Introdução à Probabilidade

Notas de Aula

Leonardo T. Rolla

7 de setembro de 2019

Prefácio

Estas notas foram produzidas a partir de notas de aula das disciplinas *Probabilidade*, do mestrado em Ciências Atuariais da PUC-Rio, ministrada em 2006, *Introdução à Probabilidade*, ministrada em 2012 e 2013 no IMPA, e *Teoria da Probabilidade*, ministrada em 2017 na NYU-Shanghai.

Para seguir estas notas não é necessário qualquer conhecimento prévio em Probabilidade. Os pré-requisitos são cálculo de derivadas e integrais em $\mathbb{R}^d$, limites de sequências, convergência de séries, e limites laterais de funções. Para seguir as demonstrações mais avançadas, o leitor deve estar familiarizado com as propriedades elementares de $\limsup$ e $\liminf$, polinômios de Taylor e supremo de conjuntos.

Descrição e Interdependência dos Capítulos

A primeira parte destas notas consiste de 4 capítulos que devem ser estudados em sequência, antes de passar para os capítulos seguintes. No Capítulo 1 introduzimos os espaços de probabilidade, probabilidade condicional e independência de eventos. Os Capítulos 2 e 3 estudam as variáveis aleatórias e vetores aleatórios, com ênfase nos casos discreto e absolutamente contínuo. No Capítulo 4 é estudada a esperança matemática, suas propriedades, momentos, variância e algumas desigualdades.

A segunda parte contém uma escolha de assuntos mais comumente abordados em um curso introdutório de Probabilidade. O Capítulo 5 trata do lema de Borel-Cantelli e da convergência de variáveis aleatórias. Os Capítulos 6 e 7 apresentam a Lei dos Grandes Números e o Teorema Central do Limite. O Capítulo 8 introduz a função geradora de momentos e a função característica, incluindo convergência em distribuição. No Capítulo 9 estudamos a esperança condicional dada uma partição e a esperança condicional regular. Os capítulos desta segunda parte são basicamente independentes entre si, exceto que os Capítulos 6, 7 e 8 que dependem em maior ou menor medida do Capítulo 5.

Na terceira parte estudamos tópicos menos canônicos para um curso introdutório. No Capítulo 10 estudamos teoremas de convergência da esperança, no Capítulo 11 estudamos passeios aleatórios na rede hipercúbica, e finalmente no Capítulo 12 estuamos o Princípio dos Grandes Desvios. Os capítulos da terceira parte pressupõem que o leitor passou pelos Capítulos 5, 6 e 7.

Rigor Matemático

A primeira parte é auto-contida e matematicamente rigorosa, inclusive na construção da Esperança Matemática como supremo sobre funções simples, sua fórmula para os casos discreto e contínuo, e suas propriedades fundamentais.

Há uma omissão importante: sem demonstrar, assumimos implicitamente a existência de variáveis aleatórias contínuas, ou de uma sequência infinita de variáveis aleatórias com determinada distribuição conjunta.

Uma omissão secundária é o significado de integral. As variáveis aleatórias absolutamente contínuas são definidas e estudadas em termos de uma integral, sem discutir o que significa a integral em si. Em todos os exemplos que vamos considerar, a integral que conhecemos do Cálculo é suficiente.

Na segunda parte, algumas demonstrações que dependem de Teoria da Medida serão omitidas com um aviso correspondente. As principais são: existência e propriedades da distribuição condicional regular e da esperança condicional, a mudança de variáveis no plano complexo para obtenção da função característica de uma gaussiana, equivalência entre convergência em distribuição e convergência da esperança de funções-teste suaves e limitadas.

Tópicos Omitidos

Alguns tópicos importantes são omitidos, dentre eles: quantil de uma variável aleatória; estatística de ordem, método do Jacobiano sem bijeção, distribuição normal multivariada, função geradora e função característica para vetores aleatórios, distribuição condicional de vetores aleatórios.

Erros e Omissões

Estas notas contêm inúmeras imprecisões e omissões. A quem faça uso deste texto, peço que me enviem todos os comentários, críticas e correções que venham a surgir.

7 de setembro de 2019.

Sumário

Capítulo 1

Espaço de Probabilidade

O objetivo deste texto é introduzir o estudo formal dos Espaços de Probabilidade, as variáveis aleatórias e suas propriedades. A Teoria da Probabilidade estuda eventos *aleatórios*, i.e., eventos que não possuem *regularidade determinística*, mas possuem *regularidade estatística.* A ausência de regularidade determinística significa que observações feitas nas mesmas condições não dão o mesmo resultado, enquanto a regularidade estatística se manifesta na estabilidade estatística de frequências.

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

Um *Espaço de Probabilidade*, ou *Modelo Probabilístico*, ou ainda *Modelo Estatístico*, é uma abstração matemática, é uma idealização que busca representar os fenômenos aleatórios.

1.1 Espaço de Probabilidade

Um modelo probabilístico tem três componentes básicas:

1. Um conjunto Ω formado por todos os resultados possíveis do experimento,

chamado *espaço amostral*.

2. Uma classe apropriada $\mathcal{F}$ de subconjuntos do espaço amostral, chamada classe de *conjuntos mensuráveis* ou *eventos aleatórios*.

3. Uma função P que associa a cada conjunto mensurável um número real, que representa a ideia de chance, verossimilhança, confiança, credibilidade, ou probabilidade. Esta função é chamada de *probabilidade*, *medida*, ou *medida de probabilidade*.

Resultados equiprováveis Num modelo em que os resultados são equiprováveis, o espaço amostral é um conjunto finito Ω e a medida de probabilidade é proporcional à quantidade de resultados que fazem parte de um dado evento:

$$P(B) = \frac{\#B}{\#\Omega},$$

onde $\#B$ denota a cardinalidade do conjunto $B \subseteq \Omega$, isto é, a quantidade de elementos que pertencem a B.

Exemplo 1.1. Imagine o sorteio de uma carta em um baralho francês com 52 cartas (numeradas $A, 2, 3, \dots, 9, 10, J, Q, K$ e de naipes $\clubsuit, \heartsuit, \spadesuit, \diamondsuit$). Queremos saber a probabilidade de um jogador tirar $4\clubsuit$, $7\heartsuit$, $A\spadesuit$ ou $7\diamondsuit$, evento que será denotado por B. Temos então:

$$P(B) = \frac{\#B}{\#\Omega} = \frac{4}{52} = \frac{1}{13} \approx 8\%.$$

Exemplo 1.2. Imagine o lançamento de um dado em que um jogador precisa obter 5 ou 6. Neste caso temos $\Omega = \{1, 2, 3, 4, 5, 6\}$, $B = \{5, 6\}$ e

$$P(B) = \frac{\#B}{\#\Omega} = \frac{2}{6} = \frac{1}{3} \approx 33\%.$$

Espaço discreto Outro exemplo um pouco mais complicado é quando o espaço amostral Ω é discreto, isto é, pode ser escrito como uma sequência $\Omega = \{x_1, x_2, x_3, \dots\}$. Neste caso não faz sentido que todos os elementos sejam igualmente prováveis.

A cada possível resultado x_n é associada uma probabilidade $p(x_n)$ de forma que

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

Para um subconjunto $B \subseteq \Omega$ definimos

$$P(B) = \sum_{x \in B} p(x).$$

Exemplo 1.3. Imagine que lançamos um dado em sequência até obter o número 3, e contamos o número de lançamentos necessários, ou seja, o resultado desse experimento é o número de lançamentos efetuados. Então caso o espaço amostral Ω é dado pelo conjunto $\mathbb{N}$ dos números naturais

$$\mathbb{N} = \{1, 2, 3, \dots\}.$$

Neste caso, $p(n) = \frac{1}{6}\left(\frac{5}{6}\right)^{n-1}$. Seja $A =$ "obter um 3 em no máximo 5 tentativas" e $B =$ "não se obter o 3 nas primeiras 10 tentativas". Temos

$$P(A) = \tfrac{1}{6} + \tfrac{1}{6} \times \tfrac{5}{6} + \cdots + \tfrac{1}{6} \times \left(\tfrac{5}{6}\right)^4 = \frac{\frac{1}{6} - \left(\frac{5}{6}\right)^5 \frac{1}{6}}{1 - \frac{5}{6}} = 1 - \left(\tfrac{1}{6}\right)^5 = \tfrac{4651}{7776} \approx 60\%.$$

e

$$P(B) = \tfrac{1}{6}\left(\tfrac{5}{6}\right)^{10} + \tfrac{1}{6}\left(\tfrac{5}{6}\right)^{11} + \tfrac{1}{6}\left(\tfrac{5}{6}\right)^{12} + \cdots = \frac{\frac{1}{6}\left(\frac{5}{6}\right)^{10}}{1 - \left(\frac{5}{6}\right)} = \left(\tfrac{5}{6}\right)^{10} \approx 16\%.$$

A seguir veremos uma formulação mais precisa desses conceitos.

Espaço Amostral

Um conjunto não-vazio Ω, cujos elementos representam todos os *resultados possíveis* de um determinado experimento, é chamado de *espaço amostral*. O experimento é dado pela escolha de algum dos possíveis $\omega \in \Omega$, e dizemos que o ω escolhido representa a *realização* do experimento.

Exemplo 1.4. Se o experimento consiste em lançar uma moeda, então

$$\Omega = \{0, 1\},$$

onde 1 representa a face "cara" e 0 representa a face "coroa".

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

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

onde cada número representa o possível valor da face observada.

Exemplo 1.6. Se o experimento consiste em lançar duas moedas, então

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

onde a primeira coordenada representa o valor observado na primeira moeda, e a segunda coordenada, o da segunda moeda.

Exemplo 1.7. Se o experimento consiste em lançar dois dados e observar as faces superiores, então

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

Exemplo 1.8. Lançar uma moeda infinitas vezes em sequência.

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

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

Eventos Aleatórios

Qualquer subconjunto A do espaço amostral Ω, isto é, $A \subseteq \Omega$, ao qual atribuímos uma probabilidade, é dito um *evento aleatório*.

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

A união $A \cup B$ é o conjunto de todos os $\omega \in \Omega$ tais que ω pertence a A ou ω pertence a B, ou seja, é o conjunto das realizações ω tais que algum dos eventos A ou B ocorrem, portanto $A \cup B$ é o evento "A ou B".

Exemplo 1.10. No lançamento de um dado ($\Omega = \{1, 2, 3, 4, 5, 6\}$) considere os eventos A = "par" = $\{2, 4, 6\}$ e B = "múltiplo de 3" = $\{3, 6\}$. O evento "A ou

B" contém todos os resultados que sejam pares ou múltiplos de 3 (ou ambos!), e é dado por $C = A \cup B = \{2, 3, 4, 6\}$.

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

Exemplo 1.11. No lançamento de um dado ($\Omega = \{1, 2, 3, 4, 5, 6\}$) considere os eventos $A =$ "par" $= \{2, 4, 6\}$ e $B =$ "múltiplo de 3" $= \{3, 6\}$. O evento "A e B" contém todos os resultados que sejam pares e ao mesmo tempo múltiplos de 3, e é dado por $C = A \cap B = \{6\}$.

Denotamos por A^c o complementar do conjunto A, dado por $A^c = \{\omega \in \Omega : \omega \notin A\}$, ou seja, o conjunto das realizações ω para as quais o evento A não ocorre, portanto A^c é o evento "não A".

Exemplo 1.12. No lançamento de um dado ($\Omega = \{1, 2, 3, 4, 5, 6\}$) considere o evento $A =$ "par" $= \{2, 4, 6\}$. O evento "não A" contém todos os resultados que não sejam pares, ou seja, que são ímpares, e é dado por $C = A^c = \{1, 3, 5\}$.

Dois eventos A e B são ditos *mutuamente exclusivos* ou *incompatíveis* se $A \cap B = \emptyset$, isto é, se o evento "A e B" é impossível. O conjunto vazio $\emptyset$ é denominado *evento impossível*. Por outro lado, suponha que, para dois eventos A e B dados, pelo menos um dos dois necessariamente ocorre. Isso quer dizer que $A \cup B = \Omega$. O conjunto Ω também é um evento denominado *evento certo*.

Exemplo 1.13. No lançamento de um dado ($\Omega = \{1, 2, 3, 4, 5, 6\}$) considere os eventos $A =$ "par" $= \{2, 4, 6\}$ e $B =$ "ímpar" $= \{1, 3, 5\}$. O evento "A e B" é o evento impossível porque nenhum número é par e ímpar ao mesmo tempo. Em termos de conjuntos, temos que $A \cap B = \emptyset$. O evento "A ou B" é o evento certo, porque todo número é par ou ímpar. Em termos de conjuntos, temos que $A \cup B = \Omega$.

Se $\omega \in \Omega$, o evento $\{\omega\}$ é dito *elementar*. A relação $A \subseteq B$ significa que todo $\omega \in A$ satisfaz $\omega \in B$, ou seja, para qualquer realização ω, se o evento A ocorre então necessariamente o evento B ocorre. Portanto, $A \subseteq B$ significa que a ocorrência do evento A implica a ocorrência do evento B.

Quando o espaço amostral Ω é um conjunto finito ou enumerável, é natural tomar a classe de eventos aleatórios $\mathcal{F}$ como $\mathcal{F} = \mathcal{P}(\Omega)$, isto é, o conjunto de todos os subconjuntos de Ω, dado por

$$\mathcal{P}(\Omega) = \{A : A \subseteq \Omega\}$$

e chamado o *conjunto das partes*. Porém há casos em que Ω não é enumerável, como no Exemplo 1.8, e não é possível construir um modelo probabilístico em toda essa classe $\mathcal{P}(\Omega)$. Em todo caso, faremos algumas suposições naturais sobre a classe $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ de eventos aleatórios. Mais precisamente, vamos assumir que $\mathcal{F}$ satisfaz as seguintes propriedades:

(F1) $\Omega \in \mathcal{F}$;

(F2) Para todo $A \in \mathcal{F}$, tem-se que $A^c \in \mathcal{F}$;

(F3) Se $A_1, A_2, A_3, \dots \in \mathcal{F}$, então $(\cup_{i=1}^{\infty} A_i) \in \mathcal{F}$.

Chamaremos de *σ-álgebra* a uma classe de subconjuntos de Ω satisfazendo as três propriedades acima.

Espaço de Probabilidade

Seja Ω um espaço amostral e $\mathcal{F}$ uma σ-álgebra para um dado experimento. Uma *medida de probabilidade* P é uma aplicação $P : \mathcal{F} \to \mathbb{R}$ satisfazendo as seguintes propriedades:

(P1) $P(A) \geqslant 0$ para todo $A \in \mathcal{F}$.

(P2) $P(\Omega) = 1$.

(P3) Se $A_1, A_2, \dots \in \mathcal{F}$ e $A_i \cap A_j = \emptyset \ \forall i \neq j$, então $P\left(\cup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$.

A maneira usual de definir medida de probabilidade é através das propriedades acima, a partir das quais podem-se demonstrar inúmeras outras. Listamos abaixo as mais comuns.

Teorema 1.14. *Toda medida de probabilidade P satisfaz:*

1. $P(\emptyset) = 0$.
2. $P(A^c) = 1 - P(A)$.
3. *Se $A, B \in \mathcal{F}$ e $A \subseteq B$ então $P(A) \leqslant P(B)$. (monotonicidade)*
4. *Se $A, B \in \mathcal{F}$ e $A \subseteq B$ então $P(B \setminus A) = P(B) - P(A)$.*

5. *Para todo* $A \in \mathcal{F}$, *temos* $0 \leqslant P(A) \leqslant 1$.

6. *Se* $A_1, A_2, \ldots, A_n \in \mathcal{F}$, *então* $P\left(\bigcup\limits_{i=1}^{\infty} A_i\right) \leqslant \sum\limits_{i=1}^{\infty} P(A_i)$. *($\sigma$-subaditividade).*

7. *Sejam* A *e* $B \in \mathcal{F}$. *Então* $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Demonstração. Feita em aula. □

Uma medida de probabilidade P também tem a propriedade de ser contínua. Dizemos que $A_n \nearrow A$ se $A_1 \subseteq A_2 \subseteq A_3 \subseteq \cdots$ e $\cup_{n=1}^{\infty} = A$. Analogamente, $A_n \searrow A$ se $A_1 \supseteq A_2 \supseteq A_3 \supseteq \cdots$ e $\cap_{n=1}^{\infty} = A$.

Teorema 1.15 (Continuidade)**.** *Se* $A_n \nearrow A$ *ou* $A_n \searrow A$, *então* $P(A_n) \to P(A)$.

Demonstração. Feita em aula. □

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

Definição 1.16 (Espaço de Probabilidade)**.** Um *espaço de probabilidade* é um trio $(\Omega, \mathcal{F}, P)$, onde

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

Exemplo 1.17. Lançamento de uma moeda. Este espaço é pequeno o suficiente para que possamos construí-lo explicitamente. Como fizemos anteriormente, as duas faces da moeda serão representadas em $\Omega = \{0, 1\}$. A σ-álgebra $\mathcal{F}$ é dada por $\mathcal{F} = \mathcal{P}(\Omega) = \big\{\{\}, \{0\}, \{1\}, \{0, 1\}\big\}$. A medida de probabilidade $P : \mathcal{F} \to \mathbb{R}$ é dada por $P(\{\}) = 0$, $P(\{0\}) = P(\{1\}) = \frac{1}{2}$, $P(\{0, 1\}) = 1$.

Exemplo 1.18. Sortear 4 cartas de um baralho francês, com reposição. Neste caso temos

$$\Omega = \Big(\{A, 2, 3, \ldots, 9, 10, J, Q, K\} \times \{\clubsuit, \heartsuit, \spadesuit, \diamondsuit\}\Big)^4$$

e

$$\#\Omega = 52^4.$$

Tomamos

$$\mathcal{F} = \mathcal{P}(\Omega)$$

e

$$P(A) = \frac{\#A}{52^4}, \quad A \in \mathcal{F}.$$

Qual a probabilidade do evento A = "as quatro cartas são valetes"? Temos $A = (\{J\} \times \{\text{qualquer naipe}\})^4$, logo $\#A = 4^4$ e portanto

$$P(A) = \frac{4^4}{52^4} = \frac{1}{13^4}.$$

Qual a probabilidade do evento B = "todas as cartas têm o mesmo naipe"? Temos 4 escolhas para o naipe, e 13 escolhas para cada uma das cartas retiradas, logo $\#B = 4 \times 13^4$ e portanto

$$P(B) = \frac{4.13^4}{52^4} = \frac{1}{4^3}.$$

Qual a probabilidade do evento C = "há um par de cartas de um naipe e um par de cartas de um outro naipe"? Temos $\binom{4}{2}$ escolhas para os naipes, onde $\binom{n}{k}$ denota o número de combinações de n, k a k, isto é, $\binom{n}{k} = \frac{n!}{k!(n-k)!}$. Escolhidos os naipes, temos $\binom{4}{2}$ combinações para quais retiradas correspondem a qual naipe. Escolhidos os naipes e as posições, há 13 escolhas de cartas para cada retirada. Assim,

$$\#C = \binom{4}{2}\binom{4}{2}13^4 = 6^2 13^4$$

e portanto

$$P(C) = \frac{6^2 13^4}{52^4} = \frac{6^2}{4^4} = \frac{9}{64}.$$

1.2 Probabilidade Condicional

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

Definição 1.19 (Probabilidade Condicional)**.** Dados $A, B \in \mathcal{F}$ em um espaço $(\Omega, \mathcal{F}, P)$, definimos a *probabilidade condicional de A dado que ocorreu B*, ou simplesmente *probabilidade condicional de A dado B*, por

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

Quando $P(B) = 0$, definimos $P(A|B) = P(A)$.

Exemplo 1.20. Um dado é lançado. Sabendo-se que o resultado é maior que 3, qual a probabilidade de que seja par? Denotamos o primeiro evento por $B = \{4, 5, 6\}$ e o segundo por $A = \{2, 4, 6\}$. Aplicando a definição, a probabilidade de ser par sabendo-se que o resultado é maior que 3 é dada por

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

Exemplo 1.21. Suponha que 30% dos genes de cor de olhos seja para olho claro c, uniformemente espelhados por toda uma população. Como essa é uma característica genética recessiva, 9% da população terá olhos claros cc, 42% terá ambos os genes diferentes Cc, e 49% terá ambos os genes correspondentes a olhos escuros CC. Os 42% com genes diferentes Cc terão o mesmo fenótipo dos 49% que têm ambos os genes para olhos escuros CC, porque este é um gene dominante. Uma pessoa de olhos escuros é selecionada ao acaso. Qual a probabilidade de que essa pessoa tenha o gene recessivo para olhos claros? Denotando $B =$ "olhos escuros"$= \{Cc, CC\}$ e $A =$ "tem um gene de olhos claros"$= \{cc, Cc\}$, temos

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(\{Cc\})}{P(\{Cc, CC\})} = \frac{42\%}{91\%} = 46,15\%.$$

Proposição 1.22. *A probabilidade condicional é uma medida de probabilidade, isto é, dado $B \in \mathcal{F}$, a função que leva A em $P(A|B)$ satisfaz (P1)–(P3).*

Demonstração. Exercício. □

Regra do produto

A regra do produto permite expressar a probabilidade da ocorrência simultânea de diversos eventos a partir do valor de cada probabilidade condicional dados os eventos anteriores.

Teorema 1.23 (Regra do Produto)**.** *Dados* $A_1, A_2, \dots, A_n$ *em* $(\Omega, \mathcal{F}, P)$*, vale*

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

Demonstração. Vamos provar por indução em n. Para $n = 1$ vale trivialmente: $P(A_1) = P(A_1)$. Para $n = 2$, temos

$$P(A_2|A_1) = \frac{P(A_2 \cap A_1)}{P(A_1)} \quad \Longrightarrow \quad P(A_1 \cap A_2) = P(A_1)P(A_2|A_1).$$

Para $n = 3$, temos

$$P(A_3|A_1 \cap A_2) = \frac{P(A_1 \cap A_2 \cap A_3)}{P(A_1 \cap A_2)}$$

e portanto

$$\begin{aligned} P(A_1 \cap A_2 \cap A_3) &= P(A_1 \cap A_2)P(A_3|A_1 \cap A_2) \\ &= P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2). \end{aligned}$$

Suponhamos a igualdade válida para $n = m$, temos

$$P(A_{m+1}|A_1 \cap \cdots \cap A_m) = \frac{P(A_1 \cap \cdots \cap A_m \cap A_{m+1})}{P(A_1 \cap \cdots \cap A_m)}$$

e portanto

$$\begin{aligned} P(A_1 \cap \cdots \cap A_{m+1}) &= \underbrace{P(A_1 \cap \cdots \cap A_m)}_{\text{usando a hipótese}} P(A_{m+1}|A_1 \cap \cdots \cap A_m) \\ &= P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdots P(A_{m+1}|A_1 \cap \cdots \cap A_m), \end{aligned}$$

completando a prova por indução. □

Exemplo 1.24. Um móvel tem 2 gavetas, a primeira gaveta contém 3 bolsas e a segunda contém 4 bolsas, a primeira bolsa da primeira gaveta contém duas bolas vermelhas e uma bola azul, e todas as demais bolsas contêm duas bolas azuis. Abre-se uma gaveta, escolhe-se uma bolsa e retira-se uma bola de dentro da bolsa, tudo ao acaso. Qual a probabilidade de que a bola retirada seja vermelha? Sejam $A =$ "abre-se a primeira gaveta", $B =$ "escolhe-se a primeira bolsa" e $C =$ "retira-se a bola vermelha", então temos

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

Exemplo 1.25. Selecionar 3 cartas de um baralho francês de 52 cartas, ao acaso e sem reposição. Qual a probabilidade de tirar 3 reis? Seja A_i = "tirar rei na i-ésima retirada" e A = "tirar 3 reis"$= A_1 \cap A_2 \cap A_3$. Temos

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

Exemplo 1.26. Continuando o Exemplo 1.21, suponha que os casais se formam e decidem ter filhos aleatoriamente. Cada vez que o casal tem um filho, cada um dos pais transmite algum dos genes com mesma probabilidade. Selecionam-se uma mãe e filha ao acaso. Qual a probabilidade de ambas mãe e filha terem olhos claros? Definindo A = "mãe tem olhos claros", B = "filha tem olhos claros", D = "o gene transmitido pelo pai é c", temos

$$P(A \cap B) = P(A)P(B|A) = P(A)P(D) = 9\% \times 30\% = 2,7\%.$$

Lei da probabilidade total

Dizemos que $B_1, B_2, B_3, \dots \in \mathcal{F}$ formam uma *partição de* Ω se $B_i \cap B_j = \emptyset \ \forall i \neq j$ e $\cup_{i=1}^{\infty} B_i = \Omega$. Partições são particularmente úteis em contextos onde determinado aspecto divide os resultados possíveis em casos, e é possível expressar determinadas relações em cada um desses casos separadamente.

Teorema 1.27 (Lei da Probabilidade Total)**.** *Sejam $A, B_1, B_2, B_3, \dots$ eventos aleatórios em $(\Omega, \mathcal{F}, P)$ tais que $B_1, B_2, B_3, \dots$ formam uma partição de Ω. Então*

$$P(A) = \sum_{i=1}^{\infty} P(B_i)P(A|B_i).$$

Demonstração. Usando a regra do produto temos

$$P(A) = P\left[\cup_{i=1}^{\infty}(A \cap B_i)\right] = \sum_{i=1}^{\infty} P(A \cap B_i) = \sum_{i=1}^{\infty} P(B_i)P(A|B_i).$$

A primeira igualdade vale pois $A = A \cap \Omega = A \cap (\cup_{i=1}^{\infty} B_i) = \cup_{i=1}^{\infty}(A \cap B_i)$. Na segunda igualdade usamos que esses eventos são disjuntos, pois $(A \cap B_i) \cap (A \cap B_j) \subseteq B_i \cap B_j = \emptyset$ para todo $i \neq j$. Na última igualdade usamos a regra do produto. $\square$

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

Exemplo 1.28. Um armário tem duas gavetas, A e B. A gaveta A tem 2 meias azuis e 3 meias pretas, e a gaveta B tem 3 meias azuis e 3 meias vermelhas. Abre-se uma gaveta ao acaso e retira-se uma meia ao acaso da gaveta escolhida.

Problema: Qual a probabilidade de escolher-se uma meia azul?

Solução: Começamos pelos valores conhecidos: $P(A) = P(B) = \frac{1}{2}$, $P(\text{azul}|A) = \frac{2}{5}$ e $P(\text{azul}|B) = \frac{3}{6}$. Assim,

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

Exemplo 1.29. São dadas duas urnas, A e B. A urna A contém 1 bola azul e 1 vermelha. A urna B contém 2 bolas vermelhas e 3 azuis. Uma bola é extraída ao acaso de A e colocada em B. Uma bola então é extraída ao acaso de B.

Problema: Qual a probabilidade de se retirar uma bola vermelha de B?

Problema: Qual a probabilidade de ambas as bolas retiradas serem da mesma cor?

Fórmula de Bayes

A fórmula de Bayes determina a probabilidade condicional de eventos que precedem aquele efetivamente observado. Mais precisamente, quando conhecemos as probabilidades de uma sequência de eventos B_j que particionam Ω e a probabilidade condicional de um evento posterior A em termos dessa partição, podemos calcular as probabilidades condicionais de ocorrência de cada B_j sabendo-se da ocorrência ou não do evento A. Os valores originais são chamados de probabilidades *a priori* dos eventos B_j, e os valores das probabilidades condicionais são chamados de probabilidades *a posteriori* desses eventos.

Teorema 1.30 (Fórmula de Bayes)**.** *Dado um espaço de probabilidade* $(\Omega, \mathcal{F}, P)$*,* *uma partição $B_1, B_2, B_3, \dots$, e um evento A, para todo $j \in \mathbb{N}$ vale a identidade*

$$P(B_j|A) = \frac{P(B_j)P(A|B_j)}{\sum_i P(B_i)P(A|B_i)}.$$

Demonstração. Feita em aula. □

Exemplo 1.31. No Exemplo 1.28, sabendo-se que uma meia azul foi retirada, qual a probabilidade de ter sido aberta a gaveta A? Pela Fórmula de Bayes temos

$$P(A|\text{azul}) = \frac{P(A)P(\text{azul}|A)}{P(A)P(\text{azul}|A) + P(B)P(\text{azul}|B)} = \frac{\frac{1}{5}}{\frac{9}{20}} = \frac{4}{9}.$$

Exercício 1.1. Num certo certo país, todos os membros de um comitê legislativo ou são trabalhistas ou são liberais. Há três comitês. O Comitê 1 tem 5 trabalhistas, o Comitê 2 tem 2 trabalhistas e 4 liberais, e o Comitê 3 consiste de 3 trabalhistas e 4 liberais. Um comitê é selecionado aleatoriamente e uma pessoa é selecionada aleatoriamente deste comitê.

(a) Ache a probabilidade de que a pessoa selecionada seja trabalhista.

(b) Dado que a pessoa selecionada é trabalhista, qual a probabilidade de ela ter vindo do comitê 1?

Exemplo 1.32. Continuando o Exemplo 1.26, selecionando-se uma criança de olhos escuros ao acaso, qual a probabilidade de que o pai tenha olhos claros? Tomando A = "uma criança tem olhos claros", B = "o pai tem olhos claros", e D = "o pai transmite o gene de olhos claros", temos

$$P(A|D) = P(A|D \cap B) = P(A|D \cap B^c) = \frac{30}{100}$$

que representam a chance de a mãe também transmitir o gene de olhos claros. Por outro lado,

$$P(D|B) = 1, \quad P(D|B^c) = \frac{1}{2}\frac{P(Cc)}{P(\{Cc, CC\})} = \frac{1}{2}\frac{42}{91} = \frac{21}{91}.$$

Logo, pela Regra do Produto,

$$P(A|B) = P(A \cap D|B) = P(D|B)P(A|D \cap B) = \frac{30}{100}$$

e

$$P(A|B^c) = P(A \cap D|B^c) = P(D|B^c)P(A|D \cap B^c) = \frac{63}{910}.$$

Finalmente, pela Fórmula de Bayes,

$$P(B|A^c) = \frac{P(B)P(A^c|B)}{P(B)P(A^c|B) + P(B^c)P(A^c|B^c)} = \frac{\frac{9}{100}\frac{70}{100}}{\frac{9}{100}\frac{70}{100} + \frac{91}{100}\frac{847}{910}} = 6.92\%.$$

1.3 Independência

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

Definição 1.33 (Eventos Independentes)**.** Os eventos aleatórios A e B são ditos *independentes* se

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

Exemplo 1.34. Uma moeda é lançada duas vezes. Sejam A = "a primeira moeda sai cara" e B = "a segunda moeda sai cara". Então A e B são independentes, pois

$$P(A) = P(\{1\} \times \{0,1\}) = \frac{2}{4} = \frac{1}{2}$$
$$P(A) = P(\{0,1\} \times \{1\}) = \frac{2}{4} = \frac{1}{2}$$
$$P(A \cap B) = P(\{1\} \times \{1\}) = \frac{1}{4} = P(A)P(B).$$

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

Exemplo 1.35. Dois dados são lançados. Consideramos os eventos A = "o primeiro dado é par" e C = "a soma dos valores dos dados é par". Então

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

Proposição 1.36. *São equivalentes:*

(*i*) *A e B são independentes,*

(*ii*) *A e B^c são independentes,*

(*iii*) *A^c e B são independentes,*

(*iv*) *A^c e B^c são independentes,*

(*v*) *$P(A|B) = P(A)$,*

(*vi*) *$P(B|A) = P(B)$.*

Demonstração. Exercício. □

Definição 1.37 (Eventos Independentes Dois a Dois)**.** Os eventos aleatórios $(A_i)_{i \in I}$, onde I é um conjunto qualquer de índices, são ditos *independentes dois a dois* se A_i e A_j são independentes para todos $i, j \in I$ com $i \neq j$.

Exemplo 1.38. Dois dados são lançados. Consideramos os eventos A = "o primeiro dado é par", B = "o segundo dado é par" C = "a soma dos valores dos dados é par". Então

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

Exemplo 1.39. Lançamento de um dado de 4 faces. Considere A = "par", B = "menor que 3", C = "1 ou 4", i.e., $A = \{2,4\}$, $B = \{1,2\}$, $C = \{1,4\}$. Então A, B e C são independentes dois a dois. De fato,

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

Definição 1.40 (Eventos Coletivamente Independentes)**.** Os eventos aleatórios $(A_i)_{i \in I}$ são ditos *coletivamente independentes* ou *estocasticamente independentes* se, dado qualquer conjunto de índices distintos $i_1, i_2, \dots, i_n \in I$, vale

$$P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_n}) = P(A_{i_1})P(A_{i_2}) \cdots P(A_{i_n}).$$

Exemplo 1.41. Lançamento de um dado de 12 faces. Seja A = "múltiplo de 3", B = "menor ou igual a 6" e C = "par", i.e., $A = \{3,6,9,12\}$, $B = \{1,2,3,4,5,6\}$ e $C = \{2,4,6,8,10,12\}$. Então A, B e C são coletivamente independentes, pois

$$\begin{aligned} P(A \cap B) &= P(\{3,6\}) = \frac{1}{6} = P(A)P(B), \\ P(B \cap C) &= P(\{2,4,6\}) = \frac{1}{4} = P(B)P(C), \\ P(A \cap C) &= P(\{6,12\}) = \frac{1}{6} = P(A)P(C), \\ P(A \cap B \cap C) &= P(\{6\}) = \frac{1}{12} = P(A)P(B)P(C). \end{aligned}$$

Contra-Exemplo 1.42. No Exemplo 1.39, os eventos A, B e C *não são* coletivamente independentes. De fato,

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

Contra-Exemplo 1.43. No Exemplo 1.38, os eventos A, B e C *não são* coletivamente independentes. De fato,

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

1.4 O Problema de Monty-Hall

Num programa de auditório, há três portas, e atrás de cada uma delas há um prêmio, sendo que uma delas esconde uma bolsa com cem mil reais, e as outras duas escondem cabras velhas sem valor. O programa sempre funciona da seguinte forma. O participante escolhe uma das portas. O apesentador, ao invés de abrir a porta escolhida e entregar o prêmio correspondente, abre uma segunda porta, revelando uma das cabras. Em seguida, ele oferece ao participante a opção de trocar sua escolha para a outra porta que ainda não foi revelada. A pergunta é:

Ao participante, o que lhe convém mais? Deve ele trocar de porta, manter sua escolha original, ou tanto faz?

Mais precisamente, como se comparam as chances de o participante ganhar o prêmio trocando de porta e mantendo a escolha original? A pergunta já gerou muita controvérsia. Até mesmo matemáticos como Erdős viram sua intuição falhar e só se convenceram depois de observar a resposta correta experimentalmente. E não precisamos ir tão longe: perguntando a professores universitários, não será difícil encontrar alguns que pensem como um dia pensou Erdős.

A resposta correta é que *sem dúvida alguma, o participante deve trocar de porta.* Mesmo que a produção do programa estabeleça uma multa de 49% sobre o prêmio obtido caso o participante decida trocar de porta, ainda assim essa é a melhor opção. Isso porque a chance de ganhar o prêmio trocando de porta é o dobro da chance de ganhar o prêmio mantendo a escolha original.

Muitas pessoas acreditam que é indiferente manter a escolha ou trocar, porque *a probabilidade condicional de o prêmio estar na terceira porta dado que não está na segunda porta é de 50%,* e portanto não há razão para trocar de porta.

Não faremos aqui as contas para justificar a proporção de 2 para 1 entre a terceira porta e a primeira, porque fazer uma outra conta não refutaria a essência da conta feita logo acima, que seguirá sólida, inquebrantável, e até mesmo correta.

O que tentaremos fazer é convencer o leitor de que *a pergunta original não trata de probabilidade condicional*, ou pelo menos não essa apresentada acima. De fato, parte do problema está na linguagem um pouco negligente que usamos quando falamos em probabilidade condicional. Por exemplo, estas expressões estão, de certa forma, incompletas: “a probabilidade de que o resultado seja par sabendo-se que é maior que 3”, “a probabilidade de ter sido aberta a primeira gaveta sabendo-se que uma meia azul foi retirada”, “a probabilidade condicional de esse casal ter dois filhos homens, sabendo-se que o casal tem um filho homem". A incompletude está na expressão "sabendo-se". Como foi que o observador soube que o resultado do dado era maior que 3, ou que a meia retirada foi azul, ou que o casal tem um filho homem? Essa pequena falha na linguagem é compreensível, porque tornar essas expressões absolutamente precisas (se é que isso é possível) teria um preço estético às vezes muito alto. Um possível compromisso entre precisão e fluidez é ir eliminando a ambiguidade nas entrelinhas dos exemplos e exercícios. Entretanto, para o problema em questão, mais precisão se faz necessária.

Probabilidade Condicional representa a nova medida de probabilidade do ponto de vista de um observador que tem acesso à ocorrência ou não de determinado evento B. Neste caso, faz-se uma observação com dois resultados possíveis: ou B ocorreu ou B não ocorreu. O fato de um terceiro trazer a notícia de que B ocorreu não se enquadra nesse tipo de situação. Isso porque não há uma dicotomia entre as possíveis observações B e B^c, já que o terceiro poderia ter decidido não revelar nada a respeito da ocorrência ou não desse evento B.

Esse é o caso do problema de Monty Hall. O participante não tinha a prerrogativa de revelar o conteúdo de uma outra porta para então decidir se queria mudar sua escolha original. Ao contrário, foi o apresentador quem decidiu, usando a informação sobre o conteúdo das portas e sobre a escolha que fez o participante, que deveria abrir a segunda porta e não a terceira. Veja também que nesse programa o apresentador nunca revela uma porta com dinheiro, o que novamente deixa claro que não se trata de uma observação que tem como resultados possíveis B ou B^c. Portanto, o cálculo de 50% mencionado acima é tão simples e tão correto quanto inútil para responder à pergunta original. A formulação pertinente seria "a probabilidade condicional de o prêmio estar na terceira porta, dado que o participante escolheu a primeira porta e, seguindo o protocolo pré-estabelecido pela produção do programa, o apresentador decidiu mostrar-lhe que o prêmio não está na segunda porta, é de ______%." Deixamos ao leitor a tarefa de formular um modelo probabilístico capaz de representar todo o processo de decisões envolvidas nesse programa, e preencher a lacuna acima com 66,666...

1.5 Exercícios

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

1.3. Considere uma população de indivíduos capazes de gerar proles do mesmo tipo. O número de indivíduos inicialmente presentes, denotado por X_0, é o tamanho da geração zero. Todos as proles da geração zero constituem a primeira geração e o seu número é denotado por X_1. Em geral, X_n denota o tamanho da n-ésima geração. Mostre que $\lim_{n\to\infty} P(X_n = 0)$ existe e interprete o seu significado.

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

(a) O casal tem um filho homem.

(b) O filho mais velho do casal é homem.

(c) O casal tem um filho homem que nasceu num sábado.

(d) O casal tem um filho homem que não nasceu num sábado.

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

1.5. Se $P(A) = P(A|B) = \frac{1}{4}$ e $P(B|A) = \frac{1}{2}$:

1. A e B são independentes?
2. A e B são mutuamente exclusivos?
3. Calcule $P(A^c|B^c)$.

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

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

(a) Calcule a probabilidade de a carta sorteada ser uma das três figuras reais (J, Q, K).

(b) Sabendo-se que foi sorteada uma figura real, calcule a probabilidade de o baralho retirado ter sido o baralho comum.

(c) Calcule a probabilidade de a carta sorteada ser de espadas $\spadesuit$.

(d) Sabendo-se que foi sorteada uma carta de espadas, calcule a probabilidade de o baralho retirado ter sido o baralho de truco.

(e) Sejam A = “Foi retirado o baralho comum”, B = “Foi sorteada uma figura real” e C = “Foi sorteada uma carta de espadas”. A e B são independentes? A e C são independentes? A, B e C são coletivamente independentes?

(f) Qual a probabilidade de se sortear uma carta de número ‘5’ ?

(g) Sabendo-se que foi sorteado um número (i.e., não foi sorteado A, J, Q nem K), qual a probabilidade de o baralho retirado ter sido o baralho de truco?

1.7. [Jam04, Capítulo 1].
Recomendados: 1, 2, 3, 4, 5, 11, 16, 18, 22.
Sugeridos: 8, 9, 10, 12, 13, 14, 15, 17, 19, 20, 21.

Capítulo 2

Variáveis Aleatórias

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

2.1 Variáveis Aleatórias

Uma variável aleatória é uma função que associa a cada resultado ω do espaço amostral Ω um número real, ou seja, uma função

$$X : \begin{array}{c} \Omega \to \mathbb{R} \\ \omega \mapsto X(\omega) \end{array}.$$

Exemplo 2.1. Joga-se um dado e observa-se a face superior. Nesse caso temos $\Omega = \{1, 2, 3, 4, 5, 6\}$ e $X(\omega) = \omega$.

Vamos colocar uma restrição sobre a função X com o intuito de poder associar probabilidade a eventos como "o valor observado de X é menor que 7". Para isso, introduzimos uma definição mais formal:

Definição 2.2 (Variável Aleatória)**.** Uma *variável aleatória* X em um espaço de probabilidade $(\Omega, \mathcal{F}, P)$ é uma função real definida no espaço Ω tal que o conjunto $\{\omega \in \Omega : X(\omega) \leqslant x\}$ é evento aleatório para todo $x \in \mathbb{R}$, isto é,

$$X : \Omega \to \mathbb{R}$$

é uma variável aleatória se $\{\omega \in \Omega : X(\omega) \leqslant x\} \in \mathcal{F}$ para todo $x \in \mathbb{R}$. Daqui para frente denotaremos por $[X \leqslant x]$ o evento $\{\omega \in \Omega : X(\omega) \leqslant x\}$.

Exemplo 2.3 (Variável aleatória constante)**.** Se $X(\omega) = c$ para todo $\omega \in \Omega$, então

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

Portanto, X é variável aleatória.

Exemplo 2.4 (Função indicadora)**.** Dado $A \subseteq \Omega$, definimos

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

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

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

Portanto, X é variável aleatória.

Contra-Exemplo 2.5. Sejam $\Omega = \{1, 2, 3, 4\}$ e $\mathcal{F} = \{\emptyset, \{1, 2\}, \{3, 4\}, \Omega\}$ e considere os conjuntos $A = \{1, 2\}$ e $B = \{1, 3\}$. Então $\mathbb{1}_A$ é variável aleatória em $(\Omega, \mathcal{F})$, mas $\mathbb{1}_B$ não é.

Espaço induzido e lei de uma variável aleatória

A *σ-álgebra de Borel* na reta $\mathbb{R}$, denotada por $\mathcal{B}$, é a menor σ-álgebra que contém todos os intervalos da reta.[1] Os conjuntos $B \subseteq \mathbb{R}$ tais que $B \in \mathcal{B}$ são chamados

[1] Equivalentemente, $\mathcal{B}$ é a menor σ-álgebra que contém todos os intervalos semi-infinitos, ou ainda, é a menor σ-álgebra que contém todos os conjuntos abertos. O leitor mais curioso pode

Borelianos. A σ-álgebra de Borel $\mathcal{B}$ é muito menor que a σ-álgebra das partes $\mathcal{P}(\mathbb{R})$, e daqui em diante, sempre que aparecer $B \subseteq \mathbb{R}$, deve-se entender $B \in \mathcal{B}$.

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

$$P_X(B) = P\Big(\{\omega : X(\omega) \in B\}\Big), \quad B \in \mathcal{B}.$$

Ou seja, o espaço amostral é o conjunto dos números reais, os eventos aleatórios são os conjuntos Borelianos, e a medida de probabilidade é aquela induzida por X. A medida de probabilidade P_X em $\mathbb{R}$ induzida por X é chamada de *lei da variável aleatória* X ou *distribuição de* X.

