante intuitivo agora iterar K (já que ele está no mesmo espaço) e obter uma medida em $\Omega = E^{\mathbb{N}}$ com a σ -álgebra canônica.

Para começar esse procedimento, seja μ_0 uma medida inicial em (E, \mathcal{A}) . Podemos então definir $\mu_1 = \mu_0 \star K$ o que é o primeiro passo da nossa construção, porém observe que não podemos escrever “ $\mu_2 = \mu_1 \star K$ ”, pois $\mu_1 \star K$ é uma medida em $(E^2, \mathcal{A}^{\otimes 2})$. Vamos com calma então.

Observe que

$$\mu_1(A_0 \times A_1) = \int_{A_0} \int_{A_1} K(x_0, dx_1) \mu_0(dx_0), \quad (2.119)$$

ou em outras palavras o valor de x_0 determina a distribuição de x_1 . Gostaríamos agora que x_1 determinasse a distribuição de x_2 via K , como por exemplo assim

$$\mu_2(A_0 \times A_1 \times A_2) = \int_{A_0} \int_{A_1} \int_{A_2} K(x_1, dx_2) K(x_0, dx_1) \mu_0(dx_0). \quad (2.120)$$

Mas essa notação fica bastante carregada à medida que iteramos.

Para tornar essa notação mais simples, definimos a projeção $\phi_n : E^n \rightarrow E$ por $\phi_n(x_0, \dots, x_{n-1}) = x_{n-1}$. Também precisamos de $K_n : E^n \times \mathcal{A} \rightarrow [0, 1]$ dado por

$$K_n(\vec{x}, A) = K(\phi_n(\vec{x}), A) \quad (= K(x_{n-1}, A)). \quad (2.121)$$

O fato de K_n ser um núcleo de transição segue imediatamente dessa propriedade para K .

Note que, nessa notação, estamos dizendo que para irmos de E^n para E^{n+1} iremos olhar apenas para a última coordenada, na qual aplicaremos o núcleo K . Isso é o ponto mais importante que caracteriza uma Cadeia de Markov: a distribuição do estado futuro da cadeia depende apenas do estado atual e não do passado. Em alguns contextos essa propriedade é chamada de ausência de memória.

Podemos finalmente definir

$$\mu_{n+1} = \mu_n \star K_n, \text{ para todo } n \geq 1. \quad (2.122)$$

Mas resta a questão sobre a existência de uma μ^∞ que será respondida com ajuda do próximo resultado.

Lema 2.10.10. *As probabilidades μ_n definidas em (2.122) são compatíveis, mais precisamente $\mu_{n+1}(A \times E) = \mu_n(A)$ para todo $A \in \mathcal{A}^{\otimes n}$.*

Demonstração. Basta observar que

$$\mu_{n+1}(A \times E) = \mu_n \star K(A \times E) = \int_A \underbrace{K_n(\vec{x}, E)}_1 \mu_n(d\vec{x}) = \mu_n(A), \quad (2.123)$$

provando o lema. \square

Logo, o Teorema da Extensão de Kolmogorov (lembre que (E, \mathcal{A}) foi suposto canônico) nos fornece uma única P em (Ω, \mathcal{F}) tal que

$$P_{(X_0, \dots, X_n)} = \mu_n, \text{ para todo } n \geq 0. \quad (2.124)$$

Lembramos que X_i denotam as projeções canônicas em $\Omega = \prod_{i=1}^{\infty} E$.

Chamamos o processo X_1, X_2, \dots sob a lei P da Cadeia de Markov com distribuição inicial μ_0 e núcleo de transição K .

Exemplo 2.10.4. Suponha que E seja enumerável. Nesse caso recordamos do Exemplo 2.9.2 que o núcleo pode ser representado por uma matriz $(p(x, y))_{x, y \in E}$ que nos retorna a probabilidade de saltar de x a y . Além disso, a distribuição inicial μ_0 é determinada por $P(\{x\}) = p_0(x)$, para alguma sequência $(p_0(x))_{x \in E}$.

Exercício 2.10.5. Mostre que no exemplo acima temos

$$P(X_0 = x_0, \dots, X_n = x_n) = p_0(x_0)p(x_0, x_1) \dots p(x_{n-1}, x_n). \quad (2.125)$$

Exercício 2.10.6. Defina $K : \mathbb{R}^2 \times \mathcal{B}(\mathbb{R}^2) \rightarrow [0, 1]$ dada por

$$K(x, A) = U_{S^1}(A - x). \quad (2.126)$$

Nesse contexto,

- a) mostre que K é um núcleo de transição e,
- b) considerando a cadeia com distribuição inicial $\mu_0 = \delta_0$ em \mathbb{R}^2 e núcleo K , mostre que X_2 tem distribuição absolutamente contínua com respeito a Lebesgue e calcule sua densidade.

Exercício 2.10.7. Mostre que para qualquer núcleo de transição K entre E e E , existe um núcleo de transição \bar{K} entre E e $\Omega = E^{\mathbb{N}}$, tal que para toda medida inicial μ_0 , temos que $\mu_0 \star K$ é a distribuição de uma Cadeia de Markov começando de μ_0 e com transição dada por K . Esse núcleo é útil se quisermos mudar a distribuição inicial μ_0 e uma notação bastante comum para esse núcleo é $P_x(\cdot) = \bar{K}(x, \cdot)$.

Vamos terminar essa seção dando uma interpretação bastante interessante para os núcleos de transição em analogia à álgebra linear. Fixe um núcleo de transição K entre E e E , uma medida inicial μ e uma função limitada $f : E \rightarrow \mathbb{R}$. Relembre a notação em (2.105) e defina $Kf : E \rightarrow \mathbb{R}$ dada por

$$Kf(x) := \int f(y)K(x, dy), \quad (2.127)$$

que é obviamente limitada e já vimos ser mensurável no Teorema de Fubini.

Então temos dois operadores definidos para núcleos, a multiplicação à esquerda por uma medida em E (μK que também é uma medida em E) e a multiplicação à direita por uma função limitada e mensurável (Kf que também é uma função limitada e mensurável). Podemos pensar em f como um vetor coluna e μ como um vetor linha, nesse caso K faria o papel de uma matriz. Essa analogia é real se E for um espaço enumerável.

Exercício 2.10.8. *No contexto de cadeias de Markov,*

- a) *mostre a relação de associatividade $\mu(Kf) = (\mu K)f$,*
- b) *defina para todo n o núcleo $K^{(n)}$ iterado (de E em E), de forma que $\mu K^{(n)}f$ ainda seja associativa.*
- c) *Mostre que a medida $\mu K^{(n)}$ é a distribuição de X_n se começamos de μ ,*
- d) *que a função $K^{(n)}f(\cdot)$ é o valor esperado de f no tempo n se começamos no zero do ponto \cdot e finalmente que*
- e) *o número real $\mu K^{(n)}$ é a esperança de f no tempo n se começamos de μ .*

Vamos agora dar um exemplo simples de Cadeia de Markov que poderemos analisar em detalhes.

Seja $E = \mathbb{Z}$ e considere $K : \mathbb{Z} \times \mathcal{P}(\mathbb{Z}) \rightarrow [0, 1]$ dado por

$$K(x, \cdot) = \frac{\delta_{x-1} + \delta_{x+1}}{2}, \quad (2.128)$$

que obviamente define um núcleo pois toda função em \mathbb{Z} é mensurável na σ -álgebra das partes.

Podemos portanto construir P em $\mathbb{Z}^{\mathbb{N}}$ que nos fornece a lei de uma Cadeia de Markov em \mathbb{Z} com distribuição inicial δ_0 e núcleo de transição K . Chamamos esse processo de passeio aleatório simples simétrico.

Poderíamos estar interessados em várias perguntas sobre esse processo, como por exemplo quão longe esperamos que o passeio aleatório pode ir depois de um determinado tempo? Para responder essa e várias outras questões, iremos mostrar outra construção do passeio simples simétrico através de uma soma de variáveis aleatórias.

Introduzimos um espaço de probabilidade \tilde{P} , variáveis Y_1, Y_2, \dots i.i.d. com distribuição $(\delta_{-1} + \delta_1)/2$ e definimos $S_0 = 0$ e $S_n = Y_1 + \dots + Y_n$.

Lema 2.10.11. *A distribuição da sequência infinita (X_0, X_1, \dots) sob a lei P do passeio aleatório simples e simétrico é igual à distribuição de (S_0, S_1, \dots) sob \tilde{P} .*

Demonstração. Observamos primeiramente que basta mostrar a igualdade de distribuições para cilindros do tipo $\{x_1\} \times \dots \times \{x_n\} \times \mathbb{Z}^{\mathbb{N}}$, pois tais eventos

compõem um π -sistema que gera a σ -álgebra produto em $\mathbb{Z}^{\mathbb{N}}$. Calculamos portanto

$$P[X_1 = x_1, \dots, X_n = x_n]$$

pela definição de Cadeia de Markov (via extensão de Kolmogorov),

$$\begin{aligned} &= \mu_n[X_1 = x_1, \dots, X_n = x_n] \\ &= \mu_{n-1} \star K_n[X_1 = x_1, \dots, X_n = x_n] \end{aligned}$$

por Fubini para núcleos (Teorema 2.9.5),

$$\begin{aligned} &= \mu_{n-1}[X_1 = x_1, \dots, X_{n-1} = x_{n-1}] K_n((x_1, \dots, x_{n-1}), \{x_n\}) \\ &= \mu_{n-1}[X_1 = x_1, \dots, X_{n-1} = x_{n-1}] K(x_{n-1}, \{x_n\}) \\ &= \frac{1}{2} \mu_{n-1}[X_1 = x_1, \dots, X_{n-1} = x_{n-1}] \mathbb{1}_{\{|x_{n-1} - x_n| = 1\}} \\ &= \dots = 2^{-n} \prod_{i=1}^n \mathbb{1}_{\{|x_{i-1} - x_i| = 1\}}. \end{aligned}$$

Faremos agora esse cálculo para a distribuição de S_i 's:

$$\begin{aligned} &\tilde{P}[S_1 = x_1, \dots, S_n = x_n] \\ &= \mu_n[Y_1 = x_1 - x_0, Y_2 = x_2 - x_1, \dots, Y_n = x_n - x_{n-1}] \\ &= \prod_{i=1}^n \tilde{P}[Y_i = x_i - x_{i-1}] = 2^{-n} \prod_{i=1}^n \mathbb{1}_{\{|x_{i-1} - x_i| = 1\}}. \end{aligned}$$

Isso mostra o enunciado do lemma. □

Podemos agora por exemplo estimar

$$P[|X_n| \geq \varepsilon n] = \tilde{P}[|S_n| \geq \varepsilon n] \leq 2 \exp\{-\psi_{(\delta_{-1} + \delta_1)/2}(\varepsilon)n\}, \quad (2.129)$$

que responde nossa pergunta sobre a probabilidade de um passeio aleatório se distanciar muito da origem.

Tópico: Urna de Pólya

Um excelente exemplo de como Cadeias de Markov podem gerar interessantes modelos de situações reais são as chamadas Urnas de Pólya. Esse processo modela sistemas de física, biologia, computação e economia que apresentam o que chamamos de reforço.

Tome por exemplo duas empresas que competem pelo mercado de aviões. Inicialmente, não temos nenhuma razão para escolher uma em detrimento da outra, portanto compramos nosso primeiro avião de cada empresa com probabilidade meio. Porém, depois que já compramos diversos aviões de uma determinada empresa, ela já recebeu bastante dinheiro que pode ser reinvestido para gerar melhor tecnologia e aumentar as chances que ela seja escolhida novamente no futuro. Isso é o que chamamos de reforço.

Vamos agora apresentar rigorosamente um modelo para situações desse tipo. O nosso modelo começa com uma urna contendo duas bolas, uma vermelha e uma azul. No cada passo do processo, escolheremos uma bola da urna ao acaso, olharemos sua cor e retornaremos essa bola para dentro urna junto com mais uma bola da mesma cor. Isso pode ser formalizado à seguir.

Vamos construir uma medida em $\{0, 1\}^{\mathbb{N}}$, dotado da σ -álgebra produto. Fixada uma sequência finita w_1, \dots, w_n em $\{0, 1\}$, definimos

$$N_x(w_1, \dots, w_n) = \#\{j \in \{1, \dots, n\} : w_j = x\} + 1, \quad (2.130)$$

que nada mais é que o número de bolas do tipo x que se encontram na urna no tempo n . Quando tivermos uma sequência infinita de w_i 's, escreveremos N_x^n para denotar $N_x(w_1, \dots, w_n)$.

Para cada $n \geq 1$, definimos $K_n : \{0, 1\}^n \rightarrow \mathcal{P}(\{0, 1\})$ por

$$K_n(w_1, \dots, w_n) = \text{Ber}\left(\frac{N_1}{n}\right). \quad (2.131)$$

Ou seja, dadas cores w_1, \dots, w_n , escolheremos uma bola de cor 1 proporcionalmente ao número N_1 de bolas de cor 1 que já foram sorteadas.

Exercício 2.10.9. *Mostre que todos K_n acima definem núcleos de transição. Além disso a seguinte sequência de medidas é compatível no sentido de Kolmogorov:*

- $P_1 = \text{Ber}(1/2)$,
- $P_2 = P_1 \star K_1$,
- $P_3 = P_2 \star K_2, \dots$

Conclua que existe a medida P em $\{0, 1\}^{\mathbb{N}}$ que define o modelo de Pólya.

Podemos agora fazer perguntas como por exemplo: será que escolheremos bolas de ambas as cores para sempre, ou a partir de um certo momento escolheremos bolas de apenas uma cor com certa probabilidade. Mais precisamente, qual é a probabilidade de $[X_i = 1, \text{ infinitas vezes}]$?

Para responder perguntas desse tipo, iremos mostrar algo muito curioso, que pode ser entendido como uma outra maneira de representar o modelo descrito acima. Mas antes, vamos colecionar alguns fatos sobre o modelo da Urna de Pólya.

Primeiramente vamos olhar para os seguintes eventos. Fixamos $n \geq 1$ e uma sequência $w_1, \dots, w_n \in \{0, 1\}$ e seja A o evento $\{w_1\} \times \dots \times \{w_n\} \times \{0, 1\} \times \dots$. Note que os eventos desse tipo (junto com o evento \emptyset) formam um π -sistema que gera a σ -álgebra canônica de $\{0, 1\}^{\mathbb{N}}$, portanto essa coleção é bastante completa para identificar a distribuição da Urna de Pólya.

Podemos calcular a probabilidade do evento A acima

$$\begin{aligned} P(A) &= \frac{N_{w_1}^1}{2} \frac{N_{w_1}^2}{3} \dots \frac{N_{w_n}^n}{n+1} = \frac{1}{(n+1)!} \prod_{i=1}^n N_{w_i}^i \\ &= \frac{N_1^n! (n - N_1^n)!}{(n+1)!} = \frac{1}{(n+1)} \binom{n}{N_1^n}^{-1}. \end{aligned} \quad (2.132)$$

O que é muito interessante sobre a equação acima é que ela nos remete a problemas combinatórios ao notarmos o fator binomial acima.

Vamos portanto construir um processo completamente diferente que apresenta as mesmas probabilidades que o anterior. Seja \mathcal{S}_N o conjunto de todas as permutações σ de $\{1, \dots, N\}$. É fácil ver que

$$\frac{1}{(n+1)} \binom{n}{j}^{-1} = U_{\mathcal{S}_{n+1}} \left[\sigma(n+1) = j+1, \sigma(i) \leq j \text{ se e só se } i \leq j \right].$$

Um método muito interessante de se produzir uma permutação uniforme é dado pelos seguintes exercícios.

Exercício 2.10.10. *Seja $n \geq 1$ um inteiro, P uma probabilidade em (E, \mathcal{A}) , σ uma permutação fixa em \mathcal{S}_n . Então*

$$(X_1, \dots, X_n) \sim_d (X_{\sigma(1)}, \dots, X_{\sigma(n)}), \quad (2.133)$$

onde X_i como sempre representam as coordenadas canônicas em $(E^n, \mathcal{A}^{\otimes n}, P^{\otimes n})$.

Ou em outras palavras, aplicar uma permutação fixa a uma sequência i.i.d. não altera sua distribuição. Sequências de elementos aleatórios (não necessariamente i.i.d. 's) que satisfazem (2.133) são ditas intercambiáveis.

Um outro exercício interessante nesse tópico é o seguinte

Exercício 2.10.11. *Seja $n \geq 1$ e $F : [0, 1]^n \rightarrow \mathcal{S}_n$ dada por*

$$F(x_1, \dots, x_n) = \begin{cases} (1, 2, \dots, n), & \text{se existe } i \neq j \text{ com } x_i = x_j, \\ \text{o único } \sigma \text{ tal que } x_{\sigma(1)} < \dots < x_{\sigma(n)}, & \text{caso contrário.} \end{cases}$$

Mostre que $F_(U_{[0,1]}^{\otimes n}) = U_{\mathcal{S}_n}$.*

Ou seja, ordenar uma sequência de uniformes independentes nos fornece uma permutação uniforme. Como prometido, isso nos dá uma maneira de construir uma permutação uniforme de $\{1, \dots, n\}$ à partir de uma sequência i.i.d. (que é algo que já estamos começando a entender melhor).

Podemos agora escrever nossa probabilidade de observar uma sequência no modelo da Urna de Pólya em termos de uma sequência i.i.d. de variáveis aleatórias.

$$\begin{aligned} \frac{1}{(n+1)} \binom{n}{N_1^n}^{-1} &= F_* U_{[0,1]}^{\otimes n+1} \left[\sigma(n+1) = N_1^n + 1, \sigma(i) \leq N_1^n \text{ se e só se } i \leq N_1^n \right] \\ &= U_{[0,1]}^{\otimes n+1} \left[X_i < X_{n+1}, \text{ para } i \leq N_1^n \text{ e } X_i > X_{n+1}, \text{ para } i \geq N_1^n + 1 \right]. \end{aligned}$$

Agora estamos prontos para provar o resultado principal que nos ajudará a calcular probabilidades no modelo da Urna de Pólya.

Dado $u \in [0, 1]$, seja $P_u = \text{Ber}(u)^{\otimes \mathbb{N}}$, ou seja a probabilidade que nos dá uma sequência infinita de moedas independentes com probabilidade u de sucesso. Definimos agora $\bar{K} : [0, 1] \times (\mathcal{P}(\{0, 1\})^{\otimes \mathbb{N}}) \rightarrow [0, 1]$ dada por

$$\bar{K}(u, A) = P_u(A). \quad (2.134)$$

Lema 2.10.12. *A função \bar{K} definida acima é um núcleo entre $[0, 1]$ e $\{0, 1\}^{\mathbb{N}}$.*

Demonstração. Usando a Proposição 2.9.2, basta ver que

$$\begin{aligned} &\text{para todo } k \geq 1 \text{ e } w_1, \dots, w_k \in \{0, 1\}, \text{ temos que} \\ P_u(X_1 = w_1, \dots, X_k = w_k) &\text{ é uma função mensurável de } u \in [0, 1]. \end{aligned} \quad (2.135)$$

Mas é fácil ver que

$$P_u(X_1 = w_1, \dots, X_k = w_k) = u^{N_1(w_1, \dots, w_k)} (1 - u)^{N_0(w_1, \dots, w_k)}, \quad (2.136)$$

que obviamente é mensurável, provando assim o lema. \square

O resultado muito curioso a qual nos referimos é o seguinte.

Lema 2.10.13. *A lei P definida no Exercício 2.10.9 é igual a $U_{[0,1]} \bar{K}$.*

Em outras palavras, digamos que realizamos os seguintes experimentos. Primeiramente João realiza o processo da Urna de Pólya e anota a sequência das cores obtidas. Depois Maria sorteia uma variável aleatória X de distribuição uniforme em $[0, 1]$ e depois joga infinitas vezes uma moeda com probabilidade X de obter vermelho e $(1 - X)$ de obter azul, anotando também quais cores foram obtidas. Finalmente, não seríamos capazes de distinguir essas duas sequências (mesmo que pudéssemos repetir várias vezes esse experimento) pois elas tem a mesma distribuição em $\{0, 1\}^{\mathbb{N}}$.

Demonstração. Já sabemos que basta mostrar a igualdade para eventos do tipo $A = \{w_1\} \times \cdots \times \{w_n\} \times \{0, 1\}^{\mathbb{N}}$. Sabemos pelo Teorema de Fubini para Núcleos que

$$U_{[0,1]} \bar{K}(A) = \int_0^1 K(u, A) \, du \stackrel{(2.136)}{=} \int_0^1 u^{N_1(w_1, \dots, w_k)} (1-u)^{N_0(w_1, \dots, w_k)} \, du. \quad (2.137)$$

Por outro lado, sabemos (usando simetria entre 0 e 1) que

$$P(A) = U_{[0,1]}^{\otimes n+1} \left[X_i < X_{n+1}, \text{ para } i \leq N_0^n \text{ e } X_i > X_0, \text{ para } i \geq N_0^n + 1 \right] \quad (2.138)$$

Se definirmos $\tilde{K} : [0, 1] \times \mathcal{B}([0, 1]^n)$, dado por $\tilde{K}(u, B) = U_{[0,1]}^{\otimes n}$, sabemos que isso define um núcleo pelo Exercício 2.9.7. Mais ainda, esse mesmo exercício nos diz que $U_{[0,1]} \star \tilde{K} = U_{[0,1]}^{\otimes}$, de forma que

$$\begin{aligned} P(A) &= U_{[0,1]} \star \tilde{K} \left[X_i < X_0, \text{ para } i \leq N_0^n \text{ e } X_i > X_0, \text{ para } i \geq N_0^n + 1 \right] \\ &= \int_0^1 U_{[0,1]}^{\otimes n} \left[X_i < u, \text{ para } i \leq N_0^n \text{ e } X_i > u, \text{ para } i \geq N_0^n + 1 \right] \, du \\ &= \int_0^1 u^{N_0^n} (1-u)^{n-N_0^n} \, du, \end{aligned}$$

que coincide com $U_{[0,1]} \bar{K}(A)$, provando o lema. \square

Exercício 2.10.12. *Mostre que a probabilidade, segundo o modelo da Urna de Pólya, de que observemos infinitas bolas de ambas as cores é um.*

Capítulo 3

Somas de variáveis independentes

Nesse capítulo introduziremos várias técnicas e resultados que serão úteis em geral, mas que aparecem naturalmente no estudo de somas de variáveis aleatórias independentes, que por sua vez é um assunto de extrema importância em teoria e aplicações de probabilidade.

3.1 Esperança

Definição 3.1.1. Se X é uma variável aleatória com $\int_{\Omega} |X| d\omega < \infty$, dizemos que X é integrável e definimos

$$E(X) = \int_{\Omega} X(\omega) P(d\omega), \quad (3.1)$$

a chamada esperança de X . Nesse caso também dizemos que $X \in \mathcal{L}^1$.

Quando $X \geq 0$, também podemos supor que $E(X)$ está bem definida, mesmo que possivelmente tomando valor infinito.

Não demonstraremos algumas propriedades conhecidas da integração de Lebesgue, tais como

- a) $E(X + \alpha Y) = E(X) + \alpha E(Y)$ (se estiverem bem definidas),
- b) Valem os Teoremas de Convergência (Monótona e Limitada).

Exercício 3.1.1. Mostre que se $X \in \mathcal{L}^1$ e $P[X > x] = 0$, então $E(X) \leq x$.

Lema 3.1.2. A esperança de uma variável aleatória $X \in \mathcal{L}^1$ depende somente de sua distribuição. Mais precisamente

$$E(X) = \int x P_X(dx). \quad (3.2)$$

Demonstração. Vamos mostrar que

$$E(f(X)) = \int f(x)(X \circ P)(dx), \quad (3.3)$$

para toda $f : \mathbb{R} \rightarrow \mathbb{R}$ mensurável tal que $f(X) \in \mathcal{L}^1$.

Para $f = \mathbb{1}_A$, temos

$$E(f(X)) = P[X \in A] = (X \circ P)(A), \quad (3.4)$$

por definição de $X \circ P$.

