Probabilidade

Notas de Aula - Versão 0.96

Daniel Miranda
Rafael Grisi
UFABC

24 de abril de 2023

"It is an old maxim of mine that when you have excluded the impossible, whatever remains, however improbable, must be the truth."
- Sherlock Holmes

Sumário

Observações

Essas notas correspondem aos apontamentos para o curso de Probabilidade ministradas na graduação e na pós-graduação em 2017, 2018, 2019, 2022 e 2023. Essas notas apesar de terem sido revisadas algumas vezes ainda estão incompletas e não polidas.

Dito de outra forma essas notas podem apresentar erros ou seções incompletas.

Nós ficaríamos muito gratos se nos fossem enviadas sugestões de melhorias ou que nos fossem apontados erros porventura encontrados.

Agradecimentos

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

Notação

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

Alterações

Versão 0.96

- Nova demonstração do TLC usando Stone-Weierstrass e logo o capítulo 11 foi reescrito. ☺

Versão 0.95

- Melhoria substancial nos textos e em algumas demonstrações dos Capítulos 1,2, 3, 4, 6, 7 e 9. ☺

Versão 0.94

- 5 correções.

Versão 0.93

- Mais de 200 pequenas correções.

Versão 0.9

- Inclusão de novas seções. Melhoria no texto dos primeiros 5 capítulos. Várias pequenas correções. Novo visual.

Versão 0.83b

- Pequenas correções nos capítulos 1,2,3, 6, 7 e 10.

demonstrações

Capítulo 1

Introdução

Existem muitos exemplos de fenômenos na natureza para os quais os modelos determinísticos são apropriados. Por exemplo, a lei da gravitação universal descreve com bastante precisão o que acontece com um corpo em queda sob certas condições. O modelo estipula que as condições em que certos fenômenos ocorrem determinam o valor de certas variáveis observáveis: a magnitude da velocidade e da aceleração, a área varrida por um corpo gravitando um segundo durante um certo período de tempo, etc.

No entanto, também existem muitos fenômenos que requerem um modelo matemático diferente para sua descrição. Estes são o que chamaremos de modelos probabilísticos. Suponha que um pedaço de material radioativo esteja emitindo partículas. Com a ajuda de um dispositivo de contagem, podemos registrar o número de tais partículas emitidas durante um intervalo de tempo. É claro que não podemos prever com precisão o número de partículas emitidas, mesmo que soubéssemos a forma exata, massa do objeto em consideração. Em outras situações, como no lançamento de uma moeda, mesmo que uma descrição determinística seja possível na teoria, na prática as equações são altamente sensíveis as condições iniciais e uma previsão de fato não é factível. Assim, parece não haver nenhum modelo determinístico razoável que forneça o número de partículas emitidas como uma função de várias características pertinentes do material de origem ou ou que nos permita prever o lançamento de uma moeda como uma função das condições nas quais a moeda foi lançada.

Apesar disso em ambos os casos conseguimos construir modelos que descrevem o comportamento probabilístico –mais especificamente, a frequência, a lei probabilística– do resultado observável.

Nesse livro vamos tratar desses modelos. Para descrever esses modelos precisamos supor que apesar de não conhecermos a priori qual será o resultado do experimento, ao menos conheçamos qual é o conjunto de todas as possíveis saídas. Então podemos definir o **espaço amostral**, que em geral denotaremos por Ω como o conjunto dos resultados possíveis de um experimento aleatório.

1.1 Exemplos

1. *Para o experimento aleatório de lançar uma moeda podemos definir o espaço amostral como* $\Omega = \{cara, coroa\}$ *ou até mesmo* $\Omega = \{0, 1\}$, *onde* 0 *representa cara e* 1 *representa coroa.*

2. *Se queremos modelar as incertezas associadas ao total de pessoas que entram em uma agência bancária ao longo do dia, podemos usar como espaço amostral o conjunto* $\Omega = \{0, 1, 2, 3, \ldots\}$.

3. *Considere o experimento de lançar um dardo em um alvo circular de raio* $r > 0$. *Neste caso podemos definir o espaço amostral como qualquer círculo de raio* r, *como por exemplo*

$$\Omega = \{(x,y) : x^2 + y^2 \leq 1\}.$$

4. *Se quisermos sortear uniformemente um número no intervalo* $[0,1]$, *então o espaço amostral seria tomado como sendo* $\Omega = [0,1]$. $\triangleleft$

Outro conceito importante é o conceito de **evento**. Ao sortearmos um carta em um baralho, podemos não estar interessados no valor exato da carta sorteada, mas sim em alguma propriedade desta como por exemplo se a carta é ou não da cor vermelha.