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

Exemplo 2.6. Um dado é lançado três vezes. Seja X o valor obtido no primeiro lançamento. Esse experimento pode ser modelado por $\Omega = \{1,2,3,4,5,6\}^3$, $\mathcal{F} = \mathcal{P}(\Omega)$ e $P(A) = \frac{\#A}{216}$ para todo $A \in \mathcal{F}$, nesse caso X é dado por

$$\begin{aligned} X : \Omega &\to \mathbb{R} \\ \omega &\mapsto \omega_1, \end{aligned}$$

onde cada $\omega \in \Omega$ é identificado como uma tripla $(\omega_1, \omega_2, \omega_3)$. O espaço induzido por X é dado por $(\mathbb{R}, \mathcal{B}, P_X)$, com P_X dado por

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

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

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

---

ver [Jam04, Exercício 1.6] a respeito da existência e unicidade da menor σ-álgebra contendo uma classe de conjuntos qualquer.

ou

$$P_X([1,5\ ,\ 3,4]) = \frac{\#\{2,3\}}{6} = \frac{1}{3}.$$

Função de Distribuição

Definição 2.7 (Função de Distribuição)**.** A *função de distribuição*, ou *função de distribuição acumulada* da variável aleatória X, denotada por F_X, é definida como

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

A função de distribuição determina o comportamento estatístico da variável aleatória, e vice-versa. Mais precisamente, dadas X e Y variáveis aleatórias, $F_X(t) = F_Y(t)$ para todo $t \in \mathbb{R}$ se e somente se P_X e P_Y em $(\mathbb{R}, \mathcal{B})$ são iguais. Neste caso escrevemos $X \sim Y$. Por isso a função de distribuição é uma característica fundamental da variável aleatória.

Exemplo 2.8. Duas moedas honestas são lançadas. Seja a variável X que conta o número de caras observadas. Temos que

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

Observe que o salto da função de distribuição acumulada corresponde à probabilidade de a variável aleatória assumir aquele valor, como se vê na Figura 2.1.

Exemplo 2.9. Seja um experimento que consiste em selecionar um ponto ao acaso do intervalo $[a, b]$ com $a < b$. Seja X a variável aleatória que representa a coordenada do ponto.

Primeiro observamos que, ao selecionar um ponto ao acaso em um intervalo, estamos dizendo implicitamente que quaisquer subintervalos de mesmo tamanho têm a mesma probabilidade de conter o ponto escolhido. Isso implica que que, dado $[c, d] \subseteq [a, b]$, temos que $P(X \in [c, d]) = \frac{d-c}{b-a}$. Para $t \in [a, b]$, tomando $c = a$ temos que $P(X \leqslant t) = \frac{t-a}{b-a}$. Para $t < a$ temos que $P(X \leqslant t) = 0$, e para $t > a$

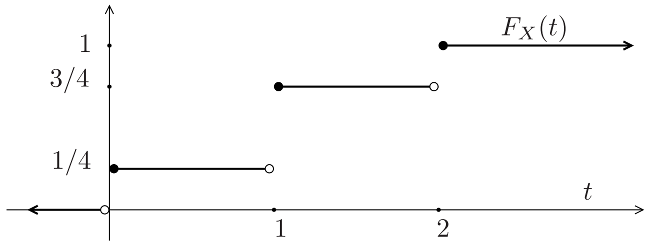

Figura 2.1: Gráfico de uma função de distribuição acumulada.

temos que $P(X \leqslant t) = 1$. Portanto,

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

cujo gráfico está ilustrado na Figura 2.2.

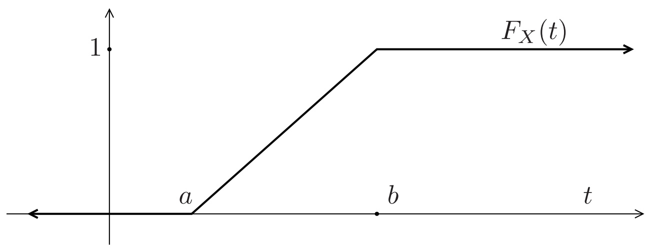

Figura 2.2: Gráfico de uma função de distribuição acumulada.

Proposição 2.10 (Propriedades da Função de Distribuição). *Se X é uma variável aleatória, sua função de distribuição F_X satisfaz as seguintes propriedades:*

1. *F_X é não-decrescente, i.e., $x \leqslant y \Rightarrow F_X(x) \leqslant F_X(y)$.*
2. *F_X é contínua à direita, i.e., $x_n \searrow x \Rightarrow F_X(x_n) \to F_X(x)$.*
3. *$\lim_{x\to-\infty} F_X(x) = 0$ e $\lim_{x\to+\infty} F_X(x) = 1$.*

Demonstração. Feita em aula. □

De forma geral, uma *função de distribuição* é qualquer função $F(\cdot)$ satisfazendo as três propriedades acima.

Teorema 2.11. *Dada uma função de distribuição F, existe um espaço de probabilidade e uma variável aleatória cuja função de distribuição é F.*

Demonstração. Omitida. Envolve Teoria da Medida. □

Não demonstraremos esse fato.

Proposição 2.12. *A função de distribuição de uma variável aleatória X satisfaz:*

1. $P(X > a) = 1 - F_X(a)$.
2. $P(a < X \leqslant b) = F_X(b) - F_X(a)$.
3. $P(a < X < b) = F_X(b-) - F_X(a)$.
4. $P(a \leqslant X < b) = F_X(b-) - F_X(a-)$.
5. $P(a \leqslant X \leqslant b) = F_X(b) - F_X(a-)$.
6. $P(X = a) = F_X(a) - F_X(a-)$.
7. $P(X = a) = 0$ *se e somente se* F_X *é contínua em* a.

Demonstração. Feita em aula. □

Exercício 2.1. Seja $F(x)$ a função

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

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

(a) $P(X > \frac{1}{8})$

(b) $P(\frac{1}{8} < X < \frac{2}{5})$

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

2.2 Variáveis Aleatórias Discretas

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

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

$$p_X(x) = P(X = x).$$

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

$$P_X(B) = \sum_{x \in B} p_X(x) \quad \forall B \in \mathcal{B}$$

e sua função de distribuição é dada por

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

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

$$\sum_{x \in B} p_X(x) = P_X(B \cap A) \leqslant P_X(B) \leqslant P_X(B \cap A) + P_X(A^c) = P_X(B \cap A),$$

onde a primeira igualdade vale pois $p_X(x) \leqslant P_X(A^c) = 0$ para todo $x \notin A$. $\square$

A função de distribuição de uma variável aleatória discreta tipicamente se parece à da Figura 2.1, sendo constante por partes e dando um saltos de tamanho $p_X(t)$ em cada ponto $t \in \mathbb{R}$ com $p_X(t) > 0$. Como

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

temos que a função de distribuição determina a função de probabilidade de uma variável aleatória, e pela equação anterior vale a recíproca. Portanto, para determinar o comportamento estatístico de uma variável aleatória discreta, é equivalente especificar p_X, F_X, ou P_X. A primeira normalmente é mais simples.

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

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

Dada uma função de probabilidade $p(\cdot)$, existe um espaço de probabilidade onde está definida uma variável aleatória discreta cuja função de probabilidade é $p(\cdot)$.

Exercício 2.2. A probabilidade de um indivíduo acertar um alvo é $\frac{2}{3}$. Ele deve atirar até atingir o alvo pela primeira vez. Seja X a variável aleatória que representa o número de tentativas até que ele acerte o alvo.

(a) Encontre a função de probabilidade de X.

(b) Mostre que p_X é função de probabilidade.

(c) Calcule a probabilidade de serem necessários exatamente cinco tiros para que ele acerte o alvo.

Exercício 2.3. Seja X uma variável aleatória com função de probabilidade $P(X = x) = cx^2$, onde c é uma constante e $x = 1, 2, 3, 4, 5$.

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

(b) Calcule $P(X$ ser ímpar).

Distribuição de Bernoulli Dizemos que X é Bernoulli com parâmetro $p \in [0, 1]$, o que denotamos $X \sim \text{Bernoulli}(p)$, se $p_X(1) = p$ e $p_X(0) = 1 - p$. Indicadores de eventos são Bernoulli e vice-versa. Às vezes associamos o evento $[X = 1]$ a "sucesso" e $[X = 0]$ a "fracasso".

Distribuição uniforme discreta Dado $I = \{x_1, x_2, \dots, x_k\}$, dizemos que X tem distribuição uniforme discreta em I, denotado por $X \sim U_d[I]$, se

$$p_X(x_i) = \frac{1}{k}, \quad i = 1, 2, \dots, k.$$

Exemplo 2.13. Lançamento de um dado. Temos $I = \{1, 2, 3, 4, 5, 6\}$ e $p(i) = \frac{1}{6}$, $i = 1, 2, \dots, 6$.

Distribuição binomial Considere n ensaios de Bernoulli independentes e com mesmo parâmetro p, e seja X o número de sucessos obtidos. Dizemos que X segue o modelo binomial com parâmetros n e p, o que denotamos por $X \sim b(n,p)$. A função de probabilidade de X é dada por

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

Exemplo 2.14. Lançar um dado 4 vezes e contar o número de vezes que se obtém o número 3. Temos $X \sim b(4, \frac{1}{6})$. A probabilidade de se obter 3 duas vezes é dada por

$$P(X=2) = p_X(2) = \tbinom{4}{2} \left(\tfrac{1}{6}\right)^2 \left(\tfrac{5}{6}\right)^{4-2} = \frac{4!}{2!(4-2)!} \frac{5^2}{6^4} = \frac{25}{216}.$$

Exercício 2.4. Seja X o número de caras obtidas em 4 lançamentos de uma moeda honesta. Construa a função de probabilidade e a função de distribuição de X e esboce os seus gráficos.

Distribuição geométrica Numa sequência de ensaios independentes com probabilidade de sucesso p, considere o número X de ensaios necessários para a obtenção de um sucesso. Dizemos que X segue o modelo geométrico de parâmetro p, e que denotamos por $X \sim \mathrm{Geom}(p)$, e sua função de probabilidade é dada por

$$p_X(n) = p(1-p)^{n-1}, \quad n = 1, 2, 3, 4, \dots.$$

Exemplo 2.15. Lançar um par de dados até obter números iguais. Se X denota o número de lançamentos necessários, então $X \sim \mathrm{Geom}(\frac{1}{6})$.

Distribuição hipergeométrica Suponha que numa caixa existem m bolas azuis e n bolas brancas, de onde retiramos r bolas ao acaso. Contamos o número X de bolas azuis retiradas. Se após cada retirada a bola fosse devolvida à caixa, teríamos um experimento *com reposição*, e $X \sim b(r, \frac{m}{m+n})$. No caso em que as bolas retiradas são guardadas fora da caixa, temos um experimento *sem reposição*, e nesse caso X segue o modelo hipergeométrico com parâmetros m, n e r, denotado por $X \sim \mathrm{Hgeo}(m,n,r)$. A função de probabilidade de X é dada por

$$p_X(k) = \frac{\binom{m}{k}\binom{n}{r-k}}{\binom{m+n}{r}}, \quad \text{para } [0 \vee r-n] \leqslant k \leqslant [r \wedge m].$$

Denotamos por $a \vee b$ e $a \wedge b$ o máximo e o mínimo entre a e b, respectivamente.

Exemplo 2.16. Num jogo de bingo com 50 pedras, conta-se o número X de pedras pares sorteadas nas 10 primeiras retiradas. Neste caso, $X \sim \text{Hgeo}(25, 25, 10)$.

Exemplo 2.17. No jogo de buraco um jogador recebe 11 cartas de um baralho francês de 52 cartas. Conta-se o número X de cartas de espadas $\spadesuit$. Neste caso, $X \sim \text{Hgeo}(13, 39, 11)$.

Distribuição de Poisson Imagine uma grande quantidade de determinados objetos (estrelas, chamadas telefônicas, uvas-passas, etc.) uniformemente distribuídas em uma certa região (o céu, a linha do tempo, uma massa de panetone, etc.) também muito grande, sendo λ a proporção entre a quantidade de objetos e o tamanho dessa região. Se contamos o número X de objetos encontrados em uma unidade de volume dessa região, temos que X segue o modelo de Poisson com parâmetro λ, denotado por $X \sim \text{Poisson}(\lambda)$, com função de probabilidade

$$p_X(k) = \frac{e^{-\lambda}\lambda^k}{k!}, \quad k = 0, 1, 2, 3, \dots.$$

De fato, se temos n grande e $p_n = \frac{\lambda}{n}$, então para cada k fixo temos

$$P(X = k) = \tbinom{n}{k} \left(\tfrac{\lambda}{n}\right)^k \left(1 - \tfrac{\lambda}{n}\right)^{n-k} = \tfrac{\lambda^k}{k!} \left(\tfrac{n}{n} \tfrac{n-1}{n} \cdots \tfrac{n-k+1}{n}\right) \left(1 - \tfrac{\lambda}{n}\right)^{n-k} \to \frac{e^{-\lambda}\lambda^k}{k!}.$$

Exemplo 2.18. Se em 1.000 horas de serviço uma operadora recebe 50.000 chamadas, essas chamadas acontecendo em instantes independentes e uniformemente distribuídas ao longo dessas 1.000 horas, então a distribuição da quantidade X de chamadas recebidas em 1 hora é bem aproximada por $X \sim \text{Poisson}(50)$.

2.3 Variáveis Aleatórias Contínuas

Definição 2.19. Uma variável aleatória X é dita *contínua* se $P(X = a) = 0$ para todo $a \in \mathbb{R}$, ou seja, se F_X for uma função contínua.

Definição 2.20. Dizemos que uma variável aleatória X, sua função de distribuição

F_X e sua lei P_X são *absolutamente contínuas* se existe $f_X(\cdot) \geqslant 0$ tal que

$$P_X(B) = P(X \in B) = \int_B f_X(x)\,\mathrm{d}x \qquad \forall B \in \mathcal{B}.$$

Neste caso, dizemos que f_X é a *função de densidade de probabilidade de* X, ou simplesmente *densidade* de X.

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

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

Exemplo 2.21. Sortear um número em $[0, 1]$. Definimos

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

e neste caso temos

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

A densidade f_X pode ser obtida por

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

para "quase" todo $x \in \mathbb{R}$, isto é, para todo x exceto talvez em conjunto pequeno.[2] Portanto, para especificar a distribuição ou a lei de uma variável

---

[2] Dizemos que um conjunto $A \in \mathcal{B}$ é *pequeno*, isto é, tem *medida zero*, se, para todo $\varepsilon > 0$, existe uma sequência de intervalos (a_n, b_n) cuja união contenha A e cujo tamanho total seja pequeno, isto é, $\sum_{n=1}^{\infty}(b_n - a_n) \leqslant \varepsilon$. Por exemplo, se $A = \{x_1, x_2, \dots\}$ é enumerável, então podemos tomar a sequência de intervalos $(x_n - 2^{-n-1}\varepsilon, x_n + 2^{-n-1}\varepsilon)$, que contém A e cujo tamanho total é exatamente ε. Podemos modificar a densidade f_X em um conjunto pequeno de pontos e ainda teremos uma densidade para X, pois um conjunto pequeno não altera o valor da integral $\int_B f_X(x)\,\mathrm{d}x$.

aleatória absolutamente contínua, é suficiente saber sua função de densidade, e vice-versa.

Observação 2.22. Uma função $f(\cdot)$ satisfazendo

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

é chamada *função de densidade*.

Exercício 2.5. Seja X uma variável aleatória absolutamente contínua tal que sua função de densidade é par, isto é, $f_X(x) = f_X(-x)$. Mostre que

(a) $F_X(x) = 1 - F_X(-x)$;

(b) $F_X(0) = \frac{1}{2}$;

(c) $P(-x < X < x) = 2F_X(x) - 1$, $x > 0$;

(d) $P(X > x) = \frac{1}{2} - \int_0^x f_X(t)\mathrm{d}t$, $x > 0$.

Exercício 2.6. Seja Z uma variável aleatória contínua com função de densidade de probabilidade

$$f_Z(z) = \begin{cases} 10\, e^{-10z}, & z > 0 \\ 0, & z \leqslant 0 \end{cases}$$

Obtenha a função de distribuição de Z e esboce o seu gráfico.

Distribuição uniforme Dizemos que a variável aleatória X tem distribuição uniforme no intervalo $[a, b]$, denotado por $X \sim U[a, b]$, se todos os subintervalos de $[a, b]$ com mesmo comprimento tiverem a mesma probabilidade. Sua densidade é

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

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

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

Exemplo 2.23. O ponto de ruptura X de algum cabo numa rede elétrica de 5 km pode ser modelado por uma variável aleatória com distribuição uniforme em $[0, 5]$. Neste caso temos que $f_X = \frac{1}{5}\mathbb{1}_{[0,5]}$. A probabilidade de um determinado cabo se romper nos primeiros 800m da rede é igual a $\int_0^{0,8} \frac{1}{5} \mathrm{d}x = 16\%$.

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

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

A distribuição exponencial se caracteriza por ter uma função de taxa de falha constante, o que chamamos de perda de memória.

Exemplo 2.24. Quando se diz que uma lâmpada incandescente de uma determinada marca tem vida média de 1.000 horas, isso quer dizer que seu tempo de vida T satisfaz $T \sim \exp(\frac{1}{1000})$.

A distribuição exponencial pode ser pensada como como o limite de distribuições geométricas com pequenos intervalos de tempo. Isto é, se $X \sim \frac{1}{n} \operatorname{Geom}(\frac{\lambda}{n})$ com n muito grande, então a distribuição de X se aproxima da distribuição exponencial com parâmetro λ. Essa é a distribuição adequada para modelar a vida útil de uma lâmpada, ou de inúmeros outros materiais, como óleos isolantes, porque estes deixam de funcionar não por deterioração ao longo do tempo mas sim porque um determinado evento passível de causar a falha pode ocorrer a qualquer instante com uma probabilidade muito pequena.

Distribuição gama A distribuição gama tem dois parâmetros, α e β, e inclui como casos particulares a distribuição exponencial e as chamadas qui-quadrado e Erlang. Dizemos que X tem distribuição gama com parâmetros positivos α e β, denotado por $X \sim \operatorname{Gama}(\alpha, \beta)$, se X tem densidade dada por

$$f_X(x) = \begin{cases} \dfrac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)}, & x \geqslant 0, \\ 0, & x < 0, \end{cases}$$

onde

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x}\,\mathrm{d}x.$$

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

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

A distribuição $\mathcal{N} = \mathcal{N}(0,1)$ é chamada *normal padrão*.

Denotamos por Φ a função de distribuição acumulada de uma normal padrão $\mathcal{N}$, dada por

$$\Phi(t) = F_{\mathcal{N}}(t) = P(\mathcal{N} \leqslant t) = \int_{-\infty}^{t} \frac{e^{-x^2/2}}{\sqrt{2\pi}}\,\mathrm{d}x.$$

Em geral, a solução de problemas numéricos envolvendo a distribuição normal inclui a consulta de uma tabela de valores de $(\Phi(t); t \geqslant 0)$ com os valores de t apropriados. Na Tabela 2.1 exibimos os valores de $\Phi(t)$ para $t = 0,00,\ 0,01,\ 0,02,\ \ldots,\ 3,49$.

Para $t < 0$ usa-se a identidade

$$\Phi(-t) = 1 - \Phi(t).$$

Consequentemente,

$$\begin{aligned}
P(+a < \mathcal{N} < +b) &= \Phi(b) - \Phi(a) \\
P(-a < \mathcal{N} < -b) &= \Phi(-b) - \Phi(-a) = \Phi(a) - \Phi(b) \\
P(-a < \mathcal{N} < +b) &= \Phi(b) - \Phi(-a) = \Phi(b) + \Phi(a) - 1.
\end{aligned}$$

Em particular,

$$P(-a < \mathcal{N} < a) = 2\Phi(a) - 1.$$

Exemplo 2.25. Calculemos as seguintes probabilidades:

(a) $P(0 < \mathcal{N} < 1) = \Phi(1) - \Phi(0) \approx 0,8413 - 0,5000 = 0,3413$.

(b) $P(-1.93 < \mathcal{N} < 3) = \Phi(1.93) + \Phi(3) - 1 \approx 0,9732 + 0,9988 - 1 = 0,9720$.

(c) $P(-1.8 < \mathcal{N} < 1.8) = 2\Phi(1.8) - 1 \approx 2 \times 0,9641 - 1 = 0,9282$.

Tabela 2.1: $\Phi(x+y)$, onde x são os valores das linhas e y os das colunas.

	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,0	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990
3,1	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,2	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995
3,3	0,9995	0,9995	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997
3,4	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998

(d) Para qual x tem-se $P(-x < \mathcal{N} < x) = 0,90$?
$2\Phi(x) - 1 = 0,90 \;\Rightarrow\; \Phi(x) = 0,95 \;\Rightarrow\; x \approx 1,645.$

(e) Para qual x tem-se $P(-x < \mathcal{N} < x) = 0,6826$?
$2\Phi(x) - 1 = 0,6826 \;\Rightarrow\; \Phi(x) = 0,8413 \;\Rightarrow\; x \approx 1,000.$

Exercício 2.7. Mostre que, se $Y = aX + b$ com $a > 0$ e $b \in \mathbb{R}$, então $f_Y(y) = \frac{1}{a} f_X(\frac{y-b}{a})$. Sugestão: determine $F_Y(y)$, $y \in \mathbb{R}$, em termos de $f_X(x)$, $x \in \mathbb{R}$, sabendo que $F_X(t) = \int_{-\infty}^{t} f_X(x)\,\mathrm{d}x$, e depois tome a derivada.

Exercício 2.8. Mostre que se $X \sim \mathcal{N}(\mu, \sigma^2)$ então a variável aleatória $\frac{X-\mu}{\sigma}$ tem distribuição normal padrão.

2.4 Distribuições Mistas e Singulares

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

Uma variável aleatória absolutamente contínua X vive em $\mathbb{R}$, sua distribuição em cada intervalo $(n, n+1]$ é similar à de uma distribuição uniforme, apenas seu peso é ponderado pela função f_X. Nesse contexto, tudo se expressa em termos de integrais com $f_X(x)\,\mathrm{d}x$.

Existem variáveis aleatórias que são misturas dos tipos discreto e absolutamente contínuo. Neste caso, a variável pode ser decomposta, separando-se as suas partes discreta e absolutamente contínua, e suas propriedades serão determinadas por combinações de somatórios e integrais. Mais precisamente, dizemos que X é uma *variável aleatória mista com componentes discreta e absolutamente contínua* se existem p_X e f_X tais que

$$P(X \in B) = \sum_{x \in B} p_X(x) + \int_B f_X(x)\,\mathrm{d}x, \quad B \in \mathcal{B}.$$

Além desses casos, existem variáveis aleatórias cuja parte contínua não é absolutamente contínua. Por um lado, nenhum ponto em particular tem probabilidade positiva de ocorrer, o que afasta o tratamento por somatórios do caso discreto. Por outro lado, sua distribuição não é similar à de uma distribuição uniforme, pois tais variáveis aleatórias vivem conjuntos pequenos da reta, não sendo aplicável tampouco o uso de integrais em $f(x)\mathrm{d}x$ para nenhuma f. A tais variáveis chamamos de *singulares*. Toda variável aleatória pode ser decomposta em suas partes discreta, absolutamente contínua, e singular. O leitor pode ler mais a respeito em [Jam04, pp. 44-48], e nas referências ali citadas.

Exemplo 2.26. Imagine uma pessoa sempre se desperta num horário S com distribuição normal de média 07:00 e desvio-padrão de 1 hora. Representemos como números reais os horários medidos em horas, relativos à meia-noite, de maneira que $S \sim \mathcal{N}(7,1)$. Colocamos um alarme para despertar essa pessoa às 08:00, que funciona com probabilidade $\frac{1}{2}$. Nesse caso, ela vai se despertar às 08:00 com probabilidade 7,94%, ou antes das 08:00 seguindo a densidade original, ou depois das 08:00 seguindo uma densidade normal atenuada. Mais precisamente, essa pessoa vai se despertar em um horário T cuja distribuição é dada por

$$f_T(t) = \begin{cases} \frac{1}{\sqrt{2\pi}} e^{(t-7)^2/2}, & t < 8, \\ \frac{1}{\sqrt{8\pi}} e^{(t-7)^2/2}, & t > 8, \end{cases} \qquad p_T(t) = \begin{cases} 0,0794, & t = 8, \\ 0, & t \neq 8. \end{cases}$$

A probabilidade de essa pessoa se despertar entre 06:00 e 09:00 é dada por

$$\begin{aligned} P(6 \leqslant T \leqslant 9) &= \sum_{6 \leqslant t \leqslant 9} p_T(t) + \int_6^9 f_T(t)\,\mathrm{d}t \\ &= 0,0794 + \int_6^8 \frac{1}{\sqrt{2\pi}} e^{(t-7)^2/2} \mathrm{d}t + \int_8^9 \frac{1}{\sqrt{8\pi}} e^{(t-7)^2/2} \mathrm{d}t \\ &= 0,0794 + [\Phi(1) - \Phi(-1)] + \tfrac{1}{2}[\Phi(2) - \Phi(1)] \\ &= 0,0794 + \tfrac{3}{2}\Phi(1) + \tfrac{1}{2}\Phi(2) - 1 = 82,99\% \end{aligned}$$

com quatro algarismos significativos.

2.5 Distribuição Condicional dado um Evento

Dado um evento A com $P(A) > 0$, definimos a *função de distribuição condicional de X dado A*

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

Exemplo 2.27. Considere dois lançamentos de uma moeda honesta e seja X o número de "caras" obtidas. Temos

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

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

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

Se X é discreta, definimos ainda a *função de probabilidade condicional de X dado A*, $p_X(\,\cdot\,|A)$ ou $p_{X|A}(\,\cdot\,)$, como a função de probabilidade associada à função de distribuição $F_X(\,\cdot\,|A)$. No exemplo acima, temos

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

Se X é absolutamente contínua, definimos a *função de densidade condicional de X dado A*, $f_X(\,\cdot\,|A)$ ou $f_{X|A}(\,\cdot\,)$, como a densidade associada à função de distribuição $F_X(\,\cdot\,|A)$.

2.6 Exercícios

2.9. Mostre que, se duas variáveis aleatórias X e Y são iguais quase certamente, isto é, $P(X = Y) = 1$, então $F_X = F_Y$.

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

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

2.11. Seja X o número de caras obtidas em 4 lançamentos de uma moeda honesta. Determine a função de probabilidade de X. Desenhe o gráfico da função de distribuição da variável aleatória X.

2.12. Se

$$f(t) = \begin{cases} e^{-3t} + c\, e^{-t}, & t > 0, \\ 0, & t \leqslant 0, \end{cases}$$

é função de densidade, ache c.

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

2.14. Mostre que a função de probabilidade do modelo de Poisson é de fato uma função de probabilidade.

2.15. Perda de memória do modelo geométrico.

1. Mostre que $P(X \geqslant m + n | X > n) = P(X \geqslant m)$ para inteiros não-negativos, se X segue o modelo geométrico.
2. Se X segue o modelo geométrico, prove que a distribuição de X dado que $X > n$ é igual à distribuição de $X + n$.

2.16. Mostre que a densidade do modelo uniforme contínuo é de fato uma função de densidade.

2.17. Mostre que a distribuição do modelo exponencial é de fato uma distribuição. Calcule a densidade associada.

2.18. Seja X uma variável aleatória em $(\Omega, \mathcal{F}, P)$ com distribuição exponencial de parâmetro $\lambda > 0$. Considere $N = \lceil X \rceil$, o menor inteiro maior ou igual a X. Encontre a distribuição de N.

2.19. Uma pesquisa eleitoral determinou que a intenção de voto do Candidato A é de 46%, com margem de erro de 3%, para mais ou para menos. Ou seja, a intenção de voto desse candidato tem distribuição normal com média $\mu = 46\%$ e desvio-padrão $\sigma = 3\%$. Calcule a probabilidade de o Candidato A ter mais de 50% das intenções de voto.

2.20. Uma caixa contém 10 parafusos, cujos tamanhos são normais independentes, com média $21,4\,\text{mm}$ e desvio-padrão $0,5\,\text{mm}$. Calcule a probabilidade de que nenhum dos parafusos tenha mais de $22\,\text{mm}$.

2.21. Perda de memória do modelo exponencial.

1. Mostre que $P(X > t + s | X > s) = P(X > t)$ para $t, s \geqslant 0$ se X tem distribuição exponencial.
2. Mostre que a distribuição de X dado que $X > s$ é igual à distribuição de $X + s$.

2.22. Se $X \sim \exp(\lambda)$ e $Y = 5X$, ache a distribuição acumulada de Y. Ache a função de distribuição condicional e a densidade condicional de Y dado que $X > 3$.

2.23. [Jam04, Capítulo 2]. Recomendados: 1, 5, 6, 7, 9, 10, 13, 14.

Capítulo 3

Vetores Aleatórios

Imagine que queremos produzir duas variáveis aleatórias com distribuição Bernoulli($\frac{1}{2}$). A forma mais natural seria lançar uma moeda duas vezes e considerar o par $\boldsymbol{X} = (Z, W)$. Uma outra forma de fazê-lo seria, por exemplo, lançar a moeda apenas uma vez e copiar o resultado: $\boldsymbol{Y} = (Z, Z)$.

Em ambos os casos, produziu-se um par de variáveis aleatórias distribuídas como Bernoulli($\frac{1}{2}$). Entretanto, o comportamento *conjunto* dessas variáveis aleatórias é bem diferente nos dois casos.

Neste capítulo vamos estudar as principais propriedades dos *vetores aleatórios*, isto é, a combinação de muitas variáveis aleatórias em que se considera seu comportamento estatístico conjunto.

3.1 Vetores Aleatórios

Começamos com um pouco de notação vetorial. $\boldsymbol{x} \in \mathbb{R}^d$ representa uma d-upla de números reais, $\boldsymbol{x} = (x_1, x_2, \dots, x_d)$. Uma função $\boldsymbol{X}$ em Ω associa a cada ω uma d-upla, i.e., um vetor $\boldsymbol{X}(\omega) = (X_1(\omega), X_2(\omega), \dots, X_d(\omega))$.

Denotamos por $\boldsymbol{x} \leqslant \boldsymbol{y}$ o conjunto de desigualdades $x_i \leqslant y_i, i = 1, \dots, d$, isto é, $\boldsymbol{x} \leqslant \boldsymbol{y}$ se, e somente se, vale a desigualdade para todas as coordenadas simultaneamente. Analogamente denotamos por $\boldsymbol{x} < \boldsymbol{y}$ o conjunto de desigualdades $x_i < y_i,\ i = 1, \dots, d$. Dados $\boldsymbol{a} \leqslant \boldsymbol{b}$, denotamos por $[\boldsymbol{a}, \boldsymbol{b}]$ o conjunto $\{x \in \mathbb{R}^d : \boldsymbol{a} \leqslant$

$\boldsymbol{x} \leqslant \boldsymbol{b}\}$. Analogamente para $(\boldsymbol{a}, \boldsymbol{b}]$, etc.

Definição 3.1 (Vetor aleatório)**.** Um *vetor aleatório* $\boldsymbol{X} = (X_1, \dots, X_d)$ é uma função $\boldsymbol{X} : \Omega \to \mathbb{R}^d$ tal que cada coordenada X_i é uma variável aleatória.

Espaço de probabilidade induzido e lei de um vetor aleatório Como na reta, a σ-álgebra de Borel no espaço Euclidiano $\mathbb{R}^d$, denotada por $\mathcal{B}^d$, é a menor σ-álgebra que contém todos os octantes $\{\boldsymbol{x} \in \mathbb{R}^d : \boldsymbol{x} \leqslant \boldsymbol{t}\}$, $\boldsymbol{t} \in \mathbb{R}^d$. Dado um espaço de probabilidade $(\Omega, \mathcal{F}, P)$ e um vetor aleatório $\boldsymbol{X}$, definimos o *espaço de probabilidade induzido por* $\boldsymbol{X}$ como $(\mathbb{R}^d, \mathcal{B}^d, P_{\boldsymbol{X}})$, onde

$$P_{\boldsymbol{X}}(B) = P\big(\{\omega : \boldsymbol{X}(\omega) \in B\}\big), \quad B \in \mathcal{B}^d.$$

Ou seja, o espaço amostral é o conjunto dos vetores d-dimensionais, os eventos aleatórios são os conjuntos Borelianos, e a medida de probabilidade é aquela induzida por $\boldsymbol{X}$. Chamaremos de *lei do vetor aleatório* $\boldsymbol{X}$ a medida de probabilidade $P_{\boldsymbol{X}}$ em $\mathbb{R}^d$ induzida por $\boldsymbol{X}$.

Função de Distribuição Conjunta

Definição 3.2 (Função de Distribuição Conjunta)**.** A *função de distribuição conjunta* de um vetor aleatório $\boldsymbol{X}$, denotada por $F_{\boldsymbol{X}}$, é uma função $F_{\boldsymbol{X}} : \mathbb{R}^d \to \mathbb{R}$ dada por

$$F_{\boldsymbol{X}}(\boldsymbol{t}) = P\big(\boldsymbol{X} \leqslant \boldsymbol{t}\big).$$

Exemplo 3.3. Lançamos duas moedas honestas e consideramos $X_1 =$ quantidade de caras, $X_2 = 1$ se os resultados forem iguais, 0 se forem diferentes, e $\boldsymbol{X} = (X_1, X_2)$. Temos então

$$P(\boldsymbol{X} \leqslant \boldsymbol{t}) = \begin{cases} 0, & t_1 < 0 \text{ ou } t_2 < 0, \quad \text{pois } [\boldsymbol{X} \leqslant \boldsymbol{t}] = \emptyset; \\ 0, & t_1, t_2 \in [0,1), \quad \text{pois } [\boldsymbol{X} \leqslant \boldsymbol{t}] = [X_1 = 0, X_2 = 0] = \emptyset; \\ \frac{1}{2}, & t_1 \geqslant 1, t_2 \in [0,1), \quad \text{pois } [\boldsymbol{X} \leqslant \boldsymbol{t}] = [X_1 = 1, X_2 = 0]; \\ \frac{1}{4}, & t_1 \in [0,1), t_2 \geqslant 1, \quad \text{pois } [\boldsymbol{X} \leqslant \boldsymbol{t}] = [X_1 = 0, X_2 = 0]; \\ \frac{3}{4}, & t_1 \in [1,2), t_2 \geqslant 1, \quad \text{pois } [\boldsymbol{X} \leqslant \boldsymbol{t}] = [X_1 = 0 \text{ ou } 1]; \\ 1, & t_1 \geqslant 2, t_2 \geqslant 1, \quad \text{pois } [\boldsymbol{X} \leqslant \boldsymbol{t}] = \Omega. \end{cases}$$

Os valores de $F_{\boldsymbol{X}}$ são ilustrados na Figura 3.1.

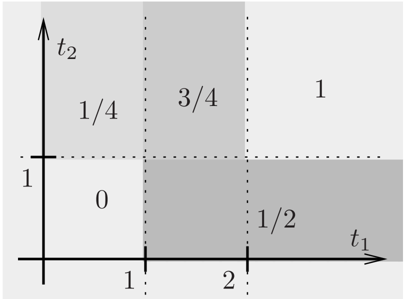

Figura 3.1: Valores assumidos por $F_{\boldsymbol{X}}(t_1, t_2)$ para cada $(t_1, t_2) \in \mathbb{R}^2$.

Considere o operador $\Delta^i_{a,b}$ sobre funções de $\mathbb{R}^d$ em $\mathbb{R}$, dado por

$$\left(\Delta^i_{a,b} F\right)(\boldsymbol{x}) = F(x_1, \dots, x_{i-1}, b, x_{i+1}, \dots, x_d) - F(x_1, \dots, x_{i-1}, a, x_{i+1}, \dots, x_d).$$

Note que a função $\Delta^i_{a,b} F$ não depende da i-ésima coordenada de $\boldsymbol{x}$.

Proposição 3.4. *Para* $\boldsymbol{a} \leqslant \boldsymbol{b} \in \mathbb{R}^d$, $\Delta^1_{a_1,b_1} \cdots \Delta^d_{a_d,b_d} F_{\boldsymbol{X}} = P(\boldsymbol{a} < \boldsymbol{X} \leqslant \boldsymbol{b})$.

Demonstração. Para quaisquer $\boldsymbol{x}, \boldsymbol{a} \leqslant \boldsymbol{b}$, temos

$$\begin{aligned} \Delta^d_{a_d,b_d} F_{\boldsymbol{X}}(\boldsymbol{x}) &= P(X_1 \leqslant x_1, \dots, X_{d-1} \leqslant x_{d-1}, X_d \leqslant b_d) - \\ &\quad - P(X_1 \leqslant x_1, \dots, X_{d-1} \leqslant x_{d-1}, X_d \leqslant a_d) = \\ &= P(X_1 \leqslant x_1, \dots, X_{d-1} \leqslant x_{d-1}, a_d < X_d \leqslant b_d), \end{aligned}$$

e sucessivamente obtemos

$$\begin{aligned} &\Delta^j_{a_j,b_j} \cdots \Delta^d_{a_d,b_d} F_{\boldsymbol{X}}(\boldsymbol{x}) = \left[\Delta^j_{a_j,b_j} \left(\Delta^{j+1}_{a_{j+1},b_{j+1}} \cdots \Delta^d_{a_d,b_d} F_{\boldsymbol{X}}\right)\right](\boldsymbol{x}) = \\ &= P(X_1 \leqslant x_1, \dots, X_{j-1} \leqslant x_{j-1}, X_j \leqslant b_j, a_{j+1} < X_{j+1} \leqslant b_{j+1}, \dots, a_d < X_d \leqslant b_d) - \\ &- P(X_1 \leqslant x_1, \dots, X_{j-1} \leqslant x_{j-1}, X_j \leqslant a_j, a_{j+1} < X_{j+1} \leqslant b_{j+1}, \dots, a_d < X_d \leqslant b_d) = \\ &= P(X_1 \leqslant x_1, \dots, X_{j-1} \leqslant x_{j-1}, a_j < X_j \leqslant b_j, \dots, a_d < X_d \leqslant b_d). \end{aligned}$$

Tomando $j = 1$ temos

$$\Delta^1_{a_1,b_1} \cdots \Delta^d_{a_d,b_d} F_{\boldsymbol{X}}(\boldsymbol{x}) = P(a_1 < X_1 \leqslant b_1, \dots, a_d < X_d \leqslant b_d). \qquad \square$$

Proposição 3.5 (Propriedades da Função de Distribuição Conjunta)**.** *Se* $\boldsymbol{X}$ *é um vetor aleatório em* $(\Omega, \mathcal{F}, P)$*, então sua função de distribuição* $F_{\boldsymbol{X}}$ *goza das seguintes propriedades:*

1. $F_{\boldsymbol{X}}$ é não-decrescente em cada uma de suas coordenadas.
2. $F_{\boldsymbol{X}}$ é contínua à direita em cada uma de suas coordenadas.
3. Se $(\boldsymbol{x}^k)_k$ é tal que, para algum j, $x_j^k \to -\infty$, então $F_{\boldsymbol{X}}(\boldsymbol{x}) \to 0$.
4. Se $(\boldsymbol{x}^k)_k$ é tal que, para todo j, $x_j^k \to +\infty$, então $F_{\boldsymbol{X}}(\boldsymbol{x}) \to 1$.
5. Para $\boldsymbol{a} \leqslant \boldsymbol{b} \in \mathbb{R}^d$, $\Delta^1_{a_1,b_1} \cdots \Delta^d_{a_d,b_d} F_{\boldsymbol{X}} \geqslant 0$.

Demonstração. Feita em aula. □

Contra-Exemplo 3.6. Considere a seguinte função:

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

Então $\Delta^1_{0,1}\Delta^2_{0,1}F = F(1,1) - F(1,0) - F(0,1) + F(0,0) = 1 - 1 - 1 + 0 = -1 < 0$. Portanto, F não pode ser função de distribuição conjunta, ainda que satisfaça as Propriedades 1–4.

Função de distribuição marginal

A partir da função de distribuição conjunta, pode-se obter o comportamento de cada variável isoladamente.

A função de distribuição de uma das coordenadas do vetor $\boldsymbol{X}$ é denominada *função de distribuição marginal* e é obtida da seguinte forma:

$$F_{X_j}(x_j) = \lim_{\substack{x_i \to \infty \\ i \neq j}} F_{\boldsymbol{X}}(x_1, \dots, x_d),$$

em que o limite é aplicado em todas as coordenadas, exceto j.

Demonstração. Feita em aula. □

Exemplo 3.7. No Exemplo 3.3, temos

$$F_{X_1}(t) = \begin{cases} 0, & t < 0, \\ \frac{1}{4}, & 0 \leqslant t < 1, \\ \frac{3}{4}, & 1 \leqslant t < 2, \\ 1, & t \geqslant 2, \end{cases} \qquad F_{X_2}(t) = \begin{cases} 0, & t < 0, \\ \frac{1}{2}, & 0 \leqslant t < 1, \\ 1, & t \geqslant 1. \end{cases}$$

3.2 Tipos de Vetores Aleatórios

Os principais tipos de vetores aleatórios são o discreto, o absolutamente contínuo, e o misto com componentes discreta e absolutamente contínua. Porém, há muitos exemplos de vetores aleatórios que não são de nenhum desses tipos, e esses exemplos não são tão artificiais como as variáveis aleatórias singulares.

Vetores Aleatórios Discretos

Definição 3.8. Dizemos que um vetor aleatório $\boldsymbol{X}$, sua função de distribuição $F_{\boldsymbol{X}}$ e sua lei $P_{\boldsymbol{X}}$ são *discretos* se existem $\{\boldsymbol{x}^1, \boldsymbol{x}^2, \boldsymbol{x}^3, \dots\}$ tais que $P\big(\boldsymbol{X} \in \{\boldsymbol{x}^1, \boldsymbol{x}^2, \boldsymbol{x}^3, \dots\}\big) = 1$. Neste caso, a *função de probabilidade* de $\boldsymbol{X}$ é dada por

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

Um vetor aleatório $\boldsymbol{X}$ é discreto se e somente se suas coordenadas $X_1, \dots, X_d$ são discretas. Uma função $p(\cdot)$ satisfazendo

$$\sum_{\boldsymbol{x}} p(\boldsymbol{x}) = 1 \qquad \text{e} \qquad p(\boldsymbol{x}) \geqslant 0, \quad \forall \boldsymbol{x} \in \mathbb{R}^d$$

é chamada *função de probabilidade conjunta.*

Função de probabilidade marginal A *função de probabilidade marginal* de uma variável X_i é obtida somando-se nas demais variáveis:

$$p_{X_i}(x_i) = P(X_i = x_i) = \sum_{x_1} \cdots \sum_{x_{i-1}} \sum_{x_{i+1}} \cdots \sum_{x_d} p(x_1, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_d).$$

Demonstração. Feita em aula. □

Exercício 3.1. No Exemplo 3.3, obtenha a função de probabilidade de $\boldsymbol{X}$, e as funções de probabilidade marginais de X_1 e X_2.

Vetores Aleatórios Absolutamente Contínuos

Dizemos que um vetor aleatório $\boldsymbol{X}$, sua função de distribuição $F_{\boldsymbol{X}}$ e sua lei $P_{\boldsymbol{X}}$ são *absolutamente contínuos* se existe $f_{\boldsymbol{X}}(\cdot) \geqslant 0$ tal que

$$P(\boldsymbol{X} \in B) = \int_B f_{\boldsymbol{X}}(\boldsymbol{x})\,\mathrm{d}^d\boldsymbol{x} \qquad \forall B \in \mathcal{B}^d.$$

Neste caso, dizemos que $f_{\boldsymbol{X}}$ é a *função de densidade conjunta de* $\boldsymbol{X}$, ou simplesmente *densidade* de $\boldsymbol{X}$.