Agora podemos estender o teorema para funções f simples por linearidade, depois para funções positivas usando o Teorema da Convergência Monótona e finalmente escrevemos $x = x\mathbb{1}_{[0,\infty)} - (-x)\mathbb{1}_{(-\infty,0)}$. \square

Vamos mostrar uma fórmula bastante simples de integração de variáveis tomando valores em um conjunto enumerável. Se $X \in \{x_1, x_2, \dots\}$ P -quase certamente, então

$$\begin{aligned} E(X) &= \int_{\Omega} XP(d\omega) = \int \sum_i \mathbb{1}_{[X=x_i]} XP(d\omega) + \int_{\{x_1, x_2, \dots\}^c} XP(d\omega) \\ &= \sum_i \int_{[X=x_i]} x_i P(d\omega) + 0 = \sum_i x_i P[X = x_i]. \end{aligned} \quad (3.5)$$

Para nos acostumar à notação de probabilidade, vamos agora mostrar o mesmo resultado da seguinte forma

$$\begin{aligned} E(X) &= E\left(\sum_i X\mathbb{1}_{[X=x_i]}\right) + E(X\mathbb{1}_{\{x_1, x_2, \dots\}^c}) \\ &= \sum_i E[X; X = x_i] + 0 = \sum_i x_i P[X = x_i]. \end{aligned} \quad (3.6)$$

Que é certamente muito útil quando nos habituamos a ela.

Observe que acima usamos a notação $E[X; \mathcal{Q}] = E(X\mathbb{1}_{[\mathcal{Q}]})$. Também utilizaremos $E[X; \mathcal{Q}_1, \mathcal{Q}_2, \dots] = E(X\mathbb{1}_{[\mathcal{Q}_1, \mathcal{Q}_2, \dots]})$

Exemplo 3.1.2. Se $X \stackrel{d}{\sim} \text{Ber}(p)$, então $E(X) = 0 \cdot P[X = 0] + 1P[X = 1] = 0 + p = p$.

Exemplo 3.1.3. Seja $X \stackrel{d}{\sim} \text{Bin}(n, p)$, então, para calcular $E(X)$, basta calcular $E(Y)$ onde $X \stackrel{d}{\sim} Y$. Como vimos anteriormente, se Z_1, Z_2, \dots, Z_n são variáveis i.i.d. (relembrando: independentes e identicamente distribuídos) com $Z_1 \stackrel{d}{\sim} \text{Ber}(p)$, então $Y = \sum_i Z_i \stackrel{d}{\sim} \text{Bin}(n, p)$. Logo

$$E(X) = E(Y) = \sum_i E(Z_i) = np. \quad (3.7)$$

Se $d(X \circ P) = \rho(x) dx$ (com $\rho \geq 0$ e $\int \rho(x) dx = 1$), então

$$E(X) = \int x(X \circ P)(dx) = \int x\rho(x) dx. \quad (3.8)$$

Exemplo 3.1.4. Se $X \stackrel{d}{\sim} U_{[0,1]}$, então sua densidade com respeito a Lebesgue é dada por $d(X \circ P) = \mathbb{1}_{[0,1]} dx$, donde $E(X) = \int_0^1 x dx = 1/2$.

Proposição 3.1.3. Se $X \geq 0$ P-q.c., então

$$E(X) = \int_0^\infty P[X > x] dx = \int_0^\infty 1 - F(x) dx. \quad (3.9)$$

Demonstração.

$$\begin{aligned} E(X) &= E\left(\int_0^X 1 dx\right) = E\left(\int_0^\infty \mathbb{1}_{[x < X]} dx\right) \\ &\stackrel{\text{Fubini}}{=} \int_0^\infty E(\mathbb{1}_{[x < X]}) dx = \int_0^\infty P[x < X] dx. \end{aligned} \quad (3.10)$$

□

Exemplo 3.1.5. Se $X \stackrel{d}{\sim} \text{Exp}(\lambda)$, então

$$P[X \geq x] = \int_x^\infty \lambda e^{-\lambda t} dt = e^{-\lambda x}, \quad (3.11)$$

donde

$$E(X) = \int_0^\infty e^{-\lambda x} dx = \frac{1}{\lambda}. \quad (3.12)$$

Exercício 3.1.6. Se $X \in \mathcal{L}^1$ e $P[X \geq x] = P[X \leq -x]$ para todo $x \geq 0$, então $E(X) = 0$.

Exercício 3.1.7. Marcelo coleciona figurinhas de futebol. O álbum completo conterá N figurinhas. No i -ésimo dia, ele compra uma nova carta $X_i \in \{1, \dots, N\}$. A coleção $(X_i)_{i \geq 0}$ é distribuída de maneira i.i.d. e uniforme nas figurinhas.

a) Para $j = 1, \dots, N$, seja T_j o tempo passado até a aquisição da j -ésima nova figurinha, i.e.

$$T_1 = 1 \quad e \quad T_j = \inf\{i, X_i \notin \{X_{T_{j'}}; j' < j\}\}. \quad (3.13)$$

Mostre que T_j é finito quase certamente, para todo $j \leq N$.

b) Calcule a distribuição conjunta de $(T_1, T_2 - T_1, \dots, T_N - T_{N-1})$.

c) Calcule a esperança de T_N (o dia em que Marcelo completa seu álbum).

Exercício 3.1.8. Sejam X_1, X_2, \dots variáveis aleatórias i.i.d. e defina o primeiro tempo de récorde como

$$R = \inf\{i \geq 2; X_i \geq X_1\}. \quad (3.14)$$

Supondo que X_1 é absolutamente contínua com respeito à medida de Lebesgue, encontre $E(R)$.

3.1.1 Desigualdade de Markov

Teorema 3.1.4. Se $X \geq 0$ P-q.c., então para todo $x > 0$,

$$P[X \geq x] \leq \frac{E(X)}{x}. \quad (3.15)$$

Demonstração. Sabemos que $X \geq x \mathbb{1}_{[X \geq x]}$, logo

$$E(X) \geq xE(\mathbb{1}_{[X \geq x]}) = xP[X \geq x], \quad (3.16)$$

que termina a prova. \square

O próximo exemplo serve muito bem para mostrar porque estamos interessados em desigualdades como a do Teorema 3.1.4 acima.

Em vários exemplos importantes, podemos ter dificuldade de calcular probabilidades explicitamente. Nesses casos, poderíamos gastar nossas energias tentando calculá-las a qualquer custo, ou podemos nos contentar em obter cotas superiores e inferiores para as probabilidades nas quais estamos interessados.

Em vários casos, a segunda estratégia tem uma grande vantagem sobre a primeira, por possibilitar que estudemos problemas mais complexos (e consequentemente mais importantes/interessantes) e muitas vezes sem nos afastarmos da realidade (em vários exemplos as cotas superiores e inferiores são próximas o suficiente para que não nos preocupemos).

Exemplo 3.1.9. Sejam n patos e m caçadores. Cada caçador escolhe um pato aleatoriamente e uniformemente e atira (abatendo-o com probabilidade p). Seja $X = \#\{\text{patos vivos}\}$, que pode ter uma distribuição complicada de calcular, mas

$$\begin{aligned} E(X) &= E\left(\sum_{i=1}^n \mathbb{1}_{[\text{pato } i \text{ vive}]}\right) = \sum_{i=1}^n P[\text{pato } i \text{ vive}] \\ &= nP[\text{pato } 1 \text{ vive}] = P\left(\bigcap_{j=1}^m [\text{caçador } j \text{ não mata pato } 1]\right) \\ &= nP[\text{caçador } j \text{ não mata pato } 1]^m = n\left(1 - \frac{p}{n}\right). \end{aligned} \quad (3.17)$$

Observe que

- a) acima obtivemos uma igualdade e
- b) $[\text{pato } i \text{ vive}]$, $i = 1, \dots, n$ não são independentes.

Finalmente estimamos (digamos para n par)

$$\begin{aligned} P[\text{patos para o jantar} \leq n/2] &= P[X \geq n/2] \leq \frac{E(X)}{n/2} \\ &= 2\frac{n}{n}\left(1 - \frac{p}{n}\right)^m \leq 2\exp\left\{-\frac{pm}{n}\right\}. \end{aligned} \quad (3.18)$$

3.1.2 Esperança e independência

Proposição 3.1.5. *Sejam X e Y variáveis aleatórias independentes e em \mathcal{L}^2 , então*

$$E(XY) = E(X)E(Y). \quad (3.19)$$

Demonstração. Obviamente o resultado acima é válido para funções indicadoras, pois $\mathbb{1}_A \mathbb{1}_B = \mathbb{1}_{A \cap B}$. Por linearidade, o resultado também vale para funções simples e usando o Teorema da Convergência Monótona podemos estendê-lo para funções positivas. Finalmente, decompomos $X = X_+ - X_-$ e $Y = Y_+ - Y_-$ e lembramos que ambas estão em \mathcal{L}^2 para concluir a prova. \square

Exercício 3.1.10. *Mostre que $E(XY)$, $E(X/Y)$, $E(X + Y)$... dependem apenas da distribuição de $(X, Y) \in \mathbb{R}^2$.*

Exercício 3.1.11. *Mostre que se $X, Y \in \mathcal{L}^1$, então também vale $E(XY) = E(X)E(Y)$.*

3.2 Variância

Na seção anterior, limitamos $P[X > a]$ usando $E(X)$ (se $X \geq 0$). Esse método é chamado de *método do primeiro momento*, de acordo com a seguinte

Definição 3.2.1. *Dada uma variável aleatória X , definimos o seu k -ésimo momento como $E(X^k)$, para $k = 1, 2, \dots$*

Então, por exemplo, se $X \in \mathcal{L}^k$ e $X \geq 0$, podemos estimar

$$P[X \geq x] = P[X^k \geq x^k] \leq \frac{E(X^k)}{x^k}, \text{ para quaisquer } k \geq 1. \quad (3.20)$$

Observe que quando o k -ésimo momento de X é finito, a razão acima decai mais rápido quando x diverge.

Exercício 3.2.1. *Mostre uma fórmula análoga à da Proposição 3.1.3.*

Exercício 3.2.2. *Mostre que se a distribuição de X tem densidade ρ e $E(|f(X)|) < \infty$, então*

$$E(f(X)) = \int f(x)\rho(x) dx. \quad (3.21)$$

Um caso bastante importante ocorre quando $k = 2$, por várias razões que descreveremos abaixo.

Digamos que estamos interessados em aproximar uma variável aleatória por uma constante de forma a minimizar o erro da aproximação. Uma possível formulação desse problema é encontrar a de forma a minimizar

$$E((X - a)^2) = E(X^2) - 2aE(X) + a^2. \quad (3.22)$$

Essa equação obviamente possui um único mínimo em $a = E(X)$. Ao erro da aproximação acima damos o nome de *variância*

Definição 3.2.2. Dada uma variável aleatória $X \in \mathcal{L}^2$, definimos sua variância como

$$\text{Var}(X) = E\left((X - E(X))^2\right) = E(X^2) - E(X)^2. \quad (3.23)$$

Observe pelas definições alternativas dadas acima que

- a) $\text{Var}(X) \geq 0$ e
- b) $E(X^2) \geq E(X)^2$.

Exercício 3.2.3. Mostre que se $X \in \mathcal{L}^2$, então $\text{Var}(X) = 0$ se e somente se $X = a$ quase certamente.

Obviamente

$$\text{Var}(aX) = E(a^2 X^2) - E(aX)^2 = a^2 \text{Var}(X). \quad (3.24)$$

Podemos alternativamente entender a variância da seguinte maneira. Sejam X e Y variáveis aleatórias independentes em \mathcal{L}^2 de mesma distribuição. Então,

$$E((X - Y)^2) = E(X^2) - 2E(XY) + E(Y^2) = E(X^2) - E(X)^2 = \text{Var}(X). \quad (3.25)$$

Exercício 3.2.4. Mostre que se $X \in \mathcal{L}^2$, então $\text{Var}(X + b) = \text{Var}(X)$.

Exercício 3.2.5. Calcule $\text{Var}(X)$ quando X tem distribuições $\text{Ber}(p)$, $U[0, 1]$ ou $\text{Exp}(\lambda)$.

A seguinte proposição mostra que a variância é uma maneira de estimar o quanto uma variável aleatória se desvia de sua média.

Proposição 3.2.3. Se $X \in \mathcal{L}^2$ e $a > 0$, então

$$P[|X - E(X)| > a] \leq \frac{\text{Var}(X)}{a^2}. \quad (3.26)$$

Demonstração. A desigualdade segue trivialmente da cota de Markov, ao observarmos que

- a) $|X - E(X)| \geq 0$,
- b) $|X - E(X)| > a$ se e somente se $|X - E(X)|^2 > a^2$ e
- c) $E(|X - E(X)|^2) = E((X - E(X))^2) = \text{Var}(X)$,

mostrando a proposição. □

Para variáveis aleatórias de média zero, a variância nada mais é que $E(X^2)$, ou em outras palavras $\|X\|_2^2$, o quadrado de sua norma em \mathcal{L}^2 . Isso nos motiva a olhar mais de perto para o produto interno em \mathcal{L}^2 , que se traduz a $E(XY)$. Mas para não nos restringirmos a variáveis de média zero, introduzimos a seguinte

Definição 3.2.4. Se X, Y são variáveis em \mathcal{L}^2 , definimos

$$\text{Cov}(X, Y) = E\left((X - E(X))(Y - E(Y))\right) = E(XY) - E(X)E(Y). \quad (3.27)$$

Uma observação importante é que

$$\text{se } X \text{ e } Y \text{ em } \mathcal{L}^2 \text{ são independentes, então } \text{Cov}(X, Y) = 0. \quad (3.28)$$

Exercício 3.2.6. Sejam X_1 e X_2 as coordenadas canônicas em \mathbb{R}^2 . Já vimos que elas não são independentes sob a distribuição U_{S^1} . Mostre que mesmo assim temos $\text{Cov}(X_1, X_2) = 0$.

Uma outra propriedade bastante importante da variância é que ela se comporta bem com somas, no seguinte sentido

Proposição 3.2.5. Se X_1, \dots, X_n são variáveis em \mathcal{L}^2 , então

$$\text{Var}(X_1 + \dots + X_n) = \sum_{i=1}^n \text{Var}(X_i) + \sum_{i \neq j} \text{Cov}(X_i, X_j). \quad (3.29)$$

Em particular, se as variáveis X_i forem independentes duas a duas, então

$$\text{Var}(X_1 + \dots + X_n) = \sum_{i=1}^n \text{Var}(X_i). \quad (3.30)$$

Demonstração. Basta fazer o tedioso desenvolvimento

$$\begin{aligned} \text{Var}\left(\sum_i X_i\right) &= E\left(\left(\sum_i X_i - E\left(\sum_i X_i\right)\right)^2\right) \\ &= E\left(\left(\sum_i X_i - E(X_i)\right)^2\right) \\ &= \sum_{i,j=1}^n E(X_i - E(X_i))(X_j - E(X_j)), \end{aligned} \quad (3.31)$$

o que termina a prova ao separarmos $i = j$ de $i \neq j$. \square

Exercício 3.2.7. Calcule $\text{Var}(X)$ quando $X \stackrel{d}{\sim} \text{Bin}(n, p)$.

Exercício 3.2.8. Calcule $E(X)$ quando $X \stackrel{d}{\sim} \text{Geo}(p)$.

Um dito popular muito comum no Brasil é que não devemos deixar todos os “ovos no mesmo cesto”, o que nos remete à possibilidade de perdermos todos eles caso o cesto caia. Uma outra maneira de pensar nas vantagens de se dividir nossos riscos entre várias fontes independentes de incerteza, vem da equação (3.30), melhor explicada no exercício abaixo.

Exercício 3.2.9. Imagine que X_1, \dots, X_n são variáveis i.i.d., tomando valores em $[0, 1]$ e que temos um certo valor $s \in \mathbb{R}_+$ que temos que guardar em n caixas (dividindo como quisermos em s_1, \dots, s_n). Ao fim da semana, obteremos $S = \sum_i s_i X_i$.

Calcule $E(S)$ e $\text{Var}(S)$,

a) se $s_1 = s$ e $s_i = 0$ para todo $i \geq 2$ e

b) se $s_i = s/n$ para todo i .

Compare os resultados.

Exercício 3.2.10. Calcule $\lim_{p \rightarrow 0} F_p(x)$ onde F_p é a função de distribuição acumulada de pX_p com $X_p \stackrel{d}{\sim} \text{Geo}(p)$. Você reconhece esse limite?

3.3 Lei fraca dos grandes números

Nessa seção iremos mostrar um dos resultados mais importantes da Teoria da Probabilidade. O que nossa intuição tem a nos dizer sobre a probabilidade de obtermos um resultado em um dado é $1/6$? Uma possível explicação seria por simetria, mas e o que podemos dizer no caso de um dado viciado?

Se dizemos a alguém que a probabilidade de obter 6 em um certo dado é $1/10$, naturalmente a pessoa pode se perguntar como descobrimos isso. Um bom jeito de obter tal medida seria jogar o dado várias vezes independentemente e calcular em qual proporção dos ensaios ele retornou um seis.

O objetivo desta seção é confirmar a validade desse experimento de maneira quantitativa.

Teorema 3.3.1. Se X_1, X_2, \dots são i.i.d.s em \mathcal{L}^2 e definimos

$$S_n = \sum_{i=1}^n X_i, \quad (3.32)$$

então para todo $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} P\left[\left|\frac{S_n}{n} - E(X_1)\right| > \varepsilon\right] = 0. \quad (3.33)$$

Ou seja, $\frac{S_n}{n} \rightarrow E(X_1)$ em medida (que também chamamos de “em probabilidade”).

Demonstração. Sabemos que

$$P\left[\left|\frac{S_n}{n} - E(X_1)\right| > \varepsilon\right] \leq \frac{\text{Var}\left(\frac{S_n}{n}\right)}{\varepsilon^2}, \quad (3.34)$$

pois $E(S_n/n) = 1/n E(X_1 + \dots + X_n) = E(X_1)$.

Mas como $\text{Var}(S_n/n) = 1/n^2 \text{Var}(X_1 + \dots + X_n) = (n/n^2) \text{Var}(X_1)$, temos o resultado. \square

Observe que nós apenas utilizamos que as variáveis X_i eram independentes duas a duas.

Além disso, obtivemos o seguinte resultado quantitativo que vale mesmo para valores finitos de n :

Escólio 3.3.2. Se X_1, X_2, \dots são i.i.d.s em \mathcal{L}^2 e definimos $S_n = \sum_{i=1}^n X_i$ como acima, então, para todo $\varepsilon > 0$ e $n \geq 1$, temos

$$P\left[\left|\frac{S_n}{n} - E(X_1)\right| > \varepsilon\right] \leq \frac{\text{Var}(X_1)}{\varepsilon^2 n}. \quad (3.35)$$

Corolário 3.3.3. Se A_1, A_2, \dots são eventos independentes dois a dois com $P(A_i) = p \in [0, 1]$ para todo i , então

$$\lim_{n \rightarrow \infty} P\left[\left|\frac{\#\{i \leq n; \omega \in A_i\}}{n} - p\right| > \varepsilon\right] = 0, \quad (3.36)$$

ou em outras palavras a proporção de ensaios onde o evento A_i ocorre converge em probabilidade para p .

Demonstração. Basta tomar $X_i = \mathbb{1}_{A_i}$ no Teorema 3.3.1. \square

Exercício 3.3.1. Sejam $(X_i)_{i \geq 1}$ variáveis i.i.d. com distribuição $\text{Ber}(p)$, $p \in [0, 1]$. Mostre que

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N X_i X_{i+1} = p^2, \text{ em probabilidade.} \quad (3.37)$$

Exercício 3.3.2. Sejam X_1, \dots, X_n e Y_1, \dots, Y_n variáveis independentes com distribuição $\text{Ber}(p)$. Defina agora $Z_{i,j} = X_i Y_j$, para $i, j \in \{1, \dots, n\}$ e

- calcule a esperança de $S_n = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n Z_{i,j}$ e
- estime $P[|S_n - E(S_n)| > a]$ usando o método do segundo momento. Como esse resultado se compara com o caso em que os $Z_{i,j}$ são i.i.d.?

Exercício 3.3.3. Considere uma rua infinita com casas $i \in \mathbb{Z}$. Para todo $i \in \mathbb{Z}$, existia uma rua entre as casas i e $i + 1$, mas após uma grande tempestade essas ruas foram danificadas. Mais precisamente, para cada $i \in \mathbb{Z}$, temos variáveis aleatórias X_i que são i.i.d. com distribuição $\text{Ber}(p)$, onde $X_i = 1$ indica que o trecho da rua entre as casas i e $i + 1$ foi danificado e não pode ser utilizado. Defina, para $i \in \mathbb{Z}$, R_i como sendo o número de casas que continuaram acessíveis à casa i após a tempestade. Por exemplo, se $X_{-2} = 1$ e $X_0 = 1$ e $X_{-1} = 0$, temos que a casa 0 somente pode acessar a casa -1 , logo $R_0 = 1$. Nesse contexto,

- Calcule a distribuição e a esperança de R_0 ,
- Use o método do segundo momento para estimar a probabilidade

$$P\left[\left|\frac{1}{n} \sum_{i=1}^n R_i - E(R_0)\right| > a\right]. \quad (3.38)$$

Tópico: Contando triângulos

Vimos como a Lei Fraca dos Grandes Números seguiu de uma estimativa de segundo momento (mais precisamente usando a variância).

Nessa seção iremos mostrar como esse método é mais geral, se aplicando mesmo em situações onde as variáveis não são necessariamente independentes duas a duas.

Seja $V_n = \{1, \dots, n\}$ com $n \geq 3$ e $\mathcal{E}_n = \{\{x, y\} \subseteq V_n; x \neq y\}$. Chamamos o par (V_n, \mathcal{E}_n) de grafo completo em n vértices.

Definimos em um certo espaço de probabilidade P_n , as variáveis aleatórias $(X_e)_{e \in \mathcal{E}_n}$ de maneira i.i.d. com distribuição $\text{Ber}(p)$, onde $p \in [0, 1]$. Essas variáveis induzem um subgrafo aleatório (V_n, \mathcal{E}'_n) , onde

$$\mathcal{E}'_n = \{e \in \mathcal{E}_n; X_e = 1\}. \quad (3.39)$$

Dizemos que os elos e , tais que $X_e = 1$ são abertos.

Definimos nesse espaço a variável aleatória

$$T_n = \#\{\text{triângulos em } (V_n, \mathcal{E}'_n)\}. \quad (3.40)$$

Essa variável claramente pode ser escrita como

$$T_n = \sum_{x, y, z \in V_n \text{ distintos}} \mathbb{1}_{A_{\{x, y, z\}}}, \quad (3.41)$$

onde $A_{\{x, y, z\}} = [\{x, y, z\} \text{ formam um triângulo em } (V_n, \mathcal{E}'_n)]$.

Gostaríamos de entender algo sobre a distribuição de T_n e começamos calculando

$$\begin{aligned} E^n(T_n) &= \sum_{\{x, y, z\} \text{ distintos}} P^n(A_{\{x, y, z\}}) \\ &= \binom{n}{3} p^3 = \frac{n(n-1)(n-2)}{6} p^3. \end{aligned} \quad (3.42)$$

Logo, $P[T_n > a] \leq n(n-1)(n-2)p^3/6a$. Mais ainda,

$$\begin{aligned} E^n(T_n^2) &= \sum_{\{x, y, z\} \text{ distintos}} \sum_{\{x', y', z'\} \text{ distintos}} P^n(A_{\{x, y, z\}} \cap A_{\{x', y', z'\}}) \\ &= \underbrace{\binom{n}{6} \binom{6}{3} p^6}_{\text{todos distintos}} + \underbrace{\binom{n}{5} \binom{5}{3} \binom{3}{1} p^6}_{1\text{-comum}} + \underbrace{\binom{n}{4} \binom{3}{2} \binom{4}{3} p^5}_{2 \text{ em comum}} + \underbrace{\binom{n}{3} p^3}_{\text{iguais}} \end{aligned} \quad (3.43)$$

Donde

$$\text{Var}^n(T_n) = \frac{1}{36} n^6 p^6 - \frac{1}{36} n^6 p^6 + cn^5 p^5 + \dots \leq c(n^5 p^5 + n^3 p^3), \quad (3.44)$$

para todos $p \in [0, 1]$ e $n \geq 1$ se escolhemos bem a constante $c > 0$.