Em outras palavras, ao realizarmos um experimento aleatório podemos estar interessados em medir a incerteza não apenas de resultados individuais $\omega \in \Omega$, mas também de subconjuntos $A \subset \Omega$.

Estes subconjuntos de Ω para os quais estamos interessados em associar uma medida de incerteza são chamados de **eventos**.

A definição de evento colocada acima ainda é propositalmente vaga. É fato que eventos são subconjuntos do espaço amostral, mas pensar em qualquer subconjunto como um evento é, em geral, um problema.

Afinal, o que precisamos pedir de uma Probabilidade?

A primeira exigência, bastante simples e natural, é que independente do modo que construirmos as probabilidades $\mathbf{P}[\cdot]$, precisamos que $\mathbf{P}[\Omega] = 1$ e que para qualquer evento E tenhamos que $\mathbf{P}(E) \in [0,1]$.

Outro ponto importante é que dados dois eventos disjuntos $A, B \subset \Omega$ tais que $A \cap B = \emptyset$, ou seja tais que A e B não podem ser observados simultaneamente, tenhamos que

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

Ou seja, queremos que se dois eventos não podem ser observados simultaneamente, então a probabilidade de observamos um **ou** outro deve ser a soma das probabilidades individuais.

No caso de espaços finitos, é suficiente que exigíssemos essa hipótese para a união de dois conjuntos, e consequentemente por indução para uma união de número finito de conjuntos, mas para espaços mais gerais é interessante dar um passo a mais, e pedir que essa propriedade seja satisfeita para um número enumerável de conjuntos, ou seja, se $A_1, A_2, \ldots \subset \Omega$ são eventos disjuntos (ou seja, eventos tais que $A_i \cap A_j = \emptyset$ para todo $i, j \geq 1$, $i \neq j$), então

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

Essa propriedade é denominada sigma aditividade e ao longo do texto deverá ficar claro porque essa é a hipótese correta.

Resumindo, uma probabilidade é uma função $\mathbf{P}$ que a cada evento $A \subset \Omega$ associa um valor $\mathbf{P}(A) \in [0,1]$ tal que

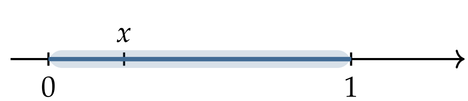
1. $\mathbf{P}[\Omega] = 1$;

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

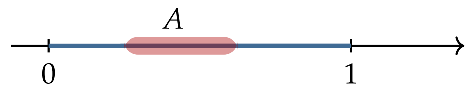
$$\mathbf{P}\left[\bigcup_{k=1}^{\infty} A_k\right] = \sum_{k=1}^{\infty} \mathbf{P}[A_k].$$

1.1 Nem sempre podemos atribuir probabilidades

Vamos tentar construir um modelo probabilístico que modele o sorteio de um ponto **uniformemente** no intervalo $[0,1)$. Nesse caso quais são as condições naturais que exigíssemos?



Claramente neste modelo não basta atribuir probabilidades somente aos pontos individuais, que por serem infinitos, terão probabilidade 0. Vamos então tentar atribuir probabilidades aos conjuntos, aos eventos. Gostaríamos de atribuir uma probabilidade à possibilidade que esse ponto pertença a um subconjunto $A \subset [0,1)$.



Ou seja, queremos uma função $\mu : \mathcal{P}([0,1)) \to [0,1)$ que seja uma probabilidade invariante por translação, isto é:

1. não seja trivial, ou equivalentemente que $\mu([0,1)) = 1$.
2. para uma sequência de conjuntos disjuntos $A_1, A_2, \ldots$ então $\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n)$.
3. Seja uniforme, o que equivale a pedir que seja invariante por translação, ou seja, se denotarmos a parte decimal de um número real x por $\operatorname{dec}(x)$, então pediremos que se $\operatorname{dec}(A') = \operatorname{dec}(A+b)$ então $\mu(A') = \mu(A)$.

Uma função não negativa e σ-aditiva é denominada medida, trataremos mais detalhadamente delas no Capítulo 2.

O problema é que isso é impossível!