A função de distribuição conjunta $F_{\boldsymbol{X}}$ pode ser calculada integrando-se a função de densidade conjunta $f_{\boldsymbol{X}}$ em cada coordenada:

$$F_{\boldsymbol{X}}(\boldsymbol{t}) = \int_{-\infty}^{t_1} \cdots \int_{-\infty}^{t_d} f_{\boldsymbol{X}}(\boldsymbol{x})\,\mathrm{d}x_d \cdots \mathrm{d}x_1$$

e, por outro lado, esta sempre pode ser calculada derivando-se aquela também em cada coordenada:

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{\partial^d}{\partial x_1 \cdots \partial x_d} F_{\boldsymbol{X}}(x_1, \dots, x_d).$$

para "quase" todo $\boldsymbol{x} \in \mathbb{R}^d$, isto é, para todo $\boldsymbol{x}$ exceto talvez em conjunto pequeno.[1]

Exemplo 3.9. Seja $G \in \mathbb{R}^d$ uma região tal que $\operatorname{Vol} G > 0$, onde $\operatorname{Vol} G$ é o volume

[1] Dizemos que um Boreliano $A \in \mathcal{B}^d$ é *pequeno*, isto é, tem *medida zero*, se, para todo $\varepsilon > 0$, existe uma sequência de paralelepípedos $(\boldsymbol{a}^j, \boldsymbol{b}^j)$ cuja união contenha A e cujo tamanho total seja pequeno, isto é, $\sum_{j=1}^{\infty}(b_1^j - a_1^j)\cdots(b_d^j - a_d^j) \leqslant \varepsilon$. Por exemplo, se $A = \{(x,y) : x > 0, y = 0\}$ é uma semi-reta no plano, então podemos tomar a sequência $(j-1, j) \times (-2^{-j-1}\varepsilon, 2^{-j-1}\varepsilon)$. Essa sequência contém a semi-reta A e seu tamanho total é exatamente ε. Podemos modificar a densidade f_X em um conjunto pequeno de pontos e ainda teremos uma densidade para X, pois um conjunto pequeno não altera o valor da integral $\int_B f_{\boldsymbol{X}}(\boldsymbol{x})\,\mathrm{d}^d\boldsymbol{x}$.

d-dimensional de G. Dizemos que $\boldsymbol{X} = (X_1, X_2, \ldots, X_d)$ com função de densidade

$$f_{\boldsymbol{X}}(x_1, \ldots, x_d) = \begin{cases} \frac{1}{\operatorname{Vol} G}, & (x_1, \ldots, x_d) \in G \\ 0, & (x_1, \ldots, x_d) \notin G \end{cases}$$

é *uniformemente distribuído em* G.

Uma função $f(\cdot)$ satisfazendo

$$f(\boldsymbol{x}) \geqslant 0, \quad \forall \boldsymbol{x} \in \mathbb{R}^d \qquad \text{e} \qquad \int_{\mathbb{R}^d} f(\boldsymbol{x})\, \mathrm{d}^d \boldsymbol{x} = 1$$

é chamada *função de densidade conjunta*.

Densidade marginal A densidade de uma variável X_i é chamada *densidade marginal*, e pode ser calculada por

$$f_{X_i}(x_i) = \underbrace{\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty}}_{d-1 \text{ vezes}} f(x_1, \ldots, x_i, \ldots, x_d) \underbrace{\mathrm{d}x_1 \cdots \mathrm{d}x_d}_{\text{exceto } x_i}.$$

Demonstração. Feita em aula. □

Exercício 3.2. Sejam três variáveis aleatórias X, Y e Z com função de densidade conjunta dada por

$$f(x, y, z) = \begin{cases} kxy^2z, & \text{se } 0 < x \leqslant 1,\ 0 < y \leqslant 1 \text{ e } 0 < z \leqslant \sqrt{2}, \\ 0, & \text{caso contrário.} \end{cases}$$

Encontre o valor de k e ache a função de densidade marginal de X.

Observação 3.10. Se um vetor aleatório $\boldsymbol{X}$ é absolutamente contínuo, então suas coordenadas $X_1, \ldots, X_d$ são absolutamente contínuas, mas não vale a recíproca! De fato, é muito fácil construir um vetor aleatório contínuo que não é absolutamente contínuo.

Exercício 3.3. Seja $X \sim U[0,1]$, $Y = 1 - X$ e $\boldsymbol{X} = (X, Y)$. Encontre a função de distribuição conjunta $F_{\boldsymbol{X}}(x, y)$. Verifique que $\frac{\partial^2}{\partial y \partial x} F_{\boldsymbol{X}}(x, y) = 0$ para todo par (x, y) no plano $\mathbb{R}^2$, exceto em algumas retas ou segmentos de reta. As coordenadas de $\boldsymbol{X}$ são absolutamente contínuas, mas o vetor $\boldsymbol{X}$ não é absolutamente contínuo!

Vetores Aleatórios Mistos

Como no caso uni-dimensional, dizemos que um vetor aleatório $\boldsymbol{X}$ é do tipo *misto com componentes discreta e absolutamente contínua* se existem $p_{\boldsymbol{X}}$ e $f_{\boldsymbol{X}}$ tais que

$$P(\boldsymbol{X} \in B) = \sum_{\boldsymbol{x} \in B} p_{\boldsymbol{X}}(\boldsymbol{x}) + \int_B f_{\boldsymbol{X}}(\boldsymbol{x})\, \mathrm{d}^d \boldsymbol{x} \qquad \forall B \in \mathcal{B}^d.$$

3.3 Independência de Variáveis Aleatórias

Definição 3.11 (Variáveis Aleatórias Independentes)**.** Dizemos que as variáveis aleatórias $X_1, X_2, \dots, X_d$ em $(\Omega, \mathcal{F}, P)$ são *coletivamente independentes*, ou simplesmente *independentes*, se

$$P(X_1 \in B_1, \dots, X_d \in B_d) = P(X_1 \in B_1) \cdots P(X_d \in B_d)$$

para quaisquer $B_1, \dots, B_d \in \mathcal{B}$. Se I é uma família qualquer de índices, dizemos que $(X_i)_{i \in I}$ são coletivamente independentes se $X_{i_1}, \dots, X_{i_n}$ são independentes para todo $n \in \mathbb{N}$ e $i_1, \dots, i_n \in I$.

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

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

Proposição 3.12 (Critério de Independência)**.** *São equivalentes:*

(i) $X_1, X_2, \dots, X_d$ *são independentes.*

(ii) $F_{\boldsymbol{X}}(\boldsymbol{t}) = F_{X_1}(t_1) F_{X_2}(t_2) \cdots F_{X_d}(t_d)$ *para todo* $\boldsymbol{t} \in \mathbb{R}^d$.

(iii) $F_{\boldsymbol{X}}(\boldsymbol{t}) = F_1(t_1) F_2(t_2) \cdots F_d(t_d)$ *para todo* $\boldsymbol{t} \in \mathbb{R}^d$*, com* $F_1, \dots, F_d$ *funções reais.*

Demonstração. $(i) \Rightarrow (ii) \Rightarrow (iii)$ são triviais. Suponha (iii). Calculando a marginal temos

$$F_{X_i}(x_i) = \lim_{\substack{x_j \to \infty \\ j \neq i}} F_{\boldsymbol{X}}(\boldsymbol{x}) = F_i(x_i) \cdot \prod_{j \neq i} \lim_{x_j \to \infty} F_j(x_j) = c_i \, F_i(x_i),$$

onde $c_i \neq 0$ pois F_{X_i} não pode ser uma função constante. Assim,

$$F_{\boldsymbol{X}}(x_1, \dots, x_d) = \frac{1}{c_1 \cdots c_d} F_{X_1}(x_1) \cdots F_{X_d}(x_d).$$

Fazendo $x_i \to \infty \, \forall i$, temos que $c_1 \cdots c_d = 1$, portanto $(iii) \Rightarrow (ii)$.

Assumindo (ii), vamos mostrar (i) supondo que os B_i são uniões de intervalos disjuntos. Observe que se $B_i = (a_i, b_i]$ para $i = 1, \dots, d$, temos

$$\begin{aligned} P(X_1 \in B_1, \dots, X_d \in B_d) &= \Delta^1_{a_1,b_1} \cdots \Delta^d_{a_d,b_d} F_{\boldsymbol{X}}(\boldsymbol{x}) \\ &= \Delta^1_{a_1,b_1} \cdots \Delta^d_{a_d,b_d} [F_{X_1}(x_1) \cdots F_{X_d}(x_d)] \\ &= [\Delta^1_{a_1,b_1} F_{X_1}(x_1)] \cdots [\Delta^d_{a_d,b_d} F_{X_d}(x_d)] \\ &= P(X_1 \in B_1) \cdots P(X_d \in B_d). \end{aligned}$$

A mesma identidade se estende para $B_i = [a_i, b_i]$ tomando-se o limite $a \to a_i-$, analogamente para intervalos abertos ou semi-infinitos, e por aditividade vale para uniões de intervalos disjuntos. A extensão a todo $B_i \in \mathcal{B}$ envolve argumentos de Teoria da Medida e será omitida. $\square$

Proposição 3.13 (Critério de Independência. Caso Discreto)**.** *Seja* $\boldsymbol{X}$ *um vetor aleatório discreto. São equivalentes:*

(i) $X_1, X_2, \dots, X_d$ *são independentes.*

(ii) $p_{\boldsymbol{X}}(\boldsymbol{t}) = p_{X_1}(t_1) p_{X_2}(t_2) \cdots p_{X_d}(t_d)$ *para todo* $\boldsymbol{t} \in \mathbb{R}^d$.

(iii) $p_{\boldsymbol{X}}(\boldsymbol{t}) = p_1(t_1) p_2(t_2) \cdots p_d(t_d)$ *para todo* $\boldsymbol{t} \in \mathbb{R}^d$, *com* $p_1, \dots, p_d$ *funções reais.*

Demonstração. $(i) \Rightarrow (ii) \Rightarrow (iii)$ são triviais. Suponha (iii). Para x_i tal que $p_{X_i}(x_i) > 0$, calculando a marginal temos

$$p_{X_i}(x_i) = \sum_{x_j} \cdots \sum_{x_{i-1}} \sum_{x_{i+1}} \cdots \sum_{x_d} p_{\boldsymbol{X}}(\boldsymbol{x}) = p_i(x_i) \cdot \prod_{j \neq i} \sum_{x_j} p_j(x_j) = c_i \, p_i(x_i),$$

onde $c_i \neq 0$. Assim,

$$p_{\boldsymbol{X}}(x_1, \dots, x_d) = \frac{1}{c_1 \cdots c_d} p_{X_1}(x_1) \cdots p_{X_d}(x_d).$$

Somando em $\boldsymbol{x}$, temos que $c_1 \cdots c_d = 1$, portanto $(iii) \Rightarrow (ii)$.

Suponha (ii). Temos que

$$P(X_1 \in B_1, \dots, X_d \in B_d) = \sum_{x_1 \in B_1} \cdots \sum_{x_d \in B_d} p_{\boldsymbol{X}}(\boldsymbol{x}) =$$
$$= \left[\sum_{x_1 \in B_1} p_{X_1}(x_1) \right] \cdots \left[\sum_{x_d \in B_d} p_{X_d}(x_d) \right] = P(X_1 \in B_1) \cdots P(X_d \in B_d),$$

e portanto $(ii) \Rightarrow (i)$. □

Proposição 3.14 (Critério de Independência. Caso Contínuo)**.** *Seja $\boldsymbol{X}$ um vetor aleatório absolutamente contínuo. São equivalentes:*

(i) *$X_1, X_2, \dots, X_d$ são independentes.*

(ii) *$f_{\boldsymbol{X}}(\boldsymbol{t}) = f_{X_1}(t_1) f_{X_2}(t_2) \cdots f_{X_d}(t_d)$ para quase todo $\boldsymbol{t} \in \mathbb{R}^d$.*

(iii) *$f_{\boldsymbol{X}}(\boldsymbol{t}) = f_1(t_1) f_2(t_2) \cdots f_d(t_d)$ para quase todo $\boldsymbol{t} \in \mathbb{R}^d$, com $f_1, \dots, f_d$ funções reais.*

Demonstração. $(i) \Rightarrow (ii) \Rightarrow (iii)$ são triviais. Suponha (iii). Para $B_i \in \mathcal{B}$ tal que $P(X_i \in B_i) > 0$,

$$P_{X_i}(B_i) = \int_{\mathbb{R}^d} \mathbb{1}_{B_i}(x_i) f_{\boldsymbol{X}}(\boldsymbol{x})\, \mathrm{d}^d\boldsymbol{x} = \int_{B_i} f_i(x_i) \mathrm{d}x_i \cdot \prod_{j \neq i} \int_{\mathbb{R}} f_j(x_j)\, \mathrm{d}x_j = c_i \int_{B_i} f_i(x_i) \mathrm{d}x_i,$$

onde $c_i \neq 0$. Logo, $f_{X_i}(x_i) = c_i\, f_i(x_i)$ para todo $x_i \in \mathbb{R}$. Assim,

$$f_{\boldsymbol{X}}(x_1, \dots, x_d) = \frac{1}{c_1 \cdots c_d} f_{X_1}(x_1) \cdots f_{X_d}(x_d).$$

Integrando em $\mathbb{R}^d$, temos que $c_1 \cdots c_d = 1$, portanto $(iii) \Rightarrow (ii)$.

Suponha (ii). Temos que

$$P(X_1 \in B_1, \ldots, X_d \in B_d) = \int_{B_1 \times \cdots \times B_d} f_{\boldsymbol{X}}(\boldsymbol{x})\, \mathrm{d}^d \boldsymbol{x} =$$
$$= \left[\int_{B_1} f_{X_1}(x_1)\, \mathrm{d}x_1\right] \cdots \left[\int_{B_d} f_{X_d}(x_d)\, \mathrm{d}x_d\right] = P(X_1 \in B_1) \cdots P(X_d \in B_d),$$

e portanto $(ii) \Rightarrow (i)$. □

Definição 3.15 (Variáveis Aleatórias Independentes Duas a Duas)**.** Se I é uma família qualquer de índices, dizemos que $(X_i)_{i\in I}$ são *duas a duas independentes* se X_i e X_j são independentes para quaisquer $i \neq j \in I$.

Segue das definições que uma família de variáveis aleatórias coletivamente independentes também é independente duas a duas. Entretanto não vale a recíproca.

Contra-Exemplo 3.16. Sejam X e Y independentes assumindo os valores -1 ou $+1$ com probabilidade $\frac{1}{2}$ cada, e tome $Z = XY$. Então temos

$$p_{X,Y,Z}(x,y,z) = \begin{cases} \frac{1}{4}, & (x,y,z) = (1,1,1), (-1,-1,1), (1,-1,-1), (-1,1,-1), \\ 0, & \text{caso contrário.} \end{cases}$$

Então X, Y e Z não são coletivamente independentes, pois

$$p_{X,Y,Z}(1,1,1) = \frac{1}{4} \neq \frac{1}{8} = p_X(1)p_Y(1)p_Z(1).$$

Entretanto, X, Y e Z são duas a duas independentes.

3.4 Método do Jacobiano

Suponha que o vetor aleatório $\boldsymbol{X}$ é absolutamente contínuo e assume valores em um domínio $G_0 \subseteq \mathbb{R}^d$, e que estamos interessados em estudar o vetor aleatório $\boldsymbol{Y}$ dado por uma transformação $\boldsymbol{Y} = \boldsymbol{g}(\boldsymbol{X})$. Vamos considerar o caso em que $\boldsymbol{g} : G_0 \to G$, $G \subseteq \mathbb{R}^d$, é bijetiva e diferenciável, com inversa $\boldsymbol{g}^{-1} = \boldsymbol{h} : G \to G_0$ também diferenciável. Escrevemos a transformada inversa como $\boldsymbol{X} = \boldsymbol{h}(\boldsymbol{Y})$ e definimos os Jacobianos:

$$J_{\boldsymbol{h}}(\boldsymbol{y}) = \det\left(\frac{\partial \boldsymbol{x}}{\partial \boldsymbol{y}}\right) = \det\begin{bmatrix} \frac{\partial x_1}{\partial y_1} & \cdots & \frac{\partial x_1}{\partial y_d} \\ \vdots & \ddots & \vdots \\ \frac{\partial x_d}{\partial y_1} & \cdots & \frac{\partial x_d}{\partial y_d} \end{bmatrix}$$

e

$$J_{\boldsymbol{g}}(\boldsymbol{x}) = \det\left(\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}}\right) = \det\begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_1}{\partial x_d} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_d}{\partial x_1} & \cdots & \frac{\partial y_d}{\partial x_d} \end{bmatrix}.$$

O Jacobiano satisfaz a seguinte identidade:

$$J_{\boldsymbol{h}}(\boldsymbol{y}) = \frac{1}{J_{\boldsymbol{g}}(\boldsymbol{x})}.$$

Proposição 3.17 (Método do Jacobiano). *Sejam $G_0, G \subseteq \mathbb{R}^d$, $\boldsymbol{g} : G_0 \to G$ uma bijeção e $\boldsymbol{h} = \boldsymbol{g}^{-1}$, e suponha que $\boldsymbol{g}$ e $\boldsymbol{h}$ sejam diferenciáveis. Se $\boldsymbol{X}$ é um vetor aleatório absolutamente contínuo assumindo valores em G_0, e $\boldsymbol{Y} = \boldsymbol{g}(\boldsymbol{X})$, então a densidade $f_{\boldsymbol{Y}}$ pode ser obtida a partir da densidade $f_{\boldsymbol{X}}$ pela relação*

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = \big|J_{\boldsymbol{h}}(\boldsymbol{y})\big| \cdot f_{\boldsymbol{X}}\big(\boldsymbol{h}(\boldsymbol{y})\big) = \frac{1}{|J_{\boldsymbol{g}}(\boldsymbol{x})|}\, f_{\boldsymbol{X}}\big(\boldsymbol{h}(\boldsymbol{y})\big).$$

Ideia da prova. Pelo cálculo de várias variáveis, sabemos que se o jacobiano for não-nulo para todo $y \in G$, então

$$\int_A f(\boldsymbol{x})\, \mathrm{d}^d\boldsymbol{x} = \int_{g(A)} f(\boldsymbol{h}(\boldsymbol{y}))\, |J_{\boldsymbol{h}}(\boldsymbol{y})|\, \mathrm{d}^d\boldsymbol{y}$$

para qualquer f integrável em A, onde $A \subseteq G_0$. Como $P(\boldsymbol{Y} \in \boldsymbol{g}(A))$ é dada por $P(\boldsymbol{X} \in A)$, e esta última é dada pelo lado esquerdo da expressão acima com $f = f_{\boldsymbol{X}}$, temos que o integrando do lado direito é necessariamente dado por $f_{\boldsymbol{Y}}(\boldsymbol{y})$. $\square$

Exemplo 3.18. Considere o vetor aleatório $\boldsymbol{X} = (X_1, X_2)$ com densidade

$$f_{\boldsymbol{X}}(x_1, x_2) = \begin{cases} 4x_1x_2, & x_1, x_2 \in [0,1], \\ 0, & \text{caso contrário}, \end{cases}$$

e o vetor $\boldsymbol{Y}$ dado por $Y_1 = X_1/X_2, Y_2 = X_1X_2$. Temos $\boldsymbol{y} = \boldsymbol{h}(\boldsymbol{x}) = (x_1/x_2, x_1x_2)$ e

$$\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}} = \begin{bmatrix} 1/x_2 & -x_1/x_2^2 \\ x_2 & x_1 \end{bmatrix}$$

e $J_{\boldsymbol{g}}(\boldsymbol{x}) = 2x_1/x_2$. Obtendo $\boldsymbol{x}$ em função de $\boldsymbol{y}$:

$$\begin{aligned} y_1 &= x_1/x_2 & y_1y_2 &= x_1^2 & y_1 &= x_1 = \sqrt{y_1y_2} \\ y_2 &= x_1x_2 & y_2/y_1 &= x_2^2 & y_2 &= x_2 = \sqrt{y_2/y_1}, \end{aligned}$$

e os valores possíveis de $\boldsymbol{y}$ são

$$G = \left\{(y_1, y_2) : 0 < y_2 < y_1, 0 < y_2 < \tfrac{1}{y_1}\right\}.$$

Agora,

$$J_{\boldsymbol{g}}(\boldsymbol{h}(\boldsymbol{y})) = \frac{2\sqrt{y_1y_2}}{\sqrt{y_2/y_1}} = 2y_1$$

e

$$f_{\boldsymbol{X}}(\boldsymbol{h}(\boldsymbol{y})) = 4\sqrt{y_1y_2}\sqrt{y_2/y_1} = 4y_2.$$

Portanto,

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = \frac{1}{|J_{\boldsymbol{g}}(\boldsymbol{x})|}\, f_{\boldsymbol{X}}\big(\boldsymbol{h}(\boldsymbol{y})\big) = \begin{cases} 2y_2/y_1, & 0 < y_2 < 1, y_2 < y_1 < 1/y_2, \\ 0, & \text{caso contrário.} \end{cases}$$

Exercício 3.4. Sejam X e Y variáveis aleatórias independentes, cada uma com distribuição exponencial com parâmetro 1, mostre que $Z = X + Y$ e $W = \frac{X}{Y}$ são também independentes com densidades

$$f_Z(z) = \begin{cases} ze^{-z}, & z > 0 \\ 0, & z \leqslant 0 \end{cases}$$

e

$$f_W(w) = \begin{cases} \frac{1}{(w+1)^2}, & w > 0 \\ 0, & w \leqslant 0 \end{cases}.$$

Exemplo 3.19. Se X e Y são independentes e distribuídas como $\mathcal{N}(0,1)$, então $X + Y$ e $X - Y$ são independentes e ambas distribuídas como $\mathcal{N}(0,2)$.

Ponha $\boldsymbol{Z} = (X, Y)$ e $\boldsymbol{W} = (X+Y, X-Y)$. Temos que $\boldsymbol{W} = \boldsymbol{g}(\boldsymbol{Z})$, onde $\boldsymbol{g}(x, y) = (x+y, x-y)$. Logo,

$$\frac{\partial \boldsymbol{w}}{\partial \boldsymbol{z}} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix},$$

assim $J_{\boldsymbol{g}}(\boldsymbol{z}) = -2$. Obtendo $\boldsymbol{z}$ como função de $\boldsymbol{w}$:

$$\begin{aligned} w_1 &= x + y & \qquad x &= \frac{w_1 + w_2}{2} \\ w_2 &= x - y & \qquad y &= \frac{w_1 - w_2}{2}. \end{aligned}$$

Ainda,

$$f_{\boldsymbol{Z}}(\boldsymbol{z}) = f_{X,Y}(x, y) = f_X(x) f_Y(y) = \frac{1}{\sqrt{2\pi}}\, e^{\frac{-x^2}{2}} \cdot \frac{1}{\sqrt{2\pi}}\, e^{\frac{-y^2}{2}},$$

logo

$$f_{\boldsymbol{Z}}(\boldsymbol{h}(\boldsymbol{w})) = \frac{1}{2\pi}\, e^{\frac{-\left(\frac{w_1+w_2}{2}\right)^2}{2}}\, e^{\frac{-\left(\frac{w_1-w_2}{2}\right)^2}{2}} = \frac{e^{-\frac{w_1^2+w_2^2+2w_1w_2+w_1^2+w_2^2-2w_1w_2}{8}}}{2\pi} = \frac{1}{2\pi}\, e^{\frac{-w_1^2}{4}}\, e^{\frac{-w_2^2}{4}}$$

e, substituindo,

$$f_{\boldsymbol{W}}(\boldsymbol{w}) = \frac{1}{|J_{\boldsymbol{g}}(\boldsymbol{h}(\boldsymbol{w}))|}\, f_{\boldsymbol{Z}}(\boldsymbol{h}(\boldsymbol{w})) = \frac{1}{4\pi}\, e^{\frac{-w_1^2}{4}}\, e^{\frac{-w_2^2}{4}} = f_{\mathcal{N}(0,2)}(w_1) f_{\mathcal{N}(0,2)}(w_2).$$

Portanto, W_1 e W_2 são independentes e distribuídas como $\mathcal{N}(0, 2)$.

Exercício 3.5. Se X e Y são independentes e distribuídas como $\mathcal{N}(0, 1)$, então $4X + 3Y$ e $3X - 4Y$ são independentes e ambas distribuídas como $\mathcal{N}(0, 25)$.

3.5 Exercícios

3.6. Considere um vetor aleatório (X, Y) absolutamente contínuo com distribuição uniforme em

$$A = \left\{(x, y) \in \mathbb{R}^2 : 0 < y < x \text{ e } x + y < 1\right\}.$$

Encontre $F_{X,Y}$.

3.7. Considere um vetor aleatório (Z, W) absolutamente contínuo com densidade

$$f_{Z,W}(z,w) = \begin{cases} c, & 0 < z < 1, 0 < w < z, \\ 0, & \text{caso contrário.} \end{cases}$$

Encontre $F_{Z,W}$.

3.8. Sejam Y e U duas variáveis aleatórias em um mesmo espaço de probabilidade, independentes e com leis $Y \sim \mathcal{N}(0,1)$ e $P(U = -1) = P(U = +1) = \frac{1}{2}$. Ache a lei de $Z = UY$.
Dica: estudar diretamente a função de distribuição acumulada.

3.9.

(a) A densidade conjunta de X e Y é dada por

$$f(x,y) = \begin{cases} \frac{c\ e^{-y}}{x^3}, & x > 1, y > 0 \\ 0, & \text{caso contrário.} \end{cases}$$

Encontre c. Diga se X e Y são independentes e por quê.

(b) Suponha que (X, Y) é um vetor aleatório absolutamente contínuo com função de distribuição conjunta dada por

$$F_{XY}(x,y) = \begin{cases} 1 - e^{-x} + e^{-x-y} - xe^{-x} - e^{-y} + xe^{-x-y}, & x, y > 0 \\ 0, & \text{caso contrário.} \end{cases}$$

Encontre a densidade conjunta f_{XY} e diga se X e Y são independentes.

(c) Com X e Y dadas no item anterior, encontre a distribuição marginal F_Y.

3.10. Sejam X e Y variáveis aleatórias discretas e independentes. Mostre que

$$p_{X+Y}(t) = \sum_s p_X(s) \cdot p_Y(t-s).$$

Sugestão: particione Ω segundo o valor de X.

3.11. Mostre por indução finita que, se $X_1, X_2, \ldots, X_n$ são variáveis aleatórias independentes com $X_i \sim b(m_i, p), i = 1, 2, \ldots, n$, então

$$\sum_{i=1}^{n} X_i \sim b\left(\sum_{i=1}^{n} m_i, p\right).$$

Dica: $\binom{a+b}{n} = \sum_{k=0}^{n} \binom{a}{k}\binom{b}{n-k}$.

3.12. Se X e Y são independentes e distribuídas respectivamente como Poisson(λ_1) e Poisson(λ_2), mostre que

$$X + Y \sim \text{Poisson}(\lambda_1 + \lambda_2).$$

Dica: $(a+b)^k = \sum_{j=0}^{k} \binom{k}{j} a^j b^{k-j}$.

3.13. Sejam X e Y variáveis aleatórias definidas no mesmo espaço de probabilidade, independentes, discretas e com distribuições Poisson(λ_1) e Poisson(λ_2), respectivamente. Mostre que, dada a ocorrência do evento $[X + Y = n]$, a probabilidade condicional de $X = k$ é

$$P(X = k | X + Y = n) = \binom{n}{k} \left(\frac{\lambda_1}{\lambda_1 + \lambda_2}\right)^k \left(\frac{\lambda_2}{\lambda_1 + \lambda_2}\right)^{n-k}.$$

Como você interpreta essa identidade?

3.14. O número X de uvas-passas encontradas em um panetone tem distribuição Poisson(λ). O panetone, estando com a data de validade vencida há alguns meses, pode ter uvas-passas estragadas. Cada uva-passa pode estar estragada independente das demais, com probabilidade p. Encontre a distribuição do número de uvas-passas estragadas e calcule a probabilidade de não haver nenhuma estragada.

3.15. Sejam X e Y variáveis aleatórias independentes, ambas com distribuição exp(1). Use o método do Jacobiano para determinar a distribuição conjunta de $X + Y$ e $\frac{X}{X+Y}$. Diga se $X + Y$ e $\frac{X}{X+Y}$ são independentes. Encontre a distribuição de $\frac{X}{X+Y}$.

3.16. Sejam X e Y i.i.d. absolutamente contínuas com densidade f. Mostre que

$$f_{X+Y}(t) = \int_{\mathbb{R}} f(t-s) f(s) \, \mathrm{d}s \qquad \forall t \in \mathbb{R}.$$

Sugestão: faça $Z = X + Y$ e $W = Y$, e calcule a densidade conjunta de Z e W.

3.17. [Jam04, Capítulo 2].
Recomendados: 2, 17, 18, 21, 30, 33, 34, 41, 46.

Capítulo 4

Esperança Matemática

A esperança EX de uma variável aleatória X é a média dos valores assumidos por X, ponderada pela probabilidade de X assumir esses valores. Podemos pensar em EX como sendo o "centro de massa" de X. A esperança de X é, em vários sentidos, a melhor aproximação determinística para a variável aleatória X. Uma das justificativas mais importantes, que veremos mais adiante, é a lei dos grandes números: se $X_1, \dots, X_n$ são independentes e têm a mesma distribuição de X, então a média amostral $\frac{1}{n}\sum_{i=1}^{n} X_i$ se aproxima de EX quando fazemos n grande.

4.1 Variáveis Aleatórias Simples

Uma variável aleatória X é dita *simples* se assume apenas finitos valores.

Definição 4.1. Dada uma variável aleatória simples X, definimos a *esperança de* X, ou *média de* X, ou ainda o *valor esperado de* X, denotada por EX, por

$$EX = \sum_{x} x \cdot P(X = x).$$

A esperança de X pode ser pensada como o "centro de massa" da variável aleatória X, como ilustrado na Figura 4.1.

Outra interpretação de EX vem dos jogos em cassinos. Sejam X o resultado que se ganha em um dado jogo, e $a_1, \dots, a_k$ os possíveis valores. Suponhamos também

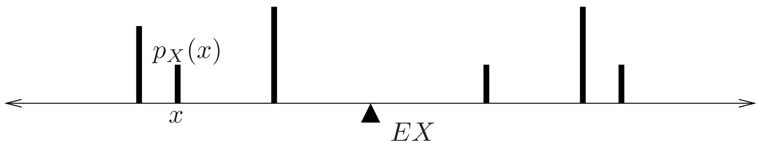

Figura 4.1: A esperança de X como o centro de massa de p_X.

que jogaremos esse jogo n vezes, e denotamos o resultado de cada jogada por $X_1, \dots, X_n$, independentes e com a mesma distribuição de X. A noção intuitiva de probabilidade como frequência relativa diz que a proporção dentre essas n repetições em que o resultado é a_i se aproxima de $P(X = a_i)$ para n grande, ou seja,

$$\frac{1}{n}\sum_{j=1}^{n} \mathbb{1}_{[X_j = a_i]} \approx P(X = a_i).$$

Dessa forma, para o ganho total dividido pelo número de jogadas, temos

$$\frac{1}{n}\sum_{j=1}^{n} X_j = \frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{k} a_i \mathbb{1}_{[X_j = a_i]} =$$
$$= \sum_{i=1}^{k} a_i \left(\frac{1}{n}\sum_{j=1}^{n} \mathbb{1}_{[X_j = a_i]} \right) \approx \sum_{i=1}^{k} a_i \cdot P(X = a_i) = EX.$$

Exemplo 4.2. Lançar um dado e observar sua face superior. Temos

$$EX = 1 \times P(X = 1) + \cdots + 6 \times P(X = 6) = \frac{1}{6} + \frac{2}{6} + \cdots + \frac{6}{6} = \frac{21}{6} = \frac{7}{2}.$$

Exemplo 4.3. Lançar uma moeda 4 vezes e contar quantas vezes saem cara. Temos

$$EX = 0 \times \frac{1}{16} + 1 \times \frac{4}{16} + 2 \times \frac{6}{16} + 3 \times \frac{4}{16} + 4 \times \frac{1}{16} = \frac{32}{16} = 2.$$

Exemplo 4.4. Seja X dada por $X = \mathbb{1}_A$ para algum $A \in \mathcal{F}$. Nesse caso temos $EX = 0 \times P(A^c) + 1 \times P(A) = P(A)$. Ou seja, $\mathbb{1}_A$ tem distribuição Bernoulli e

$$E\mathbb{1}_A = P(A).$$

Por outro lado, se $X \sim \text{Bernoulli}(p)$ então podemos considerar o evento A dado por $A = \{\omega : X(\omega) = 1\}$, de forma que $X = \mathbb{1}_A$ e $EX = p$.

Exemplo 4.5. Lançar um dado duas vezes e somar os valores observados. Temos

$$\begin{aligned} EX = 2 \times \frac{1}{36} + 3 \times \frac{2}{36} + 4 \times \frac{3}{36} + 5 \times \frac{4}{36} + 6 \times \frac{5}{36} + 7 \times \frac{6}{36} + \\ + 8 \times \frac{5}{36} + 9 \times \frac{4}{36} + 10 \times \frac{3}{36} + 11 \times \frac{2}{36} + 12 \times \frac{1}{36} = \frac{252}{36} = 7. \end{aligned}$$

Exemplo 4.6. Retirar 3 cartas de um baralho francês e contar quantas são reis.

$$EX = 0 \times \frac{48.47.46}{52.51.50} + 1 \times \frac{3.48.47.4}{52.51.50} + 2 \times \frac{3.48.4.3}{52.51.50} + 3 \times \frac{4.3.2}{52.51.50} = \frac{30600}{132600} = \frac{3}{13}.$$

Exemplo 4.7. Em geral, se $X \sim b(n,p)$, então

$$\begin{aligned} EX &= \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=1}^{n} k \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\ &= np \sum_{k=1}^{n} \frac{(n-1)!}{(k-1)!(n-k)!} p^{k-1} (1-p)^{n-k} \\ &= np \sum_{j=0}^{n-1} \frac{(n-1)!}{j!(n-1-j)!} p^j (1-p)^{n-1-j} = np[p + (1-p)]^{n-1} = np. \end{aligned}$$

Exemplo 4.8. Lançar um dado duas vezes e multiplicar os valores observados.

$$\begin{aligned} EX = \frac{1}{36}\big(1 \cdot 1 + 2 \cdot 2 + 3 \cdot 2 + 4 \cdot 3 + 5 \cdot 2 + 6 \cdot 4 + 8 \cdot 2 + 9 \cdot 1 + 10 \cdot 2 + 12 \cdot 4 + \\ + 15 \cdot 2 + 16 \cdot 1 + 18 \cdot 2 + 20 \cdot 2 + 24 \cdot 2 + 25 \cdot 1 + 30 \cdot 2 + 36 \cdot 1\big) = \frac{441}{36} = \frac{49}{4}. \end{aligned}$$

Linearidade

Teorema 4.9. *Sejam X e Y variáveis aleatórias simples.*

(i) *Se $X \geqslant 0$ então $EX \geqslant 0$.*

(ii) *Se $X = c$ então $EX = c$.*

(iii) $E[aX + bY] = aEX + bEY$.

(iv) *Se* $X \geqslant Y$ *então* $EX \geqslant EY$.

Vejamos alguns exemplos do uso da linearidade.

Exemplo 4.10. No Exemplo 4.3, temos $X = X_1+X_2+X_3+X_4$, onde X_i representa o lançamento da i-ésima moeda. Logo $EX = EX_1+EX_2+EX_3+EX_4 = 4\cdot\frac{1}{2} = 2$.

Exemplo 4.11. No Exemplo 4.5, observamos que $X = Y + Z$, onde Y e Z representam o primeiro e segundo lançamento do dado. Logo

$$EX = EY + EZ = \frac{7}{2} + \frac{7}{2} = 7.$$

Exemplo 4.12. No Exemplo 4.6, observamos que $X = X_1 + X_2 + X_3$, onde X_i é o indicador de que a i-ésima carta retirada é rei. Ao contrário dos exemplos anteriores, aqui X_1, X_2 e X_3 não são independentes. Ainda assim, cada uma individualmente satisfaz $EX_i = \frac{1}{13}$, e podemos calcular

$$EX = EX_1 + EX_2 + EX_3 = 3\frac{1}{13}.$$

Exemplo 4.13. No Exemplo 4.7, observamos que X tem a mesma distribuição de $X_1 + \cdots + X_n$, com X_i i.i.d. Bernoulli(p), e portanto

$$EX = EX_1 + \cdots + EX_n = (p + \cdots + p) = np.$$

Nos exemplos anteriores, é mais curto calcular a esperança usando linearidade do que usando a distribuição de X diretamente. Existem muitos outros casos em que descrever a distribuição de X é muito difícil ou até mesmo impossível, mas ainda assim é possível calcular a esperança usando linearidade.

Exemplo 4.14. Uma gaveta contém 10 pares de meias, todos diferentes. Abre-se a gaveta no escuro e retiram-se 6 meias. Qual é o valor esperado do número de pares X formados pelas meias retiradas? Se numeramos as meias retiradas, e contamos a quantidade N de meias cujo par também foi retirado, estaremos contando cada par duas vezes, portanto $N = 2X$. A probabilidade de que o par da primeira meia retirada também seja retirado é $\frac{5}{19}$. Somando sobre as 6 meias, temos $E[2X] = EN = 6 \times \frac{5}{19}$, e portanto $EX = \frac{15}{19}$.

Proposição 4.15 (Inclusão e exclusão). *Sejam* $A_1, \dots, A_n$ *eventos aleatórios.*

Então

$$P(A_1 \cup \cdots \cup A_n) = \sum_{1 \leqslant i \leqslant n} P(A_i) - \sum_{1 \leqslant i < j \leqslant n} P(A_i \cap A_j) + \\ + \sum_{1 \leqslant i < j < k \leqslant n} P(A_i \cap A_j \cap A_k) - \cdots \pm P(A_1 \cap \cdots \cap A_n).$$

Demonstração. Como $(\cup_i A_i)^c = \cap_i A_i^c$, temos $1 - \mathbb{1}_{\cup_i A_i} = \mathbb{1}_{\cap_i A_i^c} = \prod_i \mathbb{1}_{A_i^c}$, logo $1 - \mathbb{1}_{\cup_i A_i} = \prod_i (1 - \mathbb{1}_{A_i})$. Expandindo o produto e simplificando, obtemos

$$\mathbb{1}_{\cup_i A_i} = \sum_i \mathbb{1}_{A_i} - \sum_{1 \leqslant i < j \leqslant n} \mathbb{1}_{A_i \cap A_j} + \sum_{1 \leqslant i < j < k \leqslant n} \mathbb{1}_{A_i \cap A_j \cap A_k} - \cdots \pm \mathbb{1}_{A_1 \cap \cdots \cap A_n}.$$

Finalmente, tomando a esperança de ambos os lados e usando linearidade, obtemos a identidade desejada. □

Independência

Teorema 4.16. *Se X e Y são simples e independentes, então*

$$E[XY] = EX \cdot EY.$$

Exemplo 4.17. No Exemplo 4.8, observamos que $X = YZ$, onde Y e Z representam o primeiro e segundo lançamento do dado. Logo,

$$EX = EY \cdot EZ = \frac{7}{2} \cdot \frac{7}{2} = \frac{49}{4}.$$

Demonstrações

Concluímos esta seção com a demonstração dos teoremas acima.

Observe que, se $A_1, \dots, A_k$ formam uma partição de Ω, então cada $\omega \in \Omega$ pertence a um, e somente um, dos $A_1, \dots, A_k$. Portanto,

$$\sum_{j=1}^{k} \mathbb{1}_{A_j}(\omega) = 1 \quad \forall \omega \in \Omega.$$

Sejam $a_1, \dots, a_n$ os valores assumidos por X e considere os eventos aleatórios A_j dados por $[X = a_j]$. Observe que, exceto por permutações dos índices j, existe uma única forma de escrever

$$X = \sum_{j=1}^{n} a_j \cdot \mathbb{1}_{A_j}$$

com os a_j distintos e $A_1, \dots, A_n$ formando uma partição de Ω. Nessa notação,

$$EX = \sum_{j=1}^{n} a_j \cdot P(A_j).$$

Lema 4.18. *Se $X = \sum_{j=1}^{n} a_j \cdot \mathbb{1}_{A_j}$, onde $A_1, \dots, A_n$ formam uma partição, então $EX = \sum_{j=1}^{n} a_j \cdot P(A_j)$, mesmo que alguns a_j coincidam.*

Demonstração. Com efeito, primeiro escrevemos $X = \sum_{j=1}^{k} c_j \cdot \mathbb{1}_{C_j}$ onde os c_j são distintos e $C_1, \dots, C_k$ formam uma partição. Observe que, para todo $j = 1, \dots, k$,

$$C_j = \bigcup_{i:a_i=c_j} A_i.$$

Agrupando corretamente os termos dos somatórios, obtemos

$$EX = \sum_{j=1}^{k} c_j \cdot P(C_j) = \sum_{j=1}^{k} c_j \sum_{i:a_i=c_j} P(A_i) = \sum_{j=1}^{k} \sum_{i:a_i=c_j} a_i \cdot P(A_i) =$$
$$= \sum_{j=1}^{k} \sum_{i=1}^{n} \mathbb{1}_{a_i=c_j}\, a_i P(A_i) = \sum_{i=1}^{n} a_i \cdot P(A_i),$$

concluindo a prova. $\square$

Demonstração do Teorema 4.9. Os itens (i) e (ii) são triviais, e (iv) segue de (iii) e (i) se tomamos $Z = X - Y$. Resta provar (iii). Sejam X e Y variáveis aleatórias simples. Escrevemos X e Y como

$$X = \sum_{i=1}^{n} a_i \cdot \mathbb{1}_{A_i} \qquad e \qquad Y = \sum_{j=1}^{k} b_j \cdot \mathbb{1}_{B_j},$$

onde $A_1, \dots, A_n$ particionam Ω e $B_1, \dots, B_k$ também. Temos

$$aX + bY = a\sum_{i=1}^{n} a_i \mathbb{1}_{A_i} \sum_{j=1}^{k} \mathbb{1}_{B_j} + b\sum_{j=1}^{k} b_j \mathbb{1}_{B_j} \sum_{i=1}^{n} \mathbb{1}_{A_i} =$$
$$= \sum_{i=1}^{n}\sum_{j=1}^{k} (aa_i + bb_j)\mathbb{1}_{A_i}\mathbb{1}_{B_j} = \sum_{i=1}^{n}\sum_{j=1}^{k} (aa_i + bb_j)\mathbb{1}_{A_i \cap B_j}.$$

Mas a família $\{A_i \cap B_j\}_{i=1,\dots,n;j=1,\dots,k}$ forma uma partição de Ω, e pelo lema anterior temos

$$\begin{aligned} E[aX + bY] &= \sum_{i=1}^{n}\sum_{j=1}^{k} (aa_i + bb_j)P(A_i \cap B_j) \\ &= \sum_{i=1}^{n}\sum_{j=1}^{k} aa_i P(A_i \cap B_j) + \sum_{i=1}^{n}\sum_{j=1}^{k} bb_j P(A_i \cap B_j) \\ &= \sum_{i=1}^{n} aa_i \sum_{j=1}^{k} P(A_i \cap B_j) + \sum_{j=1}^{k} bb_j \sum_{i=1}^{n} P(A_i \cap B_j) \\ &= \sum_{i=1}^{n} aa_i P(A_i) + \sum_{j=1}^{k} bb_j P(B_j) \\ &= a\sum_{i=1}^{n} a_i P(A_i) + b\sum_{j=1}^{k} b_j P(B_j) = aEX + bEY. \end{aligned}$$ □

Demonstração do Teorema 4.16. Sejam $a_1, \dots, a_n$ os valores assumidos por X e $b_1, \dots, b_k$ os valores assumidos por Y. Tomando $A_i = [X = a_i]$ e $B_j = [Y = b_j]$, temos

$$
\begin{aligned}
E[XY] &= E\left[\left(\sum_{i=1}^{n} a_i \mathbb{1}_{A_i}\right)\left(\sum_{j=1}^{k} b_j \mathbb{1}_{B_j}\right)\right] = E\left[\sum_{i=1}^{n}\sum_{j=1}^{k} a_i b_j \mathbb{1}_{A_i}\mathbb{1}_{B_j}\right] \\
&= E\left[\sum_{i=1}^{n}\sum_{j=1}^{k} a_i b_j \mathbb{1}_{A_i\cap B_j}\right] = \sum_{i=1}^{n}\sum_{j=1}^{k} a_i b_j \, E[\mathbb{1}_{A_i\cap B_j}] \\
&= \sum_{i=1}^{n}\sum_{j=1}^{k} a_i b_j P(A_i\cap B_j) = \sum_{i=1}^{n}\sum_{j=1}^{k} a_i b_j P(A_i)P(B_j) \\
&= \left[\sum_{i=1}^{n} a_i P(A_i)\right]\left[\sum_{j=1}^{k} b_j P(B_j)\right] = EX\cdot EY.
\end{aligned}
$$

Na quarta e sexta igualdades foram usadas a linearidade da esperança e independência de X e Y, respectivamente. □

4.2 Esperança Matemática

Nesta seção definimos a esperança de uma variável aleatória qualquer, e estudamos suas principais propriedades. A esperança de uma variável aleatória não-negativa é definida aproximando-se por variáveis aleatórias simples.