Isso nos permite por exemplo estimar o que acontece em alguns regimes, como por exemplo, se $p = 1/2$, então

$$E^n(T_n) = \frac{n(n-1)(n-2)}{48}, \quad (3.45)$$

que cresce como n^3 , e $\text{Var}^n(T_n) \leq cn^5$, logo

$$P^n \left[\left| T_n - E^n(T_n) \right| > \varepsilon n^3 \right] \leq \frac{\text{Var}^n(T_n)}{\varepsilon^2 n^6} \leq \frac{c}{\varepsilon^2 n}. \quad (3.46)$$

3.4 Lei forte dos grandes números

Teorema 3.4.1 (Lei Forte dos Grandes Números). *Sejam X_1, X_2, \dots i.i.d. em \mathcal{L}^1 , com $m = E(X_1)$. Então,*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n X_i = m, \text{ } P\text{-quase certamente.} \quad (3.47)$$

Antes de começar a prova, buscando inspiração no Teorema das Três Séries, mostraremos que basta considerar versões truncadas das variáveis X_i . Isso é feito no próximo

Lema 3.4.2. *Sejam $Y_i = X_i \mathbb{1}_{[|X_i| \leq i]}$. Então, para demonstrar o Teorema 3.4.1, basta provar que*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n Y_i = m, \text{ } P\text{-quase certamente.} \quad (3.48)$$

Prova do Lema 3.4.2. Consideramos os eventos $A_i = [X_i \neq Y_i]$. Obviamente,

$$\sum_i P(A_i) = \sum_i P[|X_i| \geq i] \leq \int_0^\infty P[|X_i| \geq t] dt = E(|X_i|) < \infty. \quad (3.49)$$

Logo, pelo Lema de Borel-Cantelli, temos que P -quase certamente A_i acontece apenas finitas vezes. Digamos que A_i não acontece para $i > N(\omega)$. Dessa forma, para qualquer $n \geq 1$,

$$\left| \frac{1}{n} \sum_{i=1}^n (X_i - Y_i) \right| \leq \frac{1}{n} \sum_{i=1}^n |X_i - Y_i| \leq \frac{1}{n} \sum_{i \leq N(\omega)} |X_i|, \quad (3.50)$$

que converge para zero P -quase certamente, mostrando o resultado. \square

O próximo passo para a prova da Lei Forte dos Grandes Números é cuidar da esperança das novas variáveis Y_i .

Lema 3.4.3. *Sejam $Z_i = Y_i - E(Y_i)$, para $i \geq 1$ como acima. Então, para demonstrar o Teorema 3.4.1, basta mostrar que*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n Z_i = 0, \text{ } P\text{-quase certamente.} \quad (3.51)$$

Demonstração. Supondo a convergência em (3.51), sabemos que

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n Y_i - E(Y_i) = 0, \text{ } P\text{-quase certamente.} \quad (3.52)$$

Mas $E(Y_i) = E(X_i \mathbb{1}_{[|X_i| \leq i]})$ que converge a $E(X_i) = m$, pelo Teorema da Convergência Dominada, donde concluímos que

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n E(Y_i) = m. \quad (3.53)$$

Dessa forma, obtemos que $\frac{1}{n} \sum_{i=1}^n Y_i$ converge quase certamente a m , donde concluímos a prova do Teorema 3.4.1 por meio do Lema 3.4.2. \square

Gostaríamos de utilizar os teoremas das séries para mostrar a convergência de $\frac{1}{n} \sum_n Z_n$, mas obviamente, o fator $\frac{1}{n}$ que precede a soma nos impede de fazê-lo. O próximo resultado é um simples exercício de análise real, que nos permite reduzir a prova de (3.51) para uma simples convergência de uma série sem pré-fatores.

Lema 3.4.4 (Lema de Kronecker). *Suponha que $x_n \in \mathbb{R}$ e $b_n > 0$ sejam tais que $b_n \uparrow \infty$ e $\sum_{i=1}^{\infty} \frac{x_i}{b_i}$ convirja a $s \in \mathbb{R}$. Então*

$$\lim_{n \rightarrow \infty} \frac{1}{b_n} \sum_{i=1}^n x_i = 0. \quad (3.54)$$

Demonstração. Definindo $s_0 = 0$ e $s_n = \frac{x_1}{b_1} + \cdots + \frac{x_n}{b_n}$, temos, por integração por partes,

$$\sum_{i=1}^n x_i = \sum_{i=1}^n b_i \frac{x_i}{b_i} = \sum_{i=1}^n b_i s_i - \sum_{i=1}^n b_i s_{i-1} = b_n s_n + \sum_{i=1}^{n-1} (b_i - b_{i+1}) s_i. \quad (3.55)$$

Escolhemos agora, para qualquer $\varepsilon > 0$, um $n_0 \geq 1$ tal que $|s_n - s| < \varepsilon$ para todo $n \geq n_0$. Dessa forma,

$$\begin{aligned} \frac{1}{b_n} \sum_{i=1}^n x_i &= s_n - \frac{1}{b_n} \sum_{i=1}^{n-1} (b_{i+1} - b_i) s_i \\ &= s_n - \underbrace{\frac{1}{b_n} \sum_{i=1}^{n_0-1} (b_{i+1} - b_i) s_i}_{\Delta_{n_0}} - \frac{1}{b_n} \sum_{i=n_0}^{n-1} (b_{i+1} - b_i) s_i \\ &= \underbrace{s_n}_{\rightarrow s} - \underbrace{\frac{1}{b_n} \Delta_{n_0}}_{\rightarrow 0} - \underbrace{\frac{1}{b_n} \sum_{i=n_0}^{n-1} (b_{i+1} - b_i) s}_{=\frac{(b_n - b_{n_0})s}{b_n} \rightarrow s} - \underbrace{\frac{1}{b_n} \sum_{i=n_0}^{n-1} (b_{i+1} - b_i) (s_i - s)}_{\leq \varepsilon \frac{(b_n - b_{n_0})}{b_n} \leq \varepsilon}, \end{aligned}$$

onde os limites indicados acima representam o que acontece quando $n \rightarrow \infty$. A prova segue do fato de ε ter sido escolhido arbitrariamente. \square

Estamos agora em posição de finalizar a

Prova do Teorema 3.4.1. De acordo com o Lema de Kronecker e o Lema 3.4.3, é suficiente mostrar que

$$\sum_{i=1}^n \frac{Z_i}{i}, \text{ converge quase certamente.} \quad (3.56)$$

Por outro lado, como os Z_i 's tem média zero, o Teorema de Uma Série diz que é suficiente mostrar que

$$\sum_{i=1}^n \text{Var}\left(\frac{Z_i}{i}\right) = \sum_{i=1}^n \frac{1}{i^2} \text{Var}(Z_i) < \infty. \quad (3.57)$$

Isso segue da seguinte estimativa

$$\begin{aligned} \sum_{i=1}^n \frac{1}{i^2} \text{Var}(Z_i) &= \sum_{i=1}^n \frac{1}{i^2} \text{Var}(Y_i) \leq \sum_{i=1}^n \frac{1}{i^2} E(X_i^2 \mathbb{1}_{[|X_i| \leq i]}) \\ &= \sum_{i=1}^n \frac{1}{i^2} \sum_{k=1}^i E(X_i^2 \mathbb{1}_{[k-1 < |X_i| \leq k]}) \\ &= \sum_{k=1}^n E(X_1^2 \mathbb{1}_{[k-1 < |X_1| \leq k]}) \sum_{i=k}^n \frac{1}{i^2} \\ &\leq 2 \sum_{k=1}^n \frac{1}{k} E(X_1^2 \mathbb{1}_{[k-1 < |X_1| \leq k]}) \\ &\leq 2 \sum_{k=1}^n E(X_1 \mathbb{1}_{[k-1 < |X_1| \leq k]}) \leq 2E(X_1) < \infty. \end{aligned} \quad (3.58)$$

Isso nos permite concluir a prova de (3.51) via o Lema de Kronecker. Consequentemente, obtemos o Teorema 3.4.1 via o Lema 3.4.3. \square

Exercício 3.4.1. Sejam Y_k variáveis aleatórias independentes e com a seguinte distribuição:

$$P[Y_k = i] = \begin{cases} \frac{1}{2} - \frac{1}{k^2} & \text{se } i = 1 \text{ or } i = -1, \\ \frac{2}{k^2} & \text{se } i = 3. \end{cases} \quad (3.59)$$

Mostre que

$$P\left[\frac{1}{n} \sum_{k=1}^n Y_k \text{ converge a zero}\right] = 1. \quad (3.60)$$

Exercício 3.4.2 (Depende de **Tópico: Urna de Pólya**). Mostre que segundo a lei P construída no Exercício 2.10.9, vale que

$$P\left[\frac{1}{n} \sum_{i=1}^n X_i \text{ converge}\right] = 1. \quad (3.61)$$

Além disso calcule a distribuição do limite acima.

3.5 Lei $\{0, 1\}$ de Kolmogorov

Ao estudarmos o Lema de Borel-Cantelli, vimos que se os eventos $(A_i)_{i \geq 1}$ são independentes então a probabilidade de $[A_i \text{ infinitas vezes}]$ somente pode assumir os valores zero ou um (dependendo da somabilidade de $P(A_i)$). Nessa seção

iremos estudar outros tipos de evento que assumem apenas esses dois valores. Esperamos que esse fenômeno se torne intuitivo ao final dessa discussão.

No que se segue, consideraremos um espaço mensurável $\Omega = \times_{i=1}^{\infty} E$, com a σ -álgebra canônica \mathcal{F} , isto é a σ -álgebra gerada pelas coordenadas canônicas $(X_i)_{i=1}^{\infty}$.

Definição 3.5.1. Dizemos que um evento $A \in \mathcal{F}$ é caudal se

$$A \in \sigma(X_i; i \geq n), \text{ para todo } n \geq 1. \quad (3.62)$$

Também introduzimos a classe \mathcal{F}_{∞} de tais eventos, que claramente é uma σ -álgebra, pois pode ser escrita como

$$\mathcal{F}_{\infty} = \bigcap_{n \geq 1} \sigma(X_i; i \geq n). \quad (3.63)$$

Chamamos \mathcal{F}_{∞} de σ -álgebra caudal.

Vejamos que, dados $A_i \in \sigma(X_i)$, $i \geq 1$, temos que $[A_i \text{ infinitas vezes}]$ é caudal. Para tanto, basta observar que para todo $n \geq 1$, temos que

$$[A_i \text{ infinitas vezes}] = [\#\{i \geq 1; \omega \in A_i\} = \infty] = [\#\{i \geq n; \omega \in A_i\} = \infty],$$

que obviamente pertence a $\sigma(X_i; i \geq n)$ para todo $n \geq 1$.

Exercício 3.5.1. Mostre que em $\Omega = \mathbb{R}^{\infty}$, são caudais os seguintes eventos

- a) $[X_i \text{ converge}]$,
- b) $[\frac{1}{n} \sum_{i=1}^n X_i \text{ converge}]$ e
- c) $[\#\{i \geq 1; X_i > 0\} < \infty]$.

Podemos agora enunciar o principal teorema dessa seção

Teorema 3.5.2 (Lei $\{0, 1\}$ de Kolmogorov). Se $\Omega = E^{\infty}$, onde E é um espaço canônico, for provido de uma lei produto $P = \otimes_{i=1}^{\infty} P_i$, então todo evento caudal tem probabilidade 0 ou 1 sob P .

Quando uma σ -álgebra \mathcal{F} satisfaz $P(A) \in \{0, 1\}$ para todo $A \in \mathcal{F}$, dizemos que \mathcal{F} é trivial. Uma outra maneira de enunciar a conclusão do teorema acima é dizer que a σ -álgebra caudal \mathcal{F}_{∞} é trivial.

Demonstração. A idéia da prova, apesar de soar um pouco estranha, é mostrar que se $A \in \mathcal{F}_{\infty}$, então A é independente de si mesmo. Em outras palavras, $P(A) = P(A \cap A) = P(A)^2$, donde $P(A) \in \{0, 1\}$. Mas vamos com calma.

Fixe $k \geq 1$, $A \in \mathcal{F}_{\infty}$ e $B \in \sigma(X_1, \dots, X_k)$. Nesse caso, como o evento A pertence a $\sigma(X_{k+1}, X_{k+2}, \dots)$, temos que A e B são independentes. Fixe agora $A \in \mathcal{F}_{\infty}$ e considere a classe

$$\mathcal{B}_A = \{B \in \mathcal{F}; B \text{ é independente de } A\}. \quad (3.64)$$

Já sabemos que $\sigma(X_1, \dots, X_k) \subseteq \mathcal{B}_A$ para todo $k \geq 1$.

Obviamente Ω é independente de A , assim como $B^c \in \mathcal{B}_A$ sempre que $B \in \mathcal{B}_A$. Além disso, suponha que B_1, B_2, \dots in \mathcal{B}_A são disjuntos, então,

$$P\left(\left(\bigcup_i B_i\right) \cap A\right) = P\left(\bigcup_i (B_i \cap A)\right) \stackrel{\text{disj.}}{=} \sum_i P(B_i \cap A) \stackrel{\text{indep.}}{=} P(A)P\left(\bigcup_i B_i\right).$$

Logo \mathcal{B}_A é um λ -sistema.

Lembrando que \mathcal{B}_A contém o π -sistema $\bigcup_k \sigma(X_1, \dots, X_k)$, isto é dos eventos cilíndricos, temos que todos eventos são indepentes de A , inclusive o próprio A . Isso termina a prova do teorema. \square

Exercício 3.5.2. Dizemos que uma probabilidade P no espaço produto $\Omega = \times_{n \geq 1} E$ (com a σ -álgebra canônica) é fortemente misturadora se, para todo $k \geq 1$, temos

$$\lim_{n \rightarrow \infty} \sup |P(A \cap B) - P(A)P(B)| = 0, \quad (3.65)$$

onde o supremo acima é tomado sobre $A \in \sigma(X_1, \dots, X_k)$ e $B \in \sigma(X_n, X_{n+1}, \dots)$. Mostre que nesse caso, a σ -álgebra dos eventos caudais é trivial.

Exercício 3.5.3 (Depende de **Tópico: Percolação**). Considere o grafo $G = (\mathbb{Z}^2, E)$, onde $E = \{\{x, y\}; |x - y|_2 = 1\}$. Dotamos agora o espaço $\{0, 1\}^E$ com a σ -álgebra \mathcal{A} gerada pelas projeções canônicas $Y_e(\omega) = \omega(e)$, onde $\omega \in \{0, 1\}^E$ e $e \in E$. Definimos o conjunto $A \subseteq \{0, 1\}^E$ por

$$A = \left[\begin{array}{l} \text{existe uma sequência de distintos } x_0, x_1, \dots \in \mathbb{Z}^2, \\ \text{tais que } e_i = \{x_i, x_{i+1}\} \in E \text{ e } Y_{e_i} = 1 \text{ para cada } i \geq 0 \end{array} \right]. \quad (3.66)$$

a) Mostre que A é mensurável com respeito a \mathcal{A} .

b) Mostre que A é um evento caudal, ou seja

$$A \in \bigcap_{K \subseteq E; \text{finito}} \sigma(Y_e; e \notin K). \quad (3.67)$$

c) Conclua que $P(A) \in \{0, 1\}$.

Exercício 3.5.4. Seja $\Omega = E^{\mathbb{Z}}$ um espaço produto infinito, dotado da σ -álgebra \mathcal{A} gerada pelas projeções canônicas $(X_i)_{i \in \mathbb{Z}}$. Consideramos agora em (Ω, \mathcal{A}) a medida produto $\mathbb{P} = P^{\otimes \mathbb{Z}}$, onde P é uma probabilidade fixada no espaço polonês $(E, \mathcal{B}(E))$.

a) Mostre que para qualquer evento $A \in \mathcal{A}$ e qualquer $\varepsilon > 0$, existe um $k \in \mathbb{Z}_+$ e um evento $A_k \in \sigma(X_i, |i| \leq k)$ tais que $\mathbb{P}[(A \setminus A_k) \cup (A_k \setminus A)] < \varepsilon$.

b) Considere o shift $\theta : \Omega \rightarrow \Omega$ dado por $\theta(\omega)(i) = \omega(i - 1)$ e mostre que se $A = \theta(A)$, então $P(A) \in \{0, 1\}$.

3.6 Momentos exponenciais

Nessa seção desenvolveremos uma outra técnica para estimar a probabilidade de uma variável aleatória se desviar de sua esperança.

Já vimos o método do primeiro, segundo e quarto momento para controlar uma soma de variáveis independentes. Um exemplo disso foi visto na estimativa

$$P\left[\sum_{i=1}^n (X_i - E(X_i)) \geq a\right] \leq \frac{\sum_i \text{Var}(X_i)}{a^2}. \quad (3.68)$$

Em geral, quanto maior o momento, melhor a estimativa do decaimento para a probabilidade de que uma variável se desvie de sua esperança. Nessa seção iremos para momentos exponenciais, que em um certo sentido produzem estimativas ótimas para o comportamento assintótico da probabilidade de desvio.

Note que se quisermos uma pequena probabilidade de erro (como por exemplo ~ 0.01), o método do segundo momento é muito bom, como veremos posteriormente. Mas se quisermos uma probabilidade de erro minúscula (em situações concretas, algo como 10^{-12} por exemplo), certamente teremos que aumentar bastante o valor de n , mas quanto? As cotas de segundo momento são muito ruins para esse tipo de estimativa, nos levando a escolher um n maior que o necessário. Abaixo, desenvolveremos um método mais eficiente para responder a essa pergunta, obviamente sob certas hipóteses na distribuição das variáveis aleatórias.

Definição 3.6.1. Dada uma variável aleatória X , definimos sua transformada de Laplace como

$$\phi_X(s) = E(e^{sX}) \in (0, \infty], \quad (3.69)$$

para todos $s \in \mathbb{R}$. Essa transformada também é chamada função geradora de momentos de X .

Exercício 3.6.1. Calcule a função geradora de momentos das distribuições $\text{Ber}(p)$, $\text{Exp}(\lambda)$ e $U_{[0,1]}$.

Proposição 3.6.2. Se $E(e^{\delta|X|}) < \infty$, então

- a) $X \in \mathcal{L}^p$ para todo $1 \leq p < \infty$,
- b) $\phi_X(s) < \infty$ para todo $s \in (-\delta, \delta)$,
- c) $\phi_X(s)$ é C^∞ em $(-\delta, \delta)$ e
- d) $\phi_X^{(n)}(s) = E(X^n e^{sX})$.

A última conclusão da proposição acima justifica a nomenclatura função geradora de momentos pois $\phi_X^{(n)}(0) = E(X^n)$.

Demonstração. Obviamente, para todo $p \geq 1$ existe $c > 0$ tal que $e^{\delta|x|} \geq c|x|^p$, donde $X \in \mathcal{L}^p$. Além disso, para todo $s \in (-\delta, \delta)$, temos $\phi_X(s) = E(e^{sX}) \leq E(e^{\delta|X|}) < \infty$, donde 2. segue imediatamente.

Fixando $s \in \mathbb{R}$, vamos agora calcular

$$\frac{\phi_X(s+h) - \phi_X(s)}{h} = \frac{E(e^{(s+h)X} - e^{sX})}{h} = E\left(e^{sX} \frac{e^{hX} - 1}{h}\right). \quad (3.70)$$

Lembrando que $|\frac{1}{y}(e^y - 1)| \leq e^{|y|}$, para todo $y \in \mathbb{R}$, temos que para todos os $h < (\delta - |s|)/2$, o integrando acima é dominado por $|X|e^{(|s|+h)|X|} \leq |X|e^{\frac{\delta+|s|}{2}|X|}$ que pertence a \mathcal{L}^1 . Logo podemos usar o Teorema da Convergência Dominada para trocar o limite $h \rightarrow 0$ com a esperança, obtendo

$$\phi'_X(s) = E(Xe^{sX}). \quad (3.71)$$

Note que para todo $\varepsilon > 0$ e $k \geq 1$, $|x|^k \leq c(k)e^{\varepsilon|x|}$, isso nos permite repetir o argumento acima indutivamente para obter c) e d). \square

Lembramos que ao usar o método do segundo momento, nos foi bastante útil o fato que a variância se comporta bem com relação a somas independentes. Mais precisamente, $\text{Var}(X_1 + \dots + X_k) = \text{Var}(X_1) + \dots + \text{Var}(X_k)$.

Uma outra propriedade importante da função geradora de momentos é que ela também se comporta bem com respeito à somas independentes.

Proposição 3.6.3. *Se X_1, \dots, X_n são variáveis independentes com $\phi_{X_i}(s) < \infty$ para todo $i \leq k$ e $|s| < \delta$, então*

$$\phi_{X_1 + \dots + X_k}(s) = \phi_{X_1}(s) \cdots \phi_{X_k}(s), \text{ para todos } |s| < \delta. \quad (3.72)$$

Demonstração. Basta observar que

$$\begin{aligned} E(\exp\{s(X_1 + \dots + X_k)\}) &= E(e^{sX_1} \cdots e^{sX_k}) \\ &= E(e^{sX_1}) \cdots E(e^{sX_k}) = \phi_{X_1}(s) \cdots \phi_{X_k}(s), \end{aligned} \quad (3.73)$$

usando Fubini. \square

Consideraremos agora uma sequência X_1, X_2, \dots de variáveis i.i.d. com $\phi_{X_1}(s) < \infty$ para $|s| < \delta$. Então podemos tentar estimar, para $a > 0$ e $|s| < \delta$,

$$\begin{aligned} P\left[\frac{X_1 + \dots + X_n}{n} - E(X_1) \geq a\right] &= P[X_1 + \dots + X_n \geq (a + E(X_1))n] \\ &= P[e^{s(X_1 + \dots + X_n)} \geq e^{s(a + E(X_1))n}] \\ &\leq \phi_{X_1 + \dots + X_n}(s) e^{-s(a + E(X_1))n} = \phi_{X_1}^n(s) e^{-s(a + E(X_1))n}. \end{aligned}$$

O primeiro fator na estimativa acima pode crescer exponencialmente com n , enquanto o segundo decresce. Gostaríamos que o comportamento do segundo predominasse, o que podemos concluir do seguinte argumento.

Sabemos que $\phi_{X_1}(s)$ é diferenciável em zero e que $\phi'_{X_1}(0) = E(X_1)$. Logo, existe $s > 0$ tal que $\phi_{X_1}(s) < 1 + (E(X_1) + \frac{a}{2})s$, donde

$$\begin{aligned} P\left[\frac{X_1 + \cdots + X_n}{n} - E(X_1) \geq a\right] &\leq \phi_{X_1}^n(s) e^{-s(a + E(X_1))n} \\ &\leq \left(1 + (E(X_1) + \frac{a}{2})s\right)^n e^{-s(E(X_1) + a)n} \\ &\leq \exp\left\{s\left(E(X_1 + \frac{a}{2}) - E(X_1) - a\right)n\right\} = e^{-san/2}. \end{aligned}$$

Isso nos garante um decaimento exponencial da probabilidade da média dos X_i se desviar da esperança.

Exercício 3.6.2. *Aplique o método acima para variáveis X_i i.i.d. com distribuição $\text{Ber}(1/2)$ e encontre $s(a)$ que otimize o decaimento da probabilidade $P[\sum_{i=1}^n X_i > (1/2 + a)n]$.*

Poderíamos nos perguntar se a cota acima é suficientemente boa. Talvez pudéssemos esperar um decaimento ainda melhor que exponencial. Para responder a essa pergunta, vamos considerar o seguinte exemplo. Sejam $(X_i)_{i \geq 1}$ variáveis i.i.d. com $X_1 \sim_d \text{Ber}(1/2)$. Nesse caso temos por exemplo

$$P\left[\left|\frac{X_1 + \cdots + X_n}{n} - \frac{1}{2}\right| \geq \frac{1}{4}\right] \geq P[X_i = 1, \forall i \leq n] = 2^{-n}. \quad (3.74)$$

Dessa forma, sabemos que não podemos esperar um decaimento melhor que exponencial, mesmo para variáveis bem simples (como Bernoulli) que satisfazem $\phi_X(s) < \infty$ para todo $s \in \mathbb{R}$.