1.2 Teorema (Conjunto de Vitali) *Suponha uma medida μ não nula e invariante por translações em $[0,1)$. Então existe um subconjunto de $[0,1)$ que não é μ mensurável.*

Demonstração. *Um conjunto de Vitali é um subconjunto V do intervalo $[0,1)$, tal que para todo número real r, existe exatamente um $v \in V$ tal que $v - r$ é um numero racional.*

Conjuntos de Vitali existem: *Os conjuntos Vitali existem porque os números racionais $\mathbb{Q}$ formam um subgrupo normal dos números reais $\mathbb{R}$ sob adição, e isso permite a construção do grupo quociente aditivo $\mathbb{R}/\mathbb{Q}$.*

O grupo $\mathbb{R}/\mathbb{Q}$ consiste em "cópias deslocadas" disjuntas de $\mathbb{Q}$ no sentido de que cada elemento desse grupo quociente é um conjunto da forma $\mathbb{Q}+r$ para algum r em $\mathbb{R}$. Cada classe de $\mathbb{R}/\mathbb{Q}$ intersecta $[0,1)$ e o axioma de escolha garante a existência de um subconjunto de $[0,1)$ contendo exatamente um representante de cada classe de $\mathbb{R}/\mathbb{Q}$. Um conjunto formado desta forma é denominado de um conjunto Vitali, e será denotado por $\mathcal{V}$.

Conjuntos de Vitali não são mensuráveis: *O argumento usual para verificar que o conjunto de Vitali não é mensurável é observar que se enumerarmos os racionais em $[-1,1]$ como $(q_i)_{i=1}^{\infty}$ e definirmos $\mathcal{V}_i = (\mathcal{V} + q_i)$ então os $\mathcal{V}_i$' são disjuntos dois a dois e $[0,1) \subset \bigcup_{i=1}^{\infty} \mathcal{V}_i \subset [-1,2]$.*

Antes de seguirmos observamos que dado uma medida μ não nula e invariante por translações em $[0,1)$ ela pode ser facilmente estendida para uma medida não nula e invariante por translações em $[-1,2]$.

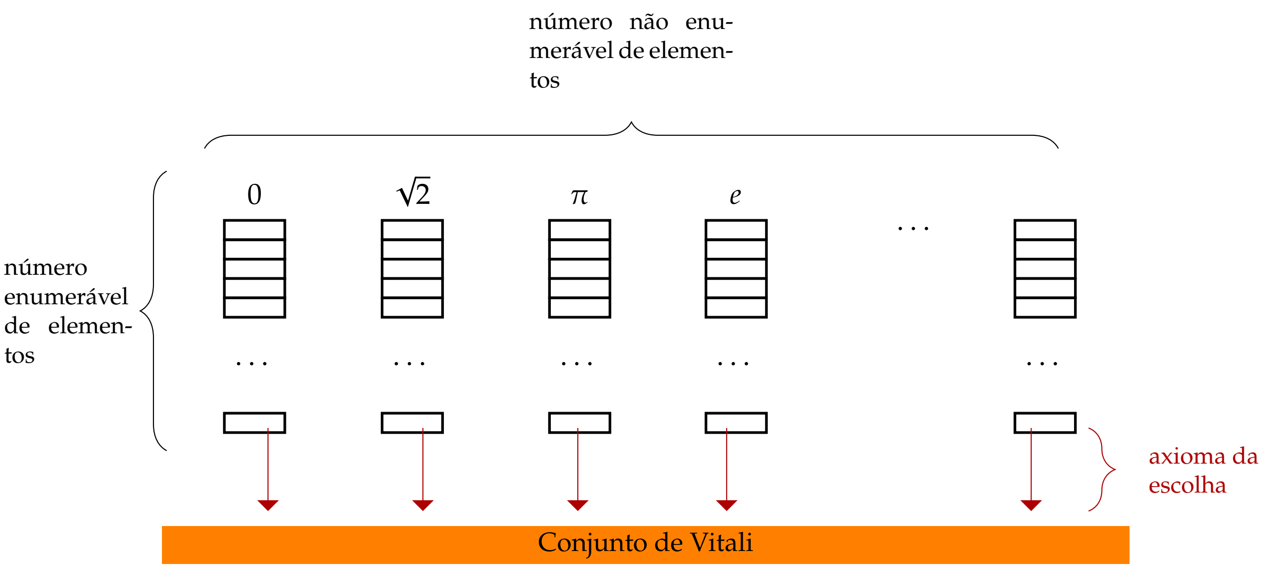
Feito essa observação, suponha que $\mathcal{V}$ é mensurável. Então teríamos

$$1 = \mu([0,1)) \leq \mu\left(\bigcup_{i=1}^{\infty} \mathcal{V}_i\right) = \sum_{i=1}^{\infty} \mu(\mathcal{V}_i) = \sum_{i=1}^{\infty} \mu(\mathcal{V}) \leq 3.$$