Definição 4.19. Seja X uma variável aleatória tal que $X \geqslant 0$. Definimos a *esperança de* X por

$$EX = \sup\{EZ : Z \text{ variável aleatória simples com } 0 \leqslant Z \leqslant X\}.$$

Para definir a esperança no caso geral, observe que uma variável aleatória sempre pode ser decomposta em suas partes positiva e negativa. De fato, temos

$$X = X^+ - X^-,$$

onde

$$X^+ = \begin{cases} X, & X \geqslant 0, \\ 0, & X \leqslant 0, \end{cases} \qquad X^- = \begin{cases} -X, & X \leqslant 0, \\ 0, & X \geqslant 0, \end{cases}$$

satisfazem $X^+ \geqslant 0$ e $X^- \geqslant 0$. Observe também que $|X| = X^+ + X^-$.

Definição 4.20 (Esperança de uma Variável Aleatória)**.** Seja X uma variável aleatória. Definimos a *esperança de* X por

$$EX = EX^+ - EX^-$$

sempre que EX^+ ou EX^- for finita.

Observe que as três definições são consistentes entre si, ou seja, se X é não-negativa esta última definição dá o mesmo valor que a definição anterior, e se X é simples teremos o mesmo valor que pela definição dada no início do capítulo.

A seguir veremos como obter EX no caso de X ser discreta, contínua, ou mista, bem como a esperança de funções de variáveis ou vetores aleatórios desses tipos.

Teorema 4.21 (Variáveis Aleatórias Discretas)**.** *Seja X uma variável aleatória discreta não-negativa. Então*

$$EX = \sum_x x \cdot p_X(x).$$

Demonstração. Segue direto do Teorema 4.27 com $h(x) = x$. □

Exemplo 4.22 (Poisson)**.** Se $X \sim \text{Poisson}(\lambda)$, então

$$EX = \sum_{n=0}^{\infty} n\frac{\lambda^n e^{-\lambda}}{n!} = \sum_{n=1}^{\infty} \frac{\lambda^n e^{-\lambda}}{(n-1)!} = \lambda e^{-\lambda} \sum_{n=1}^{\infty} \frac{\lambda^{n-1}}{(n-1)!} = \lambda e^{-\lambda} \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} = \lambda e^{-\lambda} e^{\lambda} = \lambda.$$

Portanto, o valor esperado de uma variável aleatória que segue o modelo de Poisson com parâmetro λ é o próprio λ.

Proposição 4.23. *Se X assume valores em $\{0, 1, 2, 3, \dots\}$, então*

$$EX = \sum_{n=1}^{\infty} P(X \geqslant n).$$

Demonstração. Introduzimos um indicador de $n \leqslant k$ para inverter as somas:

$$\begin{aligned} EX = \sum_{k=1}^{\infty} k \cdot P(X=k) &= \sum_{k=1}^{\infty} \sum_{n=1}^{k} P(X=k) \\ &= \sum_{k=1}^{\infty} \sum_{n=1}^{\infty} \mathbb{1}_{n \leqslant k} \, P(X=k) \\ &= \sum_{n=1}^{\infty} \sum_{k=1}^{\infty} \mathbb{1}_{n \leqslant k} \, P(X=k) \\ &= \sum_{n=1}^{\infty} \sum_{k=n}^{\infty} P(X=k) = \sum_{n=1}^{\infty} P(X \geqslant n). \end{aligned}$$ □

Exemplo 4.24 (Geométrica)**.** Se $X \sim \text{Geom}(p)$ então

$$EX = \sum_{n=1}^{\infty} P(X \geqslant n) = \sum_{n=1}^{\infty} (1-p)^{n-1} = \sum_{j=0}^{\infty} (1-p)^j = \frac{1}{1-(1-p)} = \frac{1}{p}.$$

Exercício 4.1. Sejam $X_1, X_2, X_3, \dots$ uma sequência de variáveis independentes com distribuição $U[0,1]$ e tome a variável aleatória N como sendo o menor n tal que $X_1 + X_2 + \cdots + X_n \geqslant 1$. Mostre que $EN = e$.

Teorema 4.25 (Variáveis Aleatórias Absolutamente Contínuas)**.** *Seja X uma variável aleatória absolutamente contínua não-negativa. Então*

$$EX = \int_{\mathbb{R}} x \cdot f_X(x) \, \mathrm{d}x.$$

Demonstração. Segue direto do Teorema 4.27 com $h(x) = x$. □

Exemplo 4.26 (Exponencial)**.** Se $X \sim \exp(\lambda)$, vale

$$EX = \int_{-\infty}^{+\infty} x \cdot f_X(x) \, \mathrm{d}x = \int_0^{+\infty} x \lambda e^{-\lambda x} \mathrm{d}x = \left[-x e^{-\lambda x} \Big|_0^{\infty} \right] - \int_0^{\infty} [-e^{-\lambda x}] \, \mathrm{d}x = \frac{1}{\lambda}.$$

Portanto, o valor esperado de uma variável aleatória que segue o modelo exponencial com parâmetro λ é $\frac{1}{\lambda}$.

Suponha que queremos calcular a esperança da variável aleatória Y dada por

$$Y = h(\boldsymbol{X}),$$

onde h é uma função real contínua, ou uma função contínua por partes. Temos pelo menos duas alternativas. Uma é calcular $F_Y(t)$ para todo t, a partir da distribuição acumulada $F_{\boldsymbol{X}}$ de $\boldsymbol{X}$, e depois calcular a esperança usando os Teoremas 4.21 e 4.25. Entretanto, existe outra maneira, que pode ser mais conveniente:

Teorema 4.27 (Mudança de Variável)**.** *Seja $\boldsymbol{X}$ um vetor aleatório misto com componentes discreta e absolutamente contínua. Seja $h : \mathbb{R}^d \to \mathbb{R}_+$ uma função contínua por partes, e considere $Y = h(\boldsymbol{X})$. Então*

$$EY = \sum_{\boldsymbol{x}} h(\boldsymbol{x}) \cdot p_{\boldsymbol{X}}(\boldsymbol{x}) + \int_{\mathbb{R}^d} h(\boldsymbol{x})\, f_{\boldsymbol{X}}(\boldsymbol{x})\, \mathrm{d}^d\boldsymbol{x}.$$

Exemplo 4.28. Seja $X \sim \exp(\lambda)$. Vamos calcular EX^2. Temos

$$EX^2 = \int_0^\infty x^2 \lambda e^{-\lambda x} \mathrm{d}x = \frac{2}{\lambda} \int_0^\infty x \lambda e^{-\lambda x} \mathrm{d}x = \frac{2}{\lambda^2} \int_0^\infty \lambda e^{-\lambda x} \mathrm{d}x = \frac{2}{\lambda^2},$$

integrando por partes duas vezes.

Definição 4.29. Dizemos que X é *integrável* se ambas EX^+ e EX^- são finitas, ou seja, se EX é um número finito.

Teorema 4.30. *Sejam X e Y variáveis aleatórias em $(\Omega, \mathcal{F}, P)$. Então valem as seguintes propriedades:*

($E1$) *Unitariedade. Se $X = \mathbb{1}_A$, então $EX = P(A)$.*

($E2$) *Monotonicidade. Se $X \leqslant Y$, então $EX \leqslant EY$.*

($E3$) *Linearidade. $E[aX + bY] = aEX + bEY$ para $a, b \in \mathbb{R}$.*

Em ($E2$) basta que $EY < +\infty$ ou $EX > -\infty$ para que ambas as esperanças estejam definidas e valha a desigualdade. A igualdade em ($E3$) vale se EX e EY estão definidas e $aEX + bEY$ está definido, isto é, não resulta em $+\infty - \infty$. Em ambos casos, é suficiente que X seja integrável.

Demonstração. Ver na página 85. □

Todas as propriedades da Esperança decorrem dessas três propriedades.

Proposição 4.31.

1. *Se* $X = c$ *então* $EX = c$.
2. *Se* $P(X = Y) = 1$ *e* EX *está definida, então* $EY = EX$.
3. *Se e* EX *está definida, então* $E[aX + b] = aEX + b$.
4. X *é integrável se e somente se* $E|X| < \infty$.
5. *Se* X *integrável então* $E[X - EX] = 0$.
6. *Se* $a \leqslant X \leqslant b$, *então* $a \leqslant EX \leqslant b$.
7. *Se e* EX *está definida, então* $|EX| \leqslant E|X|$.
8. *Se* $0 \leqslant |X| \leqslant Y$ *e* Y *é integrável, então* X *é integrável.*
9. *Se e* EX *está definida, então* $E[X\mathbb{1}_A]$ *está definida para todo* $A \in \mathcal{F}$.
10. *Se* X *é integrável, então* $X\mathbb{1}_A$ *é integrável para todo* $A \in \mathcal{F}$.
11. *Se* $X \geqslant 0$ *então* $EX \geqslant 0$.
12. *Se* $X \geqslant 0$ *e* $EX = 0$ *então* $P(X = 0) = 1$.

Demonstração. Vamos mostrar apenas a última, deixando as demais como exercício. Para $k \in \mathbb{N}$, temos $0 = EX \geqslant E(X\mathbb{1}_{[X\geqslant\frac{1}{k}]}) \geqslant E(\frac{1}{k}\mathbb{1}_{[X\geqslant\frac{1}{k}]}) = \frac{1}{k}P(X \geqslant \frac{1}{k})$. Logo, $P(X \geqslant \frac{1}{k}) = 0$ para todo k, portanto $P(X > 0) = \lim_k P(X \geqslant \frac{1}{k}) = 0$. □

Proposição 4.32 (Esperança de Variáveis Aleatórias Independentes)**.** *Se* X *e* Y *são independentes e integráveis, então* XY *é integrável e*

$$E[XY] = EX \cdot EY.$$

A mesma identidade vale se X *e* Y *são não-negativas sem supor integrabilidade.*

Demonstração. Ver na página 87. □

Existem outras formas de se definir a esperança, todas elas equivalentes. Isso também se reflete em distintas notações, que o leitor poderá encontrar em diferentes bibliografias:

$$EX = \int_\Omega X \, \mathrm{d}P, \qquad EX = \int_\mathbb{R} x \, \mathrm{d}P_X, \qquad EX = \int_\mathbb{R} x \, \mathrm{d}F_X(x).$$

A definição que usamos aqui corresponde à primeira, que se refere à Integral de Lebesgue sobre o Espaço de Probabilidade Ω visto como um Espaço de Medida.

Essa nomenclatura não é coincidência, e de fato a definição de esperança é parecida com a definição de integral. A área sob a curva do gráfico de uma função $g : \mathbb{R} \to \mathbb{R}_+$ constante por partes é dada pela soma de áreas de retângulos, e cada uma dessas áreas é dada pelo comprimento da base do respectivo retângulo multiplicado por sua altura. De forma análoga, a esperança de uma variável aleatória simples $X : \Omega \to \mathbb{R}_+$ é dada pela soma da contribuição de cada um dos seus valores, e a contribuição de cada valor é dada pelo próprio valor multiplicado por sua respectiva probabilidade. Quando consideramos uma função $g : \mathbb{R} \to \mathbb{R}_+$ qualquer (ou seja, não necessariamente constante por partes), a integral $\int_{\mathbb{R}} g(x)\mathrm{d}x$ equivale à noção de área sob a curva do seu gráfico, e é definida a partir de aproximações em que o domínio é dividido em pequenas partes. Para a esperança de uma variável aleatória $X : \Omega \to \mathbb{R}_+$ qualquer, a ideia é também de usar aproximações. Porém, como não há uma forma razoável de dividir o domínio em pequenas partes, o que se faz é dividir o contra-domínio, como ilustrado na Figura 4.2.

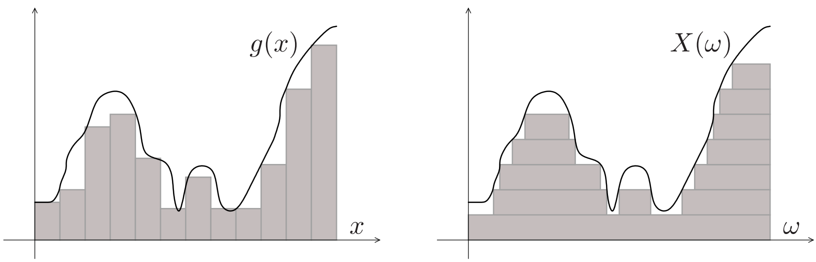

Figura 4.2: Comparação entre a integral de Riemann na reta e a esperança matemática em um espaço de probabilidade.

4.3 Demonstrações

Nesta seção demonstramos as propriedades da Esperança apresentadas na seção anterior. Primeiro observamos que qualquer variável aleatória não-negativa X pode ser aproximada por variáveis aleatórias simples. De fato, considere $g_k : \mathbb{R}_+ \to \mathbb{R}_+$ dada por

$$g_k(x) = 2^{-k} \cdot \max \left\{ j \in \{0, 1, \dots, 2^k k\} \,\middle|\, 2^{-k} j \leqslant x \right\},$$

ilustrada na Figura 4.3.

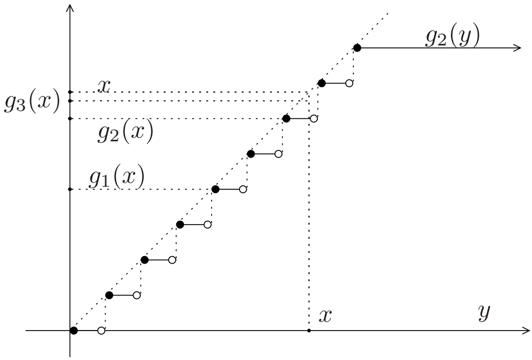

Figura 4.3: Gráfico de $g_2(y)$ e aproximação de $g_k(x) \nearrow x$ para um x fixo.

Observe que g_k assume no máximo $2^k k + 1$ valores. Além disso,

$$g_k(x) \geqslant g_{k-1}(x)$$

e

$$x - 2^{-k} < g_k(x) \leqslant x \quad \text{para todo } k \geqslant x.$$

Portanto, para todo $x \geqslant 0$,

$$g_k(x) \nearrow x \text{ quando } k \to \infty.$$

Tomando $X_k = g_k(X)$, temos que X_k é uma variável aleatória simples e $X_k \nearrow X$ para todo $\omega \in \Omega$. Veja a Figura 4.4.

Lema 4.33. *Sejam X e Y variáveis aleatórias não-negativas definidas em $(\Omega, \mathcal{F}, P)$, e tome $X_k = g_k(X)$, $Y_k = g_k(Y)$. Então, quando $k \to \infty$,*

$$EX_k \to EX \quad e \quad E[X_k + Y_k] \to E[X + Y].$$

Demonstração. Seja Z uma variável aleatória simples com $0 \leqslant Z \leqslant X + Y$. Tomando $M = \max_\omega Z(\omega)$, temos que $X_k + Y_k > Z - 2^{-k+1}$ para $k \geqslant M$. Daí segue que $E[X_k + Y_k] \geqslant E[Z] - 2^{-k+1}$, logo $\liminf_k E[X_k + Y_k] \geqslant EZ$. Tomando o supremo em Z, temos que $\liminf_k E[X_k + Y_k] \geqslant E[X + Y]$ e portanto $E[X_k + Y_k] \nearrow E[X + Y]$. Tomando $Y = 0$ temos $EX_k \nearrow EX$. □

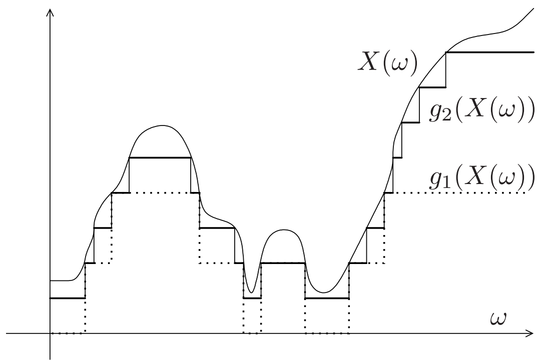

Figura 4.4: Aproximação de X por $g_1(X)$ e $g_2(X)$.

Demonstração do Teorema 4.30. A unitariedade segue da Definição 4.1.

Para a monotonicidade, primeiro supomos que $0 \leqslant X \leqslant Y$. Dada $Z \leqslant X$ simples, temos $Z \leqslant Y$, e pela definição de EY, temos $EZ \leqslant EY$. Tomando o supremo em Z, pela definição de EX, temos $EX \leqslant EY$. Para o caso geral, observe que $X \leqslant Y$ implica $X^+ \leqslant Y^+$ e $X^- \geqslant Y^-$.

Para a linearidade, primeiro observamos que da definição de esperança segue que $E[aX] = aEX$, restando apenas mostrar que $E[X+Y] = EX + EY$.

Suponha inicialmente que X e Y sejam não-negativas. Usando o Teorema 4.9 e o Lema 4.33, temos que

$$E[X+Y] = \lim_k E[X_k + Y_k] = \lim_k [EX_k + EY_k] = EX + EY.$$

Finalmente, sejam X e Y duas variáveis aleatórias quaisquer. Temos que

$$(X+Y)^+ - (X+Y)^- = X + Y = X^+ - X^- + Y^+ - Y^-,$$

logo

$$(X+Y)^+ + X^- + Y^- = (X+Y)^- + X^+ + Y^+.$$

Como todas as variáveis aleatórias acima são não-negativas, pelo caso anterior temos

$$E[(X+Y)^+] + EX^- + EY^- = E[(X+Y)^-] + EX^+ + EY^+.$$

Supondo que $EX+EY$ está definido, necessariamente temos que $EX^-+EY^- < \infty$ ou $EX^+ + EY^+ < \infty$. Consideramos sem perda de generalidade o primeiro caso. Como $(X+Y)^- \leqslant X^- + Y^-$, temos $E[(X+Y)^-] \leqslant EX^- + EY^- < \infty$, e portanto podemos subtrair, obtendo

$$E[(X+Y)^+] - E[(X+Y)^-] = (EX^+ - EX^-) + (EY^+ - EY^-). \qquad \square$$

Demonstração do Teorema 4.27. Tome $Y_k = g_k(Y) = g_k(h(\boldsymbol{X}))$. Temos que

$$EY_k = \sum_{\boldsymbol{x}} g_k(h(\boldsymbol{x})) \cdot p_{\boldsymbol{X}}(\boldsymbol{x}) + \int_{\mathbb{R}^d} g_k(h(\boldsymbol{x}))\, f_{\boldsymbol{X}}(\boldsymbol{x})\, \mathrm{d}^d\boldsymbol{x}.$$

Portanto,

$$EY = \lim_{k\to\infty} EY_k \leqslant \sum_{\boldsymbol{x}} h(\boldsymbol{x}) \cdot p_{\boldsymbol{X}}(\boldsymbol{x}) + \int_{\mathbb{R}^d} h(\boldsymbol{x})\, f_{\boldsymbol{X}}(\boldsymbol{x})\, \mathrm{d}^d\boldsymbol{x}.$$

Por outro lado,

$$\begin{aligned} EY_k &= E[Y_k \cdot \mathbb{1}_{h(\boldsymbol{X})\leqslant k}] + E[Y_k \cdot \mathbb{1}_{h(\boldsymbol{X})>k}] \\ &= \sum_{\boldsymbol{x}\in A_k} g_k(h(\boldsymbol{x})) \cdot p_{\boldsymbol{X}}(\boldsymbol{x}) + \int_{A_k} g_k(h(\boldsymbol{x}))\, f_{\boldsymbol{X}}(\boldsymbol{x})\, \mathrm{d}^d\boldsymbol{x} + E[Y_k \cdot \mathbb{1}_{h(\boldsymbol{X})>k}] \\ &\geqslant \sum_{\boldsymbol{x}\in A_k} h(\boldsymbol{x}) \cdot p_{\boldsymbol{X}}(\boldsymbol{x}) + \int_{A_k} h(\boldsymbol{x})\, f_{\boldsymbol{X}}(\boldsymbol{x})\, \mathrm{d}^d\boldsymbol{x} - 2^{-k}, \end{aligned}$$

onde $A_k = \{\boldsymbol{x} \in \mathbb{R}^d : h(\boldsymbol{x}) \leqslant k\} \nearrow \mathbb{R}^d$. Fazendo $k \to \infty$, temos que[1]

$$\lim_{k\to\infty} EY_k \geqslant \sum_{\boldsymbol{x}} h(\boldsymbol{x}) \cdot p_{\boldsymbol{X}}(\boldsymbol{x}) + \int_{\mathbb{R}^d} h(\boldsymbol{x})\, f_{\boldsymbol{X}}(\boldsymbol{x})\, \mathrm{d}^d\boldsymbol{x}.$$

Portanto, pelo Lema 4.33 temos

$$EY = \lim_{k\to\infty} EY_k = \sum_{\boldsymbol{x}} h(\boldsymbol{x}) \cdot p_{\boldsymbol{X}}(\boldsymbol{x}) + \int_{\mathbb{R}^d} h(\boldsymbol{x})\, f_{\boldsymbol{X}}(\boldsymbol{x})\, \mathrm{d}^d\boldsymbol{x},$$

como queríamos demonstrar. $\square$

[1] Ao invés de justificar que a integral de uma função não-negativa sobre A_k converge para a integral sobre $\mathbb{R}^d$, diremos que a integral deve ser calculada tomando esse limite.

Demonstração da Proposição 4.32. Suponha que X e Y são variáveis aleatórias não-negativas. Usando o Teorema 4.16, temos que

$$E[XY] = \lim_k E[g_k(X)g_k(Y)] = \lim_k [Eg_k(X) \cdot Eg_k(Y)] = EX \cdot EY,$$

onde os limites serão justificados pelo Teorema 10.1.

Suponha agora que X e Y são integráveis. Usando o caso anterior, temos

$$\begin{aligned} E[XY] &= E[X^+Y^+ - X^+Y^- - X^-Y^+ + X^-Y^-] \\ &= EX^+EY^+ - EX^+EY^- - EX^-EY^+ + EX^-EY^- = EX \cdot EY. \end{aligned}$$ □

4.4 Momentos, Variância e Covariância

Definição 4.34. Dado $k = 1, 2, 3, \dots$, definimos o *momento de ordem* k, ou o k*-ésimo momento* da variável aleatória X como EX^k. Se X é integrável, definimos o k*-ésimo momento central* por $E(X - EX)^k$. O *momento absoluto de ordem* k é definido como $E|X|^k$.

Exemplo 4.35. Se $X \sim U[0, 1]$, temos

$$EX = \int_0^1 x \, \mathrm{d}x = \frac{1}{2}, \quad EX^2 = \int_0^1 x^2 \, \mathrm{d}x = \frac{1}{3}, \quad EX^k = \int_0^1 x^k \, \mathrm{d}x = \frac{1}{k+1},$$

e o segundo momento central é dado por

$$E\left[\left(X - \tfrac{1}{2}\right)^2\right] = \int_0^1 \left(x - \tfrac{1}{2}\right)^2 \mathrm{d}x = \frac{1}{12}.$$

O segundo momento central recebe o nome de variância.

Definição 4.36 (Variância)**.** Seja X uma variável aleatória integrável. Define-se a *variância* da variável aleatória X, denotada por VX ou $\sigma^2(X)$, como

$$VX = E[(X - EX)^2].$$

Exemplo 4.37. Pelo exemplo anterior, se $X \sim U[0, 1]$, então $EX = \frac{1}{2}$ e $VX = \frac{1}{12}$.

Proposição 4.38 (Propriedades da Variância)**.** *Seja X uma variável aleatória integrável. Então:*

1. *$VX \geqslant 0$.*
2. *$VX = EX^2 - (EX)^2$.*
3. *$VX = 0$ se e somente se $P(X = c) = 1$ para algum $c \in \mathbb{R}$, neste caso $X = EX$.*
4. *$VX \leqslant EX^2$.*
5. *$V(X - b) = VX$.*
6. *$V(aX) = a^2 VX$.*

Demonstração. Feita em aula. □

Exemplo 4.39. Se $X \sim \text{Bernoulli}(\frac{1}{2})$, temos

$$EX = \frac{1}{2}, \quad EX^2 = \frac{1}{2}, \quad VX = EX^2 - (EX)^2 = \frac{1}{4}.$$

Definição 4.40 (Desvio-Padrão)**.** O *desvio-padrão* $\sigma(X)$ é dado pela raiz quadrada da variância

$$\sigma(X) = \sqrt{VX},$$

e mede a dispersão de X em torno de sua média. O desvio-padrão tem a mesma unidade de medida de X.

Exemplo 4.41. Se $X \sim \text{Bernoulli}(\frac{1}{2})$, temos

$$\sigma(X) = \sqrt{VX} = \sqrt{1/4} = \frac{1}{2}.$$

Ou seja, uma variável Bernoulli($\frac{1}{2}$) varia em média $\sigma = \frac{1}{2}$ unidade em torno de seu valor esperado $\mu = \frac{1}{2}$.

As propriedades do desvio-padrão são análogas:

1. $\sigma(X) \geqslant 0$.
2. $\sigma(X) = 0$ se e somente se $P(X = c) = 1$ para algum $c \in \mathbb{R}$.
3. $\sigma(X) \leqslant \sqrt{EX^2}$.
4. $\sigma(X - b) = \sigma(X)$ para todo $b \in \mathbb{R}$.
5. $\sigma(aX) = |a|\, \sigma(X)$ para todo $a \in \mathbb{R}$.

Definição 4.42 (Covariância)**.** Dadas duas variáveis aleatórias X e Y com segundo momento finito, uma forma de medir a dependência linear da dispersão dessas variáveis é através da sua *covariância* $\operatorname{Cov}(X,Y)$, dada por

$$\operatorname{Cov}(X,Y) = E\left[(X-EX)(Y-EY)\right].$$

Proposição 4.43 (Propriedades da Covariância)**.** *Dadas X e Y com segundo momento finito:*

1. $\operatorname{Cov}(X,Y) = E[XY] - EX\cdot EY$.
2. $\operatorname{Cov}(X,Y) = 0$ *se e somente se* $E[XY] = EX\cdot EY$.
3. $\operatorname{Cov}(cX,Y) = c\operatorname{Cov}(X,Y)$.
4. $\operatorname{Cov}(X,Y) = \operatorname{Cov}(Y,X)$.
5. $\operatorname{Cov}(X,X) = VX$.
6. $\operatorname{Cov}(X,c) = 0$ *para todo* $c\in\mathbb{R}$.
7. $\operatorname{Cov}(X,Y+Z) = \operatorname{Cov}(X,Y)+\operatorname{Cov}(X,Z)$.
8. $\operatorname{Cov}(\sum_i a_iX_i, \sum_j b_jY_j) = \sum_i\sum_j a_ib_j\operatorname{Cov}(X_i,Y_j)$.

Demonstração. Feita em aula. □

Exemplo 4.44. Se $f_{XY}(x,y) = \mathbb{1}_{[0,1]}(x)\mathbb{1}_{[0,1]}(y)$, $Z = X\wedge Y$, $W = X\vee Y$, então:

$$\begin{aligned}
E[ZW] = E[XY] &= \int_0^1\int_0^1 xy\,\mathrm{d}x\mathrm{d}y = \frac{1}{4}\\
EZ &= \int_0^1\left[\int_0^x y\mathrm{d}y + \int_x^1 x\mathrm{d}y\right]\mathrm{d}x = \int_0^1 (\tfrac{x^2}{2}+x-x^2)\mathrm{d}x = \frac{1}{2}-\frac{1}{6} = \frac{1}{3}\\
EW &= \int_0^1\left[\int_0^x x\mathrm{d}y + \int_x^1 y\mathrm{d}y\right]\mathrm{d}x = \int_0^1 (x^2+\tfrac{1}{2}-\tfrac{x^2}{2})\mathrm{d}x = \frac{1}{3}+\frac{1}{2}-\frac{1}{6} = \frac{2}{3}\\
\operatorname{Cov}(Z,W) = E[ZW]-EZ\cdot EW &= \frac{1}{4}-\frac{2}{9} = \frac{1}{36}.
\end{aligned}$$

Observação 4.45. Se as variáveis aleatórias X e Y são independentes e integráveis então X e Y são não-correlacionadas, i.e., $\operatorname{Cov}(X,Y)=0$. Entretanto, nem sempre vale a recíproca, isto é, $E[XY] = EX\cdot EY$ não implica X e Y independentes.

Contra-Exemplo 4.46. Sejam X e Y variáveis aleatórias tomando valores $-1,0,1$ com distribuição conjunta dada por $p(-1,-1) = p(-1,1) = p(1,-1) = p(1,1) =$

$p(0,0) = \frac{1}{5}$. Então $EXY = EX \cdot EY$, mas X e Y não são independentes, pois $P(X = 0, Y = 0) \neq P(X = 0)P(Y = 0)$.

Definição 4.47 (Coeficiente de Correlação)**.** Dadas X e Y com variâncias finitas e positivas, o *coeficiente de correlação* $\rho(X, Y)$ de X e Y é uma medida padronizada da dependência linear entre X e Y:

$$\rho(X,Y) = \operatorname{Cov}\left(\frac{X}{\sigma(X)}, \frac{Y}{\sigma(Y)}\right).$$

O coeficiente de correlação não tem unidade de medida.

Proposição 4.48 (Propriedades do Coeficiente de Correlação)**.** *Dadas X e Y com variâncias finitas e positivas, valem:*

1. $\rho(X, Y) = \rho(Y, X)$.
2. $\rho(X, Y) = \frac{\operatorname{Cov}(X,Y)}{\sigma(X)\sigma(Y)}$.
3. $\rho(X, X) = 1$.
4. $\rho(X, \pm aY + b) = \pm\rho(X, Y)$ *se* $a > 0$ *e* $b \in \mathbb{R}$.

Demonstração. Feita em aula. □

Exemplo 4.49. No Exemplo 4.44, temos

$$\begin{aligned}
EZ^2 &= \int_0^1 \left[\int_0^x y^2 \mathrm{d}y + \int_x^1 x^2 \mathrm{d}y\right] \mathrm{d}x = \int_0^1 (\tfrac{x^3}{3} + x^2 - x^3)\mathrm{d}x = \frac{1}{3} - \frac{1}{6} = \frac{1}{6} \\
VZ &= EZ^2 - (EZ)^2 = \frac{1}{6} - \frac{1}{9} = \frac{1}{18} \\
VW &= \cdots \text{exercício} \cdots = \frac{1}{18} \\
\rho(Z,W) &= \frac{\operatorname{Cov}(Z,W)}{\sigma(Z)\sigma(W)} = \frac{1/36}{\sqrt{1/18}\sqrt{1/18}} = \frac{1}{2}.
\end{aligned}$$

Dada uma variável aleatória X com $EX^2 < \infty$, definimos a *padronização de* X, ou a *normalização de* X, como

$$\frac{X - EX}{\sigma(X)}.$$

A padronização de uma variável aleatória não tem unidade de medida.

Exercício 4.2. Mostre que:

1. $EZ = 0$ e $VZ = 1$, onde Z é a padronização de X.
2. X e $(aX + b)$ têm a mesma padronização para $a > 0$ e $b \in \mathbb{R}$.
3. Se Z é a padronização de X e W é a padronização de Y, então

$$\rho(Z, W) = \operatorname{Cov}(Z, W) = E(ZW) = \rho(X, Y).$$

Proposição 4.50. *Sejam X e Y variáveis aleatórias com variâncias finitas e positivas. Então:*

1. $\rho(X, Y) \in [-1, +1]$.
2. $|\operatorname{Cov}(X, Y)| \leqslant \sigma(X)\sigma(Y)$.
3. $\rho(X, Y) = \pm 1 \Leftrightarrow \operatorname{Cov}(X, Y) = \pm\sigma(X)\sigma(Y) \Leftrightarrow P(Y = \pm aX + b) = 1$, $a > 0$.

Veremos a demonstração na próxima seção, como corolário da Desigualdade de Cauchy-Schwarz.

4.5 Desigualdades Básicas

Definição 4.51 (Funções côncavas e convexas)**.** Seja $B \subseteq \mathbb{R}$ um intervalo. Dizemos que $g : B \to \mathbb{R}$ é *convexa* se satisfaz às seguintes condições equivalentes:

(*i*) Para todos $a < x < b$ em B, $g(x) \leqslant g(a) + \frac{x-a}{b-a}[g(b) - g(a)]$.

(*ii*) Para todo $a \in B$, existe $c \in \mathbb{R}$ tal que $g(x) \geqslant g(a) + c(x - a)$ para todo $x \in B$.

(*iii*) g' é não-decrescente em B (caso g seja diferenciável).

(*iv*) $g''(x) \geqslant 0$ para todo $x \in B$ (caso g tenha segunda derivada).

Dizemos que g é *côncava* se $-g$ é convexa.

Exemplo 4.52. São funções convexas: $g(x) = x^2$, $g(x) = e^x$, $g(x) = |x|$, $g(x) = x^{-1}$. São funções côncavas: $g(x) = \log x$ em $(0, \infty)$, $g(x) = x$.

Proposição 4.53 (Desigualdade de Jensen)**.** *Seja $g : B \to \mathbb{R}$ uma função convexa e X uma variável aleatória integrável assumindo valores em B. Então*

$$E[g(X)] \geqslant g(EX).$$

Demonstração. Tomando $a = EX$ e c tal que $g(x) \geqslant g(a)+c(x-a)$ para todo $x \in I$, temos $E[g(X)] \geqslant E[g(EX) + c(X - EX)] = g(EX) + cEX - cEX = g(EX)$. $\square$

Corolário 4.54. *Se g é uma função côncava, então $E[g(X)] \leqslant g(EX)$.*

Corolário 4.55. *Seja X uma variável aleatória integrável. Então*

(a) $E|X| \geqslant |EX|$.

(b) $EX^2 \geqslant (EX)^2$.

(c) $E\left(\frac{1}{X}\right) \geqslant \frac{1}{EX}$ *se* $X \geqslant 0$.

(d) $E\,|X|^p \geqslant (E\,|X|)^p \geqslant |EX|^p$ *para* $p \geqslant 1$.

(e) $(E|X|^t)^{\frac{1}{t}} \geqslant (E|X|^s)^{\frac{1}{s}}$ *se* $0 < s \leqslant t$.

Demonstração. (a), (b) e (c) são imediatos. Para (d) usamos $g(x) = x^p$ em $(0, \infty)$ e depois (a). Para (e), tomamos $Y = |X|$ e $g(y) = y^{t/s}$ em $(0, \infty)$. Temos que g é convexa pois $g'(y) = \frac{t}{s} y^{\frac{t}{s}-1}$ é não-decrescente. Logo, $E|X|^t = E[(Y^s)^{t/s}] \geqslant [EY^s]^{t/s}$. Elevando todos os temos a $1/t$, temos a desigualdade desejada. $\square$

Proposição 4.56 (Desigualdade Básica de Tchebyshev)**.** *Seja X uma variável aleatória não-negativa e seja $\lambda > 0$ uma constante. Então*

$$P(X \geqslant \lambda) \leqslant \frac{E(X)}{\lambda}.$$

Demonstração. Tome $Y = \lambda \mathbb{1}_{[X \geqslant \lambda]}$. Temos que $Y \leqslant X$, logo

$$EX \geqslant EY = \lambda P(X \geqslant \lambda),$$

donde segue a desigualdade. $\square$

Exemplo 4.57. Se uma empresa recebe em média 100 chamadas telefônicas por dia, queremos estimar a probabilidade de, num certo dia, receber mais de 300 chamadas. Temos

$$P(X \geqslant 300) \leqslant \frac{EX}{300} = \frac{1}{3}.$$

Ou seja, esse evento ocorre com probabilidade igual a, no máximo, $\frac{1}{3}$.

Exercício 4.3. Suponha que X seja uma variável aleatória tal que $P(X \geqslant 0) = 1$ e $P(X \geqslant 10) = \frac{1}{5}$. Mostre que $E(X) \geqslant 2$.

Proposição 4.58 (Desigualdade de Markov)**.** *Seja X uma variável aleatória qualquer e seja $\lambda > 0$ uma constante. Então para todo $t > 0$,*

$$P(|X| \geqslant \lambda) \leqslant \frac{E\,|X|^t}{\lambda^t}.$$

Demonstração. Defina $Y = |X|^t$ e use a desigualdade básica com Y e λ^t:

$$P(|X| \geqslant \lambda) = P(Y \geqslant \lambda^t) \leqslant \frac{EY}{\lambda^t} = \frac{E|X|^t}{\lambda^t}. \qquad \square$$

Proposição 4.59 (Desigualdade Clássica de Tchebyshev)**.** *Seja X uma variável aleatória integrável e seja $\lambda > 0$ uma constante. Então*

$$P\big(|X - E(X)| \geqslant \lambda\big) \leqslant \frac{VX}{\lambda^2}.$$

Demonstração. Tomando $Y = (X - EX)^2$, temos $EY = VX$ e, aplicando a desigualdade básica,

$$P\big(|X - EX| \geqslant \lambda\big) = P(Y \geqslant \lambda^2) \leqslant \frac{EY}{\lambda^2} = \frac{VX}{\lambda^2}. \qquad \square$$

Exemplo 4.60. Estimar a probabilidade de uma variável aleatória X não diferir de sua média μ por mais que duas vezes o valor do seu desvio-padrão σ. Temos

$$P(\mu - 2\sigma < X < \mu + 2\sigma) = 1 - P\big(|X - EX| \geqslant 2\sigma\big) \geqslant 1 - \frac{VX}{(2\sigma)^2} = 1 - \frac{\sigma^2}{4\sigma^2} = \frac{3}{4}.$$

Exercício 4.4. Suponha que X seja uma variável aleatória tal que $E(X) = 10$, $P(X \leqslant 7) = 0,2$ e $P(X \geqslant 13) = 0,3$. Prove que $VX \geqslant \frac{9}{2}$.

Teorema 4.61 (Desigualdade de Cauchy-Schwarz)**.** *Se $EX^2 < \infty$ e $EY^2 < \infty$, então XY é integrável e*

$$E[XY] \leqslant \sqrt{EX^2}\,\sqrt{EY^2}.$$

Ainda, se $E[XY] = \sqrt{EX^2}\,\sqrt{EY^2}$, então existe $c \geqslant 0$ tal que $P(Y = cX) = 1$, ou então $P(X = 0) = 1$.

Demonstração. Sejam $a = \sqrt{EX^2}$ e $b = \sqrt{EY^2}$. Se $a = 0$ ou $b = 0$, o teorema vale trivialmente. Assumimos então que $0 < a < \infty$ e $0 < b < \infty$. Observamos que

$$0 \leqslant E\left(\frac{X}{a} - \frac{Y}{b}\right)^2 = E\left(\frac{X^2}{a^2} - 2\frac{XY}{ab} + \frac{Y^2}{b^2}\right) = 2 - \frac{2\,E[XY]}{ab},$$

donde

$$E[XY] \leqslant ab = \sqrt{EX^2}\,\sqrt{EY^2}.$$

Reciprocamente, suponha que $E[XY] = ab$. Temos que $E\left(\frac{X}{a} - \frac{Y}{b}\right)^2 = 0$, logo $P(\frac{X}{a} - \frac{Y}{b} = 0) = 1$ e portanto $P(Y = cX) = 1$ com $c = \frac{a}{b}$. □

Demonstração da Proposição 4.50. Tomamos

$$Z = \frac{X - EX}{\sigma(X)} \qquad \text{e} \qquad W = \frac{Y - EY}{\sigma(Y)}$$

Pela Desigualdade de Cauchy-Schwarz temos que

$$\rho(X,Y) = E[ZW] \leqslant \sqrt{EZ^2}\,\sqrt{EW^2} = +1,$$

donde

$$\operatorname{Cov}(X,Y) = \sigma(X)\sigma(Y)\rho(X,Y) \leqslant +\sigma(X)\sigma(Y).$$

Ainda, se $\rho(X,Y) = +1$, então $W = +cZ$ com $c \geqslant 0$. Mas $1 = EW^2 = c^2 EZ^2 = c^2$, logo $c = +1$ e portanto

$$Y = EY + \sigma(Y) \cdot Z = EY + \sigma(Y)\frac{X - EX}{\sigma(X)}.$$

As propriedades análogas com -1 no lugar de $+1$ seguem do caso anterior tomando-se $-X$ no lugar de X:

$$\begin{aligned}
\rho(X,Y) &= -\rho(-X,Y) \geqslant -1,\\
\operatorname{Cov}(X,Y) &= -\operatorname{Cov}(-X,Y) \geqslant -\sigma(X)\sigma(Y),\\
Y &= EY + \sigma(Y)\frac{-X - E[-X]}{\sigma(-X)} = EY - \sigma(Y)\frac{X - EX}{\sigma(X)}.
\end{aligned}$$

□

4.6 Esperança Condicional dado um Evento

A informação sobre a ocorrência de um certo evento $A \in \mathcal{F}$ com $P(A) > 0$ leva à definição de uma nova medida P' em $(\Omega, \mathcal{F})$, dada pela relação $P'(B) = P(B|A)$, $B \in \mathcal{F}$. A distribuição de qualquer variável aleatória X também é afetada neste caso. Como vimos no Capítulo 2, X passa a ter uma nova função de distribuição $F_{X|A}(t)$, $t \in \mathbb{R}$, uma nova lei $P_{X|A}(B)$, $B \in \mathcal{B}$.

Nesta situação, X também terá um novo valor esperado $E(X|A)$. No caso de X ser mista com componentes discreta e absolutamente contínua, sua esperança condicional dado A será dada por

$$E(X|A) = \sum_x x \cdot p_{X|A}(x) + \int_{\mathbb{R}} x\, f_{X|A}(x)\,\mathrm{d}x.$$

No caso discreto, escolhemos a forma mais conveniente entre calcular

$$F_{X|A}(t) = P(X \leqslant t \,|\, A)\ \forall t$$

ou

$$p_{X|A}(x) = P(X = x \,|\, A)\ \forall x.$$

Exemplo 4.62. Seja X a variável aleatória que representa o resultado do lançamento de um dado, isto é, $X \sim U_d\{1, 2, 3, 4, 5, 6\}$. Vamos calcular $E(X \mid X \text{ par})$. Primeiro encontramos a função de probabilidade condicional:

$$p_{X|A}(x) = P(X = x|A) = \frac{1}{3}\,\mathbb{1}_{\{2,4,6\}}(x)$$

e em seguida a esperança

$$E(X|A) = \sum_x x \cdot p_{X|A}(x) = 4.$$

No caso contínuo, em geral calculamos

$$F_{X|A}(t) = P(X \leqslant t \,|\, A)\ \forall t$$

e depois fazemos

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

Exemplo 4.63. Seja X uma variável aleatória com distribuição $X \sim U[0,1]$. Vamos calcular $E(X \mid X < \frac{1}{2})$. Primeiro encontramos a função de distribuição condicional

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

logo a densidade condicional

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

e finalmente a esperança condicional

$$E(X|A) = \int_{\mathbb{R}} x\, f_{X|A}(x)\,\mathrm{d}x = \frac{1}{4}.$$

4.7 Exercícios

4.5. Calcular EX, onde:

1. $X \sim \mathrm{Geom}(p)$.
2. $X \sim \mathcal{N}(\mu, \sigma^2)$.

4.6. Considere o seguinte jogo de azar. Uma urna contém 18 bolas, sendo 9 azuis e 9 brancas. Retiram-se 3 bolas da urna ao acaso. As bolas retiradas são descartadas e o jogador marca 1 ponto se pelo menos 2 dessas 3 bolas forem azuis. Em seguida retiram-se outras 3 bolas da urna ao acaso, as bolas retiradas são descartadas e o jogador marca 1 ponto se pelo menos 2 dessas 3 bolas forem azuis. Repete-se o procedimento até que a urna esteja vazia. Ao final, o jogador recebe um prêmio X igual ao total de pontos marcados. Calcule EX.

4.7. Dada X variável aleatória, defina

$$Y = \begin{cases} X, & X \leqslant a, \\ a, & \text{caso contrário}, \end{cases}$$

onde a é uma constante positiva. Mostre que $EY \leqslant EX$.