Note que para variáveis com distribuição $\text{Ber}(1/2)$, obtivemos acima cotas exponenciais em n (superior e inferior), mas elas possuem expoentes diferentes. Resta agora tentar entender qual é o expoente correto para o decaimento da probabilidade $P[X_1 + \cdots + X_n \geq n(E(X_1) + a)]$, o que será feito na próxima seção.

3.7 Princípio de Grandes Desvios

A primeira tarefa nossa será otimizar a estimativa grosseira feita na seção anterior. Essas estimativas são chamadas de *estimativas de grandes desvios*, pois se referem a probabilidades que a média empírica de X_i se desvie de sua esperança por um valor constante a . Futuramente no curso estudaremos as probabilidades de que esse desvio seja de ordem $a_n \rightarrow 0$ que são chamados de *desvios moderados* ou *flutuações*, dependendo se a probabilidade de desvio converge a zero ou não.

Teorema 3.7.1 (Princípio de Grandes Desvios - cota superior). *Consideramos variáveis aleatórias i.i.d. X_1, X_2, \dots tais que $\phi_{X_1}(s) < \infty$, para todo $s \in (-\delta, \delta)$. Então, para $a > 0$,*

$$P[X_1 + \cdots + X_n \geq (m + a)n] \leq e^{-\psi_{X_1}(m+a)n}, \quad (3.75)$$

onde $m = E(X_1)$ e

$$\psi_{X_1}(x) = \sup_{s \geq 0} \{xs - \log(\phi_{X_1}(s))\} \quad (3.76)$$

é chamada função taxa.

É importante observar que para estimar $P[X_1 + \dots + X_n \leq (m - a)n]$, basta considerarmos $X'_i = -X_i$ ao utilizar o teorema acima.

Demonstração. Já sabemos que, para todo $s \geq 0$,

$$\begin{aligned} P[X_1 + \dots + X_n \geq (m + a)n] &\leq \phi_{X_1}^n(s) e^{-s(m+a)n} \\ &= \exp \{ \log(\phi_{X_1}(s))n - s(m+a)n \} \\ &= \exp \{ -((m+a)s - \log(\phi_{X_1}(s)))n \} \end{aligned} \quad (3.77)$$

O que termina a prova do teorema se tomamos o ínfimo em $s \geq 0$. \square

Exercício 3.7.1. Calcule $\psi_X(a)$ quando X é distribuída como $\text{Ber}(p)$, $U_{[0,1]}$ e $\text{Exp}(\lambda)$.

Exercício 3.7.2. Na Nova Caledônia, temos k habitantes. Seja $f : \{1, \dots, k\} \rightarrow \{0, 1\}$ uma função que indica a intenção de voto de cada cidadão. Mais precisamente, para cada habitante $i \in \{1, \dots, k\}$, se $f(i) = 0$, então i vota no candidato 0, enquanto se $f(i) = 1$, o cidadão i vota no candidato 1. Para estimar o número $k_1 = \#f^{-1}(\{1\})$ de pessoas que votam em 1, nós escolhemos variáveis aleatórias Y_i i.i.d. com distribuição uniforme em $\{1, \dots, k\}$ e queremos estimar

$$\text{Err}_n(\epsilon) = P \left[\left| \frac{1}{n} \sum_{i=1}^n f(Y_i) - \frac{k_1}{k} \right| > \epsilon \right]. \quad (3.78)$$

Sabendo que k é par e $k_1 = k/2$, então

- a) use o método do segundo momento para obter um n tal que $\text{Err}_n(0.01) < 0.02$ e um n tal que $\text{Err}_n(0.01) < 10^{-12}$,
- b) use o método do momento exponencial para obter resolver o item acima.

Compare os quatro resultados obtidos acima.

Vamos agora tomar um exemplo concreto para análise. Sejam X_1, X_2, \dots variáveis aleatórias i.i.d. com distribuição $\text{Ber}(1/2)$, donde

$$\phi_{X_1}(s) = \frac{1}{2}(1 + e^s) \quad \text{e} \quad \psi_{X_1}(x) = \sup_{s \geq 0} \{xs - \log(1 + e^s) + \log(2)\}. \quad (3.79)$$

Um cálculo simples nos mostra que, se $x < 1$, o mínimo acima é atingido no único ponto $s_{\max} = \log(\frac{x}{1-x})$. Portanto, podemos concluir do Teorema 3.7.1 que

$$\begin{aligned} P[X_1 + \dots + X_n > 1/2 + a] &\leq e^{-\psi_{X_1}(s_{\max})n} \\ &= \exp \left\{ -n \left(b \log(b) + (1-b) \log(1-b) + \log(2) \right) \right\} \end{aligned} \quad (3.80)$$

Note que $P[X_1 + \dots + X_n = n] = 2^{-n} = e^{-\log(2)n} = e^{-\psi_{X_1}(1-)n}$. Isso nos dá um forte indício de que talvez nossas cotas superiores não estejam tão longe de ser precisas. Para confirmar essa hipótese, precisamos obter cotas inferiores parecidas.

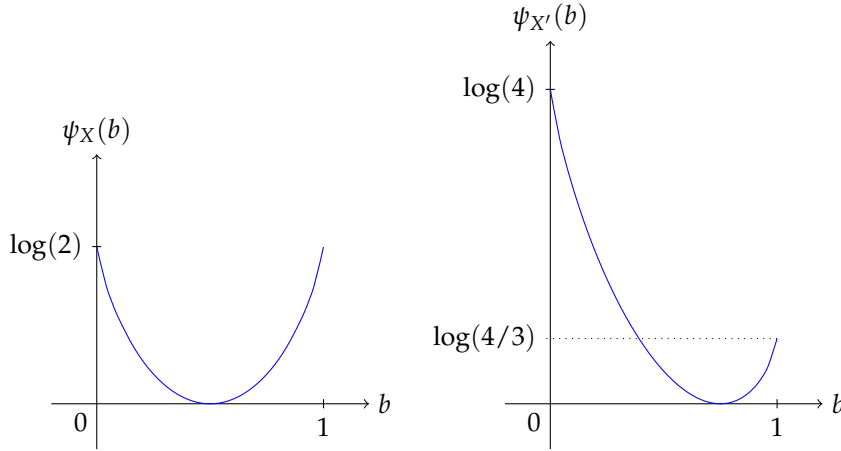


Figura 3.1: Funções taxa $\psi_X(b)$ de uma variável X com distribuição $\text{Ber}(1/2)$, e $\psi_{X'}(b)$ de uma variável com distribuição $\text{Ber}(3/4)$, para $b \in (0, 1)$.

Antes de buscar cotas inferiores para as probabilidades de desvio, vamos estabelecer algumas propriedades da função $\psi_X(b)$. Primeiramente, quando podemos dizer que o supremo na definição de ψ_X é atingido em algum s_{\max} ? Certamente, esse nem sempre é o caso, por exemplo se $X = m$ quase certamente, então $\phi_X(s) = e^{sm}$ e o supremo definindo $\psi_X(b)$ não é atingido se $b \neq m$.

Lema 3.7.2. *Seja X uma variável aleatória tal que $\phi_X(s) < \infty$ para todo $s \in (-\delta, \delta)$. Supondo $a \geq 0$ é tal que $P[X > m + a] > 0$, então existe $s_{\max} \geq 0$ tal que*

$$\psi_X(m + a) = (m + a)s_{\max} - \log(\phi_X(s_{\max})). \quad (3.81)$$

Demonstração. Por hipótese, existe $x > m + a$ tal que $p = P[X \geq x] > 0$, donde $\phi_X(s) \geq pe^{s(m+a)}$. Dessa forma, $(m + a)s - \log(\phi_X(s)) \leq (m + a - x)s - \log(p)$, que converge a menos infinito quando s diverge. Isso, junto com a continuidade de ϕ_X implica a existência do s_{\max} desejado. \square

Lema 3.7.3. *Seja X uma variável aleatória tal que $\phi_X(s) < \infty$ para todo $s \in (-\delta, \delta)$. Então o conjunto onde a função $\psi_X(s)$ é finita é um intervalo, na qual ψ_X é convexa e portanto contínua.*

Demonstração. Primeiramente, supomos que $a < b$ são tais que $\psi_X(a)$ e $\psi_X(b)$ são finitas. Logo, para todo $c \in (a, b)$, temos que a função linear cs é menor ou

igual a $as \vee bs$, daí

$$\begin{aligned}\psi_X(c) &= \sup_{s \geq 0} \{cs - \log(\phi_X(s))\} \leq \sup_{s \geq 0} \{(as \vee bs) - \log(\phi_X(s))\} \\ &\leq \sup_{s \geq 0} \{as - \log(\phi_X(s))\} \vee \sup_{s \geq 0} \{bs - \log(\phi_X(s))\} < \infty.\end{aligned}\quad (3.82)$$

Para mostrar que ψ_X é convexa, observe que $\psi_X(x)$ é dada pelo supremo (para $s \geq 0$) das funções afins $x \mapsto xs - \psi_X(s)$. Como o supremo de funções convexas é também convexo, obtemos o enunciado do lemma. \square

Exercício 3.7.3. Suponha que se $\phi_X(s)$ é finita para todo $s \in (-\delta, \delta)$ e mostre que

- a) na definição de $\psi_X(a)$, poderíamos tomar o ínfimo em todos $s \in \mathbb{R}$ (ao invés de $s \geq 0$) sem mudar o valor de $\psi_X(a)$,
- b) a função $\psi_X(s)$ é não negativa, semi-contínua inferior e convexa em seu domínio
- c) $\psi_X(a)$ se anula somente em $a = 0$ e ψ_X é crescente no seu domínio.

Buscaremos agora cotas inferiores para a probabilidade de obter um grande desvio. Gostaríamos que essas estimativas fossem o mais próximas possíveis das estimativas superiores obtidas acima. Certamente não podemos obter algo como

$$P[X_1 + \cdots + X_n \geq (m + a)n] \geq \exp\{-\psi_{X_1}(a)n\}, \quad (3.83)$$

pois senão isso nos daria uma igualdade o que é impossível, pois perdemos um pouco de precisão ao utilizar a desigualdade de Markov na cota superior.

Contudo, gostaríamos de entender se ao menos o expoente $\psi_{X_1}(a)$ na cota superior também possui algum papel na cota inferior. Isso é confirmado no seguinte resultado.

Teorema 3.7.4 (Princípio de Grandes Desvios - cota inferior). *Sejam X_1, X_2, \dots variáveis aleatórias i.i.d. com $\phi_{X_1}(s) < \infty$, para todo $s \in \mathbb{R}$. Então, para todo $a > 0$,*

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log P[X_1 + \cdots + X_n \geq (m + a)n] \geq -\psi_{X_1}(m + a), \quad (3.84)$$

onde novamente $m = E(X_1)$ e $\psi_{X_1}(x)$ é definida como no Teorema 3.7.1.

Note que o resultado do teorema acima é mais fraco que o que vemos na equação (3.83), mas mostra que $\psi_{X_1}(a)$ é realmente o expoente correto no decaimento da probabilidade de grandes desvios.

Um corolário dos Teoremas 3.7.1 e 3.7.4 é o seguinte

Corolário 3.7.5. *Se X_1, X_2, \dots variáveis aleatórias i.i.d. com $\phi_{X_1}(s) < \infty$, para todo $s \in \mathbb{R}$, então*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log P[X_1 + \cdots + X_n \geq (m + a)n] = -\psi_{X_1}(m + a). \quad (3.85)$$

A idéia da prova é transformar a distribuição de X_i , usando uma exponencial como derivada de Radon-Nikodim. Essa nova distribuição possuirá esperança maior que $m + a$, de forma que se tomamos a média de variáveis i.i.d. X'_1, \dots, X'_n distribuídas dessa forma, obteremos algo que se concentra acima de $m + a$. Finalmente, o preço pago para que as variáveis X_i se comportem como as X'_i será aproximadamente $\exp\{-\psi_{X_1}(m + a)\}$, como desejado para nossa cota inferior.

Demonstração. Primeiramente, consideraremos o caso $P[X_1 \leq m + a] = 1$, que se assemelha ao caso que analisamos acima ($\text{Ber}(1/2) \leq 1$). Nesse caso, temos

$$\begin{aligned} P[X_1 + \dots + X_n \geq (m + a)n] &= P[X_i = m + a, \text{ para todo } i \leq n] \\ &= P[X_1 = m + a]^n. \end{aligned}$$

Donde o limite acima é igual a $\log(P[X_1 = m + a])$. Mas por outro lado,

$$\begin{aligned} -\psi_{X_1}(m + a) &= \inf_{s \geq 0} \{ \log(E(e^{s(X_1)})) - (m + a)s \} = \inf_{s \geq 0} \{ \log(E(e^{s(X_1 - m - a)})) \} \\ &\leq \liminf_{s \rightarrow \infty} \log(E(e^{s(X_1 - m - a)})) = \log(P[X_1 = m + a]), \end{aligned}$$

pelo Teorema da Convergência Dominada, demonstrando o teorema nesse caso especial.

Suponhamos agora que $P[X_1 > m + a] > 0$, o que implica que para $b > m + a$ suficientemente próximo de $m + a$, temos $P[X_1 > b] > 0$. Observe que basta mostrar que para todo $b > a$ satisfazendo $P[X_1 > b] > 0$ e para todo $\delta > 0$, temos

$$\liminf_n \frac{1}{n} \log \left(P \left[\frac{X_1 + \dots + X_n}{n} \in (b - \delta, b + \delta) \right] \right) \geq -\psi_{X_1}(b), \quad (3.86)$$

pois a função $\psi_{X_1}(x)$ é convexa, portanto contínua.

Vamos definir uma nova distribuição ν com derivada de Radon-Nikodim

$$\frac{d\nu}{dP_{X_1}} = \frac{1}{Z_\sigma} e^{\sigma x}. \quad (3.87)$$

Observamos primeiramente que o valor de σ ainda não foi escolhido. Além disso após escolhido σ , teremos que calcular a constante de normalização Z_σ de forma que ν seja uma probabilidade.

Escolheremos $\sigma \geq 0$ como no Lema 3.7.2, isto é, tal que $\psi_{X_1}(b) = b\sigma - \log(\phi_{X_1}(\sigma))$. Isso nos dá imediatamente que $Z_\sigma = E[e^{\sigma X_1}] = \phi_{X_1}(\sigma)$ por definição.

Por diferenciabilidade de ϕ_{X_1} , o máximo deve ser assumido em um ponto de derivada zero para a função ψ_{X_1} , ou seja

$$b = \frac{\phi'_{X_1}(\sigma)}{\phi_{X_1}(\sigma)} \stackrel{\text{Prop. 3.6.2}}{=} \frac{E(Xe^{\sigma X})}{E(e^{\sigma X})} = \frac{E(Xe^{\sigma X})}{Z_\sigma} = \int x \nu(dx). \quad (3.88)$$

Isso implica que se uma variável aleatória tem distribuição ν , sua esperança é b . É possível verificar que uma tal variável aleatória X' satisfaz obrigatoriamente $\phi_{X'}(s) < \infty$ para todo $s \geq 0$, donde $X' \in \mathcal{L}^p$ para todo $p > 1$.

Como prometido, consideramos variáveis X'_1, X'_2, \dots i.i.d. com distribuição ν . Pela lei fraca dos grandes números, para qualquer $\delta > 0$,

$$\lim_n P\left[\frac{X'_1 + \dots + X'_n}{n} \in (b - \delta, b + \delta)\right] = 1. \quad (3.89)$$

Finalmente vamos relacionar essa probabilidade à probabilidade definida em termos de X_i , na qual estamos interessados.

$$\begin{aligned} P\left[\frac{X_1 + \dots + X_n}{n} \in (b - \delta, b + \delta)\right] &= \int_{x_i: \left|\frac{1}{n} \sum_{i=1}^n x_i - b\right| < \delta} \bigotimes_{i=1}^n (X_1 \circ P)(dx_i) \\ &= Z_\sigma^n \int_{x_i: \left|\frac{1}{n} \sum_{i=1}^n x_i - b\right| < \delta} e^{-\sigma \sum_{i=1}^n x_i} \bigotimes_{i=1}^n (X'_1 \circ P)(dx_i) \\ &\geq Z_\sigma^n \exp\{-(b + \delta)\sigma n\} P\left[\frac{X'_1 + \dots + X'_n}{n} \in (b - \delta, b + \delta)\right]. \end{aligned}$$

Tomando o logaritmo, dividindo por n e tomando o liminf quando n vai a infinito, recuperamos

$$\begin{aligned} \lim_n \frac{1}{n} \log \left(P\left[\frac{X_1 + \dots + X_n}{n} \in (b - \delta, b + \delta)\right] \right) &\geq \log(Z_\sigma) - (b + \delta)\sigma \\ &= \log(\phi_{X_1}(\sigma)) - (b + \delta)\sigma = -\psi_{X_1}(\sigma) - \delta\sigma. \end{aligned} \quad (3.90)$$

Como isso vale para todo $\delta > 0$, provamos (3.86) o que conclui a prova do teorema. \square

Exercício 3.7.4. Mostre o Teorema 3.7.4 no caso em que $\phi_{X_1}(s) < \infty$, para todo $s \in (-\delta, \delta)$.

Tópico: Funções características ¹

Esta seção trata da função característica de uma variável aleatória, que pode ser vista como um análogo complexo da transformada de Laplace, ou também como a transformada de Fourier de uma distribuição em \mathbb{R} . Vamos estudar suas principais propriedades e demonstrar que a função características determinam unicamente a distribuição da variável aleatória.

Definição 3.7.6. Dada uma variável aleatória X , a função característica de X , $\bar{\phi}_X : \mathbb{R} \rightarrow \mathbb{C}$, é definida por

$$\bar{\phi}_X(t) = \mathbb{E}(e^{itX}), \quad t \in \mathbb{R}. \quad (3.91)$$

Vamos começar estudando as propriedades básicas de $\bar{\phi}_X$.

Exercício 3.7.5. Prove que a função $\bar{\phi}_X$ é absolutamente contínua.

Exercício 3.7.6. Suponha que $\mathbb{E}(|X|^n) < +\infty$. Prove que a função $\bar{\phi}_X$ é n vezes diferenciável em $t = 0$ e que $\bar{\phi}_X^{(n)}(0) = i^n \mathbb{E}(X^n)$.

Exercício 3.7.7. Se X_1, X_2, \dots, X_n são independentes e $a_1, a_2, \dots, a_n \in \mathbb{R}$, então

$$\bar{\phi}_{a_1 X_1 + a_2 X_2 + \dots + a_n X_n}(t) = \bar{\phi}_{X_1}(a_1 t) \bar{\phi}_{X_2}(a_2 t) \cdots \bar{\phi}_{X_n}(a_n t). \quad (3.92)$$

Como vamos ver agora, a função característica nos permite recuperar a distribuição de X :

Exercício 3.7.8. Use a seguinte igualdade

$$\lim_{T \rightarrow +\infty} \int_0^T \frac{\sin(tz)}{t} dz = \begin{cases} 1 & \text{se } z > 0 \\ 0 & \text{se } z = 0 \\ -1 & \text{se } z < 0 \end{cases} \quad (3.93)$$

para provar que se $a < b$ são pontos de continuidade da função de distribuição de X , F_X , então

$$F_X(b) - F_X(a) = \lim_{T \rightarrow +\infty} \frac{1}{2\pi} \int_{-T}^T \frac{e^{-itb} - e^{-ita}}{-it} \bar{\phi}_X(t) dt. \quad (3.94)$$

Conclua que a distribuição de X é determinada por $\bar{\phi}_X$.

O próximo exercício consiste em calcular algumas funções características.

Exercício 3.7.9. Calcule as funções características das seguintes distribuições:

- i. $X \sim \text{Ber}(p)$;
- ii. $X \sim \text{Poisson}(\lambda)$;
- iii. $X \sim N(0, 1)$. Dica: fixe $z \in \mathbb{R}$, calcule $\mathbb{E}(e^{zX})$ e use o Princípio da continuação analítica.

¹Somos gratos a Rangel Baldasso por escrever essa seção.

3.8 O Teorema Central do Limite

Até o presente momento, já sabemos por exemplo que médias de variáveis aleatórias i.i.d. , suficientemente regulares convergem para sua esperança quase certamente. Vamos fazer contudo um experimento para visualizar esse fenômeno.

Nesse experimento, jogamos 100 moedas e contamos quantas caras obtivemos. Pelo que discutimos anteriormente, esperamos que esse número se encontre por volta de 50, que é a esperança desta soma de variáveis i.i.d. . Vamos portanto repetir esse experimento mil vezes e observar quantas vezes obtemos algo próximo de 50, veja Figura 3.2.

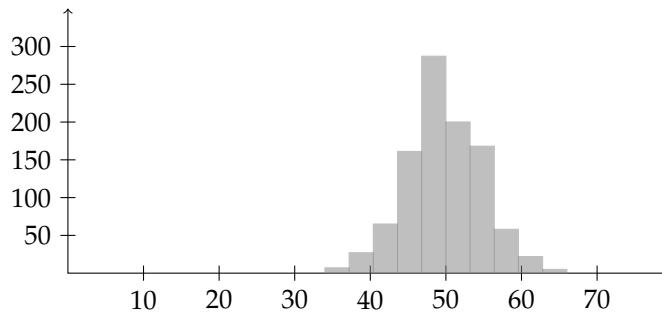


Figura 3.2: Vários ensaios de uma variável $\text{Bin}(100, 0.5)$, pra ser mais preciso 1000 ensaios. Cada barra representa o número de ensaios que caíram no intervalo determinado pela base da barra. Note que apesar dos experimentos se concentrarem em torno da média, alguns se afastam um pouco (obviamente pois o experimento é aleatório). Nessa seção estudaremos esses desvios espontâneos, que são chamados de flutuações.

Nosso objetivo nessa seção será obter qual é o tamanho típico das flutuações em torno da média dessa soma de variáveis aleatórias. Ao contrário do que fizemos ao estudar Grandes Desvios, nós agora estamos buscando flutuações menores, que acontecem espontaneamente e não com baixa probabilidade.

Note também que apesar de observarmos uma aleatoriedade na Figura 3.2, também notamos uma certa regularidade que muitas vezes é chamada de ‘forma de sino’ no histograma apresentado.

3.8.1 A distribuição normal

Começaremos estudando qual poderia ser uma possível forma limite para o histograma da Figura 3.2.

Como uma primeira tentativa, suponha que $\sum_{i=1}^{\infty} Z_i$ possui uma certa distribuição μ (veremos posteriormente que isso somente pode acontecer em casos triviais). Mas se esse fosse o caso, poderíamos dividir a soma nos termos pares e ímpares $X = \sum_{i \text{ par}} Z_i$ e $Y = \sum_{i \text{ ímpar}} Z_i$. Nesse caso teríamos X e Y indepen-

dentos e também distribuídos como μ (pois são dados por uma soma que tem a mesma distribuição daquela que define μ).

O seguinte lema mostra que isso somente pode acontecer na situação trivial em que $\mu = \delta_0$.

Lema 3.8.1. *Sejam X e Y variáveis aleatórias em \mathcal{L}^2 , i.i.d. com distribuição μ . Nesse caso, se $X + Y$ também tem distribuição μ , então $\mu = \delta_0$.*

Demonstração. Sabemos que

$$\begin{aligned} E(X + Y) &= E(X) + E(Y) = 2E(X) \text{ e} \\ \text{Var}(X + Y) &= \text{Var}(X) + \text{Var}(Y) = 2 \text{Var}(X). \end{aligned} \quad (3.95)$$

Mas como $X + Y$ tem a mesma distribuição de X , então $E(X) = 2E(X)$ e $\text{Var}(X) = 2 \text{Var}(X)$, donde ambas são zero. Usando o método dos segundo momento, para todo $a > 0$,

$$P[|X| \geq a] \leq \frac{\text{Var}(X)}{a^2} = 0, \quad (3.96)$$

terminando a prova de que $X = 0$ quase certamente. \square

A intuição dessa prova é que quando somamos duas variáveis não determinísticas, a incerteza da soma (medida através da variância) tende a aumentar. Dessa forma não podemos obter a mesma distribuição após a soma.