O que é impossível pois a soma $\sum_{i=1}^{\infty} \mu(\mathcal{V})$ é necessariamente igual a 0 ou a ∞. □

Outro modo de construir o conjunto de Vitali: Observamos que $\mathbb{R}$ é um espaço vetorial sobre $\mathbb{Q}$. E podemos ver $\mathbb{Q}$ como um subespaço unidimensional de $\mathbb{R}$. Considere V o complemento direto de $\mathbb{Q}$ em $\mathbb{R}$.

Dessa forma podemos escrever $\mathbb{R} = \mathbb{Q} \oplus V$. Considere $\mathcal{N} = \text{dec}(V)$. Temos que se $r \in \mathbb{R}$ então $r = q + n$ com $q \in \mathbb{Q}$ e $n \in \mathcal{N}$ e que $\mathcal{N} \subset [0,1)$. Então $\mathcal{N}$ é um conjunto de Vitali e a demonstração que o conjunto $\mathcal{N}$ é não mensurável é análoga a anterior.



1.2 A necessidade de álgebras

O exemplo de Vitali nos mostra que estamos pedimos demais do mundo. E dessa forma devemos abandonar alguma de nossas exigências:

a a sigma aditividade $\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n)$ para conjuntos disjuntos;

b a invariância por translação;

c o Axioma da Escolha;

d ou que todos os subconjuntos sejam mensuráveis.

Bem, a σ-aditividade funciona para uma classe muito grande de conjuntos para que possamos simplesmente abandoná-la. A invariância por translações pode não ser necessária em geral, mas é essencial para este exemplo, que por si só é de extrema importância. E o axioma da escolha... é o axioma da escolha!

Posto isso, a escolha parece fácil! Aceitamos que existem subconjuntos não mensuráveis aos quais não podemos atribuir probabilidades, mas mantemos a intuição de probabilidade nos casos em que sabemos atribuir probabilidade.

Ao fazermos isso teremos que nem todos os conjuntos são mensuráveis, mas faremos isso garantindo o mínimo de boas propriedades em relação aos conjuntos mensuráveis: garantiremos que eles se portam bem em relação à união e intersecção.

1.3 Álgebras, σ-álgebras e suas Amiguinhas

Como vimos até agora para definirmos uma medida de probabilidade precisaremos de um grande conjunto Ω, denominado **espaço amostral**. O conjunto Ω pode ser finito ou infinito e conterá a lista de todas as possíveis saídas de um experimento probabilístico. Definiremos como **eventos** os subconjuntos $E \subseteq \Omega$. Como vimos na seção anterior nem todas as coleções possíveis de resultados podem ser considerados eventos! (Mais coloquialmente, nem todos os eventos possíveis são permitidos). Assim, denotaremos por $\mathcal{F}$ a família de todos os eventos permitidos.

Ainda não é claro quais devem ser os eventos permitidos, mas queremos que a família satisfaça certas regras. Por exemplo, se você puder falar sobre os eventos A, B, você também pode falar de não A e A e B, ou seja, podemos falar sobre o complementar de A e da união de A e B.

Começaremos assim com uma família de conjuntos muito pouco ambiciosa.

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

$\boxed{1}$ *ao complementar*

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

$\boxed{2}$ *à união finita*

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

É importante observar que, a partir destas duas condições, segue que as álgebras também estão fechadas sob intersecções finitas:

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

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

1.4 Exemplo (Álgebra Trivial) *Seja $\mathcal{F} = \{\emptyset, \Omega\}$. Essa é a **álgebra trivial** e é a menor álgebra possível.* ◁

1.5 Exemplo (Álgebra das Partes) *Outro caso extremo $\mathcal{F} = \mathcal{P}(\Omega)$. Essa é a maior álgebra possível.* ◁

A primeira álgebra não parece muito útil. E a segunda, no caso dos reais, inclui eventos que não podemos atribuir probabilidade. A sabedoria está no caminho do meio.