4.8. Mostre que X é integrável se, e somente se, $E|X| < \infty$.

4.9. Seja X uma variável aleatória simétrica em torno de μ, isto é, $P(X \geqslant \mu + x) = P(X \leqslant \mu - x)$ para todo $x \in \mathbb{R}$. Mostre que se X é integrável, então $E(X) = \mu$.

4.10. Seja X uma variável aleatória. Mostre que X é integrável se, e somente se

$$\sum_{n=0}^{\infty} P\big(|X| \geqslant n\big) < \infty.$$

4.11. Prove que $E|X| \leqslant \sqrt{EX^2}$.

4.12. Sejam $X_1, \dots, X_n$ variáveis aleatórias satisfazendo $EX_i^2 < \infty \ \forall i$.

1. Se $\mathrm{Cov}(X_i, X_j) = 0 \ \forall i \neq j$, mostre que

$$V\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} V X_i.$$

2. A identidade acima também vale se as variáveis aleatórias forem independentes?

4.13. Calcular VX, onde:

1. $X \sim \mathrm{Geom}(\lambda)$.
2. $X \sim \mathrm{Poisson}(\lambda)$.
3. $X \sim b(n, p)$.
4. $X \sim \exp(\lambda)$.
5. $X \sim \mathcal{N}(\mu, \sigma^2)$.

4.14. Considere uma sequência de variáveis aleatórias $X_1, X_2, X_3, \dots$ i.i.d. com distribuição Bernoulli(p). Quantas realizações são suficientes para que a média amostral, dada por

$$\bar{X}_n(\omega) = \frac{1}{n} \sum_{j=1}^{n} X_n(\omega),$$

não difira de seu valor esperado p por mais de 0,01, com probabilidade mínima de 0,95? (Sugestão: Desigualdade de Tchebyshev)

4.15. Seja $X \sim U[-1,1]$ e sejam $A_1 = [X \geqslant 0]$ e $A_2 = [X < 0]$. Pede-se

1. A distribuição condicional de X dado A_1.
2. A distribuição condicional de X dado A_2.
3. $E(X|A_1)$.
4. $E(X|A_2)$.

4.16. Seja X uma variável aleatória exponencial com parâmetro λ. Encontre $E[X \mid X > 2]$.

4.17. Se $X \sim \text{Geom}(p)$, encontre $E[X \mid X > 5]$.

4.18. Se X tem função de probabilidade

$$p_X(n) = \frac{n\lambda^n e^{-\lambda}}{\lambda . n!}$$

para $n = 0, 1, 2, 3, \ldots$, calcule VX.
Dica: desenvolver $(n-1)(n-2+1) + 2(n-1) + 1$.

4.19. [Jam04, Capítulo 3].
Recomendados: 5, 6, 19, 20ab, 21, 23, 26, 28, 30, 36.

Capítulo 5

Convergência de Variáveis Aleatórias

Considere uma sequência de variáveis aleatórias $X_1, X_2, X_3, \ldots$. Em inúmeras situações teóricas e práticas, uma pergunta natural é qual o comportamento de longo prazo da sequência $(X_n)_n$. Dito de outra forma: como se comporta X_n quando n é suficientemente grande?

Tratando-se de variáveis aleatórias, o conceito de convergência é uma generalização do conceito de convergência para números reais. Entretanto, existem várias formas de se fazer essa generalização, e cada forma é a mais natural em determinado contexto. No caso de variáveis aleatórias degeneradas, todas as definições são equivalentes à convergência de números reais.

5.1 Lema de Borel-Cantelli

Começamos definindo o $\liminf$ e o $\limsup$ de uma sequência de eventos.

Definição 5.1 ($\limsup$ e $\liminf$ de eventos)**.** Dada uma sequência de eventos aleatórios A_n, definimos o evento $\limsup A_n$, denotado por $[A_n$ infinitas vezes$]$ ou

$[A_n \text{ i.v.}]$, por

$$\limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k.$$

Definimos o evento $\liminf A_n$, denotado por $[A_n \text{ eventualmente}]$, por

$$\liminf_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k.$$

É importante entender as seguintes interpretações:

- $\limsup A_n$ é o conjunto dos ω's tais que ω pertence a infinitos A_n's.
- O evento $\limsup A_n$ significa "A_n acontece infinitas vezes".
- $\liminf A_n$ é o conjunto dos ω's tais que ω pertence a todos os A_n's exceto uma quantidade finita deles.
- O evento $\liminf A_n$ significa "A_n acontece para todo n grande".

Além disso, vale que

$$\liminf A_n \subseteq \limsup A_n$$

e

$$\liminf(A_n^c) = (\limsup A_n)^c.$$

Exemplo 5.2. Exemplo: $\Omega = \mathbb{R}$, $A_n = \begin{cases} (-1/n, 1], & n \text{ ímpar}, \\ (-1, 1/n], & n \text{ par}. \end{cases}$

Temos

$$\limsup A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k = \bigcap_{n=1}^{\infty} (-1,1] = (-1,1]$$

e

$$\liminf A_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k = \bigcup_{n=1}^{\infty} \{0\} = \{0\}.$$

Exercício 5.1. Sejam um espaço de probabilidade $(\Omega, \mathcal{F}, P)$ e uma sequência de eventos aleatórios (A_n) em $\mathcal{F}$. Mostre que, se (A_n) é crescente, então

$$\limsup A_n = \liminf A_n = \cup_{n=1}^{\infty} A_n.$$

Por outro lado, se (A_n) é decrescente, então

$$\limsup A_n = \liminf A_n = \cap_{n=1}^{\infty} A_n.$$

Exercício 5.2. Considere o espaço de probabilidade $(\mathbb{R}^2, \mathcal{B}^2, P)$, no qual P é uma probabilidade arbitrária. Se $A_n = \{(x, y) \in \mathbb{R}^2 : 0 \leqslant x \leqslant n, 0 \leqslant y \leqslant \frac{1}{n}\}$, encontre $\limsup A_n$ e $\liminf A_n$.

Exercício 5.3. Considere a sequência de intervalos

$$A_n = \begin{cases} (0, 2 + \frac{1}{n}), & n \text{ par} \\ (0, 2 - \frac{1}{n}), & n \text{ ímpar.} \end{cases}$$

Encontre o $\liminf A_n$ e o $\limsup A_n$.

Teorema 5.3 (Lema de Borel-Cantelli)**.** *Seja $(\Omega, \mathcal{F}, P)$ um espaço de probabilidade e (A_n) uma sequência de eventos aleatórios. Então:*

1. *Se $\sum_{n=1}^{\infty} P(A_n) < \infty$ então*

$$P(A_n \textit{ infinitas vezes}) = 0.$$

2. *Se $\sum_{n=1}^{\infty} P(A_n) = \infty$ e os eventos A_n são independentes, então*

$$P(A_n \textit{ infinitas vezes}) = 1.$$

Demonstração. Feita em aula, seguindo [Jam04, p. 201]. □

Exemplo 5.4. Considere uma sequência de infinitos sorteios independentes e uniformes de um número $(X_n)_{n \in \mathbb{N}}$ entre 0 e 1. Então

1. $P(X_n \in [0, 1/n]$ para infinitos n's$) = 1$.
2. $P(X_n \in [0, 1/n^2]$ para infinitos n's$) = 0$.

Exemplo 5.5. Seja $X_1 \sim U[0, 1]$ e tome $X_2, X_3, X_4, \ldots$ todas iguais a X_1. Então

1. $P(X_n \in [0, 1/n]$ para infinitos n's$) = 0$.
2. $P(X_n \in [0, 1/n^2]$ para infinitos n's$) = 0$.

Podemos afirmar que vale a recíproca do Lema de Borel-Cantelli, ou seja, que $P(A_n \text{ i.v.}) = 0$ implica $\sum_n P(A_n) < \infty$, quando os (A_n) são independentes. Caso contrário, podemos ter $P(A_n \text{ i.v.}) = 0$ sem que necessariamente $\sum_n P(A_n) < \infty$. Neste caso podemos afirmar pelo menos que $P(A_n) \to 0$.

Proposição 5.6. *Se $P(A_n$ infinitas vezes$) = 0$ então $P(A_n) \to 0$.*

Demonstração. Tomando $B_k = \cup_{n \geqslant k} A_n$, temos que $B_k \searrow [A_n \text{ i.v.}]$ quando $k \to \infty$. Como $B_k \supseteq A_k$, vale $P(A_k) \leqslant P(B_k) \to P(A_n \text{ i.v.}) = 0$. $\square$

Observação 5.7 (Lei 0-1 para Infinitos Eventos Independentes)**.** Uma consequência imediata do Lema de Borel-Cantelli é a seguinte. Se $(A_n)_{n \in \mathbb{N}}$ é uma sequência de eventos independentes, então $P(A_n$ infinitas vezes$) = 0$ ou 1.

5.2 Convergência de Variáveis Aleatórias

Sejam X e $(X_n)_{n \in \mathbb{N}}$ variáveis aleatórias definidas num mesmo espaço de probabilidade $(\Omega, \mathcal{F}, P)$.

Definição 5.8 (Convergência em Probabilidade)**.** Dizemos que X_n *converge em probabilidade para* X, denotado por $X_n \xrightarrow{P} X$, se para todo $\varepsilon > 0$

$$P\big(\,|X_n - X| \geqslant \varepsilon\big) \to 0 \text{ quando } n \to \infty.$$

Exemplo 5.9. Sejam $X_1, X_2, \dots$ variáveis aleatórias independentes, tais que $X_n \sim \text{Bernoulli}(\frac{1}{n})$. Temos para $\varepsilon < 1$ que

$$P\big(\,|X_n - 0| \geqslant \varepsilon\big) = P(X_n = 1) = \frac{1}{n} \to 0,$$

e portanto $X_n \xrightarrow{P} 0$.

Exemplo 5.10. Sejam $X_1, X_2, \dots$ variáveis aleatórias independentes, identicamente distribuídas com distribuição $\exp(1)$ e tome

$$Y_n = \frac{X_n}{\log n}.$$

Então
$$P\left(\left|\tfrac{X_n}{\log n} - 0\right| \geqslant \varepsilon\right) = P(X_n \geqslant \varepsilon \log n) = n^{-\varepsilon} \to 0,$$
e portanto $\frac{X_n}{\log n} \overset{P}{\to} 0$.

Definição 5.11 (Convergência Quase Certa)**.** Dizemos que X_n *converge quase certamente para* X, denotado por $X_n \overset{\text{q.c.}}{\to} X$, se

$$P\left(X_n \to X \text{ quando } n \to \infty\right) = 1,$$

ou seja, o evento $A_0 = \{\omega : X_n(\omega) \to X(\omega)\}$ é de probabilidade 1.

Observação 5.12. A convergência quase certa é uma convergência pontual num conjunto de medida 1, ou seja, $X_n(\omega) \to X(\omega)$ para quase todo ω, exceto aqueles dentro de um conjunto de medida nula. Por outro lado convergência em probabilidade não diz respeito à convergência pontual, ela apenas afirma que para valores grandes de n as variáveis X_n e X são aproximadamente iguais com probabilidade muito alta.

Exemplo 5.13. Um ponto ω é selecionado uniformemente do intervalo $\Omega = [0, 1]$. Seja $(X_n)_n$ a sequência de variáveis aleatórias dada por

$$X_n(\omega) = \omega + \omega^n.$$

Então $X_n \overset{\text{q.c.}}{\to} X$ com $X \sim U[0, 1]$. De fato, tomando $X(\omega) = \omega$, temos que $X_n(\omega) \to X(\omega)$ para todo $\omega \in [0, 1)$. Como $P\big([0, 1)\big) = 1$, segue que $X_n \overset{\text{q.c.}}{\to} X$.

Proposição 5.14. *$X_n \overset{\text{q.c.}}{\to} X$ se, e somente se,*

$$P\left(|X_n - X| \geqslant \varepsilon \textit{ infinitas vezes}\right) = 0 \quad \forall \varepsilon > 0.$$

Demonstração. A proposição segue da seguinte cadeia de equivalências:

$$\begin{aligned}
&P(X_n \to X) = 1 \\
&P(\forall \varepsilon > 0,\ |X_n - X| < \varepsilon \text{ eventualmente}) = 1 \\
&P(\nexists \varepsilon > 0 \text{ tal que } |X_n - X| \geqslant \varepsilon \text{ i.v.}) = 1 \\
&P\left(\nexists k \in \mathbb{N} \text{ tal que } |X_n - X| \geqslant \tfrac{1}{k} \text{ i.v.}\right) = 1 \\
&P\left(\exists k \in \mathbb{N} \text{ tal que } |X_n - X| \geqslant \tfrac{1}{k} \text{ i.v.}\right) = 0 \\
&\forall k \in \mathbb{N},\ P\left(|X_n - X| \geqslant \tfrac{1}{k} \text{ i.v.}\right) = 0 \\
&\forall \varepsilon > 0,\ P\left(|X_n - X| \geqslant \varepsilon \text{ i.v.}\right) = 0.
\end{aligned}$$

As equivalências acima são: definição de convergência; negação de um evento ocorrer eventualmente; substituição de ε por $\frac{1}{k}$, que é possível porque a condição é monótona em ε; evento complementar; sub-aditividade da probabilidade; nova substituição de $\frac{1}{k}$ por ε. □

Proposição 5.15 (q.c. $\Rightarrow P$)**.** *Se* $X_n \stackrel{\text{q.c.}}{\to} X$ *então* $X_n \stackrel{P}{\to} X$.

Demonstração. Para qualquer $\varepsilon > 0$, pela Proposição 5.14 temos que

$$P(|X_n - X| \geqslant \varepsilon \text{ i.v.}) = 0,$$

e pela Proposição 5.6 segue que $P(|X_n - X| \geqslant \varepsilon) \to 0$, ou seja, $X_n \stackrel{P}{\to} X$. □

Exercício 5.4. Sejam $(X_n)_n$ variáveis aleatórias tais que

$$\sum_{n=1}^{\infty} P\left(|X_n| > \varepsilon\right) < \infty$$

para qualquer $\varepsilon > 0$. Mostre que

$$X_n \stackrel{\text{q.c.}}{\to} 0.$$

Mostre que também vale a recíproca no caso de as X_n serem independentes.

Contra-Exemplo 5.16 ($P \not\Rightarrow$ q.c.)**.** No Exemplo 5.9, temos pelo Lema de Borel-Cantelli que

$$P(X_n = 1 \text{ infinitas vezes}) = 1,$$

portanto não vale $X_n \stackrel{\text{q.c.}}{\to} 0$.

Contra-Exemplo 5.17 ($P \not\Rightarrow$ q.c.)**.** No Exemplo 5.10, temos que

$$P(\frac{X_n}{\log n} \geqslant \varepsilon \text{ infinitas vezes}) = 1$$

para $\varepsilon \leqslant 1$ e 0 para $\varepsilon > 1$. Portanto não vale que $\frac{X_n}{\log n} \stackrel{\text{q.c.}}{\to} 0$.

Definição 5.18 (Convergência em $\mathcal{L}_p$)**.** Dizemos que X_n *converge para* X *em* $\mathcal{L}_p$, que denotamos por $X_n \stackrel{\mathcal{L}_p}{\to} X$, se

$$\lim_{n\to\infty} E\left(|X_n - X|^p\right) = 0.$$

Quando $p = 2$, a convergência é dita em *média quadrática.*

Proposição 5.19 ($\mathcal{L}_p \Rightarrow P$)**.** *Se* $X_n \stackrel{\mathcal{L}_p}{\to} X$ *para algum* $p \geqslant 1$ *então* $X_n \stackrel{P}{\to} X$.

Demonstração. Pela desigualdade de Markov temos

$$P(|X_n - X| \geqslant \varepsilon) \leqslant \frac{E|X_n - X|^p}{\varepsilon^p} \to 0. \qquad \square$$

Proposição 5.20 ($\mathcal{L}_{p+s} \Rightarrow \mathcal{L}_p$)**.** *Sejam* $p \geqslant 1$ *e* $s \geqslant 0$. *Se* $X_n \stackrel{\mathcal{L}_{p+s}}{\to} X$ *então* $X_n \stackrel{\mathcal{L}_p}{\to} X$.

Demonstração. Fazendo $q = p + s$, pela Desigualdade de Jensen temos

$$\left(E\left|X_n - X\right|^p\right)^{\frac{1}{p}} \leqslant \left(E\left|X_n - X\right|^q\right)^{\frac{1}{q}} \to 0. \qquad \square$$

Contra-Exemplo 5.21 (q.c. $\not\Rightarrow \mathcal{L}_p$)**.** Suponha que $P(X_n = n^3) = \frac{1}{n^2} = 1 - P(X_n = 0)$. Então para $\varepsilon < 1$ temos $P(X \geqslant \varepsilon) = \frac{1}{n^2}$, portanto $X_n \stackrel{\text{q.c.}}{\to} 0$ e $X_n \stackrel{P}{\to} 0$. Entretanto, $EX_n = n$, logo não podemos ter $X_n \stackrel{\mathcal{L}_1}{\to} 0$, e pela proposição acima não podemos ter convergência em $\mathcal{L}_p$ para nenhum $p \geqslant 1$.

Contra-Exemplo 5.22 ($\mathcal{L}_p \not\Rightarrow$ q.c.)**.** No Exemplo 5.9, temos

$$E|X - 0|^p = EX^p = P(X = 1) = \frac{1}{n} \to 0,$$

portanto $X_n \xrightarrow{\mathcal{L}_p} 0$ para todo p e $X_n \xrightarrow{P} 0$. No entanto, não vale $X_n \xrightarrow{\text{q.c.}} 0$.

Definição 5.23 (Convergência em Distribuição)**.** Dizemos que X_n *converge em distribuição para* X, que denotamos por $X_n \xrightarrow{\text{d}} X$, se, para todo ponto t em que F_X é contínua, vale

$$\lim_{n\to\infty} F_{X_n}(t) = F_X(t).$$

Observação 5.24. Para convergência em distribuição não é necessário que as variáveis aleatórias estejam definidas no mesmo espaço de probabilidade, pois essa noção de convergência leva em conta apenas a sua distribuição.

Proposição 5.25 (Unicidade do Limite em Distribuição)**.** *O limite em distribuição é único, isto é, se $X_n \xrightarrow{\text{d}} X$ e $X_n \xrightarrow{\text{d}} Y$ então $X \sim Y$.*

Ideia da prova. Feita em aula. □

Exemplo 5.26. Seja $X_n = \frac{1}{n}$ para $n \geqslant 1$ e $X = 0$. Então $X_n \xrightarrow{\text{d}} X$, embora $\lim_{n\to\infty} F_n(0) = 0 \neq 1 = F(0)$. Mas como 0 não é ponto de continuidade de F, isto não é problema.

Exercício 5.5. Seja $(X_n)_n$ uma sequência de variáveis aleatórias independentes com distribuição uniforme em $(0, b)$, $b > 0$. Defina $Y_n = \max(X_1, X_2, \dots, X_n)$ e $Y = b$. Então verifique que $Y_n \xrightarrow{\text{d}} Y$.

Proposição 5.27 ($P \Rightarrow \text{d}$)**.** *Se $X_n \xrightarrow{P} X$ então $X_n \xrightarrow{\text{d}} X$.*

Demonstração. Seja $t \in \mathbb{R}$ tal que F_X seja contínua em t. Temos que mostrar que $F_{X_n}(t) \to F_X(t)$ quando $n \to \infty$. Dado $\varepsilon > 0$, temos que

$$\begin{aligned} X_n \leqslant t &\implies |X_n - X| \geqslant \varepsilon \text{ ou } X \leqslant t + \varepsilon \\ X \leqslant t - \varepsilon &\implies |X_n - X| \geqslant \varepsilon \text{ ou } X_n \leqslant t, \end{aligned}$$

donde

$$F_X(t-\varepsilon) - P(|X_n - X| \geqslant \varepsilon) \leqslant F_{X_n}(t) \leqslant F_X(t+\varepsilon) + P(|X_n - X| \geqslant \varepsilon).$$

Fazendo $n \to \infty$ temos

$$F_X(t-\varepsilon) \leqslant \liminf_n F_{X_n}(t) \leqslant \limsup_n F_{X_n}(t) \leqslant F_X(t+\varepsilon),$$

e fazendo $\varepsilon \to 0$, como t e ponto de continuidade de F_X, temos

$$\lim_n F_{X_n}(t) = F_X(t),$$

concluindo a demonstração. □

Exercício 5.6. Mostre que, se $X_n \stackrel{\text{q.c.}}{\to} Y$ e $X_n \stackrel{\text{d}}{\to} Z$, então $Y \sim Z$.

A convergência em probabilidade implica a convergência em distribuição, mas a recíproca é falsa (de fato, a recíproca nem faz sentido já que convergência em distribuição nem necessita que as variáveis estejam definidas no mesmo espaço de probabilidade). Além disso, como vimos nos exemplos acima, em geral convergência quase certa não implica convergência em $\mathcal{L}_p$, e vice-versa. Entretanto, sob condições particulares, podemos garantir mais implicações entre as diferentes definições de convergência.

Proposição 5.28. *Se $X_n \stackrel{\text{d}}{\to} c$ para $c \in \mathbb{R}$ constante, então $X_n \stackrel{P}{\to} c$.*

Demonstração. Convergência em distribuição a uma variável constante quer dizer que $\lim_n F_{X_n}(t) = 0$ se $t < c$ e $\lim_n F_{X_n}(t) \to 1$ se $t > c$. Seja $\varepsilon > 0$. Veja que

$$P(|X_n - c| \geqslant \varepsilon) \leqslant P(X_n \leqslant c - \varepsilon) + P(X_n > c + \tfrac{\varepsilon}{2}) =$$
$$= F_{X_n}(c - \varepsilon) + 1 - F_{X_n}(c + \tfrac{\varepsilon}{2}) \to 0 \text{ quando } n \to \infty.$$

Como ε é arbitrário, isso conclui a demonstração. □

Proposição 5.29 (Convergência por Subsequências)**.** *Se $X_n \stackrel{P}{\to} X$ então existe uma subsequência $n_k \to \infty$ tal que $X_{n_k} \stackrel{\text{q.c.}}{\to} X$.*

Demonstração. Como $P(|X_n - X| \geqslant \varepsilon) \to 0$ para todo $\varepsilon > 0$, podemos tomar $n_1 > 0$ tal que $P(|X_{n_1} - X| \geqslant 1) < \frac{1}{2}$. Novamente, podemos tomar $n_2 > n_1$ tal que $P(|X_{n_2} - X| \geqslant \frac{1}{2}) < \frac{1}{4}$. Sucessivamente, podemos tomar $n_k > n_{k-1}$ tal que $P(|X_{n_k} - X| \geqslant \frac{1}{k}) < \frac{1}{2^k}$.

Vamos ver que essa sequência n_k satisfaz $X_{n_k} \stackrel{\text{q.c.}}{\to} X$. Seja $\varepsilon > 0$. Temos que $P(|X_{n_k} - X| \geqslant \varepsilon) \leqslant P(|X_{n_k} - X| \geqslant \frac{1}{k})$ para todo $k \geqslant \varepsilon^{-1}$. Por outro lado temos que $\sum_k P(|X_{n_k} - X| \geqslant \frac{1}{k}) < \infty$, logo $\sum_k P(|X_{n_k} - X| \geqslant \varepsilon) < \infty$. Pelo Exercício 5.4 temos que $X_{n_k} - X \stackrel{\text{q.c.}}{\to} 0$, ou seja, $X_{n_k} \stackrel{\text{q.c.}}{\to} X$. □

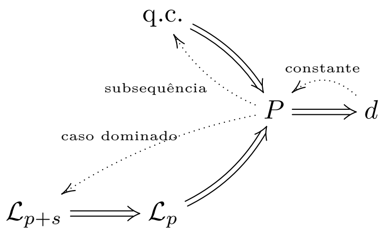

Figura 5.1: Diagrama de implicações entre os tipos de convergência.

Corolário 5.30 (Unicidade do Limite em Probabilidade)**.** *Se* $X_n \xrightarrow{P} X$ *e* $X_n \xrightarrow{P} Y$ *então* $P(X = Y) = 1$*.*

Demonstração. Tome uma subsequência n_k tal que

$$X_{n_k} \xrightarrow[k\to\infty]{\text{q.c.}} X$$

e uma subsequência n_{k_j} tal que

$$X_{n_{k_j}} \xrightarrow[j\to\infty]{\text{q.c.}} Y.$$

Para todo ω na interseção de $A = [X_{n_k} \to X]$ e $B = [X_{n_{k_j}} \to Y]$ temos que $[X = Y]$. Como $P(A) = P(B) = 1$, temos que $P(A \cap B) = 1$, e portanto $P(X = Y) \geqslant P(A \cap B) = 1$. $\square$

Proposição 5.31 (Caso Dominado)**.** *Seja* $p \geqslant 1$*. Se* $X_n \xrightarrow{P} X$ *e existe* Y *tal que* $EY^p < \infty$ *e* $|X_n| \leqslant Y$ *q.c. para todo* n*, então* $X_n \xrightarrow{\mathcal{L}_p} X$*.*

Demonstração. Ver na página 160. $\square$

Completamos assim o diagrama de implicações da Figura 5.1.

5.3 Exercícios

5.7. Sejam $(X_n)_{n\in\mathbb{N}}$ variáveis aleatórias independentes, distribuídas respectivamente como $\exp(\lambda_n)$, onde $\lambda_n = n^3$. Prove que $P\left(\sum_{n=1}^{\infty} X_n < \infty\right) = 1$.

5.8. [Jam04, Capítulo 5]. Recomendados: 5, 6, 7, 9, 10.

5.9. Seja $(A_n)_n$ uma sequência de eventos em $(\mathbb{1}_{A_n})_n$ a sequência de variáveis aleatórias indicadoras das ocorrências dos eventos correspondentes. Encontre uma condição sobre as probabilidades $P(A_n)$ para que $\mathbb{1}_{A_n} \xrightarrow{P} 0$.

5.10. Considere o espaço de probabilidade $([0,1], \mathcal{B}, P)$ com P dado pela medida de comprimento, e a sequência de variáveis aleatórias $(X_n)_n$ dadas por

$$X_n(\omega) = \begin{cases} n, & w < \frac{1}{n}, \\ 0, & w \geqslant \frac{1}{n}. \end{cases}$$

Verifique respectivamente se $X_n \overset{\mathrm{d}}{\to} X$, $X_n \overset{P}{\to} X$, $X_n \overset{\text{q.c.}}{\to} X$, $X_n \overset{\mathcal{L}_2}{\to} X$, $X_n \overset{\mathcal{L}_1}{\to} X$, para alguma variável aleatória X.

5.11. Seja $(X_n)_n$ uma sequência de variáveis aleatórias independentes com distribuição uniforme em $[0,1]$, e $Y_n = \max\{X_1, \dots, X_n\}$. Encontre a função de distribuição de Y_n e o limite em distribuição desta sequência.

5.12. Sejam X_n, $n \in \mathbb{N}$, variáveis aleatórias independentes tais que $X_n \sim \text{Bernoulli}(p_n)$. Estude as condições sobre (p_n) para que:

1. $X_n \overset{P}{\to} 0$.
2. $X_n \overset{\text{q.c.}}{\to} 0$.

5.13. Seja $(X_n)_n$ uma sequência i.i.d. Mostre que

$$\frac{X_n}{n} \overset{\text{q.c.}}{\to} 0$$

se e somente se $E|X_1| < \infty$.

5.14. Seja $(X_n)_n$ uma sequência i.i.d. Mostre que

$$\frac{X_n}{\sqrt{n}} \overset{\text{q.c.}}{\to} 0$$

se e somente se $E|X_1|^2 < \infty$.

5.15. Seja $(X_n)_n$ uma sequência i.i.d. com distribuição $\exp(1)$. Mostre que

$$P\left(X_n \geqslant 2\log n \text{ i.v.}\right) = 0.$$

5.16. Seja $(X_n)_n$ uma sequência i.i.d. com distribuição Poisson(λ). Mostre que

$$\frac{X_n}{\log n} \stackrel{\text{q.c.}}{\to} 0.$$

Sugestão: mostre antes que $Ee^{X_1/\varepsilon} < \infty$.

5.17. Seja $(X_n)_n$ uma sequência i.i.d. de variáveis aleatórias não-negativas com $EX_1^2 < \infty$. Mostre que

$$P\left\{\sum_{n=1}^{\infty} \frac{X_n}{n^2} < \infty\right\} = 1$$

5.18. [Jam04, Capítulo 6]. Recomendados: 15, 19.

Capítulo 6

Lei dos Grandes Números

A Lei dos Grandes Números é a primeira aproximação para a soma de muitas variáveis aleatórias. Como discutido no início Capítulo 4, a média de uma grande quantidade de variáveis i.i.d. deveria convergir para um valor determinístico, dado pela esperança dessas variáveis aleatórias. Neste capítulo estudaremos essa convergência e algumas de suas implicações.

6.1 Lei Fraca

Sejam $X_1, X_2, \dots$ variáveis aleatórias integráveis em $(\Omega, \mathcal{F}, P)$ e $S_1, S_2, \dots$ suas somas parciais dadas por

$$S_n = X_1 + X_2 + \cdots + X_n.$$

Definição 6.1 (Lei Fraca dos Grandes Números)**.** Dizemos que a sequência $(X_1, X_2, \dots)$ satisfaz a *Lei Fraca dos Grandes Números* se, para todo $\varepsilon > 0$, vale

$$P\left(\left|\frac{S_n - ES_n}{n}\right| \geqslant \varepsilon\right) \to 0, \text{ quando } n \to \infty,$$

ou seja, se

$$\frac{S_n - ES_n}{n} \xrightarrow{P} 0.$$

Teorema 6.2 (Lei dos Grandes Números de Bernoulli, 1713)**.** *Considere uma sequência de ensaios binomiais independentes tendo a mesma probabilidade p de sucesso em cada ensaio. Se S_n é o número de sucessos nos primeiros n ensaios, então*

$$\frac{S_n}{n} \xrightarrow{P} p.$$

Não veremos a demonstração original de Bernoulli. A Lei dos Grandes Números de Tchebyshev provada logo abaixo é mais geral.

A Lei dos Grandes Números de Bernoulli tem uma importância histórica inestimável. De certa forma, esse teorema justifica o conceito de probabilidade como sendo a frequência relativa de ocorrência de um evento, isto é,

$$p \approx \frac{\text{quantidade de experimentos em que o evento e observado}}{\text{quantidade total de experimentos realizados}},$$

onde a ideia de aproximação passa a ter um significado mais preciso, o da convergência em probabilidade. O ano de 2013 foi considerado o Ano Internacional da Estatística em comemoração dos 300 anos do teorema de Bernoulli.

Teorema 6.3 (Lei dos Grandes Números de Tchebyshev, 1867)**.** *Sejam $X_1, X_2, \dots$ variáveis aleatórias duas a duas não-correlacionadas e com variâncias finitas e uniformemente limitadas, isto é, existe M finito, tal que $VX_n < M$ para todo n. Então $(X_1, X_2, \dots)$ satisfaz a Lei Fraca dos Grandes Números:*

$$\frac{S_n - ES_n}{n} \xrightarrow{P} 0.$$

Demonstração. Pela Desigualdade Clássica de Tchebyshev, temos

$$P\left(\left|\frac{S_n - ES_n}{n}\right| \geqslant \varepsilon\right) \leqslant \frac{V(\frac{S_n}{n})}{\varepsilon^2} = \frac{VS_n}{\varepsilon^2 n^2} = \frac{\sum_{i=1}^n VX_i}{\varepsilon^2 n^2} \leqslant \frac{n \cdot M}{\varepsilon^2 n^2} \to 0. \qquad \square$$

Teorema 6.4 (Lei dos Grandes Números de Khintchine, 1929)**.** *Sejam $X_1, X_2, \dots$ variáveis aleatórias independentes, identicamente distribuídas e integráveis, com média μ. Então $(X_1, X_2, \dots)$ satisfaz a Lei Fraca dos Grandes Números:*

$$\frac{S_n}{n} \xrightarrow{P} \mu.$$

A demonstração original de Khintchine foi feita usando o método de truncamento, aparentemente introduzido por Markov, e utilizado em seguida por Kolmogorov na prova da Lei Forte dos Grandes Números. Uma prova alternativa usando funções características será dada no Capítulo 8, página 135.

6.2 Lei Forte

Definição 6.5 (Lei Forte dos Grandes Números)**.** Dizemos que $(X_1, X_2, \dots)$ satisfaz a *Lei Forte dos Grandes Números* se

$$P\left(\lim_{n\to\infty} \frac{S_n - ES_n}{n} = 0\right) = 1,$$

ou seja, se

$$\frac{S_n - ES_n}{n} \stackrel{\text{q.c.}}{\to} 0.$$

Teorema 6.6 (Lei dos Grandes Números de Borel, 1909)**.** *Considere uma sequência de ensaios binomiais independentes tendo a mesma probabilidade p de sucesso em cada ensaio. Se S_n é o número de sucessos nos primeiros n ensaios, então*

$$\frac{S_n}{n} \stackrel{\text{q.c.}}{\to} p.$$

Não veremos a demonstração original de Borel. A Lei dos Grandes Números de Cantelli é mais geral.

Teorema 6.7 (Lei dos Grandes Números de Cantelli, 1917)**.** *Sejam $X_1, X_2, \dots$ variáveis aleatórias independentes e identicamente distribuídas, com quarto momento finito e média μ. Então $(X_1, X_2, \dots)$ satisfaz a Lei Forte dos Grandes Números:*

$$\frac{S_n}{n} \stackrel{\text{q.c.}}{\to} \mu.$$

Demonstração. Podemos supor que $\mu = 0$, ou tomar $\tilde{X}_n = X_n - \mu$. Observe que

$$\begin{aligned} S_n^4 = (X_1 + \cdots + X_n)^4 = \sum_{i,j,k,l} X_i X_j X_k X_l = \sum_i X_i^4 + \frac{4!}{2!2!} \sum_{i<j} X_i^2 X_j^2 + \\ + \frac{4!}{3!} \sum_{i \neq k} X_i^3 X_k + \frac{4!}{2!} \sum_{\substack{j<k \\ i \neq j,k}} X_i^2 X_j X_k + 4! \sum_{i<j<k<l} X_i X_j X_k X_l. \end{aligned}$$

Por independência, temos que

$$
\begin{aligned}
ES_n^4 = \sum_i EX_i^4 + 6 \sum_{i<j} E[X_i^2 X_j^2] + \\
+ \sum_k E\left[4 \sum X_i^3 + 12 \sum X_i^2 X_j + 24 \sum X_i X_j X_l\right] EX_k.
\end{aligned}
$$

Como assumimos que $EX_k = 0$, a segunda linha é igual a zero. Além disso, como as X_i têm a mesma distribuição, obtemos

$$
\begin{aligned}
ES_n^4 &= nEX_1^4 + 6\tbinom{n}{2} E(X_1^2 X_2^2) \\
&= nEX_1^4 + 3(n^2 - n) E(X_1^2 X_2^2) \\
&\leqslant nEX_1^4 + 3(n^2 - n)\sqrt{EX_1^4}\,\sqrt{EX_2^4} \\
&= (3n^2 - 2n) EX_1^4 \\
&\leqslant 3n^2 EX_1^4.
\end{aligned}
$$

Pela Desigualdade de Markov, temos

$$
P\left(\left|\frac{S_n}{n}\right| \geqslant \varepsilon\right) \leqslant \frac{ES_n^4}{\varepsilon^4 n^4} \leqslant \frac{3EX_1^4}{\varepsilon^4 n^2},
$$

e pelo Lema de Borel-Cantelli segue que $\frac{S_n}{n} \overset{\text{q.c.}}{\to} 0$. □

Teorema 6.8 (Lei dos Grandes Números de Kolmogorov, 1933)**.** *Sejam $X_1, X_2, \dots$ variáveis aleatórias independentes, identicamente distribuídas e integráveis, com $EX_n = \mu$. Então $(X_1, X_2, \dots)$ satisfaz a Lei Forte dos Grandes Números:*

$$
\frac{S_n}{n} \overset{\text{q.c.}}{\to} \mu.
$$

Demonstração. O leitor interessado pode consultar [Jam04, pp. 204–214]. □

6.3 Exercícios

Observação. As questões sobre a Lei Forte dos Grandes Números, por tratarem de eventos que devem acontecer com probabilidade 1, em geral envolvem o uso do

Lema de Borel-Cantelli.

6.1. Seja $(X_n)_n$ uma sequência de variáveis aleatórias independentes com funções de probabilidade p_n dadas por $p_n(n^2) = \frac{1}{n^3} = 1 - p_n(0)$. Essa sequência satisfaz a Lei dos Grandes Números?

6.2. Seja $(X_n)_n$ uma sequência de variáveis aleatórias independentes com funções de probabilidade p_n dadas por $p_n(n^2) = \frac{1}{n^2} = 1 - p_n(0)$. Essa sequência satisfaz a Lei dos Grandes Números?

6.3. [Jam04, Capítulo 5]. Recomendados: 2, 3, 14.

Capítulo 7

Teorema Central do Limite

Seja $(X_n)_n$ uma sequência i.i.d. de variáveis aleatórias. Pela Lei dos Grandes Números sabemos que a média amostral $\frac{S_n}{n}$ se aproxima do valor esperado μ para valores grandes de n, isto é,

$$\frac{S_n}{n} \approx \mu.$$

Porém, não é razoável esperar que $\frac{S_n}{n}$ seja exatamente igual a μ. Então a primeira pergunta que surge é sobre a flutuação $\frac{S_n}{n} - \mu$ da média amostral em torno do seu valor esperado. Tipicamente, essa diferença ocorre em qual escala? Nessa escala, qual é seu comportamento estatístico?

Não é difícil adivinhar a escala em que ocorre essa flutuação. De fato, sabemos que $ES_n = nEX_1 = n\mu$ e $VS_n = nVX_1 = n\sigma^2$, ou seja, $\sigma(S_n) = \sigma\sqrt{n}$. Assim temos que a esperança da média amostral é μ e seu desvio-padrão é $\frac{\sigma}{\sqrt{n}}$. Isso é uma indicação de que tipicamente as flutuações assumem valores da ordem $\frac{\sigma}{\sqrt{n}}$ (de fato, pela desigualdade de Tchebyshev, as flutuações não podem ser muito maiores do que o desvio-padrão, porém o valor numérico da variância poderia ser resultado de uma flutuação atipicamente grande, enquanto os valores típicos fossem na verdade muito menores). Vamos supor que esse argumento está correto para tentar entender qual poderia ser o comportamento estatístico das flutuações nessa escala.

Escrevemos $\frac{S_n}{n} = \mu + \frac{\sigma}{\sqrt{n}} Y_n$, onde Y_n satisfaz $EY_n = 0$ e $VY_n = 1$. Será que o comportamento estatístico de Y_n se aproxima de alguma distribuição Y que não depende de n? Suponhamos que sim, e busquemos um candidato para essa distribuição. Observamos que $S_{2n} = S_n + \tilde{S}_n$, onde separamos os $2n$ termos da soma em dois blocos independentes com tamanho n. Assim obtemos a relação $\frac{S_{2n}}{n} = 2\mu + \frac{\sigma}{\sqrt{n}}(Y_n + \tilde{Y}_n)$, onde $\tilde{Y}_n$ é independente e com a mesma distribuição de Y_n. Por outro lado, $\frac{S_{2n}}{2n} = \mu + \frac{\sigma}{\sqrt{2n}} Y_{2n}$, donde chegamos finalmente a $Y_n + \tilde{Y}_n = \sqrt{2} \cdot Y_{2n}$, ou seja, $Y + \tilde{Y} \sim \sqrt{2} \cdot Y$. A única distribuição que satisfaz essa relação é a distribuição normal.

Esses argumentos *ad hoc* são confirmados pelo Teorema Central do Limite:

$$\frac{S_n - ES_n}{\sqrt{VS_n}} \xrightarrow{\mathrm{d}} \mathcal{N}(0,1).$$

Reescrevendo, temos

$$\frac{S_n}{n} \approx \mu + \frac{\sigma}{\sqrt{n}} \mathcal{N}(0,1).$$

Ou seja, a escala em que a média amostral $\frac{S_n}{n}$ flutua em torno de seu valor esperado μ é de fato dada por $\frac{\sigma}{\sqrt{n}}$. Ademais, seu comportamento nessa escala possui forte regularidade estatística, e sua distribuição se aproxima de uma normal padrão. Dito de outra forma, a distribuição da soma parcial S_n pode ser aproximada por uma normal com mesma média e variância de S_n:

$$S_n \approx \mathcal{N}(n\mu, n\sigma^2).$$

7.1 Teorema de De Moivre-Laplace

O exemplo mais simples da aproximação $S_n \approx \mathcal{N}(n\mu, n\sigma^2)$ é quando lançamos uma moeda honesta n vezes e contamos o número S_n de caras. Neste caso S_n tem distribuição $b(n, \frac{1}{2})$. Na Figura 7.1 vemos como essa distribuição, devidamente normalizada, se aproxima da distribuição normal padrão.

Teorema 7.1 (Teorema de De Moivre-Laplace, 1730, 1812)**.** *Seja* $(X_n)_{n\in\mathbb{N}}$ *uma sequência de variáveis aleatórias independentes, com distribuição* Bernoulli(p), *onde* $p = 1 - q \in (0,1)$, *e tome* $S_n = X_1 + \cdots + X_n$. *Então para todos* $a < b$

$$P\left(a < \frac{S_n - np}{\sqrt{npq}} \leqslant b\right) \to \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}}\,\mathrm{d}x.$$

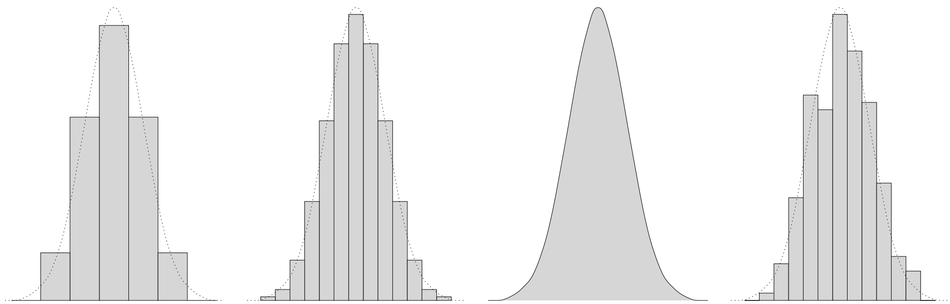

Figura 7.1: Função de probabilidade de $\frac{S_n - n\mu}{\sigma\sqrt{n}}$ para S_n com distribuições $b(4, \frac{1}{2})$ e $b(16, \frac{1}{2})$ para valores entre -3 e 3. A área de cada retângulo é dada pela função de probabilidade. O terceiro gráfico é a função de densidade de uma normal padrão, assim como as linhas pontilhadas. O quarto gráfico representa as frequências relativas de $\frac{S_n - n\mu}{\sigma\sqrt{n}}$ para S_n com distribuição $b(16, \frac{1}{2})$, em um experimento real com 200 amostras.

Ou seja,

$$S_n \approx \mathcal{N}(np, npq).$$

O teorema foi provado por De Moivre supondo que $p = \frac{1}{2}$ e por Laplace para $0 < p < 1$. De fato, ele segue de uma aproximação muito mais fina:

$$\frac{n!}{k!\,(n-k)!}\; p^k \; q^{n-k} \asymp \frac{1}{\sqrt{2\pi npq}} e^{-\frac{x_k^2}{2}}, \qquad (*)$$

onde

$$x_k = x_{n,k} = \frac{k - np}{\sqrt{npq}}$$

e $\asymp$ significa que a razão entre ambos os termos tende a 1 quando n tende a infinito. O limite em $(*)$ é uniforme se restrito a $|x_k| < M$ com qualquer $M < \infty$ fixo.

Essa aproximação é muito mais fina porque diz não apenas que a probabilidade de a flutuação estar dentro de um certo intervalo é bem aproximada pela normal, mas também que a função de probabilidade de cada um dos possíveis valores dentro de um intervalo fixo é aproximado pela densidade da normal.