Mas existe uma maneira simples de tornar esse problema interessante novamente. Digamos que X e Y pertencem a \mathcal{L}^2 e são i.i.d. Então

$$\text{Var}\left(\frac{X + Y}{\sqrt{2}}\right) = 2 \text{Var}\left(\frac{X}{\sqrt{2}}\right) = \text{Var}(X). \quad (3.97)$$

Então podemos nos perguntar se

Questão 3.8.2. *Existe alguma distribuição não trivial μ em \mathcal{L}^2 tal que, se X e Y são independentes e distribuídas de acordo com μ , temos*

$$\frac{X + Y}{\sqrt{2}} \sim_d \mu ? \quad (3.98)$$

Pelo menos sabemos agora que a variância não se altera através dessa operação.

Ou em outras palavras, queremos saber se existe algum ponto fixo para o operador Γ que toma uma distribuição μ em \mathbb{R} e retorna

$$\Gamma(\mu) = \left(\frac{X_1 + X_2}{\sqrt{2}}\right) \circ \mu \otimes \mu. \quad (3.99)$$

Para tentar responder a essa questão, vamos estudar mais a fundo qual é a distribuição da soma de duas variáveis aleatórias independentes. Para isso, considere a distribuição $(X, Y) \circ P$ do par, que coincide com $\mu \otimes \mu$, nos dando

$$P\left[\frac{X + Y}{\sqrt{2}} \leq z\right] = \mu \otimes \mu\left(\{(x, y); \frac{x+y}{\sqrt{2}} \leq z\}\right). \quad (3.100)$$

Note também que a transformação linear $(x, y) \mapsto \frac{1}{\sqrt{2}}(x + y, x - y)$ é uma rotação rígida em \mathbb{R}^2 , o que nos motiva a propor a pergunta mais simples.

Questão 3.8.3. *Existe alguma distribuição não trivial μ em \mathcal{L}^2 tal que, se X e Y são independentes e distribuídas de acordo com μ , a distribuição do par (X, Y) é invariante por rotações?*

Ainda estamos numa busca não rigorosa de tal distribuição, então vamos supor algumas outras propriedades, como por exemplo que μ seja absolutamente contínua com respeito a Lebesgue, isto é $d\mu = f(x) dx$. Nesse caso, já vimos que $(X, Y) \sim_d f(x)f(y) dx dy$ e no fundo estamos procurando uma função f tal que

$$f(x)f(y) = h(x^2 + y^2), \text{ para todo } x, y \in \mathbb{R} \text{ e alguma } h : \mathbb{R}_+ \rightarrow \mathbb{R}_+. \quad (3.101)$$

Para transformar o produto $f(x)f(y)$ em uma soma, definimos $g = \log f$ e $k = \log h$ e o que gostaríamos que acontecesse é $g(x) + g(y) = k(x^2 + y^2)$. Como ainda não estamos preocupados com unicidade de μ e apenas com a existência, já podemos encontrar nossa resposta para nossa pergunta, escolhendo uma função quadrática, tal como $g(x) = \alpha x^2 - \beta$.

Mas temos ainda que cuidar para que $f(x) = \exp\{\alpha x^2 - \beta\}$ seja uma densidade, ou seja $\int f dx = 1$. Para isso, precisamos que α seja negativo e, fixado α , o valor de β já estará determinado por normalização. Tudo isso motiva finalmente a seguinte definição.

Definição 3.8.4. *Dizemos que X tem distribuição normal canônica, se*

$$X \sim_d \frac{1}{\sqrt{2\pi}} \exp\{-x^2/2\} dx. \quad (3.102)$$

Além disso, para $m \in \mathbb{R}$ e $\sigma \geq 0$, dizemos que $Y \sim_d \mathcal{N}(m, \sigma^2)$ se Y tem a mesma distribuição de $\sigma X + m$, onde X tem distribuição normal canônica $\mathcal{N}(0, 1)$. Note que $\mathcal{N}(m, 0) = \delta_m$. Muitas vezes chamamos essa distribuição de gaussiana, obviamente em homenagem a Gauss.

Vamos rapidamente observar que a definição acima realmente descreve uma distribuição de probabilidade, ou seja que a integral dessa densidade é um. Para tanto, vamos usar um truque conhecido, que consiste em retornar ao plano. Obviamente,

$$\begin{aligned} \left(\int \exp\{-x^2/2\} dx \right)^2 &= \int \int \exp\{-(x^2 + y^2)/2\} dx dy \\ &= \int_0^{2\pi} \int_0^\infty \exp\{-r^2/2\} r dr d\theta \stackrel{2s=r^2}{=} 2\pi. \end{aligned} \quad (3.103)$$

Donde a constante em (3.102) está de fato correta.

Exercício 3.8.1. *Mostre que a distribuição $\mathcal{N}(m, \sigma^2)$, tem densidade*

$$\frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-m)^2}{2\sigma^2}\right\}. \quad (3.104)$$

Exercício 3.8.2. Mostre que $Y \sim_d \mathcal{N}(m, \sigma^2)$ tem esperança m e variância σ^2 .

Para confirmar que de fato as distribuições normais se comportam bem com respeito a somas independentes, apresentamos o seguinte resultado.

Proposição 3.8.5. Se $X \sim_d \mathcal{N}(m, \sigma^2)$ e $Y \sim_d \mathcal{N}(\bar{m}, \bar{\sigma}^2)$ são independentes, então $X + Y$ tem distribuição $\mathcal{N}(m + \bar{m}, \sigma^2 + \bar{\sigma}^2)$. Em particular, μ é um ponto fixo do operador Γ definido em (3.99).

Demonstração. O caso em que σ ou $\bar{\sigma}$ se anulam é trivial, portanto vamos considerar que ambas são positivas. Não é difícil ver que podemos também supor que $m = \bar{m} = 0$. Podemos então calcular

$$P[X + Y \leq a] = P[\sigma W + \bar{\sigma} Z \leq a], \quad (3.105)$$

onde W e Z são independentes com distribuição $\mathcal{N}(0, 1)$. Assim, a probabilidade acima pode ser escrita como

$$\mathcal{N}(0, 1) \otimes \mathcal{N}(0, 1) \left(\{ (w, z) \in \mathbb{R}^2; \sigma w + \bar{\sigma} z \leq a \} \right). \quad (3.106)$$

Agora aplicaremos a rotação rígida $A : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ dada por

$$A(w, z) = \frac{1}{\sqrt{\sigma^2 + \bar{\sigma}^2}} (\sigma w + \bar{\sigma} z, \bar{\sigma} w - \sigma z). \quad (3.107)$$

Como sabemos que a densidade f de (W, Z) é invariante por A , ou seja $f \circ A = f$, então podemos escrever (3.106) como

$$\begin{aligned} & \mathcal{N}(0, 1) \otimes \mathcal{N}(0, 1) \left(A \left(\{ (w, z) \in \mathbb{R}^2; \sigma w + \bar{\sigma} z \leq a \} \right) \right) \\ &= \mathcal{N}(0, 1) \otimes \mathcal{N}(0, 1) \left(\left\{ (w, z); \frac{1}{\sqrt{\sigma^2 + \bar{\sigma}^2}} w \leq a \right\} \right) \\ &= \mathcal{N}(0, 1) ((-\infty, a \sqrt{\sigma^2 + \bar{\sigma}^2}]) = \mathcal{N}(0, \sigma^2 + \bar{\sigma}^2) ((-\infty, a]), \end{aligned}$$

terminando a prova da proposição. \square

Podemos obter um corolário interessante sobre a soma de normais i.i.d.

Corolário 3.8.6. Sejam X_1, X_2, \dots variáveis i.i.d. com distribuição $\mathcal{N}(m, \sigma^2)$, então

$$X_1 + \dots + X_n \sim_d \mathcal{N}(nm, n\sigma^2). \quad (3.108)$$

Como consequência

$$\frac{\sum_{i=1}^n X_i - nE(X_1)}{\sigma\sqrt{n}} \sim_d \mathcal{N}(0, 1). \quad (3.109)$$

Lembrando da Lei dos Grandes Números, se dividimos a soma dos $X_i - E(X_i)$ por n , essa fração vai a zero quase certamente. O que concluímos acima é

que ao dividir por \sqrt{n} obtemos um limite não trivial (nem zero, nem infinito) e aleatório (não determinístico).

Mais uma observação curiosa: nossa motivação para a definição da distribuição normal passou por invariância por rotações e podemos estender essa invariância para n normais independentes. Note que somar as coordenadas canônicas é equivalente a tomar o produto escalar com o vetor $(1, 1, \dots, 1)$, que tem norma euclidiana \sqrt{n} .

Uma outra maneira de entender o corolário acima é que a normal é um ponto fixo da operação seguinte

- a) tome uma distribuição $\mu \in \mathcal{L}^2$,
- b) considere X_1, \dots, X_n i.i.d. com distribuição μ e
- c) retorne a distribuição de

$$\frac{X_1 + \dots + X_n - nE(X_1)}{\sqrt{n}}. \quad (3.110)$$

Na Questão 3.8.2, nos perguntamos quais seriam os outros possíveis pontos fixos de Γ e isso será considerado depois. Mas uma outra questão bastante importante é se o ponto fixo $\mathcal{N}(0, 1)$ é atrator, ou seja se começando com outras distribuições poderíamos nos aproximar de $\mathcal{N}(0, 1)$ à medida que iteramos Γ .

Isso é estudado no Teorema Central do Limite (TCL) que provaremos posteriormente. Mas antes, precisamos desenvolver uma boa definição de convergência para distribuições, ou em outras palavras definir uma topologia. Esse será o nosso próximo tópico.

3.8.2 Convergência fraca

Em muitos casos é importante termos bem definida uma noção de convergência de medidas de probabilidade. Supondo por exemplo no espaço mensurável (E, \mathcal{A}) , tenhamos uma sequência de probabilidades μ_n e gostaríamos de saber se ela converge a uma determinada μ .

Um candidato natural para dar sentido a essa convergência poderia ser a distância de variação total entre duas medidas

$$d_{VT}(\mu, \nu) = \sup_{A \in \mathcal{A}} |\mu(A) - \nu(A)|. \quad (3.111)$$

Não é difícil mostrar que a definição acima induz uma métrica, mas ela possui alguns problemas que descreveremos a seguir.

Exercício 3.8.3. *Mostre que d_{VT} define uma métrica.*

Exercício 3.8.4. *Sejam μ e ν absolutamente contínuas com respeito a uma medida fixa η , tendo densidades ρ e π respectivamente. Encontre uma fórmula para $d_{VT}(\mu, \nu)$ em termos das densidades. Essa fórmula nos remete a qual distância entre funções?*

Digamos que o espaço amostral E já seja provido de uma métrica d e \mathcal{A} seja a σ -álgebra dos borelianos em E . Qualquer que seja a noção de convergência que iremos considerar, gostaríamos de dizer que δ_{x_n} converge a δ_x sempre que $x_n \rightarrow x$ em E . Esse porém não é o caso para d_{VT} , pois se $x_n \neq x$ para todo n e $\{x\} \in \mathcal{A}$, teríamos

$$d_{VT}(\delta_{x_n}, \delta_x) \geq |\delta_{x_n}(\{x\}) - \delta_x(\{x\})| = |0 - 1| = 1. \quad (3.112)$$

Aqueles que já viram o conceito de convergência fraca acharão natural que a convergência de μ_n para μ seja definida em termos da convergência das integrais $\int f d\mu_n$ para $\int f d\mu$. Porém, como mencionamos no exemplo das medidas δ_{x_n} acima, gostaríamos também de a convergência respeitasse a topologia original do espaço E , o que torna natural o seguinte conceito.

Definição 3.8.7. Dizemos que uma sequência de medidas de probabilidade μ_n converge fracamente (ou converge em distribuição) para uma probabilidade μ se

$$\lim_{n \rightarrow \infty} \int f d\mu_n = \int f d\mu, \text{ para toda } f : E \rightarrow \mathbb{R} \text{ contínua e limitada.} \quad (3.113)$$

Essa convergência muitas vezes é denotada por $\mu_n \Rightarrow \mu$.

Essa definição fica ainda mais natural para aqueles que conhecem o Teorema da Representação de Riesz. Com isso em mente, podemos relacionar a convergência em distribuição com a convergência fraca- \star no espaço de medidas finitas.

Exercício 3.8.5. Mostre que em $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, temos que $\frac{1}{n} \sum_{i=1}^n \delta_{i/n} \Rightarrow U_{[0,1]}$.

Exercício 3.8.6. Considere a função ϕ do espaço de medidas em $([0, 1], \mathcal{B}([0, 1]))$ nele mesmo, dada por:

$$\phi(\mu)(A) = \frac{1}{2}(\mu(3A) + \mu(3A - 2)). \quad (3.114)$$

Identifique o limite em distribuição de $\phi^{(n)}(\delta_0)$. Mostre que

- a) a função de distribuição acumulada associada ao limite é contínua,
- b) o limite não é absolutamente contínuo com respeito à medida de Lebesgue.

Exercício 3.8.7. Sejam X_1, X_2, \dots i.i.d. distribuídas como $\text{Exp}(1)$ e defina

$$M_n = \max_{i=1, \dots, n} X_i. \quad (3.115)$$

Mostre que $M_n - \log(n)$ converge fracamente e identifique o limite. Observe que não precisamos dividir $M_n - \log(n)$ por nada para obter a convergência.

Nós algumas vezes denotamos $X_n \Rightarrow X$ quando X_n e X são elementos aleatórios de (Ω, \mathcal{F}, P) para descrever a convergência fraca de suas respectivas distribuições. Mais precisamente, $X_n \circ P \Rightarrow X \circ P$.

3.8.3 Convergência fraca em \mathbb{R}

No caso especial em que $E = \mathbb{R}$, temos várias outras maneiras de caracterizar convergência em distribuição. A primeira é dada pela seguinte

Proposição 3.8.8. *Se $\int g d\mu_n$ converge para $\int g d\mu$ para toda $g \in C^3$ limitada e com as três primeiras derivadas limitadas, então $\mu_n \Rightarrow \mu$.*

Demonstração. Primeiramente, vamos ver que podemos nos concentrar em um conjunto compacto da reta.

Para isso fixe um $\varepsilon > 0$ e tome M' tal que $\mu([-M', M']) > 1 - \varepsilon/3$. Tomando uma função g satisfazendo as hipóteses do teorema e tal que

$$\mathbf{1}[-M', M'] \leq g \leq \mathbf{1}[-M' - 1, M' + 1], \quad (3.116)$$

concluimos que

$$\mu_n([-M' - 1, M' + 1]) \geq 1 - \varepsilon/2, \quad (3.117)$$

para todo n suficientemente grande. Se tomamos $M \geq M'$ suficientemente grande, podemos obter a cota acima para todo n (com M no lugar de $M' + 1$ e ε no lugar de $\varepsilon/2$).

Fixamos agora uma $f : \mathbb{R} \rightarrow \mathbb{R}$ contínua e limitada. Sabemos que é possível aproximar f por uma função $g \in C^3$ de suporte compacto, com $\|g\|_\infty \leq 2\|f\|_\infty$ e $|g - f| \leq \varepsilon/M$ uniformemente no intervalo $[-M, M]$. Essa g certamente satisfaz as hipóteses do teorema.

Portanto,

$$\begin{aligned} \left| \int f d\mu_n - \int f d\mu \right| &\leq 2\varepsilon\|f\|_\infty + \left| \int_{-M}^M f d\mu_n - \int_{-M}^M f d\mu \right| \\ &\leq 2\varepsilon\|f\|_\infty + \frac{\varepsilon}{M}2M + \left| \int_{-M}^M g d\mu_n - \int_{-M}^M g d\mu \right| \\ &\leq 2\varepsilon\|f\|_\infty + 2\varepsilon + \left| \int g d\mu_n - \int g d\mu \right|. \end{aligned}$$

Como o último termo converge a zero e ε foi escolhido arbitrariamente, isso conclui a prova da proposição. \square

3.8.4 O TCL para uma sequência i.i.d.

Teorema 3.8.9 (Teorema Central do Limite). *Considere em (Ω, \mathcal{F}, P) , uma sequência X_1, X_2, \dots de variáveis aleatórias i.i.d. em \mathcal{L}^3 . Nesse caso, se definimos $m = E(X_1)$ e $\sigma^2 = \text{Var}(X_1)$, temos*

$$\frac{\sum_{i=1}^n (X_i - m)}{\sigma\sqrt{n}} \Rightarrow \mathcal{N}(0, 1). \quad (3.118)$$

Demonstração. Primeiramente, observe que podemos supor que $m = 0$, pois de qualquer forma iremos subtrair a média da distribuição na qual nos interessamos. Uma outra observação importante é que podemos supor $\sigma = 1$, pois no caso geral de qualquer forma estamos somando X_i/σ no enunciado.

Como vimos na Proposição 3.8.8, basta mostrar a convergência das integrais de funções $g \in C^3$, que possuam todas as três primeiras derivadas limitadas. Considerando a função

$$\phi^n(x_1, \dots, x_n) := g\left(\frac{x_1 + \dots + x_n}{\sqrt{n}}\right), \quad (3.119)$$

nos basta provar a convergência das sequências de números reais

$$\lim_n \int \phi^n(X_1, \dots, X_n) dP = \int g(s) \mathcal{N}(0, 1)(ds). \quad (3.120)$$

Vale lembrar que no Corolário 3.8.6 já estabelecemos algo mais forte para variáveis normais. Mais precisamente, suponha que extendemos nosso espaço de probabilidade para $(\Omega', \mathcal{F}', P')$, onde exista uma sequência Y_1, Y_2, \dots de variáveis aleatórias i.i.d. com distribuição $\mathcal{N}(0, 1)$ independente de X_1, X_2, \dots . Então, para todo $n \geq 1$,

$$\int \phi^n(Y_1, \dots, Y_n) dP' = \int g(s) \mathcal{N}(0, 1)(ds), \quad (3.121)$$

o que tornaria o limite em (3.120) trivial para tais variáveis. A nossa estratégia será aproximar $\phi^n(X_1, \dots, X_n)$ por $\phi(Y_1, \dots, Y_n)$, e faremos isso trocando uma variável de cada vez.

Para entender o que acontece quando trocamos uma das variáveis X_i por Y_i , temos que expandir g em série de potências, isto é, escrever

$$g(s) = g(s_0) + g'(s_0)(s - s_0) + g''(s_0)(s - s_0)^2/2 + r_{s_0}(s - s_0), \quad (3.122)$$

onde $r_{s_0}(h)/h^3$ é limitada por M , uniformemente em h e s_0 em consequência das nossas suposições sobre g .

Denotando $z_i = (y_1, \dots, y_{i-1}, x_i, \dots, x_n)$, $z_i^0 := (y_1, \dots, y_{i-1}, 0, x_{i+1}, \dots, x_n)$ e $s_i^0 = y_1 + \dots + y_{i-1} + x_{i+1} + \dots + x_n$, temos

$$\phi^n(z_i) = \phi^n(z_i^0) + g'\left(\frac{s_i^0}{\sqrt{n}}\right) \frac{x_i}{\sqrt{n}} + g''\left(\frac{s_i^0}{\sqrt{n}}\right) \frac{x_i^2}{2n} + r_{\frac{s_i^0}{\sqrt{n}}}\left(\frac{x_i}{\sqrt{n}}\right), \quad (3.123)$$

Nós propositalmente expandimos ϕ^n até ordem dois, pois X_i e Y_i possuem os mesmos momentos de ordem um ($m = 0$) e dois ($\sigma^2 = 1$).

Integrando os dois lados da igualdade acima com respeito a $Z_i \circ P$ (denotamos como antes, $Z_i = (Y_1, \dots, Y_{i-1}, X_i, \dots, X_n)$ e Z_i^o, S_i^o analogamente), teremos

$$\int \phi^n(Z_i) dP' = \int \phi^n(Z_i^o) dP' + \frac{1}{2n}v_i + k_i, \quad (3.124)$$

onde as quantidades v e k , se escrevem como

$$v_i = \int g''\left(\frac{S_i^o}{\sqrt{n}}\right) dP' \quad \text{e} \quad k_i = \int r_{S_i^o/\sqrt{n}}\left(\frac{X_i}{\sqrt{n}}\right) dP'. \quad (3.125)$$

Note que v_i não depende de X_i e que

$$|k_i| \leq \left| \int \left(\frac{X_i^3}{n^{3/2}}\right) \left(\frac{n^{3/2}}{X_i^3}\right) r_{S_i^o/\sqrt{n}}\left(\frac{X_i}{\sqrt{n}}\right) dP' \right| \leq \frac{M}{n^{3/2}} E(|X_i^3|). \quad (3.126)$$

As observações acima são o ponto mais importante da prova de que essa aproximação funciona e uma outra maneira de colocá-las é a seguinte. Como X_i e Y_i possuem os dois primeiros momentos iguais, os dois primeiros termos de Taylor coincidem após a integração (o primeiro se anula e o segundo é v_i tanto para X_i quanto para Y_i). O resto é de ordem muito pequena para influir no limite.

De fato, se retiramos o termo Y_i de Z_{i+1} , fazendo a mesma expansão que para X_i , obtemos

$$\int \phi^n(Z_{i+1}) dP' = \int \phi^n(Z_i^o) dP' + \frac{1}{2n}v_i + k'_i, \quad (3.127)$$

com o termo de ordem superior k'_i sendo definido exatamente como k_i , mas com Y_i no lugar de X_i .

Estamos prontos agora para a computação final

$$\begin{aligned} & \left| \int \phi^n(X_1, \dots, X_n) dP - \int g(s) \mathcal{N}(0, 1)(ds) \right| \\ &= \left| \int \phi^n(Z_0) dP' - \int \phi^n(Z_n) dP' \right| \\ &\leq \sum_{i=0}^{n-1} \left| \int \phi^n(Z_i) dP' - \int \phi^n(Z_{i+1}) dP' \right| = \sum_{i=0}^{n-1} |k_i - k'_i| \\ &\leq n \frac{M}{n^{3/2}} (E(|X_1|^3) + E(|Y_1|^3)), \end{aligned}$$

que claramente converge a zero, provando o teorema. \square

Corolário 3.8.10. *A $\mathcal{N}(0, 1)$ é a única distribuição μ que possui esperança zero, variância 1 e é tal que se X, Y são i.i.d. com distribuição μ , então $(X + Y)/\sqrt{2}$ também possuem distribuição μ . Em outras palavras, $\mathcal{N}(0, \sigma^2)$, para $\sigma \geq 0$, são os únicos pontos fixos de Γ em \mathcal{L}^3 .*

3.8. O TEOREMA CENTRAL DO LIMITE

Demonstração. Usando a invariância enunciada acima, temos que

$$\frac{X_1 + \cdots + X_{2^k}}{\sqrt{2^k}} \sim_d \mu. \quad (3.128)$$

Mas pelo Teorema central do limite, a distribuição dessa combinação de X_i deve convergir a $\mathcal{N}(0, 1)$, logo temos $\mu = \mathcal{N}(0, 1)$. \square

Vamos terminar essa seção com uma aplicação do teorema acima.

Exercício 3.8.8. Digamos que jogamos 100 moedas honestas e independentes, como foi proposto no início da seção, obtendo finalmente uma variável aleatória $Y \sim_d \text{Bin}(100, 1/2)$. Usando o *Teorema Central do Limite*, estime $P[Y \geq 55]$ usando uma aproximação por uma $\mathcal{N}(0, 1)$. Calcule numericamente o valor real desta probabilidade e compare ambas as estimativas.

Tópico: O Teorema de Portmanteau

O próximo resultado é bastante útil para provar convergência fraca, pois nos fornece uma coleção de equivalências muitas vezes mais fáceis de verificar.

Teorema 3.8.11 (Teorema de Portmanteau). *Sejam $(\mu_n)_{n \geq 1}$ e μ medidas de probabilidade em (E, \mathcal{A}) . São equivalentes:*

- a) $\mu_n \Rightarrow \mu$,
- a') $\int f d\mu_n \rightarrow \int f d\mu$, para toda f uniformemente contínua e limitada,
- b) $\limsup_n \mu_n(F) \leq \mu(F)$, para todo $F \subseteq E$ fechado,
- b') $\liminf_n \mu_n(G) \geq \mu(G)$, para todo $F \subseteq E$ aberto,
- c) $\lim_n \mu_n(A) = \mu(A)$, para todo $A \in \mathcal{A}$ com $\mu(\partial A) = 0$.