Passemos a um exemplo muito mais natural e importante!

1.6 Exemplo (Álgebra de Borel) *Considere $\Omega = \mathbb{R}$. Queremos que todos os intervalos possíveis estejam na álgebra. Por causa das propriedades da Definição 1.3, temos que a menor álgebra $\mathcal{F}$ que contém os intervalos será:*

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

1.7 Exemplo (Álgebra de partição) *Dados um conjunto qualquer Ω e uma coleção de conjuntos disjuntos $\mathcal{P} = \{P_1, P_2, \ldots, P_n\}$ é uma partição se $\Omega = \cup_{j=1}^{n} P_j$. Nesse caso os conjuntos $P_1, P_2, \ldots, P_n$ são ditos átomos da partição.*

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

$$f(\mathcal{P}) = \{P_{n_1} \cup P_{n_2} \ldots P_{n_k} : n_1, n_2, n_k \in \{1, \ldots, n\}\}$$

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

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

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

Observe que a álgebra de Borel conterá apenas os conjuntos que são expressos como a união finita de intervalos em $\mathbb{R}$. O infeliz infortúnio é que o conceito de álgebra também é **fraco** para a teoria da probabilidade. Em particular, ele não nos permite falar sobre limites de conjuntos, o que obviamente queremos.

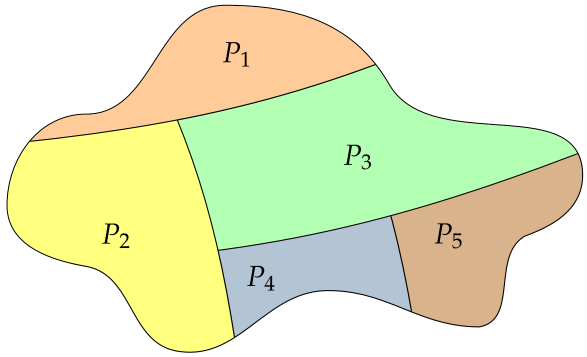


Figura 1.1: Uma partição de Ω

Por outro lado considerar uniões infinitas arbitrárias é pedir **demais** para a teoria da probabilidade, pois com uniões infinitas arbitrárias é possível construir a partir dos intervalos qualquer subconjunto dos reais pois em particular qualquer subconjunto é união de seus pontos:

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

Qual é o compromisso? Incluir apenas uniões enumeráveis. Este é o primeiro dogma da teoria da probabilidade.

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

1 *ao complementar*

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

2 *a união enumerável*

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

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

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

a A família $\{\emptyset, A, A^{\mathsf{c}}, \Omega\}$ é uma σ-álgebra simples gerada pelo subconjunto A.

b A família dos subconjuntos de X que são enumeráveis ou cujos complementos são enumeráveis é uma σ-álgebra.

1.9 Exemplo (σ-álgebra de Borel) *Considere $\Omega = \mathbb{R}$. Considere a σ-álgebra que contém todos os intervalos. Na seção 1.5 mostraremos que existe tal σ-álgebra.*

Essa σ-álgebra é denominada σ-álgebra de Borel, e veremos também que essa sigma álgebra não contém todos os subconjuntos de $\mathbb{R}$, como queremos.

Podemos demonstrar que existem $2^{2^{\aleph_0}}$ subconjuntos de $\mathbb{R}$. E que existem somente $2^{\aleph_0}$ conjuntos de Borel. Logo existem subconjuntos de $\mathbb{R}$ que não são borelianos. A demonstração desse fato é apresentada na Seção 2.5 ◁

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

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

Uma ressalva importante é que não devemos nos preocupar com σ-álgebras ao trabalhar com probabilidades em espaços enumeráveis. Nesse caso podemos considerar simplesmente o conjunto das partes como sendo a σ-álgebra.

1.4 Um Pouco mais Sobre Eventos

Agora que entendemos que σ-álgebras são a maneira como devemos organizar eventos em um modelo de probabilidade, vamos tomar um pouco mais de tempo para explorar um pouco mais esses conjuntos, e como os vemos no contexto da probabilidade.

Comecemos por lembrar que em um espaço mensurável $(\Omega,\mathcal{F})$ um evento $A \in \mathcal{F}$ é tão somente uma coleção de resultados de um experimento aleatório. Por essa razão é comum dizermos que *A foi observado* ou *a probabilidade de observarmos A* (ou qualquer sentença similar) fazendo referência ao experimento aleatório e a observação de um elemento de A em uma realização deste experimento.