Para entender de onde vem essa aproximação, primeiro precisamos de uma expressão mais palpável para $n!$. Qual a probabilidade de obtermos exatamente 60 caras se lançamos uma moeda 120 vezes? A resposta é fácil:

$$P(S_{120} = 60) = \frac{120!}{60!\,60!} \times \frac{1}{2^{120}}.$$

Essa expressão é simples e matematicamente perfeita. Porém, quanto vale essa probabilidade? Mais de 15%? Menos de 5%? Entre 5% e 10%? Uma calculadora de bolso trava ao calcular 120!. Num computador esse cálculo resulta em $7,2684979\%$. Mas e se fossem 40.000 lançamentos da moeda? E se estivéssemos interessados em calcular $P(S_{40.000} \leqslant 19.750)$, faríamos um cálculo semelhante 250 vezes para depois somar? As expressões com fatorial são perfeitas para a combinatória, mas impraticáveis para se fazer estimativas. Nosso socorro será a fórmula de Stirling

$$n! \asymp n^n\, e^{-n}\, \sqrt{2\pi n},$$

provada no Apêndice A. Essa aproximação de $n!$ pela fórmula de Stirling é muito boa mesmo sem tomar n grande. Ela aproxima 1! por $0,92$, 2! por $1,92$, 4! por $23,5$, e a partir de $9! = 362.880$, que é aproximado por 359.537, o erro é menor que 1%. À primeira vista $n^n\, e^{-n}\, \sqrt{2\pi n}$ não parece mais agradável do que $n!$. Mas vejamos como isso ajuda com o cálculo anterior. Temos:

$$\frac{(2k)!}{k!\,k!} \times \frac{1}{2^{2k}} \asymp \frac{(2k)^{2k} e^{-2k}\sqrt{4\pi k}}{(k^k e^{-k}\sqrt{2\pi k})^2} \times \frac{1}{2^{2k}} = \frac{1}{\sqrt{\pi k}}\bigg]_{k=60} = 0,0728\ldots,$$

que pode ser feito até mesmo sem calculadora. Mais do que isso, acabamos de obter a aproximação $(*)$ no caso particular em que $p = q = \frac{1}{2}$ e $x_k = 0$.

Vamos que mostrar $(*)$ para $|x_k| < M$ onde M está fixo. Aplicando a fórmula de Stirling obtemos

$$\frac{n!}{k!\,(n-k)!}\, p^k\, q^{n-k} \asymp \frac{n^n e^{-n}\sqrt{2\pi n}\, p^k\, q^{n-k}}{k^k e^{-k}\sqrt{2\pi k}(n-k)^{n-k} e^{k-n}\sqrt{2\pi(n-k)}} = \frac{(\frac{np}{k})^k (\frac{nq}{n-k})^{n-k}}{\sqrt{2\pi k(n-k)/n}}.$$

Observe que para $|x_k| < M$ vale

$$k = np + \sqrt{npq}\, x_k \asymp np \qquad \text{e} \qquad n-k = nq - \sqrt{npq}\, x_k \asymp nq,$$

donde obtemos

$$\frac{n!}{k!\,(n-k)!}\; p^k\; q^{n-k} \asymp \frac{\left(\frac{np}{k}\right)^k \left(\frac{nq}{n-k}\right)^{n-k}}{\sqrt{2\pi npq}} = \frac{f(n,k)}{\sqrt{2\pi npq}}.$$

Vamos estudar $\log f(n,k)$. Reescrevendo cada termo temos

$$\frac{np}{k} = 1 - \frac{\sqrt{npq}\,x_k}{k} \qquad \text{e} \qquad \frac{nq}{n-k} = 1 + \frac{\sqrt{npq}\,x_k}{n-k}.$$

Fazendo a expansão de Taylor de $\log(1+x)$ temos

$$\log(1+x) = x - \frac{x^2}{2} + r(x), \qquad \text{onde } \frac{r(x)}{x^2} \to 0 \text{ quando } x \to 0.$$

Assim,

$$\begin{aligned}\log f(n,k) = k\left(-\frac{\sqrt{npq}\,x_k}{k} - \frac{npqx_k^2}{2k^2} + r(\tfrac{\sqrt{npq}\,x_k}{k})\right) + \\ + (n-k)\left(\frac{\sqrt{npq}\,x_k}{n-k} - \frac{npqx_k^2}{2(n-k)^2} + r(\tfrac{\sqrt{npq}\,x_k}{n-k})\right).\end{aligned}$$

Note que os primeiros termos se cancelam e, quando $n \to \infty$,

$$\log f(n,k) \asymp -\frac{npqx_k^2}{2k} - \frac{npqx_k^2}{2(n-k)} = -\frac{n^2pqx_k^2}{2k(n-k)} \asymp -\frac{n^2pqx_k^2}{2npnq} = -\frac{x_k^2}{2},$$

donde segue que

$$f(n,k) \asymp e^{-\frac{x_k^2}{2}}$$

uniformemente em $|x_k| < M$, o que termina a prova de $(*)$.

Somando sobre os possíveis valores de S_n temos

$$P\left(a < \tfrac{S_n - np}{\sqrt{npq}} \leqslant b\right) = \sum_{a<x_k\leqslant b} P(S_n = k) = \sum_{a<x_k\leqslant b} \frac{n!}{k!\,(n-k)!}\, p^k\, q^{n-k},$$

onde os somatórios são sobre k com a condição sobre x_k, que é dado por $x_k = \frac{k-np}{\sqrt{npq}}$.

Observando que

$$x_{k+1} - x_k = \frac{1}{\sqrt{npq}},$$

e substituindo $(*)$, obtemos

$$P\left(a < \tfrac{S_n - np}{\sqrt{npq}} \leqslant b\right) \asymp \sum_{a < x_k \leqslant b} \frac{e^{-\frac{x_k^2}{2}}}{\sqrt{2\pi npq}} = \frac{1}{\sqrt{2\pi}} \sum_{a < x_k \leqslant b} e^{-\frac{x_k^2}{2}} \cdot [x_{k+1} - x_k].$$

Finalmente, observamos que o somatório acima é uma soma de Riemann que se aproxima da integral $\frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}}\,\mathrm{d}x$. Isso termina a prova do Teorema 7.1.

7.2 Teorema Central do Limite

Teorema 7.2 (Teorema Central do Limite para Variáveis Aleatórias I.I.D.)**.** *Seja $(X_n)_{n\in\mathbb{N}}$ uma sequência de variáveis aleatórias i.i.d., com média μ e variância σ^2, onde $0 < \sigma^2 < \infty$, e tome $S_n = X_1 + X_2 + \cdots + X_n$. Então*

$$\frac{S_n - ES_n}{\sqrt{VS_n}} \xrightarrow{\mathrm{d}} \mathcal{N}(0,1),$$

isto é,

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{\mathrm{d}} \mathcal{N}(0,1).$$

A demonstração será vista no Capítulo 8, como aplicação do Teorema da Continuidade de Lévy para funções características.

Teorema 7.3 (Teorema Central do Limite de Lyapunov)**.** *Sejam $(X_n)_n$ independentes com terceiros momentos finitos satisfazendo*

$$\frac{\sum_{i=1}^n E|X_i - EX_i|^3}{\left(\sum_{i=1}^n VX_i\right)^{3/2}} \to 0 \text{ as } n \to \infty.$$

Então,

$$Ef\left(\frac{S_n - ES_n}{\sqrt{VS_n}}\right) \to \int_{\mathbb{R}} f(x)\frac{e^{-x^2/2}}{\sqrt{2\pi}}\mathrm{d}x \text{ quando } n \to \infty$$

para qualquer f tal que f, f', f'' e f''' sejam limitadas.

Convergência para toda f satisfazendo as condições acima é equivalente à convergência para todas f contínua e limitada, o que por sua vez equivale a convergência em distribuição. A prova desses fatos envolve Teoria da Medida e será omitida.

Demonstração. Podemos assumir $EX_i = 0$ sem perda de generalidade. Escreva $\sigma_i = \sqrt{VX_i}$. Considere $(Y_n)_n$ independentes, e também independentes de $(X_n)_n$, com a distribuição $\mathcal{N}(0, \sigma_i^2)$. (Aceitaremos sem prova existência de tal seqüência no mesmo espaço de probabilidade.)

Escreva

$$W_n = \frac{Y_1 + \cdots + Y_n}{\sqrt{VS_n}} \sim \mathcal{N}(0,1) \quad \text{e} \quad Z_n = \frac{X_1 + \cdots + X_n}{\sqrt{VS_n}}.$$

Fixe n, defina $Z^0 = W_n$ e, para $i = 1, \ldots, n$, defina

$$V^i = Z^{i-1} - \frac{Y_i}{\sqrt{VS_n}} \quad \text{e} \quad Z^i = V^i + \frac{X_i}{\sqrt{VS_n}},$$

de modo que $Z^n = Z_n$. Agora,

$$Ef(Z_n) - Ef(W_n) = Ef(Z^n) - Ef(Z^0) = \sum_{i=1}^{n} Ef(Z^i) - Ef(Z^{i-1}).$$

Usando a expansão de Taylor de f,

$$f(Z^i) = f(V_i + \frac{X_i}{\sqrt{VS_n}}) = f(V^i) + f'(V^i)\frac{X_i}{\sqrt{VS_n}} + \frac{f''(V_i)}{2}\frac{X_i^2}{VS_n} + \frac{f'''(\theta^i)}{6}\frac{X_i^3}{(VS_n)^{3/2}},$$

$$f(Z^{i-1}) = f(V_i + \frac{Y_i}{\sqrt{VS_n}}) = f(V^i) + f'(V^i)\frac{Y_i}{\sqrt{VS_n}} + \frac{f''(V_i)}{2}\frac{Y_i^2}{VS_n} + \frac{f'''(\tilde{\theta}^i)}{6}\frac{Y_i^3}{(VS_n)^{3/2}},$$

onde θ_i e $\tilde{\theta}_i$ vêm do polinômio de Taylor e dependem de V_i, X_i e Y_i.

Observe que $EX_i = EY_i$, $EX_i^2 = EY_i^2$, $X_i \perp\!\!\!\perp V^i$ e $Y_i \perp\!\!\!\perp V^i$, e recordemos que f e suas três primeiras derivadas são limitadas. Queremos tomar a esperança acima e subtrair. Como f é limitada, $f(V^i)$ é integrável e este primeiro termo se cancela. Além disso, $f'(V^i)$ é integrável e, usando a independência,

$$E[f'(V^i)X_i] = Ef'(V^i) \cdot EX_i = Ef'(V^i) \cdot EY_i = E[f'(V^i)Y_i],$$

de forma que este termo também se cancela, bem como o terceiro termo pelo mesmo motivo. Portanto,

$$Ef(Z^i) - Ef(Z^{i-1}) = \frac{E[f'''(\theta_i)X_i^3] - E[f'''(\tilde{\theta}_i)Y_i^3]}{6(VS_n)^{3/2}}.$$

Tomando $C = \sup_x |f'''(x)|$ temos

$$\begin{aligned}\left|Ef(Z^i) - Ef(Z^{i-1})\right| &\leqslant \frac{CE|X_i|^3 + CE|Y_i|^3}{6(VS_n)^{3/2}} \\ &= C\frac{E|X_i|^3 + \sigma_i^3 E|\mathcal{N}|^3}{6(VS_n)^{3/2}} \\ &\leqslant C\frac{E|X_i|^3}{(VS_n)^{3/2}},\end{aligned}$$

onde na última desigualdade usamos que $E|\mathcal{N}|^3 = \frac{4}{\sqrt{2\pi}} < 2$ e que

$$\sigma_i = (E|X_i|^2)^{\frac{1}{2}} \leqslant (E|X_i|^3)^{\frac{1}{3}}$$

pela desigualdade de Jensen. Finalmente, somando sobre i,

$$\left|Ef(Z^i) - Ef(Z^{i-1})\right| \leqslant \sum_{i=1}^{n} \left|Ef(Z^i) - Ef(Z^{i-1})\right| \leqslant C\frac{\sum_{i=1}^n E|X_i|^3}{\left(\sum_{i=1}^n VX_i\right)^{3/2}} \to 0$$

quando $n \to \infty$ por hipótese, o que conclui a prova do teorema. □

7.3 Exercícios

7.1. Um par de dados honestos é lançado 180 vezes por hora.

1. Qual a probabilidade aproximada de que 25 ou mais lançamentos tenham tido soma 7 na primeira hora?
2. Qual a probabilidade aproximada de que entre 700 e 750 lançamentos tenham tido soma 7 durante 24 horas?

7.2. Imagine um modelo idealizado com M eleitores, dos quais M_A pretendem votar no candidato A. Suponha que seja possível sortear um desses eleitores ao acaso, e de forma equiprovável. Definimos

$$X = \begin{cases} 1, & \text{caso o eleitor sorteado vá votar no candidato } A, \\ 0, & \text{caso contrário.} \end{cases}$$

Deseja-se estimar a proporção $p = \frac{M_A}{M}$ de eleitores do candidato A, que é desconhecida. Para isso, repete-se este processo N vezes, obtendo-se $X_1, \dots, X_N$. Para estimar o valor de p considera-se

$$\widehat{p}_N = \frac{X_1 + \cdots + X_N}{N}.$$

Supomos *a priori* que p é bem próximo de $\frac{1}{2}$, de forma que $VX \approx \frac{1}{4}$. Se entrevistamos $N = 2500$ eleitores, calcule aproximadamente a probabilidade de essa pesquisa cometer um erro $|\widehat{p}_N - p|$ maior que $0,01$.

7.3. A quantidade de uvas-passas encontradas em cada panetone de uma determinada marca é independente dos demais panetones e segue a distribuição de Poisson com parâmetro $\lambda = 25$ (ou seja, têm esperança igual à variância, igual a λ). Um grupo de estudantes de férias resolve estimar o valor de λ, uma vez que o mesmo é desconhecido para eles. Para isso, vão contar as uvas-passas de uma amostra de $N = 625$ panetones e registrar o resultado de cada contagem $X_1, \dots, X_N$. A estimativa $\widehat{\lambda}_N$ para o valor de λ que os estudantes vão adotar será dada por

$$\widehat{\lambda}_N = \frac{X_1 + \cdots + X_N}{N}.$$

a) Qual é o valor aproximado da probabilidade de que o valor $\widehat{\lambda}_N$ esteja entre $24,8$ e $25,4$?

b) Para que o erro $\left|\widehat{\lambda}_N - \lambda\right|$ fosse menor que $0,075$ com probabilidade pelo menos igual a $86,64\%$, qual deveria ser o número N de panetones examinados? (Sugestão: resolve-se esse item como o anterior, porém de trás para frente.)

7.4. Use o Teorema Central do Limite para verificar que

$$\lim_{n\to\infty} e^{-n} \sum_{k=0}^{n} \frac{2n^k}{n!} = 1.$$

7.5. Se lançamos 10.000 vezes uma moeda honesta, calcule aproximadamente a probabilidade de que o número de vezes que se obtém coroa seja no mínimo 4.893 e no máximo 4.967.

7.6. [Jam04, Capítulo 7]. Recomendados: 2 e 9.

Capítulo 8

Funções Geradoras

A função geradora de momentos e a função característica estão entre os exemplos mais importantes de *transformadas*. A ideia geral de transformada é mapear certos objetos em objetos de outro tipo e outras propriedades, onde determinadas análises são possivelmente mais fáceis. Isso ficará claro nos exemplos e aplicações. A função geradora de momentos é a instância da *Transformada de Laplace* de uma distribuição em $\mathbb{R}$, e a função característica é a *Transformada de Fourier*.

8.1 Função Geradora de Momentos

Definição 8.1. Seja X uma variável aleatória. Define-se a *função geradora de momentos de* X como a função $M_X : \mathbb{R} \to [0, \infty]$ dada por

$$M_X(t) = E[e^{tX}].$$

Pelo Teorema 4.27, podemos calcular a função geradora de momentos por

$$M_X(t) = \sum_x e^{tx} \cdot P(X = x) \qquad \text{se } X \text{ é discreta}$$

e

$$M_X(t) = \int_{\mathbb{R}} e^{tx} f_X(x)\,\mathrm{d}x \qquad \text{se } X \text{ é absolutamente contínua.}$$

Exemplo 8.2 (Bernoulli)**.** Se $X \sim \text{Bernoulli}(p)$, então

$$M_X(t) = pe^t + 1 - p = 1 + p(e^t - 1).$$

Exemplo 8.3 (Binomial)**.** Se $X \sim b(n, p)$, então

$$\begin{aligned} M_X(t) &= \sum_{k=0}^{n} e^{tk} \tbinom{n}{k} p^k (1-p)^{n-k} = \sum_{k=0}^{n} \tbinom{n}{k} (e^t p)^k (1-p)^{n-k} \\ &= [e^t p + (1-p)]^n = [1 + p(e^t - 1)]^n \end{aligned}$$

Exemplo 8.4 (Geométrica)**.** Se $X \sim \text{Geom}(p)$, então

$$\begin{aligned} M_X(t) &= \sum_{n=1}^{\infty} e^{tn} p(1-p)^{n-1} = e^t p \sum_{m=0}^{\infty} [(e^t)(1-p)]^m \\ &= \frac{pe^t}{1-(1-p)e^t} = \frac{p}{e^{-t} + p - 1}, \qquad \text{para } t < \log \tfrac{1}{1-p}. \end{aligned}$$

Proposição 8.5. *Se X tem função geradora de momentos $M_X(t)$ e $Y = aX + b$, então $M_Y(t) = e^{bt} M_X(at)$.*

Demonstração. Feita em aula. □

Exemplo 8.6 (Poisson)**.** Se $X \sim \text{Poisson}(\lambda)$, então

$$M_X(t) = \sum_{n=0}^{\infty} e^{tn} \frac{e^{-\lambda} \lambda^n}{n!} = e^{-\lambda} \sum_{n=0}^{\infty} \frac{(\lambda e^t)^n}{n!} = e^{-\lambda} e^{\lambda e^t} = e^{\lambda(e^t - 1)}.$$

Proposição 8.7. *Se M_X é finita em algum intervalo $(-\varepsilon, +\varepsilon)$, então os momentos de X podem ser obtidos através das derivadas de M_X por*

$$EX^k = M_X^{(k)}(0), \quad k = 1, 2, 3, \ldots.$$

Demonstração. Para lembrar da fórmula é interessante entender a ideia da prova:

$$\tfrac{\mathrm{d}^k}{\mathrm{d}t^k} M_X(t) = \tfrac{\mathrm{d}^k}{\mathrm{d}t^k} E[e^{tX}] = E\left[\tfrac{\mathrm{d}^k}{\mathrm{d}t^k} e^{tX} \right] = E[X^k e^{tX}].$$

No caso de X ser uma variável aleatória simples, essa é a demonstração, pois a esperança é uma soma finita e podemos derivar dentro da soma. No caso geral, há que se justificar a derivada dentro da esperança. Faremos isso no Capítulo 10, página 160. □

Exemplo 8.8 (Bernoulli)**.** Se $X \sim \text{Bernoulli}(p)$, então

$$\begin{aligned} EX &= M_X'(0) = p, \\ EX^2 &= M_X''(0) = p, \\ VX &= EX^2 - (EX)^2 = p(1-p). \end{aligned}$$

Exemplo 8.9 (Binomial)**.** Se $X \sim b(n,p)$, então

$$\begin{aligned} EX &= M_X'(0) = np, \\ EX^2 &= M_X''(0) = np(1-p) - n^2p^2, \\ VX &= EX^2 - (EX)^2 = np(1-p). \end{aligned}$$

Exemplo 8.10 (Geométrica)**.** Se $X \sim \text{Geom}(p)$, então

$$\begin{aligned} EX &= M_X'(0) = \frac{1}{p}, \\ EX^2 &= M_X''(0) = \frac{2}{p^2} - \frac{1}{p}, \\ VX &= EX^2 - (EX)^2 = \frac{1-p}{p^2}. \end{aligned}$$

Exemplo 8.11 (Poisson)**.** Se $X \sim \text{Poisson}(\lambda)$, então

$$\begin{aligned} EX &= M_X'(0) = \lambda, \\ EX^2 &= M_X''(0) = \lambda^2 + \lambda, \\ VX &= EX^2 - (EX)^2 = \lambda. \end{aligned}$$

Proposição 8.12 (Unicidade)**.** *A função geradora de momentos define de forma unívoca a distribuição da variável aleatória. Mais precisamente, se $M_X = M_Y < \infty$ em algum intervalo $(-\varepsilon, +\varepsilon)$ então $F_X = F_Y$ e $M_X = M_Y$ em $\mathbb{R}$.*

A demonstração envolve aspectos técnicos acima dos nossos objetivos, e de fato é difícil encontrá-la em livros-texto.[1]

[1] Veja [Billingsley, P. *Probability and Measure.* Third edition. Wiley Series in Probability and

Exemplo 8.13. Se X é uma variável aleatória não-constante assumindo valores em $\{0, 1, 2, 3, \dots\}$ e $EX < \infty$, chamamos de *amostragem por tamanho* de X à distribuição dada por $p_Y(n) = \frac{1}{EX} \cdot n \cdot p_X(n)$. Vamos mostrar que, se $Y \sim X + 1$ e M_X é finita para todo t, então $X \sim \text{Poisson}(\lambda)$ para algum $\lambda > 0$.

Demonstração. Com efeito, observe que

$$M_{X+1}(t) = e^t M_X(t)$$

e, tomando $\lambda = EX$,

$$\frac{M_X'(t)}{\lambda} = \frac{1}{\lambda} \sum_n n e^{tn} p_X(n) = \sum_n e^{tn} \left(\frac{n p_X(n)}{\lambda} \right) = \sum_n e^{tn} p_Y(n) = M_Y(t).$$

Se $M_{X+1} = M_Y$ vale

$$\begin{aligned} e^t M(t) &= \frac{M'(t)}{\lambda} \\ \frac{M'(t)}{M(t)} &= \lambda e^t, \end{aligned}$$

integrando em t obtemos

$$\log M(t) = \lambda e^t + c$$

e, como $M(0) = 1$, temos $c = -\lambda$. Logo,

$$M(t) = e^{\lambda(e^t - 1)}$$

e portanto $X \sim \text{Poisson}(\lambda)$. □

Proposição 8.14 (Variáveis Aleatórias Independentes)**.** *Se X e Y são independentes, então*

$$M_{X+Y}(t) = M_X(t) \cdot M_Y(t)$$

para todo t onde ambas M_X e M_Y estejam definidas.

Demonstração. Feita em aula. □

Mathematical Statistics, 1995, section 30] ou [Curtiss, J. H. *A note on the theory of moment generating functions.* Ann. Math. Statistics 13:430-433, 1942].

Exemplo 8.15 (Soma de Poissons Independentes)**.** Se $X \sim \text{Poisson}(\lambda)$ e $Y \sim \text{Poisson}(\mu)$ são independentes, então

$$M_{X+Y}(t) = M_X(t) \cdot M_Y(t) = e^{\lambda(e^t-1)}e^{\mu(e^t-1)} = e^{(\lambda+\mu)(e^t-1)} = M_Z(t),$$

onde $Z \sim \text{Poisson}(\lambda + \mu)$. Portanto, $X + Y \sim \text{Poisson}(\lambda + \mu)$.

Exemplo 8.16 (Binomial)**.** Se $X \sim b(n, p)$, então X é distribuída como a soma de n variáveis $X_1, \dots, X_n$ independentes com distribuição Bernoulli(p). Portanto,

$$M_X(t) = M_{X_1}(t) \cdots M_{X_n}(t) = [1 + p(e^t - 1)]^n.$$

8.2 Função Característica

Do ponto de vista teórico, a função característica é bem mais robusta e funcional que a função geradora de momentos: está definida para qualquer distribuição; sempre determina a distribuição; determina também a convergência em distribuição; não bastasse, ainda gera momentos. Entretanto, a função característica envolve a manipulação de números complexos.[2]

Definição 8.17 (Variável Aleatória Complexa)**.** Uma *variável aleatória complexa* Z é uma função $Z : \Omega \to \mathbb{C}$ tal que $Z = X + i\,Y$, onde X e Y são variáveis aleatórias reais. Se X e Y são integráveis, dizemos que Z é integrável e definimos

$$EZ = EX + iEY.$$

A integração de funções complexas em domínios reais pode ser feita, para todos os fins práticos, como no caso real. Ou seja, se $F : \mathbb{R} \to \mathbb{C}$ satisfaz $\frac{\mathrm{d}}{\mathrm{d}x}F(x) = f(x)$ para $x \in [a, b]$, então

$$\int_a^b f(x)\,\mathrm{d}x = F(b) - F(a).$$

[2] O uso de funções características não requer conhecimentos de cálculo em uma *variável* complexa. Isso porque as integrais são calculadas em $\mathrm{d}x$ para $x \in \mathbb{R}$ e não em $\mathrm{d}z$ para caminhos $\gamma \subseteq \mathbb{C}$. As únicas situações em que teríamos que sair de $\mathbb{R}$ e usar argumentos típicos de variáveis complexas, em particular $\oint_\gamma f(z)\,\mathrm{d}z = 0$, seriam na obtenção da função característica da Normal e da distribuição de Cauchy.

Vamos utilizar a *fórmula de Euler*

$$e^{iy} = \cos(y) + i\,\mathrm{sen}(y), \qquad |e^{iy}| = 1,$$

e usaremos sem demonstração os seguintes fatos:

$$e^z = \sum_n \frac{z^n}{n!}, \qquad e^{z+w} = e^z e^w, \qquad (e^g)' = e^g g', \qquad \Big(1 + \frac{z_n}{n}\Big)^n \to e^z \text{ se } z_n \to z.$$

Proposição 8.18. *Se Z e W são variáveis aleatórias complexas integráveis, então $Z+W$ é integrável com $E[Z+W] = EZ+EW$, e para $z \in \mathbb{C}$ tem-se zW integrável com $E[zW] = zEW$. Se, além disso, Z e W são independentes, então ZW é integrável com $E[ZW] = EZ \cdot EW$.*

Demonstração. Feita em aula. □

Proposição 8.19. $|EZ| \leqslant E|Z|$

Demonstração. Fazendo $EZ = re^{i\theta}$, com $r = |EZ|$, temos $E[e^{-i\theta}Z] = e^{-i\theta}E[Z] = r \in \mathbb{R}$, logo $r = E[\Re(e^{-i\theta}Z)] \leqslant E|e^{-i\theta}Z| = E|Z|$. □

Definição 8.20 (Função Característica)**.** A *função característica* de uma variável aleatória X, denotada por φ_X, é a função $\varphi_X : \mathbb{R} \to \mathbb{C}$ definida como

$$\varphi_X(t) = E[e^{itX}] = E\cos(tX) + iE\,\mathrm{sen}(tX), \quad t \in \mathbb{R}.$$

Observação 8.21. Como $|e^{itX}| = 1$, $\varphi_X(t)$ sempre está definida para todo $t \in \mathbb{R}$.

Exemplo 8.22 (Uniforme)**.** Se $X \sim U[a,b]$, então

$$\begin{aligned}
\varphi_X(t) &= E[e^{itX}] = E[\cos(tX)] + iE[\mathrm{sen}(tX)] \\
&= \int_a^b \cos(tx)\frac{1}{b-a}\mathrm{d}x + i\int_a^b \mathrm{sen}(tx)\frac{1}{b-a}\mathrm{d}x \\
&= \frac{1}{t(b-a)}\,\mathrm{sen}(tx)\Big|_a^b - \frac{i}{t(b-a)}\cos(tx)\Big|_a^b \\
&= \frac{1}{t(b-a)}\left[\mathrm{sen}(tb) - \mathrm{sen}(ta) - i\cos(tb) + i\cos(ta)\right] \\
&= \frac{-ie^{itb} + ie^{ita}}{t(b-a)} = \frac{e^{itb} - e^{ita}}{it(b-a)}.
\end{aligned}$$

Ou, mais rápido:

$$\varphi_X(t) = \int_a^b e^{itx} \frac{1}{b-a} \mathrm{d}x = \frac{1}{b-a} \frac{1}{it} \left[e^{itx} \Big|_a^b \right] = \frac{e^{itb} - e^{ita}}{it(b-a)}.$$

Exemplo 8.23 (Poisson)**.** Se $X \sim \text{Poisson}(\lambda)$, então:

$$\varphi_X(t) = E[e^{itX}] = \sum_{n=0}^{\infty} e^{itn} \frac{e^{-\lambda}\lambda^n}{n!} = e^{-\lambda} \sum_{n=0}^{\infty} \frac{(e^{it}\lambda)^n}{n!} = e^{-\lambda} e^{e^{it}\lambda} = e^{\lambda(e^{it}-1)}.$$

Exemplo 8.24 (Geométrica)**.** Se $X \sim \text{Geom}(p)$, então

$$\varphi_X(t) = \sum_{n=1}^{\infty} e^{itn} \cdot p(1-p)^{n-1} = e^{it} p \sum_{m=0}^{\infty} [(e^{it})(1-p)]^m = \frac{p}{e^{-it} + p - 1}.$$

Proposição 8.25. *Para todo $t \in \mathbb{R}$ vale $|\varphi_X(t)| \leqslant 1$. Além disso, $\varphi(0) = 1$. Ademais, para $a, b \in \mathbb{R}$, vale $\varphi_{aX+b}(t) = e^{itb}\varphi_X(at)$.*

Demonstração. Feita em aula. □

Proposição 8.26 (Independência)**.** *Se X e Y são independentes, então*

$$\varphi_{X+Y}(t) = \varphi_X(t) \cdot \varphi_Y(t)$$

para todo $t \in \mathbb{R}$.

Demonstração. Feita em aula. □

Proposição 8.27 (Cálculo de Momentos)**.** *Se $E|X^k| < \infty$ então*

$$\frac{\mathrm{d}^k}{\mathrm{d}t^k} \varphi_X(t) \Big|_{t=0} = i^k \, EX^k.$$

Demonstração. Idêntico ao caso da função geradora de momentos, com iX no lugar de X. □

Corolário 8.28 (Expansão de Taylor)**.** *Se $E|X^k| < \infty$, então*

$$\begin{aligned}\varphi_X(t) &= \varphi_X(0) + \varphi_X'(0)\cdot t + \varphi_X''(0)\,\frac{t^2}{2} + \varphi_X'''(0)\,\frac{t^3}{6} + \cdots + \varphi_X^{(k)}\,\frac{t^k}{k!} + r_k(t)\\ &= 1 + iEX\cdot t - \frac{EX^2}{2}t^2 - i\frac{EX^3}{6}t^3 + \cdots + i^k\frac{EX^k}{k!}t^k + r_k(t),\end{aligned}$$

onde o resto $r_k(t)$ é pequeno: $\frac{r_k(t)}{t^k} \underset{t\to 0}{\longrightarrow} 0.$

Demonstração. Essa é uma propriedade básica do Cálculo em uma variável real, de que toda função com k-ésima derivada no ponto $t = 0$ admite essa expansão com resto pequeno. ☐

Exemplo 8.29 (Poisson)**.** Calculando os momentos da Poisson:

$$\begin{aligned}EX &= -i\,\varphi_X'(0) = \lambda,\\ EX^2 &= -\varphi_X''(0) = \lambda^2 + \lambda,\\ VX &= EX^2 - (EX)^2 = \lambda.\end{aligned}$$

Proposição 8.30 (Unicidade)**.** *Se $\varphi_X(t) = \varphi_Y(t)$ $\forall t \in \mathbb{R}$, então $X \sim Y$.*

Demonstração. O leitor interessado pode consultar [Jam04, pp. 226–228]. ☐

Exemplo 8.31 (Soma de Poissons Independentes)**.** Se $X \sim$ Poisson(λ) e $Y \sim$ Poisson(μ) são independentes, então

$$\varphi_{X+Y}(t) = \varphi_X(t)\cdot\varphi_Y(t) = e^{\lambda(e^{it}-1)}e^{\mu(e^{it}-1)} = e^{(\lambda+\mu)(e^{it}-1)} = \varphi_Z(t),$$

onde $Z \sim$ Poisson($\lambda+\mu$). Portanto, $X + Y \sim$ Poisson($\lambda+\mu$).

Convergência em distribuição

O Teorema de Continuidade relaciona convergência de funções características com convergência em distribuição, vista no Capítulo 5.

Teorema 8.32 (Teorema da Continuidade de Lévy)**.** *Sejam X e $(X_n)_{n\in\mathbb{N}}$ variáveis aleatórias. Então*

$$X_n \xrightarrow{\mathrm{d}} X \qquad \textit{se, e somente se,} \qquad \varphi_{X_n}(t) \to \varphi_X(t) \quad \forall t \in \mathbb{R}.$$

Demonstração. O leitor interessado pode consultar [Jam04, pp. 234–239]. $\square$

Exemplo 8.33 (Binomial Converge a Poisson)**.** Seja $\lambda > 0$ e para $n > \lambda^{-1}$ considere $X_n \sim b(n, \frac{\lambda}{n})$. Então

$$X_n \xrightarrow{\mathrm{d}} \text{Poisson}(\lambda).$$

Demonstração. Analisando a função característica das X_n obtemos

$$\varphi_{X_n}(t) = [1 + \tfrac{\lambda}{n}(e^{it} - 1)]^n \to e^{\lambda(e^{it}-1)} = \varphi_X(t) \qquad \text{com } X \sim \text{Poisson}(\lambda). \quad \square$$

Demonstração do Teorema 6.4. Como as X_n são i.i.d., temos

$$\varphi_{\frac{S_n}{n}}(t) = \varphi_{S_n}(\tfrac{t}{n}) = \varphi_{X_1}(\tfrac{t}{n}) \cdots \varphi_{X_n}(\tfrac{t}{n}) = \left[\varphi_{X_1}\left(\tfrac{t}{n}\right)\right]^n = \left[1 + i\frac{\mu t}{n} + r_1\left(\tfrac{t}{n}\right)\right]^n,$$

onde $r_1(\cdot)$ é tal que $\frac{r_1(w)}{w} \to 0$ quando $w \to 0$. Segue que $\varphi_{\frac{S_n}{n}}(t) \to e^{it\mu}$ quando $n \to \infty$, para todo $t \in \mathbb{R}$. Pelo Teorema 8.32, $\frac{S_n}{n} \xrightarrow{\mathrm{d}} \mu$. Como μ é constante, isso é o mesmo que $\frac{S_n}{n} \xrightarrow{P} \mu$. $\square$

Demonstração do Teorema 7.2. Supomos sem perda de generalidade que $\mu = 0$. Como as X_n são i.i.d., temos

$$\varphi_{\frac{S_n}{\sigma\sqrt{n}}}(t) = \varphi_{S_n}(\tfrac{t}{\sigma\sqrt{n}}) = \left[\varphi_{X_1}\left(\tfrac{t}{\sigma\sqrt{n}}\right)\right]^n = \left[1 - \frac{t^2}{2n} + r_2\left(\tfrac{t}{\sigma\sqrt{n}}\right)\right]^n,$$

onde $r_2(\cdot)$ é tal que $\frac{r_2(w)}{w^2} \to 0$ quando $w \to 0$. Segue que $\varphi_{\frac{S_n}{\sigma\sqrt{n}}}(t) \to e^{-\frac{t^2}{2}}$ quando $n \to \infty$, para todo $t \in \mathbb{R}$. Pelo Teorema 8.32, $\frac{S_n}{\sigma\sqrt{n}} \xrightarrow{\mathrm{d}} \mathcal{N}$. $\square$

8.3 Exercícios

8.1. Se $X \sim \mathcal{N}(0,1)$, mostre que $M_X(t) = e^{\frac{t^2}{2}}$. Mostre que $EX = 0$. Mostre que $VX = 1$. (Sugestão: verifique que $-(z^2 - 2tz) = t^2 - (z-t)^2$ e faça $z - t = u$.)

8.2. Sejam $X_1, X_2, X_2, \dots$ independentes,

$$S_n = X_1 + X_2 + \cdots + X_n \qquad \text{e} \qquad \bar{S}_n = \frac{X_1 + X_2 + \cdots + X_n}{n}.$$

Mostre as seguintes propriedades:

1. Se $X \sim \mathcal{N}(\mu, \sigma^2)$, então $Z = \frac{X-\mu}{\sigma} \sim \mathcal{N}(0,1)$.
2. Assim, se $X \sim \mathcal{N}(\mu, \sigma^2)$, então

$$\begin{aligned} M_X(t) &= e^{\mu t + \frac{1}{2}\sigma^2 t^2} \\ EX &= \mu \\ VX &= \sigma^2. \end{aligned}$$

3. Se $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$, então $S_n \sim \mathcal{N}(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2)$.
4. Se $X_i \sim \mathcal{N}(\mu, \sigma^2)$, então $S_n \sim \mathcal{N}(n\mu, n\sigma^2)$.
5. Se $X_i \sim \mathcal{N}(\mu, \sigma^2)$, então $\bar{S}_n \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$.

8.3. A distribuição dos comprimentos dos elos da corrente de bicicleta é normal, com média 2 cm e variância $0,01\ cm^2$. Para que uma corrente se ajuste à bicicleta, deve ter comprimento total entre 58 e 61 cm. Qual é a probabilidade de uma corrente com 30 elos não se ajustar à bicicleta?

8.4. As durações de gravidez têm distribuição normal com média de 268 dias e desvio-padrão de 15 dias.

(a) Selecionada aleatoriamente uma mulher grávida, determine a probabilidade de que a duração de sua gravidez seja inferior a 260 dias.

(b) Se 25 mulheres escolhidas aleatoriamente são submetidas a uma dieta especial a partir do dia em que engravidam, determine a probabilidade de os prazos de duração de suas gravidezes terem média inferior a 260 dias (admitindo-se que a dieta não produza efeito).

(c) Se as 25 mulheres têm realmente média inferior a 260 dias, há razão de preocupação para os médicos de pré-natal? Justifique adequadamente.

8.5. O peso de uma determinada fruta é uma variável aleatória com distribuição normal com média de 200 gramas e desvio-padrão de 50 gramas. Determine a probabilidade de um lote contendo 100 unidades dessa fruta pesar mais que 21 kg.

8.6. Um elevador pode suportar uma carga de 10 pessoas ou um peso total de 1750 libras. Assumindo que apenas homens tomam o elevador e que seus pesos são normalmente distribuídos com média 165 libras e desvio-padrão de 10 libras, qual a probabilidade de que o peso limite seja excedido para um grupo de 10 homens escolhidos aleatoriamente?

8.7. Se $X \sim U[a, b]$, calcule $M_X(t)$. Use a função geradora de momentos para calcular EX e VX.

8.8. As cinco primeiras repetições de um experimento custam R\$ $10,00$ cada. Todas as repetições subsequentes custam R\$ $5,00$ cada. Suponha que o experimento seja repetido até que o primeiro sucesso ocorra. Se a probabilidade de sucesso de uma repetição é igual a $0,9$, e se as repetições são independentes, qual é custo esperado da operação?

8.9. Se $X \sim \exp(\lambda)$, calcule $M_X(t)$. Use a função geradora de momentos para calcular EX e VX.

8.10. Seja Y uma variável aleatória absolutamente contínua com função de densidade de probabilidade dada por

$$f_Y(y) = \begin{cases} ye^{-y}, & \text{se } y > 0 \\ 0, & \text{caso contrário} \end{cases}$$

Ache a função geradora de momentos de Y e use-a para calcular EY e VY.

Exercício 8.11. Seja X uma variável aleatória. Mostre que o conjunto $\{t \in \mathbb{R} : M_X(t) < \infty\}$ é um intervalo que contém o ponto $t = 0$.

8.12. Se $X \sim \mathcal{N}(0, 1)$, mostre que $\varphi_X(t) = e^{-\frac{t^2}{2}}$.
Você pode usar o seguinte fato, da teoria do cálculo em uma variável complexa:

$$\int_{\mathbb{R}} e^{-(w+ci)^2} \, \mathrm{d}w = \int_{\mathbb{R}} e^{-w^2} \, \mathrm{d}w$$

para qualquer $c \in \mathbb{R}$.

8.13. Se $X \sim \mathcal{N}(\mu, \sigma^2)$, calcule $\varphi_X(t)$.

8.14. [Jam04, Capítulo 6].
Recomendados: 1, 2, 3, 4, 7, 9, 13a, 14, 17, 18, 21, 29.

Capítulo 9

Esperança Condicional

Muitas vezes a estrutura do espaço amostral Ω é complicada demais para estudarmos as grandezas de interesse diretamente a partir dos eventos elementares $\omega \in \Omega$, até mesmo em situações aparentemente simples.

Neste contexto, estudamos as propriedades de algumas grandezas observáveis, ou ainda, conseguimos dividir Ω em classes que podem ser estudadas como um todo. Estudar uma partição $\mathcal{D}$ de Ω quer dizer que estamos trabalhando apenas com a informação relacionada àquela partição.

Da mesma forma, e inúmeras situações queremos estudar o comportamento de uma dada variável aleatória X em termos de outra variável aleatória Y, o que em estatística significa dizer que buscamos um estimador para X sabendo-se o valor da variável Y.

9.1 Esperança Condicional dada uma Partição

Nesta seção, assumimos que todas as variáveis aleatórias são discretas e não-negativas. Caso as variáveis aleatórias sejam todas simples, pode-se somar-lhes uma constante para trabalhar com variáveis não-negativas.

Definição 9.1. Dizemos que $\mathcal{D} = \{D_1, D_2, D_3, \dots\}$ é uma *partição* de $(\Omega, \mathcal{F})$ se $D_i \in \mathcal{F}$ $\forall i$, $D_i \cap D_j = \emptyset$ $\forall i \neq j$, e $\cup_i D_i = \Omega$. Os elementos D_i da partição $\mathcal{D}$ são chamados *átomos* de $\mathcal{D}$.

Exemplo 9.2. Sejam $X_1, X_2, X_3, \dots$ variáveis aleatórias assumindo valores em $\{-1, 1\}$. O espaço Ω pode ser dividido em átomos onde X_1 e X_2 são constantes.

Definição 9.3 (Probabilidade Condicional Dada uma Partição)**.** Dada uma partição $\mathcal{D} = \{D_i\}_i$ e um evento A, definimos a *variável aleatória*

$$P(A|\mathcal{D}) = P(A|\mathcal{D})(\omega) = \sum_i P(A|D_i)\, \mathbb{1}_{D_i}(\omega).$$

Note que, em cada átomo D_i da partição $\mathcal{D}$, a variável aleatória $P(A|\mathcal{D})$ assume o valor constante $P(A|D_i)$.

Exemplo 9.4. Suponha que

$$P\,(\text{chover amanhã}|\text{chove hoje}) = 0,7,$$
$$P\,(\text{chover amanhã}|\text{não chove hoje}) = 0,1,$$

e seja $\mathcal{D} = \{\text{chove hoje}, \text{não chove hoje}\}$. Então

$$Z = P(\text{chover amanhã}|\mathcal{D}) = \begin{cases} 0,7, & \text{se no estado } \omega \text{ chove hoje}, \\ 0,1, & \text{caso contrário}. \end{cases}$$

Teorema 9.5 (Lei da Probabilidade Total)**.**

$$P(A) = E\left[P(A|\mathcal{D})\right].$$

Demonstração. $E\left[P(A|\mathcal{D})\right] = E[\sum_j P(A|D_j)\mathbb{1}_{D_j}] = \sum_j P(A|D_j)P(D_j) = P(A)$, onde a segunda igualdade é basicamente uma aplicação do Teorema 4.27 com relação à variável aleatória $J = \sum_j j\mathbb{1}_{D_j}$ que vale j em cada D_j. □

Exemplo 9.6. Se $P(\text{chover hoje}) = 0,6$, então

$$P(\text{chover amanhã}) = EZ = \sum_z z \cdot P(Z = z) = 0,7 \times 0,6 + 0,1 \times 0,4 = 0,46.$$

Definição 9.7. Seja X uma variável aleatória discreta. Definimos a *partição induzida por* X como $\mathcal{D}_X = \{D_1, D_2, D_3, \dots\}$, onde $D_j = \{\omega : X(\omega) = x_j\}$. Denotamos a variável aleatória $P(A|\mathcal{D}_X)(\omega)$ por $P(A|X)(\omega)$.