Para memorizar o teorema acima, é conveniente lembrar dos dois exemplos:

- i) se $x_n \rightarrow x$ com $x_n \neq x$, $F = \{x\}$ e $G = B(x, \delta) \setminus \{x\}$ temos, para n grande,

$$\mu_n(F) = \mu(G) = 0 < 1 = \mu(F) = \mu_n(G), \quad (3.129)$$

- ii) em $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, seja $\mu_{2n} = \delta_n$ e $\mu_{2n+1} = \mu = \delta_0$. Obviamente μ_n não converge fracamente a μ . Contudo, para todo $A \in \mathcal{B}(\mathbb{R})$,

$$\begin{aligned} \liminf_n \mu_n(A) &\leq \liminf_n \mu_{2n}(A) = \mu(A) \text{ e} \\ \limsup_n \mu_n(A) &\geq \limsup_n \mu_{2n}(A) = \mu(A). \end{aligned} \quad (3.130)$$

Prova do Teorema 3.8.11. Obviamente, $(a \Rightarrow a')$, pois a' somente supõe a convergência das integrais para funções f que sejam uniformemente contínuas, portanto é um requisito mais fraco que a .

Observamos também que $(b \Leftrightarrow b')$. De fato, basta tomarmos complementos e observar a mudança nos sinais das desigualdades.

Então, para a prova do teorema, basta mostrar que $(a' \Rightarrow b)$, $(b + b' \Rightarrow c)$ e $(c \Rightarrow a)$.

Começamos com $(a' \Rightarrow b)$ e para tanto, consideramos $F \subseteq E$ fechado. Seja $\delta > 0$ e defina a função $f_\delta : E \rightarrow \mathbb{R}$ dada por

$$f_\delta(x) = \max \left\{ 1 - \frac{d(x, F)}{\delta}, 0 \right\}. \quad (3.131)$$

Claramente, f é uniformemente contínua e vale $\mathbb{1}_F \leq f_\delta \leq \mathbb{1}_{B(F, \delta)}$. Dessa desigualdade, temos $\limsup_n \mu_n(F) \leq \limsup_n \int f_\delta d\mu_n = \int f_\delta d\mu \leq \mu(B(F, \delta))$. Tomando agora o limite com $\delta \rightarrow 0$, obtemos $b)$ por continuidade da probabilidade μ .

Para mostrar $(b + b' \Rightarrow c)$, seja $A \in \mathcal{A}$ tal que $\mu(\partial A) = 0$. Nesse caso, sabemos que

$$\begin{aligned} \limsup_n \mu_n(A) &\leq \limsup_n \mu_n(\bar{A}) \leq \mu(\bar{A}) = \mu(\mathring{A}) \\ &\leq \liminf_n \mu_n(\mathring{A}) \leq \liminf_n \mu_n(A), \end{aligned}$$

o que mostra o limite em c).

Finalmente, resta mostrar $(c \Rightarrow a)$ e, para tanto, consideramos uma função $f : E \rightarrow \mathbb{R}$ contínua e limitada. Digamos, com $\|f\|_\infty = M$.

Sabemos que os conjuntos $\{f^{-1}(\{a\})\}_{a \in \mathbb{R}}$ são disjuntos, logo os conjuntos $f^{-1}(\{a\})$ podem ter medida μ positiva apenas para uma coleção enumerável de valores $a \in \mathbb{R}$. Obtemos assim uma coleção finita $b_0 < b_1 < \dots < b_k$, tal que

$$\begin{aligned} b_0 < -M \text{ e } b_k > M, \quad b_{i+1} - b_i \leq \delta \text{ e} \\ \mu(f^{-1}(\{b_i\})) = 0 \text{ para todo } i \leq k. \end{aligned} \quad (3.132)$$

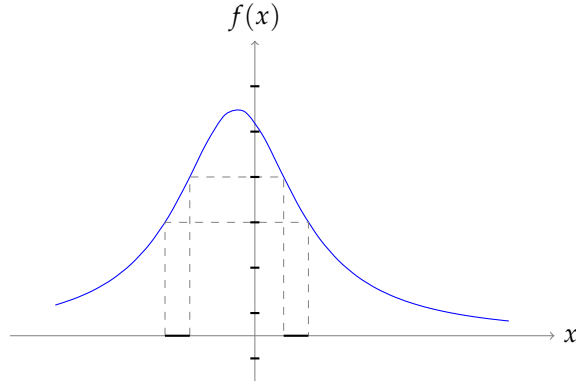


Figura 3.3: Uma função contínua e limitada f , os pontos b_i e um conjunto A_i .

Iremos aproximar f por uma função da forma $f_\delta = \sum_i b_i \mathbb{1}_{A_i}$, onde os conjuntos $A_i = f^{-1}([b_i, b_{i+1}))$ são disjuntos. Obviamente $f_\delta \leq f \leq f_\delta + \delta$, donde

$$\liminf \int f_\delta d\mu_n \leq \liminf \int f d\mu_n \leq \limsup \int f d\mu_n \leq \liminf \int f_\delta d\mu_n + \delta.$$

Mas como $\int f_\delta d\mu_n = \sum_i b_i \mu_n(A_i)$, a prova estará concluída se mostrarmos que $\mu_n(A_i) \rightarrow \mu(A_i)$ para todo $i \leq k$. Isso segue de d), pois $\partial A_i \subseteq f^{-1}(\{b_i, b_{i+1}\})$, que tem medida zero. \square

Exercício 3.8.9. Lembrando que em $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, temos $\frac{1}{n} \sum_{i=1}^n \delta_{i/n} \Rightarrow U_{[0,1]}$, use o ítem d) do Teorema 3.8.11 para dar uma caracterização dos conjuntos Riemann-mensuráveis. Mais precisamente, encontre os $A \subseteq \mathbb{R}$ tais que $\frac{1}{n} \sum_{i=1}^n \delta_{i/n}(A)$ converge para a medida de Lebesgue de A .

Capítulo 4

Esperança condicional

4.1 Esperança condicional

Como já foi dito anteriormente, a estrutura de σ -álgebra tem um papel muito importante em probabilidade. Durante o curso de Teoria da Medida, muitas vezes o conceito de σ -álgebra parece uma técnica que simplesmente dificulta nosso acesso ao conteúdo realmente interessante do curso. Em alguns momentos, chegamos a desejar que tudo fosse mensurável e não tivéssemos que nos preocupar com tais formalidades.

Contudo, no estudo que iniciaremos agora, nos restringiremos a σ -álgebras menores de maneira proposital. Ficará claro em particular, que o estudo de mensurabilidade não é uma mera técnica, mas sim uma ferramenta importante.

Esse interesse, vem da necessidade de representar situações de “informação incompleta”, onde podemos apenas observar uma parte da realidade. Isso certamente é de suma importância em diversas aplicações, desde a estatística, física e computação até a teoria de jogos. Vamos começar com um exemplo simples.

Suponha que $\Omega = \mathbb{R}^2$ é dotado da σ -álgebra de Borel e denotamos por X_1, X_2 as coordenadas canônicas. Como podemos representar matematicamente a afirmação “uma pessoa somente conhece o valor de X_1 e não de X_2 ”? Digamos por exemplo que essa pessoa deverá tomar uma decisão (por exemplo escolher um elemento de E) baseando-se apenas nessa informação incompleta. A maneira que modelamos isso matematicamente é dizendo que a decisão da pessoa deve ser uma função $f : \Omega \rightarrow E$ mensurável com respeito a $\sigma(X_1)$.

Nossa primeira utilização desse conceito será feita agora ao introduzirmos a noção de esperança condicional, que generaliza o conceito de esperança. Relem-

brando o cálculo (3.22), nós podemos pensar em $E(X)$ como uma boa maneira de aproximar X por um número real. Isso por exemplo poderia ser útil se não temos nenhuma informação sobre o que ocorreu, mas ainda sim temos que tentar adivinhar o valor de X . Mas vamos agora imaginar uma outra situação, onde temos um pouco de informação sobre o que ocorreu.

Voltando ao exemplo em que $\Omega = \mathbb{R}^2$, digamos que nós podemos observar o valor de X_1 , mas gostaríamos de estimar o valor de X_2 . De acordo com o que discutimos acima, nossa estimativa agora não precisa mais ser apenas um número real, podendo ser qualquer função mensurável com respeito a $\sigma(X_1)$.

Vamos no que segue tornar esse discussão rigorosa, mas antes lembramos um lema básico de Teoria da Medida.

Lema 4.1.1. *Se f, f' são funções mensuráveis tais que*

$$\int_A f \, d\mu = \int_A f' \, d\mu, \text{ para todo } A \in \mathcal{F}', \quad (4.1)$$

então $f = f'$ μ -quase certamente.

Demonstração. Aplicando a hipótese para $A = [f > f']$, vemos que

$$\int_A f - f' \, d\mu = 0, \quad (4.2)$$

mas no conjunto A acima, o integrando é positivo. Portanto, $f = f'$, μ -quase certamente em A . Aplicando o mesmo raciocínio para $[f < f']$ obtemos que $f = f'$ quase certamente. \square

O lema acima nos diz que se soubermos integrar f em todos os eventos A , então podemos recuperar a função f propriamente dita. O que aconteceria se soubéssemos integrar f apenas para eventos A em uma sub- σ -álgebra? É isso que estudaremos à partir de agora.

Definição 4.1.2. *Seja uma variável aleatória $X \in \mathcal{L}^1(P)$ e uma sub- σ -álgebra $\mathcal{F}' \subseteq \mathcal{F}$. Dizemos que uma variável aleatória Y é a esperança condicional de X com respeito a \mathcal{F}' (ou a esperança condicional de X dada \mathcal{F}') se*

- a) Y é \mathcal{F}' -mensurável e
- b) $E(X\mathbb{1}_A) = E(Y\mathbb{1}_A)$ para todo $A \in \mathcal{F}'$.

Nesse caso, escrevemos

$$Y = E(X|\mathcal{F}'). \quad (4.3)$$

Observe que faz sentido escrever $E(Y|\mathcal{F}')(\omega)$, pois $E(X|\mathcal{F}')$ é uma variável aleatória.

Interpretamos informalmente a definição acima como “ Y é a melhor aproximação \mathcal{F}' -mensurável de X ”. Ou Y é a melhor aproximação que podemos fazer de X se “conhecemos apenas \mathcal{F}' ”.

Exemplo 4.1.1. Se $\mathcal{F}' = \{\emptyset, \Omega\}$, então $Y = E(X)$ (uma variável aleatória constante) é esperança condicional de X dado \mathcal{F}' , pois

- a) Y é \mathcal{F}' -mensurável (por ser constante). Além disso
- b) $E(X\mathbb{1}_\emptyset) = 0 = E(Y\mathbb{1}_\emptyset)$ e $E(X\mathbb{1}_\Omega) = E(X) = E(Y\mathbb{1}_\Omega)$.

Uma propriedade muito importante que segue da Definição 4.1.2 é dada pela seguinte

Proposição 4.1.3. Se Y satisfaz as a) e b) em Definição 4.1.2, então $Y \in \mathcal{L}^1(P)$.

Demonstração. Tomamos $A = [Y \geq 0]$ e $A' = [Y < 0]$ que estão em \mathcal{F}' e estimamos

$$\int |Y| dP = \int_A Y dP + \int_{A'} Y dP = \int_A X dP + \int_{A'} X dP \leq \int |X| dP < \infty \quad (4.4)$$

O que mostra a proposição. \square

Além caso trivial dado acima pelo Exemplo 4.1.1, quando podemos esperar que existam esperanças condicionais?

Teorema 4.1.4. Dada $X \in \mathcal{L}^1(P)$ e $\mathcal{F}' \subseteq \mathcal{F}$ uma σ -álgebra, então existe a esperança condicional $E(X|\mathcal{F}')$. Além disso ela é única P -quase certamente.

Demonstração. Vamos primeiro mostrar a unicidade quase certa. Para isso, supomos que existam Y e Y' satisfazendo as condições da Definição 4.1.2 (logo em \mathcal{L}^1). Iremos proceder como no Lema 4.1.1 acima, definindo $A = [Y > Y']$, donde concluímos que

$$E((Y - Y')\mathbb{1}_A) = E(Y\mathbb{1}_A) - E(Y'\mathbb{1}_A) = 0. \quad (4.5)$$

Mas como $Y > Y'$ em A , vemos que $Y \leq Y'$ quase certamente. A prova da unicidade pode ser completa trocando os papéis de Y e Y' acima.

Vamos agora para a prova da existência. Como $X \in \mathcal{L}^1(P)$, podemos introduzir

$$\mu(A) = E(X\mathbb{1}_A), \quad (4.6)$$

que define uma medida com sinal em (Ω, \mathcal{F}) , com variação total finita.

Caso o leitor não se sinta familiarizado com o conceito de medida com sinal, poderá decompor X em partes positiva e negativa e prosseguir sem problemas.

Um passo importante da prova é observar que μ também define uma medida no espaço (Ω, \mathcal{F}') . Estamos portanto propositalmente restringindo nossa σ -álgebra. Como $P(A) = 0$ implica que $\mu(A) = 0$, temos que $\mu \ll P$ e podemos aplicar o Teorema de Radon-Nikodim para obter uma derivada $Y : \Omega \rightarrow \mathbb{R}$ tal que

- a) Y é \mathcal{F}' -mensurável e
- b) $\mu(A) = \int_A Y dP$.

Agora é só observar que as afirmações acima correspondem às condições da Definição 4.1.2. \square

Observe que a condição de \mathcal{F}' -mensurabilidade é essencial para a unicidade. De fato, X obviamente satisfaz a segunda condição da Definição 4.1.2, mas não necessariamente a primeira.

Exercício 4.1.2. *Mostre que se $X \in \mathcal{F}'$, então $E(X|\mathcal{F}') = X$ quase certamente.*

Exercício 4.1.3. *Seja P a probabilidade uniforme em $\{(x_1, x_2) \in [0, 1]^2; x_1 \geq x_2\}$. Calcule $E(X_2|X_1)$.*

4.2 Propriedades básicas da esperança condicional

Nessa seção justificaremos, em certa medida, a nomenclatura “esperança condicional”. Faremos isso mostrando que ela satisfaz várias propriedades que já conhecemos para a esperança tradicional.

Mas como podemos mostrar propriedades simples tais como a linearidade da esperança condicional? Vamos começar com um exemplo

Proposição 4.2.1. *Se $X, X' \in \mathcal{L}^1(P)$, então*

$$E(X + X'|\mathcal{F}') = E(X|\mathcal{F}') + E(X'|\mathcal{F}'), \text{ } P\text{-quase certamente.} \quad (4.7)$$

Note que a igualdade acima é uma igualdade entre variáveis aleatórias.

Demonstração. Sabemos que $Y = E(X|\mathcal{F}') + E(X'|\mathcal{F}')$ é uma variável aleatória bem definida. Mais do que isso, sabemos que ela é uma candidata muito boa a $E(X + X'|\mathcal{F}')$. Logo, por unicidade da esperança condicional, basta verificar que Y satisfaz as condições da Definição 4.1.2 com respeito a $X + X'$. De fato

- a) Y é \mathcal{F}' -mensurável, por ser uma soma de duas variáveis \mathcal{F}' -mensuráveis e
- b) por linearidade da esperança (não da esperança condicional), temos

$$\begin{aligned} E(Y\mathbb{1}_A) &= E(E(X|\mathcal{F}')\mathbb{1}_A + E(X'|\mathcal{F}')\mathbb{1}_A) \\ &= E(E(X|\mathcal{F}')\mathbb{1}_A) + E(E(X'|\mathcal{F}')\mathbb{1}_A) \\ &= E(X\mathbb{1}_A) + E(X'\mathbb{1}_A) = E((X + X')\mathbb{1}_A). \end{aligned} \quad (4.8)$$

Isso termina a prova da proposição. \square

Exercício 4.2.1. *Dados $X \in \mathcal{L}^1$ e $\alpha \in \mathbb{R}$, mostre que $E(\alpha X|\mathcal{F}') = \alpha E(X|\mathcal{F}')$.*

Uma outra propriedade bem simples da esperança condicional é a monotonicidade.

Lema 4.2.2. *Se $X \geq X'$ em $\mathcal{L}^1(P)$, então*

$$E(X|\mathcal{F}') \geq E(X'|\mathcal{F}'), \text{ } P\text{-quase certamente.} \quad (4.9)$$

Em particular, se $X \geq 0$, então $E(X|\mathcal{F}') \geq 0$ quase certamente.

4.2. PROPRIEDADES BÁSICAS DA ESPERANÇA CONDICIONAL

Demonstração. Seja $A = [E(X'|\mathcal{F}') - E(X|\mathcal{F}') > 0]$, que pertence a \mathcal{F}' . Então

$$0 \leq E((E(X'|\mathcal{F}') - E(X|\mathcal{F}'))\mathbb{1}_A) = E((X' - X)\mathbb{1}_A) \leq 0, \quad (4.10)$$

o que implica que $P(A) = 0$. \square

Proposição 4.2.3. Se $X, ZX \in \mathcal{L}^1(P)$, com $Z \in \mathcal{F}'$, temos

$$E(XZ|\mathcal{F}') = ZE(X|\mathcal{F}') \text{ } P\text{-quase certamente.} \quad (4.11)$$

Em particular, $E(\alpha X|\mathcal{F}') = \alpha E(X|\mathcal{F}')$, para todo $\alpha \in \mathbb{R}$. Uma outra consequência interessante é que $ZE(X|\mathcal{F}')$ estará automaticamente em \mathcal{L}^1 .

De maneira bastante informal, vamos dar uma intuição para o resultado acima. Ao considerarmos a esperança condicional dada \mathcal{F}' , nós já conhecemos as variáveis aleatórias \mathcal{F}' -mensuráveis, portanto elas se comportam como constantes.

Demonstração. Mais uma vez, basta verificar que $ZE(X|\mathcal{F}')$ satisfaz as condições que definem a esperança condicional. A primeira é trivial, pois $ZE(X|\mathcal{F}')$ é \mathcal{F}' -mensurável por ser um produto de funções \mathcal{F}' -mensuráveis.

Para provar a segunda condição, começamos com o caso $Z = \mathbb{1}_B$, implicando que $B \in \mathcal{F}'$, donde

$$E(ZE(X|\mathcal{F}')\mathbb{1}_A) = E(E(X|\mathcal{F}')\mathbb{1}_{A \cap B}) = E(X\mathbb{1}_{A \cap B}) = E(ZX\mathbb{1}_A).$$

Por linearidade, já sabemos que o resultado vale para funções Z simples e gostaríamos de estender para quaisquer Z positivas via Teorema da Convergência Monótona. Um problema aqui é que mesmo que Z seja positiva, não sabemos se $E(X|\mathcal{F}')$ também será positiva.

Portanto, trataremos primeiramente do caso $X \geq 0$. Para tais X , sabemos pelo Lema 4.2.2 que $E(X|\mathcal{F}') \geq 0$ quase certamente. Daí, podemos concluir que $ZE(X|\mathcal{F}') = E(ZX|\mathcal{F}')$ para toda $Z \geq 0$, podemos aproximá-la por baixo por Z_n simples e, pelo Teorema da Convergência Monótona,

$$\begin{aligned} E(ZE(X|\mathcal{F}')) &\stackrel{\text{TCM}}{=} \lim_n E(Z_n E(X|\mathcal{F}')) \\ &= \lim_n E(E(Z_n X|\mathcal{F}')) \stackrel{\text{TCM}}{=} E(E(ZX|\mathcal{F}')). \end{aligned} \quad (4.12)$$

O que mostra o resultado sempre que $X \geq 0$.

Além disso, pela Proposição 4.1.3, sabemos que $ZE(X|\mathcal{F}') \in \mathcal{L}^1$. Podemos finalmente concluir a prova por linearidade decompondo $X = X_+ - X_-$. \square

O próximo resultado tenta corroborar nossa afirmação que a esperança condicional é uma boa maneira de aproximar uma variável aleatória.

Lema 4.2.4. Se $X \in \mathcal{L}^2(P)$ e $\mathcal{F}' \subseteq \mathcal{F}$, então $E(X|\mathcal{F}')$ é a projeção ortogonal de X no espaço vetorial $H_{\mathcal{F}'}$. Onde $H_{\mathcal{F}'} = \{Y \in \mathcal{L}^2; Y \text{ é } \mathcal{F}'\text{-mensurável}\}$.

Demonstração. Temos que verificar que $X - E(X|\mathcal{F}')$ é ortogonal a $H_{\mathcal{F}'}$. Ou seja, mostrar que para todo $Z \in H_{\mathcal{F}'}$, temos

$$E(XZ - E(X|\mathcal{F}')Z) = 0. \quad (4.13)$$

Note que não é claro que essa esperança faz sentido, pois não sabemos que $ZE(X|\mathcal{F}') \in \mathcal{L}^1$. Mas isso segue facilmente da Proposição 4.2.3.

Mas $E(E(X|\mathcal{F}')Z) = ZE(E(X|\mathcal{F}')\mathbb{1}_\Omega) = ZE(X\mathbb{1}_\Omega)$, provando o resultado. \square

Vimos acima uma metodologia que se repete frequentemente. Digamos que queremos provar que uma determinada expressão nos dá a esperança condicional de algo. Podemos começar provando esse resultado para funções indicadoras, depois para funções simples usando a linearidade provada acima.

Porém ainda falta um ingrediente bastante importante para construir ou verificar que determinadas variáveis são esperanças condicionais.

Teorema 4.2.5 (Convergência Monótona para Esperanças Condicionais). *Se as variáveis X_n satisfazem $X_n \uparrow X$ e estão todas em $\mathcal{L}^1(P)$, então*

$$\lim_n E(X_n|\mathcal{F}') = E(X|\mathcal{F}'). \quad (4.14)$$

Demonstração do Teorema 4.2.5. Sabemos que $E(X_{n+1}|\mathcal{F}') \geq E(X_n|\mathcal{F}')$, donde concluímos que $E(X_n|\mathcal{F}') \uparrow Y$. Vamos demonstrar que $Y = E(X|\mathcal{F}')$.

a) Por ser um limite de funções \mathcal{F}' mensuráveis, Y é \mathcal{F}' -mensurável.

b) Dado $A \in \mathcal{F}'$, temos

$$\begin{aligned} E(Y\mathbb{1}_A) &= E(\lim_n E(X_n|\mathcal{F}')\mathbb{1}_A) \stackrel{\text{TCM}}{=} \lim_n E(E(X_n|\mathcal{F}')\mathbb{1}_A) \\ &= \lim_n E(X_n\mathbb{1}_A) \stackrel{\text{TCM}}{=} E(X\mathbb{1}_A). \end{aligned} \quad (4.15)$$

O que termina a prova do teorema. \square

No que segue, muitas vezes escreveremos $E(X|Z)$ para representar a esperança condicional $E(X|\sigma(Z))$.

Exercício 4.2.2. *Sejam X_1 e X_2 as coordenadas canônicas em $\mathbb{R} \times E$ e definimos a probabilidade $dP = \rho(x, y) d\mu_1 d\mu_2$, onde $\rho : \mathbb{R} \times E \rightarrow \mathbb{R}_+$ é uma densidade. Dê sentido à expressão abaixo e mostre que ela é $E(X_1|X_2)$:*

$$\frac{\int x\rho(x, X_2)\mu_1(dx)}{\int \rho(x, X_2)\mu_1(dx)}. \quad (4.16)$$

Exercício 4.2.3. *Seja E enumerável com uma σ -álgebra \mathcal{F}' . Mostre que*

$$\mathcal{F}' = \sigma(A_i, i \geq 1), \text{ com } A_i \subseteq E \text{ disjuntos.} \quad (4.17)$$

4.2. PROPRIEDADES BÁSICAS DA ESPERANÇA CONDICIONAL

Suponha que todos conjuntos A_i tem probabilidade positiva e mostre que

$$E(X|\mathcal{F}') = \sum_i E^i(X)\mathbb{1}_{A_i}, \quad (4.18)$$

onde E^i é a esperança com respeito à probabilidade $P(\cdot|A_i)$. Em breve estenderemos esse tipo de resultado a espaços quaisquer.

Uma outra propriedade que a esperança condicional herda da integral é a

Proposição 4.2.6 (Desigualdade de Jensen). *Se $\phi : \mathbb{R} \rightarrow \mathbb{R}$ é convexa, $X, \phi(X) \in \mathcal{L}^1(P)$, então*

$$\phi(E(X|\mathcal{F}')) \leq E(\phi(X)|\mathcal{F}'). \quad (4.19)$$

Demonstração. Se ϕ for uma função linear, o resultado segue da linearidade que já provamos para a esperança condicional. Além disso, se temos uma função $\psi : \mathbb{R} \rightarrow \mathbb{R}$ linear e tal que $\psi(x) \leq \phi(x)$ para todo $x \in \mathbb{R}$, então

$$E(\phi(X)|\mathcal{F}') \geq E(\psi(X)|\mathcal{F}') = \psi(E(X|\mathcal{F}')). \quad (4.20)$$

Tomamos finalmente o supremo em todas as ψ lineares com $\psi \leq \phi$ dos dois lados da desigualdade acima, obtendo

$$E(\phi(X)|\mathcal{F}') \geq \sup_{\substack{\psi \leq \phi \\ \psi \text{ linear}}} \psi(E(X|\mathcal{F}')) = \phi(E(X|\mathcal{F}')), \quad (4.21)$$

terminando a prova da proposição. \square

Corolário 4.2.7. *Se $X \in \mathcal{L}^1(P)$, então $|E(X|\mathcal{F}')| \leq E(|X||\mathcal{F}')$.*

Uma outra propriedade interessante da esperança condicional diz respeito a sua relação com independência.

Proposição 4.2.8. *Se $X \in \mathcal{L}^1(P)$ é independente de \mathcal{F}' , então*

$$E(X|\mathcal{F}') = E(X) \text{ P-quase certamente.} \quad (4.22)$$

Demonstração. Funções constantes são sempre mensuráveis. Além disso, se $A \in \mathcal{F}'$, então

$$E(X\mathbb{1}_A) = E(X)P(A) = E(E(X)\mathbb{1}_A), \quad (4.23)$$

concluindo a prova. \square

Terminamos essa seção com o que chamamos da propriedade de torre da esperança condicional.

Proposição 4.2.9. *Se $\mathcal{F}' \subseteq \mathcal{F}''$ são ambas sub- σ -álgebras de \mathcal{F} , então para $X \in \mathcal{L}^1(P)$, temos*

$$E(E(X|\mathcal{F}')|\mathcal{F}'') = E(X|\mathcal{F}') = E(E(X|\mathcal{F}'')|\mathcal{F}'), \quad (4.24)$$

ou em outras palavras, independentemente da ordem, prevalece a condição na menor σ -álgebra. Consequentemente, $E(E(X|\mathcal{F}')) = E(X)$.

Demonstração. Como $E(X|\mathcal{F}')$ é \mathcal{F}'' -mensurável, a Proposição 4.2.3, aplicada com $X = 1$, mostra a primeira igualdade em (4.24).

Falta mostrar que $E(E(X|\mathcal{F}'')|\mathcal{F}')$ é a esperança condicional de X dada \mathcal{F}' . Obviamente ela é \mathcal{F}' -mensurável, e nos resta verificar a segunda condição. Mas para todo $A \in \mathcal{F}'$, lembrando que A também pertence a \mathcal{F}'' e usando a definição de esperança condicional duas vezes,

$$E\left(E(E(X|\mathcal{F}'')|\mathcal{F}')\mathbb{1}_A\right) = E(E(X|\mathcal{F}'')\mathbb{1}_A) = E(X\mathbb{1}_A). \quad (4.25)$$

O que termina a prova da proposição. \square

Lema 4.2.10. *Se $X : \Omega \rightarrow E$ é um elemento aleatório e $f : \Omega \rightarrow \mathbb{R}$ é $\sigma(X)$ -mensurável, então existe uma $g : E \rightarrow \mathbb{R}$ mensurável tal que $f = g \circ X$.*

Demonstração. Como de costume, consideramos primeiramente o caso $f = \mathbb{1}_A$. Claramente A tem que pertencer a $\sigma(X)$, ou seja $A = X^{-1}(B)$ para algum $B \in \mathcal{A}$. Neste caso colocamos $g = \mathbb{1}_B$, donde obtemos $f(\omega) = 1 \Leftrightarrow \omega \in A \Leftrightarrow X(\omega) \in B \Leftrightarrow g \circ X = 1$.

No caso em que f é simples, temos $f = \sum_i a_i (g_i \circ X) = (\sum_i a_i g_i) \circ X$. Se f é positiva, então ela é um limite crescente de funções do tipo $g_n \circ X$, além disso podemos tomar g_n crescentes, pois

$$f_{n+1} = f_{n+1} \vee f_n = (g_{n+1} \circ X) \vee (g_n \circ X) = (g_n \vee g_{n+1}) \circ X. \quad (4.26)$$

Finalmente usamos a linearidade da composição novamente para resolver o caso geral $f = f_+ - f_-$. \square

Se $X : \Omega \rightarrow E$ é elemento aleatório, então $E(Y|\sigma(X))$ é obviamente $\sigma(X)$ -mensurável. Pelo lema anterior, $E(Y|\sigma(X)) = g \circ X$ para alguma $g : E \rightarrow \mathbb{R}$. Nesse caso denotamos

$$E(Y|X = x) = g(x). \quad (4.27)$$

Exercício 4.2.4. *Mostre que g é única $X \circ P$ -quase certamente.*

Gostaríamos de dizer que $E(Y|X = x)$ satisfaz alguma propriedade que justifique essa notação. Apesar de que apenas na próxima seção poderemos justificar completamente essa nomenclatura, nesse momento já podemos mostrar a seguinte relação

$$E(Y) = E(E(Y|X)) = E(E(Y|X = x) \circ X) = \int E(Y|X = x)(X \circ P)(dx).$$

Em outras palavras, para integrar Y , basta conhecermos a distribuição de X e a esperança condicional de Y , dado que $X = x$.

Exercício 4.2.5. *Sejam X e Y as coordenadas canônicas em $E_1 \times E_2$, com a probabilidade $P = \mu_1 \otimes \mu_2$ e seja $f : E_1 \times E_2 \rightarrow \mathbb{R}$ em $\mathcal{L}^1(P)$. Mostre que*

$$E(f|X = x) = \int f(x, y) \mu_2(dy). \quad (4.28)$$

4.2. PROPRIEDADES BÁSICAS DA ESPERANÇA CONDICIONAL

Exercício 4.2.6. Se K é um núcleo de transição entre E_1 e \mathbb{R} e P_1 é uma probabilidade em E_1 , mostre que em $P_1 \star K$ temos

$$E(X_2|X_1 = x_1) = \int x_2 K(x_1, dx_2). \quad (4.29)$$

Um outro resultado bastante importante é o seguinte

Teorema 4.2.11 (Teorema da Convergência Dominada para Esperanças Condicionais). Se $X_n \rightarrow X$ e existe $Y \in \mathcal{L}^1(P)$ tal que $|X_n| \leq Y$ para todo n , então

$$E(X_n|\mathcal{F}) \rightarrow E(X|\mathcal{F}) \text{ } P\text{-quase certamente.} \quad (4.30)$$

Demonstração. Seja $Z_n = \sup_{k \geq n} |X_k - X|$ o erro máximo à partir de n . Claramente, $Z_n \downarrow 0$ quase certamente e além disso

$$|Z_n| \leq \sup_{k \geq 1} |X_k| + |X| \leq 2Y, \quad (4.31)$$

donde $E(Z_n) \rightarrow E(0) = 0$, quase certamente pelo Teorema da Convergência Dominada.

Obviamente $E(Z_n|\mathcal{F})$ é uma sequência positiva e não-crescente, logo decresce quase certamente para algum Z . Daí,

$$|E(X_n|\mathcal{F}) - E(X|\mathcal{F})| \leq E(Z_n|\mathcal{F}) \downarrow Z \geq 0. \quad (4.32)$$

Mas $E(Z) \leq E(E(Z_n|\mathcal{F})) = E(Z_n)$. Como $E(Z_n)$ vai a zero pelo Teorema da Convergência Dominada, temos que $Z = 0$ quase certamente como gostaríamos. \square

Exercício 4.2.7. Sejam Z_1, Z_2, \dots variáveis aleatórias i.i.d. em $\mathcal{L}^1(P)$ com $E(Z_1) = 0$.

a) Defina $X_0 = 0$ e

$$X_n = \sum_{i=1}^n Z_i, \text{ para } n \geq 1. \quad (4.33)$$

Mostre que $E(X_{n+1}|Z_1, \dots, Z_n) = X_n$.

b) Supondo agora que $Z_1 \in \mathcal{L}^2(P)$ e $E(Z) = 0$, defina $Y_0 = 0$ e

$$Y_n = \left(\sum_{i=1}^n Z_i \right)^2 - nE(Z_1^2) \quad (4.34)$$

Mostre que $E(Y_{n+1}|Z_1, \dots, Z_n) = Y_n$.

4.3 Probabilidade Condicional Regular

Já sabemos definir por exemplo $E(\mathbb{1}_A|X = x)$. Gostaríamos porém de garantir que essa expressão definisse uma probabilidade em A , e chamaríamos essa probabilidade de $P(A|X = x)$. Mas certamente gostaríamos que $P(\cdot|X = x)$ fosse uma função σ -aditiva. Essa especulação parece promissora, por exemplo se A e B são disjuntos,

$$P(A \cup B|\mathcal{F}') = E(\mathbb{1}_{A \cup B}|\mathcal{F}') = E(\mathbb{1}_A|\mathcal{F}') + E(\mathbb{1}_B|\mathcal{F}') = P(A|\mathcal{F}') + P(B|\mathcal{F}').$$

Ótimo, mas ainda temos o seguinte problema.

Lembramos que a equação acima está bem definida apenas quase certamente. Poderíamos portanto garantir que para uma classe enumerável de conjuntos $A \in \mathcal{F}$, essa aditividade fosse satisfeita. Porém, a σ -álgebra \mathcal{F} é frequentemente não enumerável, portanto não conseguimos a σ -aditividade plena. Isso pode ser contornado se o espaço for canônico, como afirma o nosso próximo resultado.

Ele nos ajudará bastante ao fazermos cálculos usando condicionais, de maneira semelhante à Lei da Probabilidade Total. Esse é o conteúdo do seguinte resultado.

Teorema 4.3.1 (Teorema da Desintegração). *Sejam espaços mensuráveis (Ω, \mathcal{F}) e (E, \mathcal{A}) , com E canônico. Se P é uma probabilidade no espaço produto $(\Omega \times E, \mathcal{F} \otimes \mathcal{A})$ e denotamos por $P_\Omega = P \circ X_1$ a primeira distribuição marginal de P , então existe um núcleo de transição $K : \Omega \times \mathcal{A} \rightarrow [0, 1]$ satisfazendo*

$$P = P_\Omega \star K, \quad (4.35)$$

Em particular,

$$P(A \times B) = \int_A K(\omega, B) P_\Omega(d\omega) \text{ para todo } A \in \mathcal{F}, B \in \mathcal{A}. \quad (4.36)$$

Nesse caso denotamos $K(\omega, B)$ por $P[X_2 \in B|X_1 = \omega]$ (como de costume X_i denota a i -ésima coordenada canônica).

Demonstração. Como de costume, basta resolver o caso $(E, \mathcal{A}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$. De fato, se assumimos a validade do teorema para a reta, podemos usar a função bi-mensurável $\phi : E \rightarrow \mathbb{R}$ para concluir o caso geral.

Nos restringiremos agora ao espaço $(\Omega \times \mathbb{R}, \mathcal{F} \otimes \mathcal{B}(\mathbb{R}), P)$. Para cada $q \in \mathbb{Q}$, definimos $P_\Omega^q : \mathcal{F} \rightarrow [0, 1]$ por

$$P_\Omega^q(A) = P((-\infty, q] \times A). \quad (4.37)$$

Observando que P_Ω^q é absolutamente contínua com respeito a P_Ω , podemos definir

$$F(\omega, q) = \frac{dP_\Omega^q}{dP_\Omega}(\omega). \quad (4.38)$$

Observamos as seguintes propriedades de F :

- a) para cada $q \in \mathbb{Q}$, $F(\cdot, q) \in [0, 1]$, P_Ω -quase certamente, pois $P_\Omega^q(A) \leq P_\Omega(A)$ para todo $A \in \mathcal{F}$,
- b) para $q < q' \in \mathbb{Q}$, $F(\cdot, q) \leq F(\cdot, q')$, P_Ω -quase certamente, pois $P_\Omega^q(A) \leq P_\Omega^{q'}(A)$ para todo $A \in \mathcal{F}$ e
- c) $F(\cdot, n) \rightarrow 1$ (analogamente $F(\cdot, -n) \rightarrow 0$) quando n tende a infinito, P_Ω -quase certamente. Para ver isso, note que a sequência de variáveis aleatórias $F(\cdot, n)$ é quase certamente monótona não decrescente, logo converge P_Ω -quase certamente. Sendo limitada, converge em \mathcal{L}^1 e como sua integral em P_Ω converge para um, $F(\cdot, n) \rightarrow 1$, quase certamente (analogamente para $F(\cdot, -n)$).

Existe pois um conjunto $\Omega' \in \mathcal{F}$ com $P_\Omega(\Omega') = 1$ no qual as três hipóteses acima são satisfeitas. Definimos $\hat{F}(\omega, q)$ como sendo igual a $F(\omega, q)$ em Ω' e igual a $F_0(q)$ (uma função de distribuição fixa) caso contrário (que claramente será mensurável). Finalmente podemos definir $\tilde{F}(\omega, x) = \inf_{q \in \mathbb{Q}; q \downarrow x} \hat{F}(\omega, q)$, que satisfaz para todo ω as hipóteses do Teorema 2.3.4. Logo, existe para cada $\omega \in \Omega$ uma medida $K(\omega, \cdot)$ em $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ satisfazendo $K(\omega, (-\infty, q]) = F(\omega, q)$ P_Ω -quase certamente.

Precisamos mostrar que K é um núcleo, e para isso basta observar que $F(\omega, q)$ são mensuráveis e a família $\{(-\infty, q]; q \in \mathbb{Q}\}$ forma um π -sistema que gera $\mathcal{B}(\mathbb{R})$.

Finalmente, vamos verificar (4.36), notando que se $A \in \mathcal{F}$ e $B = (-\infty, q]$,

$$\int_A K(\omega, B) P_\Omega(d\omega) = \int_A F(\omega, q) P_\Omega(d\omega) = P_\Omega^q(A) = P(A \times B). \quad (4.39)$$

Como a classe B é um π -sistema gerando $\mathcal{B}(\mathbb{R})$ terminamos a prova. \square

Interpretamos $P[X_2 \in B | X_1 = \omega]$ da seguinte forma. Se alguém tiver acesso à σ -álgebra $\sigma(X_1)$, ou seja, essa pessoa é capaz de observar o valor de ω , ela pode não saber o valor de X_2 , mas já pode atualizar sua distribuição para $P(X_2 \in \cdot | X_1 = \omega)$.

Uma das grandes vantagens de ter um núcleo de transição a determinar uma distribuição conjunta, como foi feito acima, é que podemos usar a versão generalizada de Fubini. Antes, nós somente podíamos usar Fubini para espaços construídos através de um núcleo.

Exercício 4.3.1. Se $\Omega = E_1 \times E_2$ com E_2 canônico é dotado da probabilidade $dP = \rho(x_1, x_2) \mu_1 \otimes \mu_2(dx_1 dx_2)$, mostre que

$$P(X_2 \in A | X_1 = x_1) = \frac{\int_A \rho(x_1, x_2) \mu_2(dx_2)}{\int \rho(x_1, x_2) \mu_2(dx_2)}, \quad (4.40)$$

$(X_1 \circ P)$ -quase certamente.

Exercício 4.3.2. Sejam X_1 e X_2 as projeções canônicas em um espaço produto $\Omega \times E$, com E canônico. Então, se X_1 e X_2 são independentes com respeito a P , vale

$$P[X_2 \in B | X_1 = \omega] = P[X_2 \in B] \text{ para } (X_1 \circ P)\text{-quase todo } \omega. \quad (4.41)$$

Exercício 4.3.3. Considere em $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$ as projeções canônicas X_1 e X_2 . Calcule, em cada um dos exemplos abaixo, a probabilidade condicional regular $P[X_1 \in \cdot | X_2 = x_2]$, justificando sua resposta,

- a) Quando P é a medida uniforme em $T = \{(x, y) \in [0, 1]^2; x \leq y\}$ (ou seja, a medida de Lebesgue em \mathbb{R}^2 restrita a T e normalizada para ser uma probabilidade).
- b) Quando P é a medida U_{S^1} (uniforme em S^1).

4.4 Princípio da substituição

O Teorema 4.3.1 é bastante poderoso e nos permite definir e calcular diversas probabilidades, como faremos à seguir. Nessa seção construiremos nossa última versão de probabilidade condicional regular que não se restringe a espaços produtos e nos fornecerá o que chamamos de Princípio da Substituição.

Teorema 4.4.1. Sejam (Ω, \mathcal{F}, P) e (E, \mathcal{A}) espaços mensuráveis canônicos. Considere também $X : \Omega \rightarrow E$ um elemento aleatório, então existe um núcleo de transição K de E a Ω tal que

$$K(X(\omega), F) = E[\mathbb{1}_F | X], \text{ para todo } F \in \mathcal{F}. \quad (4.42)$$

Também denotamos esse núcleo como $K(x, F) = P[F | X = x]$, que é único no sentido que se K' também satisfaz (4.42), então $K(x, F) = K'(x, F)$ para $(X \circ P)$ -quase todo $x \in E$.

Além disso vale o que chamamos de Princípio da Substituição:

$$K(x, [X = x]) = 1, \quad X \circ P\text{-quase certamente}. \quad (4.43)$$

Que pode ser dito de maneira estranha: $P[X = x | X = x] = 1$, quase certamente.

Demonstração. Defina o elemento aleatório $W : \Omega \rightarrow E \times \Omega$, dado por $W(\omega) = (X(\omega), \omega)$, que percorre o gráfico de X (representado horizontalmente). Observe que a medida $P_W := W \circ P$ possui marginais $(X_1 \circ P_W) = (X \circ P)$ e $(X_2 \circ P_W) = P$. Como P_W satisfaz as condições do Teorema 4.3.1, existe um núcleo $K : E \times \mathcal{F} \rightarrow [0, 1]$ tal que para todo $A \in \mathcal{A}, F \in \mathcal{F}$,

$$P_W(A \times F) = \int_A K(x, F) P_X(dx). \quad (4.44)$$

Fixado $F \in \mathcal{F}$, $K(X(\omega), F)$ é obviamente $\sigma(X)$ mensurável, por ser uma composição de uma função mensurável em E com X . Logo, para provar (4.42),

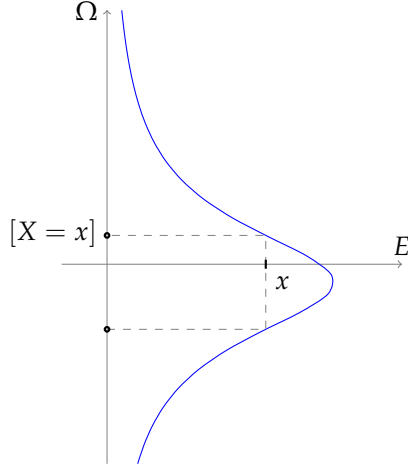


Figura 4.1: O gráfico do elemento aleatório X representado horizontalmente. Os pontos marcados no eixo vertical representam o conjunto $[X = x]$ que possui medida um segundo $P[\cdot | X = x]$ de acordo com o Teorema 4.4.1

basta mostrar a segunda propriedade de esperanças condicionais. Se $B \in \sigma(X)$, podemos escrever $B = [X \in A]$ para algum $A \in \mathcal{A}$, donde

$$\begin{aligned} E[K(X, F)\mathbb{1}_B] &= E[K(X, F)\mathbb{1}_{[X \in A]}] = \int_A K(x, F)P_X(dx) \\ &= P_W(A \times F) = E[\mathbb{1}_{X \in A}\mathbb{1}_F] = E[\mathbb{1}_B\mathbb{1}_F], \end{aligned} \quad (4.45)$$

concluindo a prova de (4.42).

Para mostrarmos o Princípio da Substituição, vamos usar o seguinte lema.

Lema 4.4.2. *Se $X : \Omega \rightarrow E$ é um elemento aleatório tomando valores em um espaço E canônico, então seu gráfico $G = \{(\omega, X(\omega)); \omega \in \Omega\}$ é mensurável na σ -álgebra produto $\mathcal{F} \otimes \mathcal{A}$.*

Demonstração. Primeiramente, consideramos o caso $(E, \mathcal{A}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Neste caso, vemos que

$$G = \bigcap_{n \geq 1} \bigcup_{j \in \mathbb{Z}} [X \in (j/2^n, (j+1)/2^n]] \times (j/2^n, (j+1)/2^n], \quad (4.46)$$

que é mensurável.

Caso E seja outro espaço canônico qualquer, existe $\phi : E \rightarrow B \in \mathcal{B}(\mathbb{R})$ bi-mensurável e $G = \Phi^{-1}(G_{\phi \circ X})$, onde $G_{\phi \circ X}$ é o gráfico de $\phi \circ X$ e $\Phi(\omega, x) = (\omega, \phi(x))$. Logo G também é mensurável nesse caso. \square

Retornando à prova de (4.43), já sabemos que $G' = \{(X(\omega), \omega); \omega \in \Omega\}$ é mensurável. Além disso, por definição $P_W(G') = P[(X(\omega), \omega) \in G'] = P(\Omega) = 1$, ou seja a medida P_W tem suporte em G' .

Logo podemos escrever

$$\begin{aligned} 1 = P_W(G') &= \int \int \mathbb{1}_{G'}(x, \omega) K(x, d\omega) (X \circ P)(dx) \\ &= \int K(x, [X = x]) (X \circ P)(dx). \end{aligned} \quad (4.47)$$

Mas como o integrando acima pertence a $[0, 1]$, essa integral só pode ser um se $K(x, [X = x]) = 1$, $(X \circ P)$ -quase certamente, como desejado. \square

Exercício 4.4.1. Sejam $X : \Omega \rightarrow E$ e $Y : \Omega \rightarrow E'$ elementos aleatórios com E canônico. Então existe um núcleo de transição K entre E e E' tal que

$$K(X(\omega), B) = E[\mathbb{1}_{Y \in B} | X], \text{ para todo } B \in \mathcal{A}'. \quad (4.48)$$

Poderíamos chamar esse núcleo de $K(x, B) = P[Y \in B | X = x]$.

Exercício 4.4.2. Mostre que se $K(x, F) = P[F | X = x]$, então

$$\int f(\omega') K(X(\omega), d\omega') = E(f | X)(\omega), \text{ para toda } f \in \mathcal{F}. \quad (4.49)$$

Exercício 4.4.3. Se Y é variável aleatória e $X : \Omega \rightarrow E$ é um elemento aleatório canônico, mostre que

$$E(Y | X) = \int y P(Y \in dy | X = \cdot) \circ X, \text{ P-q.c.} \quad (4.50)$$

Vamos agora mostrar uma aplicação do que foi feito acima, tentando justificar o nome Princípio da Substituição.

Lema 4.4.3. Se X, Y são variáveis aleatórias independentes, então a função de distribuição acumulada F de $X + Y$ é dada por

$$F(z) = P[X + Y \leq z] = \int_{-\infty}^{\infty} F_Y(z - x) (X \circ P)(dx), \quad (4.51)$$

onde $F_Y(y) = P[Y \leq y]$.

Esse lema pode ser visto como uma generalização do Exercício 2.5.15 para o caso não absolutamente contínuo. Vale a pena tentar diferenciar (não rigorosamente) a equação acima em z .

Demonstração. Vamos calcular

$$\begin{aligned} P[X + Y \leq z] &= E(E(\mathbb{1}_{[X+Y \leq z]} | X)) \\ &= E(E(\mathbb{1}_{[X+Y \leq z]} | X)) \\ &= E(P[X + Y \leq z | X = \cdot] \circ X) \\ &= E(P[X + Y \leq z, X = x | X = \cdot] \circ X) \\ &= E(P[Y \leq z - x | X = \cdot] \circ X), \end{aligned} \quad (4.52)$$

onde $P[Y + X \leq z | X = \cdot]$ representa a função $x \mapsto P[Y + X \leq z | X = x]$.