Exemplo 9.8. Se X e Y são i.i.d. Bernoulli(p), considere o evento $A = [X+Y = 1]$. Vamos calcular $P(A|Y)$:

$$P(A|Y) = p\,\mathbb{1}_{[Y=0]} + (1-p)\,\mathbb{1}_{[Y=1]},$$

ou, escrevendo explicitamente como função de Y:

$$P(A|Y) = p\,(1-Y) + (1-p)\,Y.$$

Definição 9.9 (Esperança Condicional Dada uma Partição)**.** Seja X uma variável aleatória simples. Considere $\mathcal{D}$ uma partição de $(\Omega, \mathcal{F})$. Definimos a *variável aleatória*

$$E(X|\mathcal{D})(\omega) = \sum_i E(X|D_i)\,\mathbb{1}_{D_i}(\omega).$$

Observe que, desenvolvendo a expressão acima, temos

$$E(X|\mathcal{D}) = \sum_j \left[\sum_x x \cdot P(X = x|D_j)\right]\mathbb{1}_{D_j} = \sum_x x \cdot \left(\sum_j P(X = x|D_j)\,\mathbb{1}_{D_j}\right),$$

e portanto

$$E(X|\mathcal{D}) = \sum_x x \cdot P(X = x\,|\,\mathcal{D}).$$

De forma mais geral, se $X = \sum_i x_i \mathbb{1}_{A_i}$ para uma partição $\{A_i\}_i$, então $E(X|\mathcal{D}) = \sum_i x_i \cdot P(A_i\,|\,\mathcal{D})$.

A esperança condicional $E(X|\mathcal{D})$ é a uma aproximação para X que depende apenas da informação relacionada à partição $\mathcal{D}$. Ela é grosseira o suficiente para atender à restrição de ser constante no átomos de $\mathcal{D}$, mas fina o suficiente para ser a melhor entre todas as aproximações sujeitas a essa restrição. Veja a Figura 9.1.

Exemplo 9.10. Lançamento de um dado honesto. Seja $\mathcal{D} = \{\text{ímpar}, \text{par}\}$. Temos

$$Z(\omega) = E(X|D)(\omega) = \begin{cases} E(X|X \text{ é par}), & \text{se } X(\omega) \text{ é par},\\ E(X|X \text{ é ímpar}), & \text{se } X(\omega) \text{ é ímpar}. \end{cases}$$

Assim,

$$Z(\omega) = \begin{cases} 4, & \text{se } X(\omega) \text{ é par},\\ 3, & \text{se } X(\omega) \text{ é ímpar}. \end{cases}$$

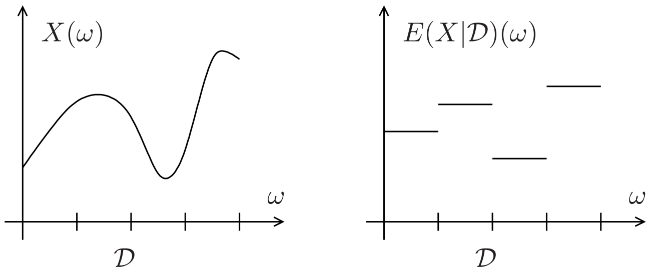

Figura 9.1: Ilustração da definição de $E(X|\mathcal{D})$.

Proposição 9.11 (Propriedades da esperança condicional)**.**

1. $E(c\,|\,\mathcal{D}) = c$
2. *Se* $X \leqslant Y$ *então* $E(X|\mathcal{D}) \leqslant E(Y|\mathcal{D})$
3. $E(aX + bY|\mathcal{D}) = aE(X|\mathcal{D}) + bE(Y|\mathcal{D})$
4. $E(X|\{\Omega\}) = EX$.

Demonstração. Feita em aula. □

Teorema 9.12 (Generalização da Lei da Probabilidade Total)**.**

$$EX = E\left[E(X|\mathcal{D})\right].$$

Demonstração. Como no Teorema 9.5,

$$\begin{aligned} E\left[E(X|\mathcal{D})\right] &= \sum_j \sum_x x \cdot P(X = x|D_j)P(D_j) = \\ &= \sum_x x \cdot \sum_j P(X = x|D_j)P(D_j) = \sum_x x \cdot P(X = x) = EX. \end{aligned}$$ □

Com o Teorema 9.12 completamos o diagrama da Figura 9.2.

Exemplo 9.13. Lançamento do dado no Exemplo 9.10. Temos

$$EX = E\big[E(X|\mathcal{D})\big] = EZ = \frac{7}{2}.$$

$$
\begin{tikzcd}
P(\cdot) \arrow[r, "EX=\sum_x x\cdot P(X=x)"] \arrow[d, "P(A|D)=\frac{P(A\cap D)}{P(D)}"] & E(\cdot) \arrow[d, "E(X|D)=\frac{E[X\mathbb{1}_D]}{P(D)}"'] \\
P(\cdot|D) \arrow[r, "E(X|D)=\sum_x x\cdot P(X=x|D)"] \arrow[d, "P(A|\mathcal{D})=\sum_i P(A|D_i)\mathbb{1}_{D_i}"] & E(\cdot|D) \arrow[d, "E(X|\mathcal{D})=\sum_i E(X|D_i)\mathbb{1}_{D_i}"'] \\
P(\cdot|\mathcal{D}) \arrow[r, "E(X|\mathcal{D})=\sum_x x\cdot P(X=x|\mathcal{D})"'] \arrow[uu, bend left, "P(A)=E[P(A|\mathcal{D})]"] & E(\cdot|\mathcal{D}) \arrow[uu, bend right, "EX=E[E(X|\mathcal{D})]"']
\end{tikzcd}
$$

Figura 9.2: Relação entre probabilidade, esperança, probabilidade condicional dado um evento, esperança condicional dado um evento, probabilidade condicional dada uma partição, e esperança condicional dada uma partição.

Se Y é uma variável aleatória discreta, denotamos

$$E(X|Y) = E(X|\mathcal{D}_Y).$$

Exercício 9.1. Se X e Y são independentes então $E(X|Y) = EX$ é constante.

Observação 9.14. Caso particular do teorema anterior: $EX = E\left[E(X|Y)\right]$.

Dizemos que $\mathcal{D}_2$ é *mais fina* que $\mathcal{D}_1$, denotado por $\mathcal{D}_2 \succcurlyeq \mathcal{D}_1$, se todo elemento de $\mathcal{D}_1$ é igual à união de elementos de $\mathcal{D}_2$, isto é, se para todo $D \in \mathcal{D}_1$ existe $\mathcal{C} \subseteq \mathcal{D}_2$ tal que $D = \cup\mathcal{C}$. Isso significa que $\mathcal{D}_2$ tem “mais informação” do que $\mathcal{D}_1$.

Exemplo 9.15. Seja $\mathcal{D}_2 = \{D_1, D_2, D_3, D_4\}$ uma partição de Ω, e sejam $D_5 = D_1 \cup D_3$, $D_6 = D_2$ e $D_7 = D_4$. Se definimos $\mathcal{D}_1 = \{D_5, D_6, D_7\}$, temos $\mathcal{D}_2 \succcurlyeq \mathcal{D}_1$.

Exemplo 9.16. Para qualquer partição $\mathcal{D}$ vale $\mathcal{D} \succcurlyeq \mathcal{D} \succcurlyeq \{\Omega\}$.

Dada uma partição $\mathcal{D} = \{D_1, D_2, D_3, \dots\}$, dizemos que X é *$\mathcal{D}$-mensurável* se $\mathcal{D} \succcurlyeq \mathcal{D}_X$, isto é, se a informação sobre $\mathcal{D}$ determina o valor de X. De forma equivalente, X é *$\mathcal{D}$-mensurável* se e somente se pode ser escrito como

$$X = \sum_i x_i \mathbb{1}_{D_i},$$

onde os x_i não são necessariamente distintos.

Observação 9.17. X sempre é $\mathcal{D}_X$-mensurável. Se $Y = g(X)$ para alguma função $g : \mathbb{R} \to \mathbb{R}$, então Y é $\mathcal{D}_X$-mensurável.

Definimos $\mathcal{D}_{X_1,X_2,\dots,X_d}$ como sendo a partição gerada pelo vetor $(X_1, X_2, \dots, X_d)$, ou seja, a partição cujos átomos são os maiores conjuntos onde todas as X_j são constantes. Mais formalmente, se $\{\boldsymbol{x}^1, \boldsymbol{x}^2, \boldsymbol{x}^3, \dots\}$ são os valores assumidos pelo vetor aleatório $\boldsymbol{X}$, definimos $D_i = [\boldsymbol{X} = \boldsymbol{x}^i]$ e $\mathcal{D} = \{D_1, D_2, D_3, \dots\}$.

Exercício 9.2. Mostre que $\mathcal{D}_{X_1,X_2} \succcurlyeq \mathcal{D}_{X_1}$.

De forma análoga a $E(X|Y)$, definimos

$$E(X|Y_1, \dots, Y_n) = E\left(X \middle| \mathcal{D}_{Y_1,\dots,Y_n}\right).$$

Proposição 9.18. *Se X é $\mathcal{D}$-mensurável, então*

$$E(XY|\mathcal{D}) = XE(Y|\mathcal{D}).$$

Em particular, $E(X|\mathcal{D}) = X$. Ademais, $E(X|X) = X$.

Demonstração. Escrevemos

$$X = \sum_i x_i \mathbb{1}_{D_i} \quad \text{e} \quad Y = \sum_j y_j \mathbb{1}_{A_j},$$

assim

$$XY = \sum_i \sum_j x_i y_j \mathbb{1}_{A_j \cap D_i}$$

e portanto

$$\begin{aligned}
E[XY|\mathcal{D}] &= \sum_i \sum_j x_i y_j P(A_j \cap D_i|\mathcal{D}) \\
&= \sum_i \sum_j x_i y_j \sum_m P(A_j \cap D_i|D_m)\mathbb{1}_{D_m} \\
&= \sum_i \sum_j x_i y_j P(A_j|D_i)\mathbb{1}_{D_i}.
\end{aligned}$$

Por outro lado,

$$\begin{aligned}
XE[Y|\mathcal{D}] &= \sum_i x_i \mathbb{1}_{D_i} \cdot \sum_j y_j P(A_j|\mathcal{D}) \\
&= \sum_i x_i \mathbb{1}_{D_i} \cdot \sum_j y_j \sum_m P(A_j|D_m)\mathbb{1}_{D_m} \\
&= \sum_i \sum_j x_i y_j \sum_m P(A_j|D_m)\mathbb{1}_{D_m \cap D_i} \\
&= \sum_i \sum_j x_i y_j P(A_j|D_i)\mathbb{1}_{D_i},
\end{aligned}$$

concluindo a prova. □

Observação 9.19. $E(X|\mathcal{D})$ sempre é $\mathcal{D}$-mensurável.

Proposição 9.20. *Se $\mathcal{C} \preccurlyeq \mathcal{D}$, então*

$$E\left[E(X|\mathcal{D})\middle|\mathcal{C}\right] = E\left[E(X|\mathcal{C})\middle|\mathcal{D}\right] = E(X|\mathcal{C}).$$

Em particular,

$$E\left[E(X|Y_1, Y_2)\middle|Y_1\right] = E(X|Y_1).$$

Demonstração. Pela observação anterior, $E[X|\mathcal{C}]$ é $\mathcal{C}$-mensurável. Como $\mathcal{D} \succcurlyeq \mathcal{C}$, $E[X|\mathcal{C}]$ também é $\mathcal{D}$-mensurável, e pela proposição anterior $E\left[E(X|\mathcal{C})\middle|\mathcal{D}\right] = E(X|\mathcal{C})$. Resta mostrar que $E\left[E(X|\mathcal{D})\middle|\mathcal{C}\right] = E(X|\mathcal{C})$.

Escrevendo $\mathcal{C} = \{C_1, C_2, \dots\}$ e $\mathcal{D} = \{D_1, D_2, \dots\}$, temos

$$\begin{aligned} E[E(X|\mathcal{D})|\mathcal{C}] &= \sum_i \mathbb{1}_{C_i} E\left[\sum_j E(X|D_j)\mathbb{1}_{D_j} \middle| C_i\right] \\ &= \sum_i \mathbb{1}_{C_i} \sum_j E(X|D_j)P(D_j|C_i), \end{aligned}$$

onde a segunda igualdade é basicamente uma aplicação do Teorema 4.27 com relação à variável aleatória $J = \sum_j j\mathbb{1}_{D_j}$ que vale j em cada D_j. Continuando,

$$\begin{aligned} E[E(X|\mathcal{D})|\mathcal{C}] &= \sum_i \mathbb{1}_{C_i} \sum_{j:D_j \subseteq C_i} \frac{E(X\mathbb{1}_{D_j})}{P(D_j)} \times \frac{P(D_j)}{P(C_i)} \\ &= \sum_i \mathbb{1}_{C_i} \frac{1}{P(C_i)} \sum_{j:D_j \subseteq C_i} E(X\mathbb{1}_{D_j}) \\ &= \sum_i \mathbb{1}_{C_i} \frac{1}{P(C_i)} E(X\mathbb{1}_{C_i}) \\ &= \sum_i \mathbb{1}_{C_i} E(X|C_i) = E(X|\mathcal{C}). \end{aligned}$$

Para justificar a terceira igualdade sem usar o Corolário 10.5, escrevemos $X = \sum_k x_k \mathbb{1}_{A_k}$ e fazemos

$$\begin{aligned} \sum_{j:D_j \subseteq C_i} E(X\mathbb{1}_{D_j}) = \sum_{j:D_j \subseteq C_i} \sum_k x_k P(A_k \cap D_j) = \\ = \sum_k x_k \sum_{j:D_j \subseteq C_i} P(A_k \cap D_j) = \sum_k x_k P(A_k \cap C_i) = E(X\mathbb{1}_{C_i}). \end{aligned}$$

Isso conclui a prova da proposição. □

Exemplo 9.21. Dada uma função g, vale

$$E\left[g(Y)E\left(X|Y\right)\right] = E\left[Xg(Y)\right].$$

Com efeito, como $Z = g(Y)$ é $\mathcal{D}_Y$-mensurável, temos

$$E\left[Xg(Y)\middle|Y\right] = g(Y)E(X|Y).$$

Tomando a esperança dos dois lados, obtemos a equação anterior.

9.2 Distribuição Condicional Regular

Quando Y é uma variável aleatória discreta assumindo valores $y_1, y_2, \ldots,$ essa variável aleatória induz uma partição $\mathcal{D}_Y$ de $(\Omega, \mathcal{F})$, e temos as seguintes relações:

$$P(X \in B) = \sum_y P(X \in B|Y = y)P(Y = y) = E\left[P(X \in B|Y)\right]$$
$$E(X) = \sum_y E(X|Y = y)P(Y = y) = E\left[E(X|Y)\right].$$

No caso de variáveis aleatórias Y que não sejam discretas, temos que dar sentido a expressões como $P(X \in B|Y = y)$ e $E(X|Y = y)$, mesmo que $P(Y = y)$ seja zero, para poder dizer que relações análogas continuam valendo.

Definição 9.22 (Distribuição Condicional Regular)**.** Sejam X e Y variáveis aleatórias definidas no mesmo espaço de probabilidade $(\Omega, \mathcal{F}, P)$. A *distribuição condicional regular de X dado que $Y = y$* é definida por

$$P\big(X \in [s,t] \,\big|\, Y = y\big) = \lim_{\Delta \to 0} \lim_{\delta \to 0} P\big(X \in [s - \Delta, t + \Delta] \,\big|\, Y \in [y - \delta, y + \delta]\big)$$

para todo $s < t$ e $y \in A$, onde A é algum conjunto tal que $P(Y \in A) = 1$. É importante tomar o limite primeiro em δ e depois em Δ. Quando $s = -\infty$, definimos a *função de distribuição condicional acumulada*

$$F_X(t|Y = y) = P(X \leqslant t|Y = y).$$

Teorema 9.23. *Para quase todo $y \in \mathbb{R}$, isto é, para todo $y \in A$ onde A é um conjunto tal que $P(Y \in A) = 1$, o duplo limite acima existe para todo $s < t$ e determina uma probabilidade em $\mathbb{R}$.*

Demonstração. Omitida. Envolve Teoria da Medida. □

Na prática, o que se faz é *encontrar um candidato ad hoc* de quem deveria ser a distribuição condicional regular de X dado Y, segundo princípios que se aplicam em diferentes casos, e *verifica-se a posteriori* que o candidato proposto satisfaz a Definição 9.22. À continuação veremos alguns desses princípios.

Caso de Y discreta Se Y é variável aleatória discreta, a distribuição condicional de X dado $Y = y$ é dada por

$$P(X \in B|Y = y) = \frac{P(X \in B, Y = y)}{P(Y = y)}$$

para todo y tal que $P(Y = y) > 0$.

Caso de X e Y independentes Se X e Y são independentes, o condicionamento em $Y = y$ não afeta em nada a variável X. Neste caso temos

$$P(X \in B|Y = y) = P(X \in B).$$

Caso de X e Y possuírem densidade conjunta Se X e Y possuem função de densidade conjunta $f_{X,Y}(x, y)$, a função de densidade condicional de X dado $Y = y$ é dada por

$$f_X(x|Y = y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}$$

para todo y tal que $f_Y(y) > 0$. Neste caso a função de distribuição condicional de X dado que $Y = y$ é dada por

$$F_X\left(t|Y = y\right) = \int_{-\infty}^{t} f_X(x|Y = y)\,\mathrm{d}x.$$

Exemplo 9.24. Sejam X e Y com densidade conjunta

$$f_{X,Y}(x, y) = \begin{cases} 6xy(2 - x - y), & 0 < x < 1, 0 < y < 1, \\ 0, & \text{caso contrário.} \end{cases}$$

Vamos determinar a distribuição condicional de X dado que $Y = y$. Temos

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{X,Y}(x, y)\mathrm{d}x = \int_0^1 6xy(2 - x - y)\mathrm{d}x = 4y - 3y^2$$

se $y \in (0,1)$ e 0 caso contrário. Assim, para $y \in [0,1]$ temos

$$f_X(x \mid Y=y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \begin{cases} \frac{6x(2-x-y)}{4-3y}, & 0<x<1 \\ 0, & \text{caso contrário.} \end{cases}$$

Para y fora desse intervalo $f_X(\cdot|Y=y)$ é irrelevante, pois $P(Y \notin [0,1]) = 0$.

Exemplo 9.25. Sejam X e Y com densidade conjunta

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{2} y e^{-xy}, & 0<x<\infty \ \text{ e } \ 0<y<2, \\ 0, & \text{caso contrário.} \end{cases}$$

Vamos determinar a distribuição condicional de X dado que $Y=y$. Temos

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)\mathrm{d}x = \frac{1}{2}\int_0^{\infty} y e^{-xy}\mathrm{d}x = \frac{1}{2}$$

para $0<y<2$. Logo $Y \sim U[0,2]$.

Assim, para $y \in (0,2]$ temos

$$f_X(x \mid Y=y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \begin{cases} y e^{-xy}, & x>0, \\ 0, & x \leqslant 0. \end{cases}$$

Caso de Y possuir densidade e X ser discreta Se X é discreta e Y tem função de densidade $f_Y(y)$, a função de probabilidade condicional de X dado $Y=y$ é dada por

$$p_X(x|Y=y) = \frac{P(X=x)f_Y(y|X=x)}{f_Y(y)}$$

para todo y tal que $f_Y(y)>0$. Neste caso a função de distribuição condicional de X dado $Y=y$ é

$$F_X\left(t|Y=y\right) = \sum_{x \leqslant t} p_X(x|Y=y).$$

Princípio da preservação das chances relativas O princípio da preservação das chances relativas diz que, dada a ocorrência de um evento, os resultados possíveis dentro desse evento mantêm as mesmas chances relativas que possuíam antes.

Exemplo 9.26. $X \sim \mathcal{N}(0,1)$ e $Y = X^2$. Qual a distribuição condicional de X dado que $Y = y$?

Como $P(Y > 0) = 1$, basta considerar valores $y > 0$. Sabendo que $Y = y$ temos duas alternativas: $X = \sqrt{y}$ ou $X = -\sqrt{y}$. Como $f_X(y) = f_X(-y)$, esses dois valores continuam tendo a mesma chance quando condicionamos a $Y = y$. Temos então $P\left(X = \sqrt{y}\,\middle|\,Y = y\right) = P\left(X = -\sqrt{y}\,\middle|\,Y = y\right) = \frac{1}{2}$, $y > 0$.

Exemplo 9.27. Seja $X \sim U[0,2]$ e $Y \sim U[-1,1]$ independentes. Vamos encontrar $F_X(x|X+Y=z)$.

Seja $Z = X + Y$. A densidade conjunta de X e Y é dada por $f_{XY}(x,y) = \frac{1}{4}\mathbb{1}_{[0,2]\times[-1,1]}(x,y)$, e a marginal de X é dada por $f_X(x) = \frac{1}{2}\mathbb{1}_{[0,2]}(x)$. Condicionando a $Z = z$, temos que o conjunto dos resultados possíveis fica restrito a uma diagonal $\{(x,y) \in [0,2]\times[-1,1] : x+y=z\}$ que corta o quadrado $[0,2]\times[-1,1]$. Pelo Princípio da Preservação das Chances Relativas, todos os pontos desse conjunto eram "equiprováveis" antes do condicionamento e devem continuar equiprováveis dentro do conjunto da restrição. Assim, para $z > 1$ devemos ter $X \sim U[z-1,2]$ e para $z < 1$ devemos ter $X \sim U[0,z+1]$, ou seja

$$f_X(X|Z=z) = \begin{cases} \frac{1}{3-z}\mathbb{1}_{[z-1,2]}(x), & 1 < z < 3, \\ \frac{1}{z+1}\mathbb{1}_{[0,z+1]}(x), & -1 < z < 1. \end{cases}$$

Princípio da substituição O princípio da substituição permite substituir Y por y sempre que se condiciona a $Y = y$. Se $W = g(X,Y)$, então

$$P(W \in B|Y=y) = P(g(X,y) \in B|Y=y) = P\left[X \in \{x : g(x,y) \in B\}\,\middle|\,Y=y\right].$$

9.3 Esperança Condicional Regular

Dada X integrável, definimos $E(X|Y=y)$ como a esperança de X com respeito à sua distribuição condicional regular dado que $Y = y$.

Teorema 9.28. *Sejam X e Y variáveis aleatórias definidas em $(\Omega,\mathcal{F},P)$ com X integrável. Então existe algum $A \in \mathcal{B}$ tal que $P(Y \in A) = 1$ e $E(X|Y=y)$ é finita para todo $y \in A$.*

Demonstração. Omitida. Envolve Teoria da Medida. □

Tomando $g : \mathbb{R} \to \mathbb{R}$ como sendo a função tal que $E(X|Y = y) = g(y)$, definimos a *variável aleatória* $E(X|Y)$ por $E(X|Y) = g(Y)$, isto é, $E(X|Y)(\omega) = g(Y(\omega))$.

Exemplo 9.29. Se $X \sim U[0,2]$ e $Y = \max\{X, 1\}$. Temos que Y assume valores em $[1,2]$. Tomando y em $(1,2]$, temos que $[Y = y] = [X = y]$ e, pelo Princípio da Substituição, $E[X|Y = y] = y$. Tomando $y = 1$, temos que $[Y = 1] = [X \leqslant 1]$. Assim,

$$F_X(x|Y=1) = F_X(x|X \leqslant 1) = \frac{P(X \leqslant x, X \leqslant 1)}{P(X \leqslant 1)} = \begin{cases} \frac{x/2}{1/2} = x, & 0 \leqslant x \leqslant 1, \\ 0, & x < 0, \\ 1, & x > 1. \end{cases}$$

Logo, $f_X(x|Y = 1) = \frac{\mathrm{d}}{\mathrm{d}x} F_X(x|Y = 1) = \mathbb{1}_{[0,1]}(x)$ e

$$E(X|Y = 1) = \int_0^1 x f_X(x|Y = 1) \mathrm{d}x = \frac{1}{2}.$$

Portanto, $E(X|Y = y) = y$ se $y \in (1,2]$ e $E(X|Y = y) = \frac{1}{2}$ se $y = 1$. Substituindo,

$$E(X|Y) = \begin{cases} \frac{1}{2}, & Y = 1, \\ Y, & 1 < Y \leqslant 2. \end{cases}$$

Teorema 9.30. *Se X é integrável então*

$$EX = E\big[E(X|Y)\big].$$

Demonstração. Omitida. Envolve Teoria da Medida. □

Exemplo 9.31. No Exemplo 9.29, temos que Y é mista com funções de densidade e probabilidade dadas por

$$p_Y(y) = \tfrac{1}{2}\mathbb{1}_{\{1\}}(y), \quad f_Y(y) = \tfrac{1}{2}\mathbb{1}_{[1,2]}(y)$$

e portanto

$$EX = E\big[E(X|Y)\big] = E[g(y)] = \tfrac{1}{2} \times \tfrac{1}{2} + \int_1^2 \tfrac{1}{2}\, y \,\mathrm{d}y = 1.$$

Teorema 9.32 (Propriedades da Esperança Condicional)**.**

1. *$E(c|Y) = c$ quase certamente.*
2. *$X \leqslant Z \Rightarrow E(X|Y) \leqslant E(Z|Y)$ quase certamente.*
3. *$E(aX + bZ|Y) = aE(X|Y) + bE(Z|Y)$ quase certamente.*
4. *Se $X = g(Y)$ então $E(X|Y) = X$ quase certamente.*
5. *Se $Z = g(Y)$, então*

$$E\left[E\left(X|Z\right)\big|Y\right] = E\left[E\left(X|Y\right)\big|Z\right] = E\left[X\big|Z\right] \text{ quase certamente.}$$

6. *Se $Z = g(Y)$, $E|X| < \infty$ e $E|XZ| < \infty$, então*

$$E\left(XZ\big|Y\right) = Z.E\left(X\big|Y\right) \text{ quase certamente.}$$

Demonstração. Omitida. Envolve Teoria da Medida. □

Exemplo 9.33. O Jogador I lança uma moeda honesta n vezes, obtendo k "caras", onde $0 \leqslant K \leqslant n$. Depois o Jogador II lança a moeda k vezes, obtendo j "coroas". Seja X o número j de "coroas" obtidas pelo Jogador II. Queremos calcular EX. (Poderíamos fazer algum esforço neste caso – nem sempre isso é possível – para mostrar que $X \sim b(n, \frac{1}{4})$ e portanto $EX = \frac{n}{4}$, mas estamos interessados apenas em saber EX.)

Seja Y o número de "caras" obtidas pelo Jogador I. É claro que $X|Y = k \sim b(k, \frac{1}{2})$, logo $E(X|Y = k) = \frac{k}{2}$. Assim, $E(X|Y) = \frac{Y}{2}$. Calculamos então

$$EX = E\left[E(X|Y)\right] = E\left[\frac{Y}{2}\right] = \frac{1}{2}EY = \frac{1}{2}\frac{n}{2} = \frac{n}{4},$$

uma vez que $Y \sim b(n, \frac{1}{2})$.

Exemplo 9.34. No Exemplo 9.24, vamos cacular $E\left[X|Y\right]$ e $E\left[X\right]$.

Substituindo a densidade obtida temos

$$E[X|Y = y] = \int_{-\infty}^{+\infty} x f_X(x \mid Y = y)\mathrm{d}x = \int_0^1 \frac{6x^2(2 - x - y)}{4 - 3y}\mathrm{d}x = \frac{5 - 4y}{8 - 6y}.$$

Então $E[X|Y] = \frac{5-4Y}{8-6Y}$ e

$$E[X] = E\big[E(X|Y)\big] = \int_0^1 \frac{5 - 4y}{8 - 6y}(4y - 3y^2)dy = \frac{15}{12} - \frac{8}{12} = \frac{7}{12}.$$

Exemplo 9.35. No Exemplo 9.25, vamos calcular $E\left[e^{X/2}\middle|Y\right]$ e $E\left[e^{X/2}\middle|Y=1\right]$. Substituindo a densidade condicional obtida, temos

$$E\left[e^{\frac{X}{2}}\middle|Y=y\right] = \int_0^\infty e^{\frac{x}{2}} y e^{xy} \mathrm{d}x = y\int_0^\infty e^{(\frac{1}{2}-y)x}\mathrm{d}x.$$

Se $y \leqslant \frac{1}{2}$ a integral vale $+\infty$. Se $y > \frac{1}{2}$, a integral vale $\frac{y}{y-\frac{1}{2}}$. Assim,

$$E\left[e^{X/2}\middle|Y\right] = \begin{cases} +\infty, & Y \leqslant \frac{1}{2}, \\ \frac{y}{y-\frac{1}{2}}, & y > \frac{1}{2}, \end{cases}$$

e $E\left[e^{X/2}\middle|Y=1\right] = \frac{1}{2}$.

Exemplo 9.36. Seja $X \sim U[0,1]$. Se $X = x$, então uma moeda com probabilidade x de sair cara é lançada n vezes independentemente. Seja Y a variável aleatória que representa o número de caras obtidas.

Temos que $Y|X = x \sim b(n,x)$ e $X \sim U(0,1)$ Se $y \in 0,1,\dots,n$ então:

$$P(Y=y) = \int_0^1 P(Y=y \mid X=x) f_X(x)\mathrm{d}x = \int_0^1 \tbinom{n}{y} x^y (1-x)^{n-y}\mathrm{d}x.$$

Portanto

$$\begin{aligned} E[Y] &= \sum_{y=0}^n yP(Y=y) = \sum_{y=0}^n \int_0^1 y\tbinom{n}{y} x^y(1-x)^{n-y}\mathrm{d}x \\ &= \int_0^1 xn\sum_{y=0}^n \tbinom{n-1}{y-1} x^{y-1}(1-x)^{n-y}\mathrm{d}x \\ &= \int_0^1 xn(x+1-x)^{n-1}\mathrm{d}x = n\int_0^1 x\mathrm{d}x = \frac{n}{2}. \end{aligned}$$

Por outro lado, $E[Y \mid X = x] = nx$, ou seja, $E[Y \mid X] = nX$, logo

$$E\big[E(Y|X)\big] = E[nX] = \frac{n}{2}.$$

Exercício 9.3. Sejam X e Y variáveis aleatórias independentes tais que $X \sim U[0,2]$ e $Y \sim U[-1,1]$.

(a) Calcule $E[X|X+Y \leqslant 2]$.

(b) Calcule $E[X|X+Y]$.

(c) Calcule $E[X|X+Y=2]$.

Exercício 9.4. Seja $X_1, X_2, \ldots$.uma sequência de variáveis aleatórias independentes e identicamente distribuídas e seja N uma variável aleatória inteira e não-negativa independente da sequência $X_1, X_2, \ldots$. Seja $Y = \sum_{i=1}^{N} X_i$. Mostre que

$$E[Y] = E[N]\,E[X]\,.$$

Exercício 9.5. Sejam $Y_1, Y_2, \ldots, Y_n$ variáveis aleatórias não-negativas i.i.d. Mostre que

$$E[Y_1 + Y_2 + \cdots + Y_k | Y_1 + Y_2 + \cdots + Y_n = y] = \frac{k}{n} y, \;\; k = 1, 2, \ldots, n.$$

Exercício 9.6. Um número não-negativo X é escolhido com densidade $f_X(x) = xe^{-x}$ para $x > 0$. Se $X = x$, um número Y é escolhido no intervalo $[0, x]$. Ache $P(X + Y \leqslant 2)$.

9.4 Exercícios

9.7. Considere X e Y i.i.d. Bernoulli(p). Calcule $E(X+Y|Y)$ e escreva essa variável aleatória como uma função da variável aleatória Y, de duas formas diferentes:

(a) usando $P(X + Y = k|Y)$ e aplicando a definição de esperança condicional dada uma partição.

(b) usando a linearidade da esperança condicional, a independência entre X e Y e o fato de que Y é $\mathcal{D}_Y$-mensurável.

9.8. Sejam X e Y variáveis aleatórias simples e i.i.d. Mostre que

$$E(X|X+Y) = E(Y|X+Y) = \frac{X+Y}{2}.$$

9.9. Seja X uma variável aleatória simples definida em $(\Omega, \mathcal{F}, P)$ e $\mathcal{D}$ uma partição de $(\Omega, \mathcal{F})$. A variância condicionada a uma partição é definida de forma análoga à variância de uma variável aleatória:

$$V(X|\mathcal{D}) = E\left\{[X - E(X|\mathcal{D})]^2 \middle| \mathcal{D}\right\}.$$

Mostre que

$$V(X|\mathcal{D}) = E\left(X^2 \middle| \mathcal{D}\right) - [E(X|\mathcal{D})]^2$$

e que

$$VX = E[V(X|\mathcal{D})] + V[E(X|\mathcal{D})].$$

9.10. Sejam X e Y variáveis aleatórias simples definidas em $(\Omega, \mathcal{F}, P)$ e $\mathcal{D}$ uma partição. Mostre que

$$E[\,X\,E(Y|\mathcal{D})\,] = E[\,Y\,E(X|\mathcal{D})\,].$$

9.11. Sejam X e Y variáveis aleatórias simples definidas em $(\Omega, \mathcal{F}, P)$ e $\mathcal{D}$ uma partição. Se

$$E\left(Y^2 \,\middle|\, \mathcal{D}\right) = X^2 \quad \text{e} \quad E(Y|\mathcal{D}) = X,$$

mostre que $P(X = Y) = 1$. Dica: desenvolva $E\left[(X - Y)^2\right]$.

9.12. Joga-se um dado, depois uma moeda, depois o dado novamente e segue-se alternando entre o dado e a moeda. Quando se obtém cara na moeda, o jogo é imediatamente interrompido e conta-se o total Z de pontos obtidos nos lançamentos do dado. Calcule EZ.

9.13. Seja $X \sim \exp(\lambda)$ e $Y = \min\{X, c\}$, onde $c > 0$ é uma constante. Encontre $E(X|Y)$.

9.14. [Jam04, Capítulo 4]. Recomendados: 1, 9, 15, 16b, 32, 40.

Capítulo 10

Convergência da Esperança

Uma questão fundamental é sobre o limite de EX_n para uma sequência qualquer $X_1, X_2, X_3, \ldots$ de variáveis aleatórias. Mas precisamente, a questão é quando podemos comutar a esperança com o limite em n, ou seja, se $E[\lim_n X_n] = \lim_n EX_n$. Vale lembrar que derivadas e integrais também são limites.

Pensemos a região abaixo do gráfico de uma função real não-negativa como tendo uma "área", "volume" ou "massa". Se a função é dada por $f(x) = n \cdot \mathbb{1}_{(0,\frac{1}{n}]}(x)$, ou por $g(x) = \mathbb{1}_{(n,n+1]}(x)$, essa massa é sempre igual a 1 e, no entanto, desaparece quando tomamos o limite em n. Podemos dizer que a massa "escapou ao infinito", no primeiro exemplo o fez verticalmente e, no segundo, horizontalmente. As três propriedades estudadas neste capítulo explicam o que pode acontecer com a massa no limite.

10.1 Teoremas de Convergência

O Teorema da Convergência Monótona diz que nada de estranho pode acontecer com a massa de uma seqüência crescente de funções, mais precisamente que não se pode ganhar massa.

Teorema 10.1 (Teorema da Convergência Monótona)**.** *Seja X_n uma sequência não-negativa não-decrescente de variáveis aleatórias estendidas, e defina $X = \lim_n X_n$. Então $EX_n \to EX$ quando $n \to \infty$.*

Demonstração. Por monotonicidade, temos que EX_n converge e $\lim_n EX_n \leqslant EX$, faltando demonstrar a desigualdade oposta. Seja $0 \leqslant Z \leqslant X$ simples, e seja $\alpha < 1$. Tomando $A_k = [X_k \geqslant \alpha Z]$, temos que

$$\begin{aligned} EX_n \geqslant E[X_n \mathbb{1}_{A_n}] \geqslant E[\alpha Z \mathbb{1}_{A_n}] = \\ = \alpha \sum_z z \cdot P\big([Z = z] \cap A_n\big) \to \alpha \sum_z z \cdot P(Z = z) = \alpha EZ. \end{aligned}$$

A soma é sobre finitos z's porque Z é simples, e para cada z o limite justifica-se pois, como $X_n \nearrow X \geqslant Z$, temos que $A_k \nearrow \Omega$. Portanto, $\lim_n EX_n \geqslant \alpha EZ$. Tomando o supremo em α e depois em Z, concluímos a prova do teorema. □

O Lema de Fatou diz que, embora os exemplos acima mostrem que é possível perder massa no limite, não é possível ganhar massa.

Teorema 10.2 (Lema de Fatou). *Seja $(X_n)_n$ uma sequência de variáveis aleatórias estendidas não-negativas. Então $E[\liminf_n X_n] \leqslant \liminf_n EX_n$.*

Demonstração. Tomando $Y_n = \inf_{k \geqslant n} X_k$ e definindo $Y = \liminf_n X_n$, temos que $0 \leqslant Y_n \nearrow Y$. Pelo Teorema da Convergência Monótona, $\lim_n EY_n = EY$. Como $Y_n \leqslant X_n$, temos $\liminf_n EX_n \geqslant \liminf_n EY_n = EY = E[\liminf_n X_n]$. □

O Teorema da Convergência Dominada diz que, se os gráficos estão todos confinados a uma região de massa finita, tampouco se pode perder massa no limite. A razão é que o gráfico de X_n divide essa região de massa total finita em duas partes, e o fato de que cada uma dessas partes não pode ganhar massa no limite implica que a outra não pode perder.

Teorema 10.3 (Teorema da Convergência Dominada). *Se $|X_n| \leqslant Y$ para todo n com Y integrável, e $X_n \xrightarrow{P} X$, então $EX_n \to EX$ quando $n \to \infty$.*

Demonstração. Primeiro supomos que $X_n \to X$ q.c. Como $X_n \leqslant Y$, temos $Y - X_n \geqslant 0$, e portanto podemos aplicar Fatou, obtendo

$$\liminf_n E[Y - X_n] \geqslant E[\liminf_n Y - X_n] = E[Y - X]$$

Como Y é integrável, vale $E[Y - X_n] = EY - EX_n$, $E[Y - X] = EY - EX$ e

$$\limsup_n EX_n \leqslant EX.$$

Aplicando o mesmo argumento com $-X_n$ no lugar de X_n, obtemos a desigualdade oposta $\liminf_n EX_n \geqslant EX$, e portanto $EX_n \to EX$.

Finalmente, suponhamos que $X_n \to X$ em probabilidade. Queremos mostrar que a sequência numérica EX_n converge para o número EX. Isto é equivalente a dizer que qualquer sub-sequência tem uma sub-sub-sequência que converge a EX. Seja X_{n_k} uma subsequência. Como $\lim_n X_n = X$ em probabilidade, temos $\lim_k X_{n_k} = X$ em probabilidade, e podemos tomar $X_{n_{k_j}}$ tal que $\lim_j X_{n_{k_j}} = X$ q.c. Aplicando o caso anterior, temos que $\lim_j EX_{n_{k_j}} = EX$, concluindo a prova. □

10.2 Corolários e Aplicações

Corolário 10.4. *Se $X_n \searrow X \geqslant 0$ q.c. e X_1 é integrável, então $EX_n \to EX$ quando $n \to \infty$.*

Demonstração. Exercício. □

Corolário 10.5. *Seja $(X_n)_n$ uma sequência de variáveis aleatórias não-negativas. Então $E[\sum_n X_n] = \sum_n EX_n$.*

Demonstração. Exercício. □

Corolário 10.6. *Sejam $(X_n)_n$ variáveis aleatórias tais que $\sum_n E|X_n| < \infty$. Então a série $\sum_n X_n$ converge quase-certamente, e $E[\sum_n X_n] = \sum_n EX_n$.*

Demonstração. Exercício. □

Corolário 10.7. *Se $X_n \overset{q.c.}{\to} X$ e exite $M \in \mathbb{R}$ tal que $|X_n| \leqslant M$ q.c. para todo n, então $EX_n \to EX$.*

Demonstração. Tome $Y = M$ e aplique o Teorema da Convegência Dominada. □

Corolário 10.8. *Seja $(X_n)_n$ uma sequência não-decrescente de variáveis aleatórias não-negativas, e tome $X = \lim_n X_n$. Suponha que eixte $M \in \mathbb{R}$ tal que $EX_n < M$ para todo n. Então X é quase-certamente finita.*

Demonstração. Exercício. □

Demonstração da Proposição 5.31. Suponha que $X_n \xrightarrow{P} X$ e $|X_n| \leqslant Y$ q.c., onde $EY^p < \infty$. Então $|X| \leqslant Y$ q.c. e temos

$$|X_n - X|^p \leqslant (|X_n| + |X|)^p \leqslant (2Y)^p,$$

que é integrável. Por outro lado, como $X_n - X \xrightarrow{P} 0$ temos $|X_n - X|^p \xrightarrow{P} 0$. Pelo Teorema da Convergência Dominada, $E|X_n - X|^p \to 0$. □

Demonstração da Proposição 8.7. Vamos mostrar que

$$\frac{\mathrm{d}^k}{\mathrm{d}t^k} M_X(t) = E[X^k e^{tX}]$$

para todo $t \in (-\varepsilon, +\varepsilon)$, por indução em k.

Para $k = 0$ temos a identidade $M_X(t) = Ee^{tX}$ que vale trivialmente. Suponhamos a identidade acima válida para $k \in \mathbb{N}$. Escrevemos $g_x(t) = x^k e^{tx}$ e $g_x'(t) = x^{k+1} e^{tx}$. Temos

$$\frac{\mathrm{d}}{\mathrm{d}t} M_X^{(k)}(t) = \frac{\mathrm{d}}{\mathrm{d}t} E g_X(t) = \lim_{h \to 0} E \frac{g_X(t+h) - g_X(t)}{h}.$$

Se pudermos comutar a esperança e o limite, teremos

$$M_X^{(k+1)}(t) = E \lim_{h \to 0} \frac{g_X(t+h) - g_X(t)}{h} = E g_X'(t) = E[X^{k+1} e^{tX}].$$

Para aplicar o Teorema da Convergência Dominada, tomamos uma sequência qualquer $h_n \to 0$, e basta dominar o termo aleatório $|\frac{g_X(t+h) - g_X(t)}{h}|$ por uma variável aleatória integrável. Ora, pelo Teorema do Valor Médio,

$$\frac{g_x(t+h) - g_x(t)}{h} = g_x'(\theta),$$

onde $\theta \in [t, t+h]$ depende de x, t e h. Tomando $\delta = \frac{\varepsilon - |t|}{3}$, para $|h| < \delta$ temos

$$|g_x'(\theta)| \leqslant |x|^{k+1} e^{(\varepsilon - 2\delta)|x|} \leqslant C e^{(\varepsilon - \delta)|x|} \quad \text{para todo } x \in \mathbb{R},$$

onde C depende de ε e t. Da hipótese segue que $Ee^{(\varepsilon - \delta)|X|} < \infty$, concluindo a demonstração. □

Proposição 10.9. *Se X é integrável e M_X é finita em algum intervalo $[0, +\varepsilon)$, então $EX = M_X'(0)$ como derivada à direita.*

Demonstração. Similar à demonstração da Proposição 8.7, porém separando $g'_x(\theta)$ em suas partes positiva e negativa. Deixamos os detalhes como exercício. $\square$

10.3 Exercícios

10.1. Seja X uma variável aleatória integrável, e seja $(A_n)_n$ uma sequência de eventos tais que $P(A_n) \to 0$. Prove que $E[X\mathbb{1}_{A_n}] \to 0$.

10.2. Prove os corolários da seção anterior.

Capítulo 11

O Passeio Aleatório

Neste capítulo vamos definir o passeio aleatório simples e simétrico em $\mathbb{Z}^d$, e mostrar que ele é recorrente para $d \leqslant 2$ e transiente para $d > 2$.

11.1 Passeio Aleatório, Recorrência e Transiência

$\mathbb{Z}^d$ é o conjunto de vetores em $\mathbb{R}^d$ com coordenadas inteiras. Cada ponto $x \in \mathbb{Z}^d$ tem $2d$ vizinhos, $x \pm e_j$, $j = 1, \dots, d$, onde $e_1 = (1, 0, 0, \dots, 0)$, $e_2 = (0, 1, 0, \dots, 0)$, etc.