Agora vamos usar a hipótese que X e Y são independentes. Isso equivale a dizer que a distribuição conjunta desse par é igual a $P_X \otimes P_Y$ e pela unicidade da probabilidade condicional regular temos que $P[Y \in F | X = x] = P[Y \in F]$, $(X \circ P)$ -quase certamente, veja Exercício 4.3.2. Portanto,

$$P[X + Y \leq z] = E(P[Y \leq z - \cdot] \circ X) = \int_{-\infty}^{\infty} F_Y(z - x)(X \circ P)(dx), \quad (4.53)$$

terminando a prova do lema. \square

Exercício 4.4.4. Considere as medidas

$$\mu_a = \frac{\delta_{-1} + \delta_1}{2}, \quad e \quad \mu_b = \mathcal{N}(0, 1). \quad (4.54)$$

e $K : \mathbb{R} \times \mathcal{B}(\mathbb{R}) \rightarrow [0, 1]$ dada por

$$K(x, A) = \begin{cases} \mu_a(A - x), & \text{se } x < 0, \\ \mu_b(A - x), & \text{se } x \geq 0, \end{cases} \quad (4.55)$$

Mostre que

- a) K define um núcleo de transição entre \mathbb{R} em \mathbb{R} .
- b) Se X_1, X_2, \dots for uma cadeia de Markov em \mathbb{R} com núcleo de transição K , então calcule

- i) $E(X_i)$, para todo $i \geq 1$ e
- ii) $\text{Var}(X_i)$, para todo $i \geq 1$.
- iii) Mostre que

$$\frac{\sum_{i=1}^n X_i}{\sqrt{n}} \Rightarrow \mathcal{N}(0, 1). \quad (4.56)$$

Tópico: Processos de Poisson em \mathbb{R}

Nessa seção aplicaremos o conceito de Probabilidade Condicional Regular e do Princípio da Substituição para estudarmos um importante processo de chegadas chamado Processo de Poisson.

O Tenente Boavista está encarregado de vigiar o Sargento Pimenta, que frequentemente dorme durante sua vigília. Para isso, Boavista tem que decidir os momentos $t_1, t_2, \dots \in \mathbb{R}$ que ele irá verificar se Pimenta está cochilando. Uma primeira estratégia poderia ser tomar intervalos igualmente espaçados, $t_1 = 1, \dots, t_k = k$, mas o Sargento certamente iria dormir nos intervalos $(k + \varepsilon, k + 1 - \varepsilon)$ sem se preocupar.

Dado esse problema, o Tenente decide escolher tempos aleatórios T_1, T_2, \dots . Mas é importante lembrar que não são todas as distribuições que funcionarão bem, por exemplo se $T_k - T_{k-1} \geq a$ quase certamente o Sargento irá se aproveitar desse intervalinho.

A primeira simplificação que o Tenente imagina para esse problema é a seguinte: dado que houve uma vistoria no instante t_k , então o que acontecerá à partir daí será o mesmo processo com o qual ele começou. Isso pode ser traduzido de maneira rigorosa como

$$P[(T_{k+1} - t_k, T_{k+2} - t_k, \dots) \in A | T_k = t_k] = P[(T_1, T_2, \dots) \in A], \quad (4.57)$$

$T_k \circ P$ -quase certamente. Não iremos entrar muito em detalhes sobre qual é essa esperança condicional, pois no momento ainda estamos trabalhando heurísticamente, mas já podemos dizer que:

$$\begin{aligned} P[T_1 \in A_1, T_2 - T_1 \in A_2] &= E[\mathbb{1}_{T_1 \in A_1} P[T_2 - T_1 \in A_2 | T_1 = t_1] \circ T_1] \\ &\stackrel{(4.57)}{=} E[\mathbb{1}_{T_1 \in A_1} P[T_1 \in A_2]] = P[T_1 \in A_1]P[T_1 \in A_2]. \end{aligned} \quad (4.58)$$

Procedendo de maneira análoga, podemos concluir que $(T_1, T_2 - T_1, T_3 - T_2, \dots)$ são uma coleção i.i.d.. Agora o Tenente Boavista somente precisa escolher a distribuição de T_1 .

Para essa escolha, ele sabe que se ele não chegar em tempo t , então o Sargento Pimenta sabe que sua próxima chegada terá distribuição $P[T_1 - t \in A | T_1 > t]$. Como o Tenente Boavista gostaria que essa informação fosse inútil para o Sargento Pimenta, ele escolherá

$$P[T_1 - t \in A | T_1 > t] = P[T_1 \in A]. \quad (4.59)$$

E sabemos que as distribuições $\text{Exp}(\lambda)$, para $\lambda > 0$ satisfazem isso, portanto já temos um candidato ao nosso processo de vistorias, mas antes vamos introduzir algumas notações.

Já podemos perceber por (4.58) que mais importante que os tempos T_k , serão os intervalos entre visitas $X_k = T_k - T_{k-1}$.

Seja $\mathcal{D}([0, \infty))$ o espaço de todas as funções càdlàg em \mathbb{N} , ou seja

$$\mathcal{D}([0, \infty)) = \{f : \mathbb{R}_+ \rightarrow \mathbb{N} : f \text{ é contínua à direita e com limite à esquerda}\}.$$

Definiremos $\Gamma : \mathbb{R}^{\mathbb{N}} \rightarrow \mathcal{D}([0, \infty))$ da seguinte forma: dados $(x_1, x_2, \dots) \in \mathbb{R}^{\mathbb{N}}$, seja $\Gamma(x_1, \dots) = N$, tal que

$$N_t = \max\{n; \sum_{i=1}^n x_i \leq t\}, \quad (4.60)$$

que conta quantas visitas ocorreram antes de t , veja Figura 4.2.

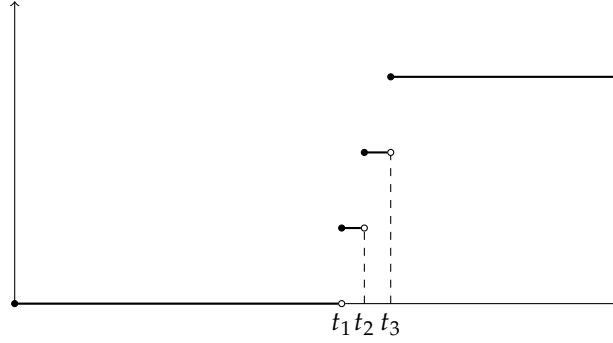


Figura 4.2: A função N_t definindo o número de chegadas do Processo de pontos de Poisson. Note que N é càdlàg.

Poderíamos nos perguntar qual é a σ -álgebra que estamos considerando no espaço $\mathcal{D}([0, \infty))$, essa é uma interessante questão que deve ser abordada em estudos mais profundos desse espaço. Mas por enquanto será suficiente considerarmos a σ -álgebra induzida pelo mapa Γ (a maior que ainda o deixa mensurável).

Estamos prontos agora pra definir o nosso processo.

Definição 4.4.4. Fixado $\lambda > 0$, definimos um Processo de Poisson em \mathbb{R} com parâmetro λ como a lei \mathbb{P}_λ em $\mathcal{D}([0, \infty))$, dada por $\Gamma \circ \text{Exp}(\lambda)^{\otimes \mathbb{N}}$. Ou em outras palavras, o processo de contagem de chegadas N_t , no qual os intervalos entre chegadas são independentes e distribuídos como $\text{Exp}(\lambda)$.

Lembramos que como de costume definimos X_1, X_2, \dots como sendo as projeções canônicas em $\mathbb{R}^{\mathbb{N}}$ onde definimos $\text{Exp}(\lambda)^{\otimes \mathbb{N}}$. Como esses representam os intervalos entre chegadas, definimos também

$$T_k = \sum_{i=1}^k X_i, \text{ para } k \geq 1. \quad (4.61)$$

Podemos agora enunciar o primeiro lema, que nos fornece a distribuição do número de chegadas em um dado tempo $t \geq 0$.

Lema 4.4.5. *Se $\lambda > 0$ e $t \geq 0$, então $N_t \sim_d \text{Poisson}(\lambda t)$ sob \mathbb{P}_λ .*

Demonstração. Vamos primeiramente ver que

$$\mathbb{P}_\lambda[N_t = 0] = \mathbb{P}_\lambda[X_1 > t] = e^{-\lambda t}, \quad (4.62)$$

que coincide com o caso poissoniano.

Para verificar o caso arbitrário $[N_t = k]$, utilizaremos indução e os resultados de esperança condicional regular que vimos anteriormente. Primeiro, observe que se $x_1 > s$, então

$$\Gamma(x_1, x_2, \dots)(r - s) = \Gamma(x_1 - s, x_2, \dots)(r). \quad (4.63)$$

Logo,

$$\begin{aligned} \mathbb{P}_\lambda[N_t = k] &= \mathbb{P}_\lambda[X_1 \leq t, \Gamma(X_2, X_3, \dots)(t - X_1) = k - 1] \\ &= \mathbb{E}_\lambda \left[\mathbb{1}_{X_1 \leq t} \mathbb{P}_\lambda[\Gamma(X_2, X_3, \dots)(t - X_1) = k - 1 | X_1] \right] \\ &\stackrel{\text{Subst.}}{=} \mathbb{E}_\lambda \left[\mathbb{1}_{X_1 \leq t} \mathbb{P}_\lambda[\Gamma(X_2, X_3, \dots)(t - x_1) = k - 1 | X_1 = x_1] \circ X_1 \right] \\ &\stackrel{\text{induc.}}{=} \mathbb{E}_\lambda \left[\mathbb{1}_{X_1 \leq t} (\text{Poisson}(\lambda(t - x_1))(\{k - 1\})) \circ X_1 \right] \\ &= \mathbb{E}_\lambda \left[\mathbb{1}_{X_1 \leq t} \frac{(\lambda(t - X_1))^{k-1} e^{-\lambda(t - X_1)}}{(k - 1)!} \right] \\ &= \int_0^t \frac{(\lambda(t - x_1))^{k-1} e^{-\lambda(t - x_1)}}{(k - 1)!} \lambda e^{-\lambda x_1} dx_1 = \frac{\lambda^k e^{-\lambda t}}{(k - 1)!} \frac{t^k}{k}, \end{aligned}$$

como queríamos demonstrar. \square

Um outro resultado importante sobre esses processos se relaciona ao fato de reiniciar o sistema em tempo $t > 0$. Isso é feito com o seguinte mapa $\theta_t : \mathcal{D}([0, \infty)) \rightarrow \mathcal{D}([0, \infty))$, que leva N em

$$\theta_t(N)(s) = N_{s+t} - N_t. \quad (4.64)$$

Exercício 4.4.5. Mostre que o mapa θ_t é mensurável.

Lema 4.4.6. Fixe $\lambda, t > 0$ e seja N um processo de Poisson de taxa λ . Então, para $k \in \mathbb{Z}_+$ e A mensurável,

$$\mathbb{P}_\lambda[N_t = k, \theta_t \circ N \in A] = \mathbb{P}_\lambda[N_t = k] \mathbb{P}_\lambda[N \in A]. \quad (4.65)$$

Em particular, isso mostra que a distribuição do processo de Poisson N é invariante pelo mapa θ_t .

Demonstração. Começamos reescrevendo o evento e condicionando em T_k como abaixo

$$\begin{aligned} \mathbb{P}_\lambda[N_t = k, \theta_t \circ N \in A] &= \mathbb{P}_\lambda[T_k \leq t, T_{k+1} > t, \theta_t \circ N \in A] \\ &= \mathbb{E}_\lambda[\mathbf{1}_{T_k \leq t} \mathbb{E}_\lambda[X_{k+1} > t - t_k, \theta_t \circ N \in A | T_k = t_k] \circ T_k] \\ &= \mathbb{E}_\lambda[\mathbf{1}_{T_k \leq t} \mathbb{E}_\lambda[X_{k+1} > t - t_k, \Gamma(X_{k+1} - (t - t_k), X_{k+2}, X_{k+3}, \dots) \in A | T_k = t_k] \circ T_k], \end{aligned}$$

que, usando que X_i são independentes e X_{k+1} não tem sem memória, é igual a

$$\begin{aligned} &= \mathbb{E}_\lambda[\mathbf{1}_{T_k \leq t} \mathbb{P}_\lambda[X_{k+1} > t - t_k | T_k = t_k] \circ T_k] \mathbb{P}_\lambda[N \in A] \\ &= \mathbb{P}_\lambda[N_t = t] \mathbb{P}_\lambda[N \in A], \end{aligned}$$

terminando a prova do lema. \square

Como corolário do lema acima, podemos deduzir que um processo de Poisson possui incrementos independentes. Mais precisamente,

Corolário 4.4.7. Seja N um Processo de Poisson N com taxa $\lambda > 0$. Considerando também tempos $0 = t_0 < t_1 < \dots < t_j$, e inteiros $k_1, \dots, k_j \geq 0$ temos

$$\mathbb{P}_\lambda[N_{t_1} = k_1, \dots, N_{t_j} - N_{t_{j-1}} = k_j] = \mathbb{P}_\lambda[N_{t_1} = k_1] \cdots \mathbb{P}_\lambda[N_{t_j - t_{j-1}} = k_j]$$

Demonstração. Basta observar que

$$[N_{t_2} - N_{t_1}, \dots, N_{t_j} - N_{t_{j-1}}] = [N_{t_2 - t_1}, \dots, N_{t_j - t_1} - N_{t_{j-1} - t_1}] \circ \theta_{t_1} \quad (4.66)$$

e aplicar o lema para obter

$$\begin{aligned} \mathbb{P}_\lambda[N_{t_1} = k_1, \dots, N_{t_j} - N_{t_{j-1}} = k_j] &= \mathbb{P}_\lambda[N_{t_1} = k_1] \mathbb{P}_\lambda[N_{t_2 - t_1} = k_2, \dots, N_{t_j - t_1} - N_{t_{j-1} - t_1} = k_j]. \end{aligned}$$

Repetindo essa operação iterativamente, obtemos o resultado desejado. \square

$$T_1, \dots, T_k \text{ under } \mathbb{P}_\lambda^{t,k} \stackrel{d}{\sim} \quad (4.67)$$

$$x_1 = t_1, x_2 = t_2 - t_1, \dots, x_k = t_k - t_{k-1}$$

$$\begin{aligned} \rho_\rho^{t,k}(t_1, \dots, t_k) &= \mathbf{1}_{0 \leq t_1 \leq \dots \leq t_k \leq t} \frac{1}{\mathbb{P}_\lambda[N_t = k]} \lambda e^{-\lambda x_1} \lambda e^{-\lambda x_2} \dots e^{-\lambda x_k} e^{-\lambda(t-t_k)} \\ &= \mathbf{1}_{0 \leq t_1 \leq \dots \leq t_k \leq t} \frac{k!}{e^{-\lambda t} (\lambda t)^k} \lambda^k e^{-\lambda t} = \mathbf{1}_{0 \leq t_1 \leq \dots \leq t_k \leq t} \frac{k!}{t^k}. \end{aligned} \quad (4.68)$$

Note que não depende de λ .

Considere variáveis uniformes U_1, \dots, U_k no intervalo $[0, t]$

$$\rho^{t,k}(u_1, \dots, u_k) = \mathbf{1}_{\tilde{u}_1, \dots, \tilde{u}_k \in [0, t]} \frac{1}{t^k} \quad (4.69)$$

Seja $\tilde{U}_1 < \dots < \tilde{U}_k$ a versão ordenada das U_i 's.

$$(\tilde{U}_1, \dots, \tilde{U}_k) \stackrel{q.c.}{=} \sum_{\sigma \text{ perm. de } \{1, \dots, k\}} (U_{\sigma_1}, \dots, U_{\sigma_k}) \mathbf{1}_{0 \leq U_{\sigma_1} \leq \dots \leq U_{\sigma_k} \leq t} \quad (4.70)$$

Então,

$$\begin{aligned} \tilde{\rho}^{t,k} &= \mathbf{1}_{0 \leq \tilde{u}_1 \leq \dots \leq \tilde{u}_k \leq t} \frac{1}{t^k} \sum_{\sigma \text{ perm. de } \{1, \dots, k\}} \rho^{t,k}(\tilde{u}_{\sigma_1}, \dots, \tilde{u}_{\sigma_k}) \\ &= \mathbf{1}_{0 \leq \tilde{u}_1 \leq \dots \leq \tilde{u}_k \leq t} \frac{k!}{t^k} \end{aligned} \quad (4.71)$$

Capítulo 5

Soluções de exercícios

Solução de 3.3.3 Primeiramente, vamos ver qual é a distribuição de R_0 . Vamos escrever $R_0 = E_0 + D_0$, onde E_0 é o número de casas acessíveis à esquerda e D_0 à direita. Note que E_0 e D_0 são independentes e identicamente distribuídas, com

$$P[D_0 = l] = P[X_l = 1, X_i = 0 \text{ para } i = 0, \dots, l-1] = p(1-p)^l. \quad (5.1)$$

Podemos agora calcular

$$P[R_0 = k] = \sum_{l=0}^k P[D_0 = l, E_0 = k-l] = \sum_{l=0}^k p^2(1-p)^k = p^2 k(1-p)^k. \quad (5.2)$$

Além disso,

$$E(R_0) = 2E(D_0) = \sum_{l=0}^{\infty} lP[D_0 = l] = 2p \sum_{l=0}^{\infty} l(1-p)^l = \frac{2(1-p)}{p} =: m. \quad (5.3)$$

O que resolve o primeiro item.

A grande dificuldade do segundo item é que as variáveis R_i não são independentes, veja por exemplo que $P[R_0 = 0, R_1 = 2, R_2 = 0] = 0$. Nesse caso, o método do segundo momento deve ser feito com atenção. Chamando de $S_n = \sum_{i=1}^n R_i$, temos

$$P\left[\left|\frac{1}{n}S_n - E(R_0)\right| > a\right] \leq \frac{\text{Var}(S_n)}{a^2 n^2}, \quad (5.4)$$

mas a variância da soma não se torna a soma das variâncias. De fato

$$\begin{aligned}\text{Var}(S_n) &= E\left(\left(\sum_{i=1}^n (R_i - E(R_i))\right)^2\right) = \sum_{i=1}^n \sum_{j=1}^n E\left((R_i - E(R_i))(R_j - E(R_j))\right) \\ &= \sum_{i=1}^n \sum_{j=1}^n \text{Cov}(R_i, R_j) = n\text{Var}(R_0) + 2 \sum_{k=1}^{n-1} (n-k)\text{Cov}(R_0, R_k).\end{aligned}\tag{5.5}$$

Aqui já temos metade da estimativa resolvida, mas ainda falta obter uma estimativa explícita.

Então precisamos estimar superiormente $\text{Cov}(R_i, R_j) = \text{Cov}(R_0, R_{j-1})$. Podemos calcular essa quantidade explicitamente, mas vamos evitar contas chatas fazendo uma estimativa do tipo

$$\text{Cov}(R_0, R_k) \leq c \exp\{-c'k\}, \text{ para todo } k \geq 1.\tag{5.6}$$

O que nos daria que

$$\text{Var}(S_n) \leq n\text{Var}(R_0) + 2 \sum_{k=1}^{n-1} (n-k)c \exp\{-c'k\} \leq c''n.\tag{5.7}$$

Donde a probabilidade que queríamos estimar é no máximo c/a^2n , como no caso independente.

Para obter a prometida cota para a covariância, observe que podemos truncar D_0 e E_k para obter independência. Definindo

$$\tilde{R}_0 = E_0 + (D_0 \wedge \lfloor k/2 \rfloor) \text{ e } \tilde{R}_k = D_k + (E_k \wedge \lfloor k/2 \rfloor),\tag{5.8}$$

temos que \tilde{R}_0 e \tilde{R}_k são independentes (pois dependem de elos disjuntos). Daí

$$\begin{aligned}\text{Cov}(R_0, R_k) &= E(R_0 R_k) - m^2 \\ &= E(\tilde{R}_0 \tilde{R}_k) + E(R_0 R_k \mathbf{1}_{[R_0 \neq \tilde{R}_0] \cup [R_k \neq \tilde{R}_k]}) - m^2 \\ &\leq E(\tilde{R}_0)^2 - m^2 + E((E_0 + D_0)(E_k + D_k) \mathbf{1}_{[R_0 \neq \tilde{R}_0] \cup [R_k \neq \tilde{R}_k]}) \\ &\leq E((E_0 + k + D_k)^2 \mathbf{1}_{[R_0 \neq \tilde{R}_0] \cup [R_k \neq \tilde{R}_k]}) \\ &= E((E_0 + k + D_k)^2) P([R_0 \neq \tilde{R}_0] \cup [R_k \neq \tilde{R}_k]) \\ &\leq (2E(E_0^2) + k^2 + 2kE(E_0) + E(E_0)^2) \cdot 2 \cdot P[R_0 \neq \tilde{R}_0] \\ &\leq ck^2(1-p)^{\lfloor k/2 \rfloor} \leq c \exp\{-c'k\}.\end{aligned}\tag{5.9}$$

Finalizando a cota para a covariância.

Referências Bibliográficas

- [BR06] B. Bollobás and O. Riordan, *Percolation*, Cambridge University Press, 2006.
- [Gri99] Geoffrey Grimmett, *Percolation*, second ed., Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 321, Springer-Verlag, Berlin, 1999. MR 1707339 (2001a:60114)

Índice Remissivo

- $[\omega \text{ satisfaz } Q]$, 3
- anel de conjuntos, 29
- bi-mensurável, 46
- Cadia de Markov, 53
- càdlàg, 18
- condição de compatibilidade, 31
- conjunto
 - livre de somas, 15
- continuidade no vazio, 29
- convergência
 - fraca, 91
- coordenadas canônicas, 30
- densidade, 17
- Desigualdade de Markov, 64
- distribuição, 9
 - binomial, 13
 - conjunta, 37
 - de Bernoulli, 13
 - de Poisson, 27
 - exponencial, 17
 - geométrica, 14, 39
 - marginal, 31
 - normal, 88
 - uniforme, 17
- $dP = \rho \, d\mu$, 17
- elemento aleatório, 8
- espaço
 - mensurável, 2
- espaço
 - amostral, 2
 - canônico, 46
 - polonês, 48
- esperança, 61
 - condicional, 100
 - aditividade, 102
 - desigualdade de Jensen, 105
 - monotonicidade, 102
 - T.C.D., 107
 - T.C.M., 104
 - torre, 105
- evento, 1, 2
- flutuações, 86
- função
 - geradora de momentos, 77
 - taxa, 80
- função de distribuição, 17
- F_X , 18
- inclusão e exclusão, 4
- independência
 - de elementos, 22
 - de eventos, 20, 21
 - de σ -álgebras, 21

- λ -sistema, 6
- Lei
 - $\{0, 1\}$ de Kolmogorov, 75
 - dos Pequenos Números, 27
 - Forte dos Grandes Números, 72
 - Fraca dos Grandes Números, 68
- Método Probabilístico, 15
- momento
 - primeiro, 65
 - segundo, 70
- $\|\mu_1 - \mu_2\|$, 24
- núcleo de transição, 41
- Paradoxo de Bertrand, 11
- passeio aleatório simples, 54
- π -sistema, 6
- Princípio
 - da Substituição, 110, 114
 - de Grandes Desvios, 79
- Princípio de Grandes Desvios, 82
- probabilidade, 3
 - condicional, 37
- Processo de Poisson, 114
- sequências
 - intercambiáveis, 57
- σ -álgebra, 2
 - caudal, 75
 - de borel, 2
 - gerada por \mathcal{G} , 2
 - trivial, 75
- Teorema
 - Central do Limite, 93
 - da Desintegração, 108
 - da Extensão de Caratheodory, 29
 - da Extensão, 31, 46
 - de Dynkin, 6
 - de Fubini para Núcleos, 43
 - de Portmanteau, 96
- transformada
 - de Laplace, 77
- variação total, 24
- variância, 66
- variável aleatória, 9
 - integrável, 61
- $X \sim_d \mu$, 10
- $X \sim_d Y$, 9