Considere uma sequência i.i.d. $(X_n)_n$ em $\mathbb{Z}^d$ com distribuição dada por

$$P(X_n = e_j) = P(X_n = -e_j) = \tfrac{1}{2d}, \quad j = 1, \dots, d.$$

O *passeio aleatório simples e simétrico em $\mathbb{Z}^d$* é definido por

$$S_0 = 0, \quad S_n = S_{n-1} + X_n.$$

Teorema 11.1.

$$P(S_n = 0 \ i.v.) = \begin{cases} 1, & d \leqslant 2, \\ 0, & d > 2. \end{cases}$$

Antes de ver a prova completa, tentaremos entender de onde vem essa diferença entre $d \leqslant 2$ e $d > 2$. Definimos $Z_n = \mathbb{1}_{S_n=0}$ e $R = \sum_{n=1}^{\infty} Z_n$, que conta o número de retornos à origem. Então

$$ER = \sum_{n=1}^{\infty} P(S_n = 0).$$

Para $d = 1$, usando a fórmula de Stirling temos, para uma constante positiva c,

$$P(S_n = 0) = \binom{n}{n/2} 2^{-n} \asymp c n^{-1/2}$$

se n é par, e $P(S_n = 0) = 0$ se n é ímpar. Para $d = 2$, nos primeiros n passos, aproximadamente $\frac{n}{2}$ são horizontais e $\frac{n}{2}$ são verticais. Além disso, se sabemos quantos passos foram feitos em cada direção, os sentidos são independentes, e é plausível que

$$P(S_n = 0) \asymp \tfrac{1}{2}\ c^2\ (\tfrac{n}{2})^{-1}$$

para n par (o termo $\frac{1}{2}$ a mais se deve à possibilidade de os números de saltos horizontais e verticais serem ambos pares ou ambos ímpares). Assim, tanto em $d = 1$ como $d = 2$, $ER = \infty$.

Para $d = 3$, nos primeiros n passos, cerca de $\frac{n}{3}$ estão na direção x, $\frac{n}{3}$ estão na direção y e $\frac{n}{3}$ na direção z. Novamente, é plausível que

$$P(S_n = 0 | \text{paridade}) \asymp c^3 (\tfrac{n}{3})^{-3/2}$$

ou 0 dependendo da paridade. Isso, por sua vez, implica que $ER < \infty$, o que certamente implica que $R < \infty$ q.c. O mesmo raciocínio também funciona para $d = 4, 5, 6, \dots$.

Há dois pontos que devemos justificar: para $d \leqslant 2$, por que $ER = \infty$ implica $P(R = \infty) = 1$; para $d > 2$, formalizar a idéia de que $\frac{n}{d}$ passos são feitos em cada direção, enquanto os sentidos dos passos permanecem independentes.

11.2 Prova da Transiência

Vamos considerar $d = 3$. Dimensões maiores são tratadas de forma análoga.

Denotamos por $J_n \in \{1, 2, 3\}$ a direção de X_n, ou seja, $J_n = 1$ se $X_n = \pm e_1$, $J_n = 2$

se $X_n = \pm e_2$, e $J_n = 3$ se $X_n = \pm e_3$, . Tome $Y_n = +1$ ou -1 de acordo com se $X_n = +e_{J_n}$ ou $-e_{J_n}$. Então Y_n é independente de J_n, e eles são distribuídos como variáveis uniformes discretas em $\{-1, +1\}$ e $\{1, 2, 3\}$, respectivamente.

Fixemos $n \in \mathbb{N}$, e denotemos por $N_1 = \sum_{k=1}^{n} \mathbb{1}_{J_k=1}$ o número de passos dados na direção x. Analogamente para $N_2 = \sum_{k=1}^{n} \mathbb{1}_{J_k=2}$ e $N_3 = \sum_{k=1}^{n} \mathbb{1}_{J_k=3}$.

Dado que $N_1 = n_1$, $N_2 = n_2$ e $N_3 = n_3$, a probabilidade condicional de $S_n = 0$ é dada por

$$P\big(S_n = 0 \big| (N_1, N_2, N_3) = (n_1, n_2, n_3)\big) = \tbinom{n_1}{n_1/2} 2^{-n_1} \tbinom{n_2}{n_2/2} 2^{-n_2} \tbinom{n_3}{n_3/2} 2^{-n_3} \asymp \tfrac{c^3}{\sqrt{n_1 n_2 n_3}}$$

se n_1, n_2, n_3 são todos pares, e 0 se algum deles for ímpar.

Portanto,

$$\begin{aligned}
P(S_n = 0) &\asymp \sum_{\substack{n_1, n_2, n_3 \\ \text{todos pares}}} \frac{c^3}{\sqrt{n_1 n_2 n_3}} \cdot P\big((N_1, N_2, N_3) = (n_1, n_2, n_3)\big) \\
&\leqslant \sum_{\substack{n_1, n_2, n_3 \\ \text{todos maiores que } \frac{n}{4}}} \frac{c^3}{\sqrt{n_1 n_2 n_3}} \cdot P\big((N_1, N_2, N_3) = (n_1, n_2, n_3)\big) + \\
&\quad + \sum_{\substack{n_1, n_2, n_3 \\ \text{algum menor que } \frac{n}{4}}} \frac{c^3}{\sqrt{n_1 n_2 n_3}} \cdot P\big((N_1, N_2, N_3) = (n_1, n_2, n_3)\big) \\
&\leqslant \frac{c^3}{(\frac{n}{4})^{3/2}} \sum_{\substack{n_1, n_2, n_3 \\ \text{todos maiores que } \frac{n}{4}}} P\big((N_1, N_2, N_3) = (n_1, n_2, n_3)\big) + \\
&\quad + \sum_{\substack{n_1, n_2, n_3 \\ \text{algum menor que } \frac{n}{4}}} P\big((N_1, N_2, N_3) = (n_1, n_2, n_3)\big) \\
&= \frac{8c^3}{n^{3/2}} P(N_1, N_2, N_3 \text{ são maiores que } \tfrac{n}{4}) + \\
&\qquad + P(\text{algum } N_1, N_2, N_3 \text{ é menor que } \tfrac{n}{4}) \\
&\leqslant \frac{8c^3}{n^{3/2}} + 3 \cdot \frac{E(N_1 - \frac{n}{3})^4}{(\frac{n}{12})^4} \\
&\leqslant (8c^3) n^{-3/2} + (3 \cdot 3 \cdot 12^4) n^{-2}.
\end{aligned}$$

Na última desigualdade usamos a mesma estimativa que na prova da Lei dos

Grandes Números de Cantelli, segundo a qual o quarto momento centrado de uma soma i.i.d. é menor que $3n^2$ vezes o quarto momento centrado de cada variável.

A partir da estimativa acima, vemos que $P(S_n = 0)$ é somável sobre n, logo R é integrável e, portanto, finito quase certamente. Isso conclui a prova de que o passeio aleatório simples e simétrico em dimensão 3 é q.c. transiente.

11.3 Prova da Recorrência

Lema 11.2. *Para todo $k \in \mathbb{N}$, $P(R \geqslant k) = P(R \geqslant 1)^k$.*

Demonstração. Quando ocorre o evento "$R \geqslant k$", definimos $T_1 < \cdots < T_k$ como os primeiros k instantes em que o passeio aleatório retorna, isto é,

$$S_0 = S_{T_1} = \cdots = S_{T_k} = 0, \quad S_n \neq 0 \text{ para } T_j < n < T_{j+1}.$$

Então

$$P(R \geqslant k) = \sum_{t_1 < t_2 < \cdots < t_k} P(T_1 = t_1, T_2 = t_2, \ldots, T_k = t_k).$$

Observe que todo caminho possível até o momento t_k tem a mesma probabilidade $\frac{1}{(2d)^{t_k}}$. Logo, a probabilidade acima é dada por

$$\frac{\#\{(x_1, \ldots, x_{t_k}) : s_{t_1} = \cdots = s_{t_k} = 0, s_n \neq 0 \text{ para } t_j < n < t_{j+1}\}}{(2d)^{t_k}} = \frac{\#A_{t_1,\ldots,t_k}}{(2d)^{t_k}},$$

onde, dada uma sequência $x_1, \ldots, x_{t_k}$, denotamos o caminho correspondente por $s_n = x_1 + \cdots + x_n$.

A principal observação é que as sequências no conjunto acima podem ser obtidas pela concatenação de subsequências correspondentes a cada tempo de retorno. Mais precisamente, escrevendo

$$A_t = \big\{(x_1, \ldots, x_t) : s_t = 0, s_n \neq 0 \text{ para } 0 < n < t\big\},$$

temos

$$\#A_{t_1,\ldots,t_k} = \#A_{t_1} \times \#A_{t_2 - t_1} \times \#A_{t_3 - t_2} \cdots \times \#A_{t_k - t_{k-1}}.$$

Para concluir, escrevemos

$$
\begin{aligned}
P(R \geqslant k) &= \sum_{r_1, r_2, \ldots, r_k > 0} P(T_1 = r_1, T_2 = r_1 + r_2, \ldots, T_k = r_1 + r_2 + \cdots + r_k) \\
&= \sum_{r_1, r_2, \ldots, r_k > 0} \frac{\# A_{r_1, r_1 + r_2, \ldots, r_1 + r_2 + \cdots + r_k}}{(2d)^{r_1 + r_2 + \cdots + r_k}} \\
&= \sum_{r_1, r_2, \ldots, r_k > 0} \frac{\# A_{r_1}}{(2d)^{r_1}} \cdot \frac{\# A_{r_2}}{(2d)^{r_2}} \cdots \frac{\# A_{r_k}}{(2d)^{r_k}} \\
&= \sum_{r_1 > 0} \frac{\# A_{r_1}}{(2d)^{r_1}} \cdot \sum_{r_2 > 0} \frac{\# A_{r_2}}{(2d)^{r_2}} \times \cdots \times \sum_{r_k > 0} \frac{\# A_{r_k}}{(2d)^{r_k}} \\
&= P(R \geqslant 1)^k.
\end{aligned}
$$

□

Corolário 11.3. $P(R = \infty) = \begin{cases} 1, & \sum_n P(S_n = 0) = \infty, \\ 0, & \sum_n P(S_n = 0) < \infty. \end{cases}$

Demonstração. Seja $\rho = P(R \geqslant 1)$. Existem apenas duas possibilidades:

Se $\rho = 1$, então $P(R \geqslant k) = 1$ para cada k, logo $P(R = \infty) = 1$, e $ER = \infty$.

Se $\rho < 1$, então $P(R \geqslant k) = \rho^k \to 0$ quando $k \to \infty$, logo $P(R = \infty) = 0$ e $ER = \sum_{n=1}^{\infty} P(R \geqslant n) = \sum_{n=1}^{\infty} \rho^n = \frac{\rho}{1-\rho} < \infty$.

Recordando que $ER = \sum_n P(S_n = 0)$, isso prova o corolário. □

Provaremos agora a recorrência para dimensões $d = 1$ e $d = 2$ mostrando que $\sum_n P(S_n = 0) = \infty$. Para $d = 1$, $P(S_n = 0) \asymp c n^{-1/2}$ para n par, que não é somável, concluindo a prova.

Para $d = 2$, usamos a definição de N_1 e N_2 do início da seção anterior. Afirmamos que para n par, $P(N_1, N_2 \text{ ambos pares}) = \frac{1}{2}$. Com efeito, considere a paridade dessas mesmas contagens no passo $n-2$: se ambas forem ímpares, então um passo em cada direção as fará pares no passo n; se ambas forem pares, então dois passos na mesma direção as manterão pares no passo n. Esses eventos ocorrem com a probabilidade $\frac{1}{2}$ cada, provando assim a afirmação.

Portanto, para n par, temos

$$
\begin{aligned}
P(S_n = 0) &\asymp \sum_{\substack{n_1, n_2 \\ \text{ambos pares}}} \frac{c^2}{\sqrt{n_1 n_2}} \cdot P\big((N_1, N_2) = (n_1, n_2)\big) \\
&\geqslant \frac{c^2}{\sqrt{\frac{n}{2}\frac{n}{2}}} \sum_{\substack{n_1, n_2 \\ \text{ambos pares}}} P\big((N_1, N_2) = (n_1, n_2)\big) \\
&= \frac{c^2}{\sqrt{\frac{n}{2}\frac{n}{2}}} \, P(N_1, N_2 \text{ são ambas pares}) \\
&= c^2 \, n^{-1},
\end{aligned}
$$

que não é somável, concluindo a prova da recorrência para $d = 2$.

Capítulo 12

Grandes Desvios

Sejam $(X_n)_n$ variáveis aleatórias i.i.d. e $S_n = X_1 + \cdots + X_n$. Neste capítulo vamos enunciar e provar a Desigualdade de Concentração de Chernoff para $\frac{S_n}{n}$ em torno μ. Vamos enunciar o Princípio dos Grande Desvios de Cramér, e prová-lo sob certas hipóteses.

Notação. O termo $o(b_n)$ denota uma função $g(n)$ satisfazendo $\frac{g(n)}{b_n} \to 0$. Cada vez que aparece, denota uma função diferente.

12.1 Desigualdade de Concentração

Seja X uma variável aleatória integrável com média μ. Sejam $(X_n)_n$ independentes e com a mesma distribuição de X, e tome $S_n = X_1 + \cdots + X_n$.

A Lei dos Grandes Números foi provada da seguinte forma: dado $a > \mu$,

$$P\left(\tfrac{S_n}{n} \geqslant a\right) \leqslant P\left[(\tfrac{S_n}{n} - \mu)^2 \geqslant (a-\mu)^2\right] \leqslant \tfrac{1}{(a-\mu)^2} E\left(\tfrac{S_n}{n} - \mu\right)^2 = \tfrac{VX}{(a-\mu)^2} \cdot \frac{1}{n}$$

A desigualdade acima diz que, quando $EX^2 < \infty$, a probabilidade de $\frac{S_n}{n}$ diferir de μ por mais que uma quantidade fixa $a - \mu$ decai pelo menos tão rápido quanto $\frac{1}{n}$. Na prova da Lei dos Grandes Números de Cantelli, vimos que, quando $EX^4 < \infty$, esta probabilidade decai pelo menos tão rápido quanto $\frac{1}{n^2}$. Em geral, se $EX^{2k} < \infty$

ela decai pelo menos tão rápido quanto $\frac{1}{n^k}$.

Tentaremos agora obter estimativas melhores usando momentos de e^{tX} ao invés de X^{2k}. Para $t \geqslant 0$,

$$P\left(\tfrac{S_n}{n} \geqslant a\right) \leqslant P\left[e^{tS_n} \geqslant e^{atn}\right] \leqslant \tfrac{1}{e^{atn}} E e^{tS_n} = e^{-atn} M(t)^n = e^{-[at - \log M(t)]n},$$

onde $M(t) = Ee^{tX}$ é a função geradora de momentos de X. Da mesma forma, para $a < \mu$ e $t \leqslant 0$,

$$P\left(\tfrac{S_n}{n} \leqslant a\right) \leqslant e^{-[at - \log M(t)]n}. \tag{12.1}$$

Portanto, se mostrarmos que a expressão entre colchetes é positiva para algum t, teremos estabelecido que essa probabilidade de fato decai pelo menos exponencialmente rápido.

Sabemos que a função geradora de momentos é finita em um intervalo que contém o ponto $t = 0$. Denotaremos os extremos desse intervalo por $\mathcal{D}_M^- \leqslant 0$ e $\mathcal{D}_M^+ \geqslant 0$.

Teorema 12.1 (Desigualdade de Concentração de Chernoff)**.** *Se $\mathcal{D}_M^+ > 0$, então para qualquer $a > \mu$ existe $t > 0$ tal que $[at - \log M(t)] > 0$. Como*

$$P\left(\tfrac{S_n}{n} \geqslant a\right) \leqslant e^{-[at - \log M(t)]n},$$

segue que essa probabilidade decai pelo menos exponencialmente rápido em n. Analogamente, se $\mathcal{D}_M^- < 0$ e $a < \mu$ então $[at - \log M(t)] > 0$ para algum $t < 0$ e a estimativa em (12.1) decai exponencialmente rápido.

Demonstração. Suponha que $\mathcal{D}_M^+ > 0$ e seja $a > \mu$. Pela Proposição 10.9, podemos tomar a derivada lateral pela direita, obtendo

$$\frac{\mathrm{d}}{\mathrm{d}t}\Big[at - \log M(t)\Big]\Big|_{t=0} = a - \frac{M'(0)}{M(0)} = a - \mu > 0,$$

de forma que para t positivo e pequeno a expressão acima é positiva.

Da mesma forma, suponha que $\mathcal{D}_M^- < 0$ e seja $a < \mu$. Tomando a derivada lateral pela esquerda, vemos que o termo entre colchetes será positivo para t negativo e pequeno. $\square$

12.2 Princípio dos Grandes Desvios

Começamos com o conceito fundamental de função taxa.

Definição 12.2 (função taxa)**.** Seja X uma variável aleatória. Definimos a *função taxa* I associada à distribuição de X, por

$$I(a) = \sup_{t\in\mathbb{R}} \big[at - \log M(t)\big] .$$

Podemos pensar na função taxa como uma tentativa de obter a melhor estimativa possível a partir de (12.1). A razão pela qual a função I merece esse nome é que, uma vez que maximizamos $[at - \log M(t)]$ sobre todo t, a desigualdade (12.1) deixa de ser apenas mais uma cota superior, sendo de fato a melhor cota superior possível. O próximo teorema torna esta afirmação precisa. Dado $A \subseteq \mathbb{R}$, para descrever a maneira mais fácil (ou menos difícil) de $\frac{S_n}{n}$ estar em A, vamos denotar $I(A) = \inf\limits_{a\in A} I(a)$.

Teorema 12.3 (Princípio dos Grandes Desvios de Cramér)**.** *Seja J um intervalo, e denote por J° e J^* o intervalos aberto e fechado correspondentes a J. Então*

$$e^{-I(J^\circ)\cdot n+o(n)} \leqslant P\left(\tfrac{S_n}{n} \in J\right) \leqslant e^{-I(J^*)\cdot n+o(n)}.$$

Em particular, quando $I(J^\circ) = I(J^*)$, temos a taxa exata de decaimento exponencial para estas probabilidades:

$$P\left(\tfrac{S_n}{n} \in J\right) = e^{-I(J)\cdot n+o(n)}.$$

Antes de provar o teorema acima, vamos discutir a relação entre M e I, e sua interpretação geométrica.

Proposição 12.4. *As funções I e $\log M$ são convexas.*

Não vamos usar essa proposição, e daremos a demonstração na página 178.

Vejamos como encontrar $I(a)$ graficamente e algebricamente. No caso de o supremo na definição de $I(a)$ ser atingido em algum $y \in \mathbb{R}$, temos

$$0 = \tfrac{\mathrm{d}}{\mathrm{d}y} \big[ay - \log M(y)\big] = a - \frac{M'(y)}{M(y)},$$

assim

$$a = \tfrac{\mathrm{d}}{\mathrm{d}y} \log M(y) = \tfrac{M'(y)}{M(y)},$$

e resolvendo y em termos de a às vezes é possível calcular I explicitamente por

$$I(a) = a \cdot y - \log M(y), \quad y = y(a).$$

Esse processo de encontrar y tal que $(\log M)'(y) = a$ e expressar $I(a) = ay - \log M(y)$ está ilustrado na Figura 12.1 e nos exemplos abaixo.

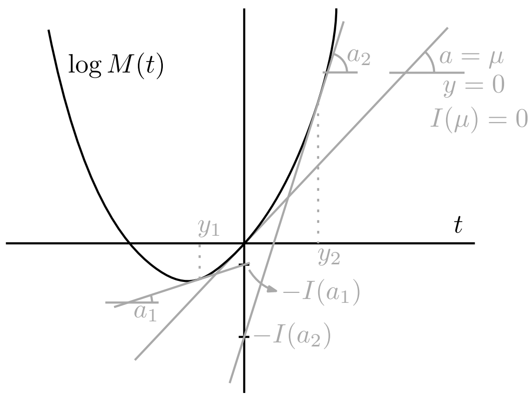

Figura 12.1: Obtenção de $I(a)$ para distintos valores de a a partir da função $\log M$.

Se $X \sim \mathcal{N}(\mu, 1)$, temos

$$\log M(t) = \frac{t^2}{2} + t\mu,$$

assim

$$a = [\log M(y)]' = y + \mu, \qquad y = a - \mu,$$

e

$$I(a) = a(a-\mu) - \left[\tfrac{(a-\mu)^2}{2} + \mu(a-\mu)\right] = \frac{(a-\mu)^2}{2}, \quad a \in \mathbb{R}.$$

Se $X \sim \text{Poisson}(\lambda)$, temos

$$\log M(t) = \lambda(e^t - 1),$$

assim

$$a = [\log M(y)]' = \lambda e^y, \qquad y = \log \tfrac{a}{\lambda},$$

e

$$I(a) = ay - \log M(y) = a \log \tfrac{a}{\lambda} - a + \lambda.$$

De fato,

$$I(a) = \begin{cases} a \log \frac{a}{\lambda} - a + \lambda, & a \geqslant 0, \\ +\infty & a < 0. \end{cases}$$

Se $X \sim \exp(1)$, temos

$$M(t) = \frac{1}{1-t}, \quad t < 1,$$

assim

$$a = \frac{M'(y)}{M(y)} = \frac{(1-y)^{-2}}{(1-y)^{-1}} = \frac{1}{1-y}, \qquad y = 1 - \frac{1}{a},$$

e

$$I(a) = ay - \log M(y) = a(1 - \tfrac{1}{a}) \log M(y) = a - 1 - \log a.$$

De fato,

$$I(a) = \begin{cases} a - 1 - \log a, & a > 0, \\ +\infty & a \leqslant 0. \end{cases}$$

Se $X \sim \text{Bernoulli}(p)$, temos

$$M(t) = pe^t + 1 - p,$$

assim

$$a = \frac{M'(y)}{M(y)} = \frac{pe^y}{pe^y + 1 - p}, \qquad y = \log(\tfrac{a}{p} \cdot \tfrac{1-p}{1-a}),$$

e

$$I(a) = ay - \log M(y) = \cdots = a \log \tfrac{a}{p} + (1-a) \log \tfrac{1-a}{1-p}.$$

De fato,

$$I(a) = \begin{cases} a \log \frac{a}{p} + (1-a) \log \frac{1-a}{1-p}, & 0 < a < 1, \\ \log \frac{1}{p}, & a = 1, \\ \log \frac{1}{1-p}, & a = 0, \\ +\infty & a < 0 \text{ ou } a > 1. \end{cases}$$

Se $X = \mu$ é constante, temos

$$\log M(t) = \mu t,$$

assim

$$a = [\log M(t)]' = \mu, \qquad y \text{ pode ser qualquer número},$$

e

$$I(a) = ay - \log M(y) = 0.$$

De fato,

$$I(a) = \begin{cases} 0, & a = \mu \\ +\infty & a \neq \mu. \end{cases}$$

12.3 Prova da Cota Inferior

Teorema 12.5. *Para qualquer $a \in \mathbb{R}$ e $\delta > 0$,*

$$P\left(\tfrac{S_n}{n} \in [a-\delta, a+\delta]\right) \geqslant e^{-I(a)\cdot n + o(n)}.$$

O teorema vale tal como enunciado, sem suposições adicionais sobre a distribuição de X. No entanto, vamos supor que o supremo em $I(a)$ é atingido, ou seja,

$$I(a) = a \cdot y - \log M(y)$$

para algum $y \in (\mathcal{D}_M^-, \mathcal{D}_M^+)$. Abandonar essa hipótese requer passos técnicos e complicados que não nos interessam.

Demonstração. Como o supremo é atingido no interior de $(\mathcal{D}_M^-, \mathcal{D}_M^+)$, temos

$$0 = \frac{\mathrm{d}}{\mathrm{d}t}\Big[at - \log M(t)\Big]\Big|_{t=y} = a - \frac{M'(y)}{M(y)},$$

e portanto

$$\frac{E[Xe^{yX}]}{E[e^{yX}]} = a.$$

A principal observação é que a expressão do lado esquerdo corresponde à esperança de uma variável aleatória Y cuja distribuição é obtida a partir da distribuição de X, distorcida por um fator da forma $f(x) = e^{yx}$, $x \in \mathbb{R}$. Ou seja, para uma variável aleatória Y cuja distribuição é dada por

$$P(Y \in B) = \frac{E[\mathbb{1}_B(X)e^{yX}]}{E[e^{yX}]} = E\left[\mathbb{1}_B(X)\frac{e^{yX}}{M(y)}\right],$$

temos $EY = a$. Portanto, para $Y_1, Y_2, \dots$ i.i.d. distribuídos como esta versão distorcida de X, a ocorrência de $\frac{Y_1+\dots+Y_n}{n} \approx a$ não é um evento raro.

A prova então consiste em controlar a razão de verossimilhança entre $(X_1, \dots, X_n)$ e $(Y_1, \dots, Y_n)$ em um subconjunto de $\mathbb{R}^n$ que é típico para este último vetor, de forma tal que tal razão não fique menor que $e^{-I(a)\cdot n - o(n)}$.

Fixe $\varepsilon \in (0, \delta]$, e defina o conjunto

$$B_n^\varepsilon = \left\{(z_1, \dots, z_n) : \left|\tfrac{z_1+\dots+z_n}{n} - a\right| \leqslant \varepsilon\right\} \subseteq \mathbb{R}^n.$$

Então

$$\begin{aligned}
P\left(\tfrac{S_n}{n} \in [a-\varepsilon, a+\varepsilon]\right) &= E\left[\mathbb{1}_{B_n^\varepsilon}(X_1, \dots, X_n)\right] \\
&= E\left[\tfrac{M(y)^n}{e^{y(X_1+\dots+X_n)}}\,\mathbb{1}_{B_n^\varepsilon}(X_1, \dots, X_n)\tfrac{e^{yX_1}}{M(y)} \cdots \tfrac{e^{yX_n}}{M(y)}\right] \\
&\geqslant E\left[\frac{M(y)^n}{e^{ayn+|y|\varepsilon n}}\mathbb{1}_{B_n^\varepsilon}(X_1, \dots, X_n)\tfrac{e^{yX_1}}{M(y)} \cdots \tfrac{e^{yX_n}}{M(y)}\right] \\
&= e^{-[ay-\log M(y)-|y|\varepsilon]\cdot n} \cdot P\left((Y_1, \dots, Y_n) \in B_n^\varepsilon\right) \\
&= e^{-[ay-\log M(y)-|y|\varepsilon]\cdot n} \cdot P\left(\tfrac{Y_1+\dots+Y_n}{n} \in [a-\varepsilon, a+\varepsilon]\right).
\end{aligned}$$

Esta última probabilidade converge para 1 pela Lei dos Grandes Números, e portanto

$$P\left(\left|\tfrac{S_n}{n} - a\right| \leqslant \delta\right) \geqslant P\left(\left|\tfrac{S_n}{n} - a\right| \leqslant \varepsilon\right) \geqslant e^{-I(a)\cdot n - 2|y|\varepsilon\cdot n}$$

para todo n suficientemente grande. Como $\varepsilon \in (0, \delta]$ é arbitrário, isso implica

$$P\left(\left|\tfrac{S_n}{n} - a\right| \leqslant \delta\right) \geqslant e^{-I(a)\cdot n + o(n)},$$

completando a prova. □

Demonstração da cota inferior no Teorema 12.3. Seja $\varepsilon > 0$. Tome $a \in J^\circ$ tal que $I(a) \leqslant I(J^\circ) + \varepsilon$. Tome $\delta > 0$ tal que $[a-\delta, a+\delta] \subseteq J$. Então, usando o

Teorema 12.5 temos

$$P\left(\tfrac{S_n}{n} \in J\right) \geqslant P\left(\tfrac{S_n}{n} \in [a-\delta, a+\delta]\right) \geqslant e^{-I(a)\cdot n + o(n)} \geqslant e^{-I(J^\circ)\cdot n - \varepsilon n + o(n)}.$$

Como ε é arbitrário, $P\left(\tfrac{S_n}{n} \in J\right) \geqslant e^{-I(J^\circ)\cdot n + o(n)}$, o que conclui a prova. □

12.4 Prova da Cota Superior

Vejamos como a cota superior no Teorema 12.3 é uma conseqüência direta do Teorema 12.1. Começamos com propriedades de monotonicidade da função taxa.

Proposição 12.6. *A função taxa I é não-crescente em $(-\infty, \mu]$ e não-decrescente em $[\mu, +\infty)$. Além disso, $I(\mu) = 0$,*

$$I(a) = \sup_{t \geqslant 0}[at - \log M(t)] \text{ para } a \geqslant \mu \quad e \quad I(a) = \sup_{t \leqslant 0}[at - \log M(t)] \text{ para } a \leqslant \mu.$$

Demonstração. Tomando $t = 0$ temos $[at - \log M(t)] = 0$, logo $I(a) \geqslant 0$ para todo a. Agora, pela desigualdade de Jensen, $M(t) = Ee^{tX} \geqslant e^{EtX} = e^{t\mu}$, donde

$$\mu t - \log M(t) \leqslant 0.$$

Isso implica que $I(\mu) = 0$. Isso também implica que, para $a > \mu$ e $t < 0$, $at - \log M(t) < 0$, assim $I(a) = \sup_{t \geqslant 0}[at - \log M(t)]$. Analogamente, para $a < \mu$ temos $I(a) = \sup_{t \leqslant 0}[at - \log M(t)]$.

Para provar monotonicidade em $[\mu, +\infty)$, vejamos que, para $a > c > \mu$,

$$I(a) = \sup_{t \geqslant 0}[at - \log M(t)] \geqslant \sup_{t \geqslant 0}[ct - \log M(t)] = I(c) \geqslant 0 = I(\mu).$$

Monotonicidade em $(-\infty, \mu]$ se prova da mesma forma. □

Demonstração da cota superior no Teorema 12.3. Escrevemos $J^* = [c, a] \subseteq \mathbb{R}$. Se $c \leqslant \mu \leqslant a$, $I(J^*) = 0$ e não há nada a provar. Podemos então assumir que $a < \mu$, pois o caso $c > \mu$ é análogo. Seja $\varepsilon > 0$. Pela Proposição 12.6, temos

$$I(J^*) = I(a) = \sup_{t \leqslant 0}[at - \log M(t)],$$

e podemos tomar $t \leqslant 0$ tal que $[at - \log M(t)] \geqslant I(J^*) - \varepsilon$. Agora, usando a estimativa (12.1) obtemos

$$P\left(\tfrac{S_n}{n} \in J\right) \leqslant P\left(\tfrac{S_n}{n} \leqslant a\right) \leqslant e^{-I(J^*)\cdot n + \varepsilon n}.$$

Como ε é arbitrário, $P\left(\tfrac{S_n}{n} \in J\right) \leqslant e^{-I(J^*)\cdot n + o(n)}$, concluindo a demonstração. □

12.5 Convexidade

Esta seção pode ser omitida. Vamos enunciar e provar a Desigualdade de Young, para então enunciar e provar a Desigualdade de Hölder, e finalmente demonstrar a Proposição 12.4.

Sejam $p \geqslant 1$, $q \geqslant 1$ tais que

$$\frac{1}{p} + \frac{1}{q} = 1.$$

Proposição 12.7 (Desigualdade de Young)**.** *Para $a, b \geqslant 0$,*

$$ab \leqslant \frac{a^p}{p} + \frac{b^q}{q}.$$

Demonstração. Primeiro veja que

$$p - 1 = \frac{1}{q-1}.$$

Considere a curva $s = r^{p-1}$, ou seja, $r = s^{q-1}$, no quadrante $\{(r,s) \in [0,\infty)^2\}$. A desigualdade em

$$\frac{a^p}{p} + \frac{b^q}{q} = \int_0^a r^{p-1} \mathrm{d}r + \int_0^b s^{q-1} \mathrm{d}s \geqslant ab$$

vale porque as integrais correspondem a áreas disjuntas do quadrante cuja união contém o retângulo $[0,a] \times [0,b]$. □

Proposição 12.8 (Desigualdade de Hölder)**.** *Se X e Y têm momentos de ordem p e q finitos, respectivamente, então XY é integrável e $E[XY] \leqslant \|X\|_p \cdot \|Y\|_q$.*

Demonstração. Podemos supor que $\|X\|_p > 0$ e $\|Y\|_q > 0$, caso contrário $XY = 0$ q.c. e a desigualdade vale trivialmente. Tomemos

$$\tilde{X} = \frac{|X|}{\|X\|_p} \qquad \text{e} \qquad \tilde{Y} = \frac{|Y|}{\|Y\|_q}$$

(observe que o numerador é uma variável aleatória e o denominador é uma constante). Usando a desigualdade de Young,

$$\frac{E[XY]}{\|X\|_p\|Y\|_q} \leqslant \frac{E|XY|}{\|X\|_p\|Y\|_q} = E[\tilde{X}\tilde{Y}] \leqslant \frac{E\tilde{X}^p}{p} + \frac{E\tilde{X}^q}{q} = \frac{1}{p} + \frac{1}{q} = 1. \qquad \square$$

Demonstração da Proposição 12.4. Começamos com a convexidade de I. Sejam $0 < \alpha < 1$ e $\beta = 1 - \alpha$. Para a_1 e $a_2 \in \mathbb{R}$,

$$\begin{aligned}
I(\alpha a_1 + \beta a_2) &= \sup_{t\in\mathbb{R}} \big[(\alpha a_1 + \beta a_2)t - (\alpha_1 + \alpha_2)M(t)\big] \\
&= \sup_{t\in\mathbb{R}} \big[\alpha(a_1 t - M(t)) + \beta(a_2 t - M(t))\big] \\
&\leqslant \sup_{t\in\mathbb{R}} \big[\alpha(a_1 t - M(t))\big] + \sup_{t\in\mathbb{R}} \big[\beta(a_2 t - M(t))\big] \\
&= \alpha I(a_1) + \beta I(a_2).
\end{aligned}$$

Passamos agora à convexidade de M. Sejam t_1 e $t_2 \in \mathbb{R}$. Usando a desigualdade de Hölder,

$$\begin{aligned}
\log M(\alpha t_1 + \beta t_2) &= \log E\left[e^{\alpha t_1 X} \cdot e^{\beta t_2 X}\right] \\
&\leqslant \log \left\{ \left[E\left(e^{\alpha t_1 X}\right)^{\frac{1}{\alpha}}\right]^{\alpha} \left[E\left(e^{\beta t_2 X}\right)^{\frac{1}{\beta}}\right]^{\beta} \right\} \\
&= \alpha \log E e^{t_1 X} + \beta \log E e^{t_2 X} \\
&= \alpha \log M(t_1) + \beta \log M(t_2). \qquad \square
\end{aligned}$$

Apêndice A

Fórmula de Stirling

Este capítulo é independente dos anteriores, e tem como objetivo demonstrar

Teorema A.1 (Fórmula de Stirling)**.** $n! \asymp n^n\, e^{-n}\, \sqrt{2\pi n}$.

A.1 Obtenção da Fórmula e Demonstração

Para entender como surge essa a fórmula, observe que

$$\log n! = \log 1 + \log 2 + \cdots + \log n = \sum_{k=1}^{n} \log k$$

é uma aproximação superior para

$$\int_0^n \log x \, \mathrm{d}x = n \log n - n = \log(n^n e^{-n}).$$

Com esse argumento de aproximação de soma por integral, pode-se mostrar que

$$\log n! = n \log n - n + r(n) \qquad \text{com } \frac{r(n)}{n} \to 0,$$

que é suficiente em muitas aplicações, mas queremos uma aproximação mais fina. De fato, queremos aproximar assintoticamente $n!$ e não apenas $\log n!$.

Admitindo uma correção polinomial, busquemos aproximar $n!$ por um múltiplo de $n^n e^{-n} n^\alpha$ com $\alpha \in \mathbb{R}$. Tomando

$$c_n = \log\left(\frac{n^n e^{-n} n^\alpha}{n!}\right),$$

temos

$$\begin{aligned} c_{n+1} - c_n = \log\left[(n+1)\left(\tfrac{n+1}{n}\right)^n \tfrac{e^{-n-1}}{e^{-n}} \left(\tfrac{n+1}{n}\right)^\alpha \tfrac{n!}{(n+1)!}\right] = \\ = \left[n\log(1+\tfrac{1}{n}) - 1\right] + \alpha\log(1+\tfrac{1}{n}). \end{aligned}$$

Fazendo a expansão de Taylor de $\log(1+x)$ para $x \in [0,1]$ temos

$$\log(1+x) = x - \frac{x^2}{2} + r(x)$$

onde $r(x)$ é igual a $\frac{2}{(1+\tilde{x})^3}\frac{x^3}{6}$ para algum $\tilde{x} \in [0,x]$ e satisfaz $0 \leqslant r(x) \leqslant \frac{x^3}{3}$.
Continuando o desenvolvimento de $c_{n+1} - c_n$, temos Temos que

$$\begin{aligned} c_{n+1} - c_n &= \left[n\left(\tfrac{1}{n} - \tfrac{1}{2n^2} + r(\tfrac{1}{n})\right) - 1\right] + \alpha\left(\tfrac{1}{n} - \tfrac{1}{2n^2} + r(\tfrac{1}{n})\right) \\ &= \tfrac{\alpha}{n} - \tfrac{1}{2n} + n\,r(\tfrac{1}{n}) - \tfrac{\alpha}{2n^2} + \alpha\,r(\tfrac{1}{n}) \\ &= n\,r(\tfrac{1}{n}) + \tfrac{1}{2}r(\tfrac{1}{n}) - \tfrac{1}{4n^2} \end{aligned}$$

se escolhemos $\alpha = \frac{1}{2}$ para cancelar os termos de ordem $\frac{1}{n}$.
Finalmente, combinando a última identidade e expansão de Taylor temos

$$|c_{n+1} - c_n| \leqslant \tfrac{1}{2n^2},$$

que é somável, logo $c_n \to c$ para algum $c \in \mathbb{R}$. Portanto,

$$\frac{n!}{n^n e^{-n}\sqrt{n}} \to e^{-c} = \sqrt{2\lambda}$$

para algum $\lambda > 0$, ou seja

$$n! \asymp n^n e^{-n}\sqrt{2\lambda n}.$$

Resta mostrar que a constante é dada por $\lambda = \pi$.

A.2 Cálculo da Constante

A fórmula de Stirling foi provada primeiro por De Moivre, e Stirling encontrou o valor da constante. Vamos provar que $\lambda = \pi$ de duas formas diferentes.

Usando a demonstração do teorema de De Moivre A primeira prova supõe que o leitor viu a demonstração do teorema de De Moivre-Laplace na Seção 7.1. Pela desigualdade clássica de Tchebyshev,

$$1 - \tfrac{1}{m^2} \leqslant P\left(-m \leqslant \tfrac{S_n - np}{npq} \leqslant +m\right) \leqslant 1.$$

Agora observe que a demonstração do teorema de De Moivre-Laplace funciona assumindo a fórmula de Stirling com uma constante desconhecida λ no lugar de π. Assim, fazendo $n \to \infty$,

$$1 - \tfrac{1}{m^2} \leqslant \int_{-m}^{m} \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\lambda}}\, \mathrm{d}x \leqslant 1.$$

Fazendo agora $m \to \infty$ obtemos

$$\int_{\mathbb{R}} \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\lambda}}\, \mathrm{d}x = 1,$$

e portanto $\lambda = \pi$.

Usando o produto de Wallis O produto de Wallis é dado por

$$\frac{\pi}{2} = \prod_{n=1}^{\infty} \left(\frac{2n}{2n-1} \cdot \frac{2n}{2n+1}\right) = \lim_{n\to\infty} \frac{2}{1} \cdot \frac{2}{3} \cdot \frac{4}{3} \cdot \frac{4}{5} \cdot \frac{6}{5} \cdot \frac{6}{7} \cdots \frac{2n}{2n-1} \cdot \frac{2n}{2n+1},$$

o que será demonstrado mais abaixo.

Tomando a raiz quadrada e usando que $\frac{2n}{2n+1} \to 1$ obtemos

$$\sqrt{\frac{\pi}{2}} = \lim_{n\to\infty} \frac{2 \cdot 4 \cdot 6 \cdots (2n-2)}{3 \cdot 5 \cdot 7 \cdots (2n-1)} \cdot \sqrt{2n}.$$

Multiplicando pelo numerador chegamos a

$$\begin{aligned}\sqrt{\frac{\pi}{2}} &= \lim_{n\to\infty} \frac{2\cdot 2\cdot 4\cdot 4\cdot 6\cdot 6\cdots(2n-2)\cdot(2n-2)}{2\cdot 3\cdot 4\cdot 5\cdot 6\cdot 7\cdots(2n-2)\cdot(2n-1)}\cdot\frac{2n\cdot 2n}{2n}\cdot\frac{\sqrt{2n}}{2n}\\ &= \lim_{n\to\infty}\frac{2^{2n}\left(1^2\cdot 2^2\cdot 3^2\cdots n^2\right)}{(2n)!}\cdot\frac{1}{\sqrt{2n}} = \lim_{n\to\infty}\frac{2^{2n}(n!)^2}{(2n)!\sqrt{2n}}.\end{aligned}$$

Finalmente, substituindo na fórmula de Stirling chegamos a

$$\sqrt{\frac{\pi}{2}} = \lim_{n\to\infty}\frac{2^{2n}n^{2n}e^{-2n}2\lambda n}{(2n)^{2n}e^{-2n}\sqrt{4\lambda n}\sqrt{2n}} = \sqrt{\frac{\lambda}{2}},$$

e portanto $\lambda = \pi$.

Demonstração do produto de Wallis Daremos a demonstração em forma de exercício. Seja

$$I_n = \int_0^{\pi/2} \operatorname{sen}^n x\,\mathrm{d}x, \qquad n \geqslant 0.$$

a) Mostre que para todo $n \geqslant 2$ vale

$$I_n = \frac{n-1}{n} I_{n-2}.$$

Sugestão: integrando $\operatorname{sen}^n x = \operatorname{sen}^{n-1} x \cdot \operatorname{sen} x$ por partes, mostre que $\int \operatorname{sen}^n x\,\mathrm{d}x = (n-1)\int(\operatorname{sen}^{n-2} x)(\cos^2 x)\,\mathrm{d}x = (n-1)[I_{n-2} - I_n]$.

b) Verifique que para todo $n \geqslant 1$ vale

$$\frac{I_{2n-2}}{I_{2n-1}} = \frac{2n}{2n-1}\cdot\frac{2n}{2n+1}\cdot\frac{I_{2n}}{I_{2n+1}}.$$

c) Verifique que $I_0 = \frac{\pi}{2}$ e $I_1 = 1$.

d) Mostre por indução que para todo $n \geqslant 0$ vale

$$\frac{\pi}{2} = \frac{2}{1}\cdot\frac{2}{3}\cdot\frac{4}{3}\cdot\frac{4}{5}\cdot\frac{6}{5}\cdot\frac{6}{7}\cdots\frac{2n}{2n-1}\cdot\frac{2n}{2n+1}\cdot\frac{I_{2n}}{I_{2n+1}}.$$

e) Mostre que $\frac{2n}{2n+1} = \frac{I_{2n+1}}{I_{2n-1}} \leqslant \frac{I_{2n+1}}{I_{2n}} \leqslant 1$, e portanto $\frac{I_{2n}}{I_{2n+1}} \to 1$.

Lista de Figuras

Lista de Tabelas

Notação

$o(\cdot)$	ordem pequena; qualquer função satisfazendo $\frac{o(w)}{\lvert w\rvert} \to 0$	169
$\mathcal{P}(\Omega)$	Conjunto das partes: $\mathcal{P}(\Omega) = \{A : A \subseteq \Omega\}$	18
Poisson(λ)	Distribuição de Poisson	42
$U[a, b]$	Distribuição uniforme	44
$U_d[I]$	Distribuição uniforme discreta	40
$\boldsymbol{X}$	Vetor aleatório	54
$\boldsymbol{x}$	Um vetor com d coordenadas, $\boldsymbol{x} \in \mathbb{R}^d$	53
$X \sim Y$	X e Y têm a mesma distribuição	36
$\boldsymbol{x} \leqslant \boldsymbol{y}$	Desigualdade de vetores, $x_1 \leqslant y_1, \dots, x_d \leqslant y_d$	54
X, Y	Variáveis aleatórias	33

Índice Remissivo

Referências Bibliográficas

[CA03] K. L. Chung, F. AitSahlia. *Elementary probability theory*. Undergraduate Texts in Mathematics. Springer-Verlag, New York, 4 edn., 2003.

[Jam04] B. R. James. *Probabilidade: Um Curso em Nível Intermediário*. IMPA, Rio de Janeiro, 3 edn., 2004.

[Shi96] A. N. Shiryaev. *Probability*, vol. 95 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2 edn., 1996.