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

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

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

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

1.11 Exemplo *Consideremos o experimento de lançar infinitas moedas. Para simplificar as notações, vamos representar coroas por 0 e caras por 1. Isso nos permite definir os espaço amostral por*

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

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

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

Deste modo as seguintes descrições são validas.

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

Mas estes não são os únicos eventos que estamos interessados.

As σ-álgebras são mais adequadas para trabalharmos com probabilidade pois no contexto de σ-álgebras podemos definir o conceito de **limite de uma sequência de eventos**.

1.12 Definição (Limite Superior e Inferior) *Suponha que $\mathcal{F}$ é uma σ-álgebra. E considere $A_1, A_2, \ldots \subseteq \Omega$.*

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

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

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

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

1.13 Heurística (Mendigos) *Seja A o conjunto de desempregados de uma cidade. Por falta de alternativas eles se tornam mendigos e têm que viver na rua. Um dia, a igreja local decide começar a dar*

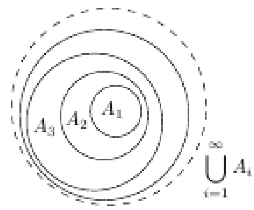


Figura 1.2: Sequência de intervalos crescentes.

comida gratuita semanalmente para essas pessoas. Denotaremos por A_n as pessoas que receberam comida na n-ésima semana.

Suponha também que os mendigos nunca morrem.

Algumas pessoas conseguem finalmente um novo emprego e dessa forma nunca mais retornam à igreja. Outros são muito orgulhosos e tentam não ser vistos o tempo todo, mas eles precisam comer por isso eles sempre voltam, eventualmente. Finalmente, há pessoas que aceitam a sua situação e começam a ir na igreja todas as semanas.

- $\limsup A_n$ *são todas as pessoas que não conseguem outro emprego.*
- $\liminf A_n$ *são as pessoas que se tornam frequentadores regulares semanais.*

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

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

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

1.15 Exemplo *Seja* $A_1 = A_3 = A_5 = \cdots = A$ *e* $A_2 = A_4 = A_6 = \cdots = B$ *com* B *um subconjunto próprio de* A*. Então* $\limsup A_n = A$ *e* $\liminf A_n = B$. ◁

1.16 Definição (Eventos não Crescentes e não Decrescentes) *A sequência de eventos* $A_1, A_2, \ldots$ *é dita* ***não crescente*** *se* $A_1 \supseteq A_2 \supseteq A_3 \supseteq \cdots$ *. Nesse caso escreveremos que* $A_n \downarrow$ *.*

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

1.17 Proposição

$\boxed{1}$ *Se $A_n \uparrow$, então A_n converge, e nesse caso temos que*

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

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

$\boxed{2}$ *Se $A_n \downarrow$, então A_n converge, e nesse caso temos que*

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

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

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

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

Suponha que os conjuntos A_n sejam não crescentes, $A_{n+1} \subset A_n$. Seja $x \in A = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k$. Para cada n, $\bigcup_{k=n}^{\infty} A_k = A_n$ por serem não crescentes, então $A = \bigcap_{n=1}^{\infty} A_n$. Em particular, $x \in A$ implica $x \in A_n$ para cada n, implicando que x está em todos, com exceção de um número finito de A_n , então $x \in B$ e consequentemente $A \subset B$. ◻

1.5 Classes de Conjuntos

Pelo que dissemos anteriormente, sempre que desejarmos modelar um fenômeno probabilístico temos que escolher uma uma σ-álgebra e uma medida.

Nessa seção apresentaremos algumas ferramentas para a construção de σ-álgebras. Começamos com algumas definições.

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

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

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

$\boxed{3}$ $\underbrace{A_1, A_2, \ldots}_{disjuntos} \in \mathcal{F} \Longrightarrow \bigcup_{k=1}^{\infty} A_k \in \mathcal{F}$.

Observe que a única razão pela qual nós exigimos $\Omega \in \mathcal{F}$ na definição acima é forçar a classe $\mathcal{F}$ ser não vazia.

1.19 Proposição *Um λ-sistema é fechado em relação a diferenças próprias. Isto é se $A, B \in L$ com $B \subset A$, então $A \setminus B \in L$.*

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

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

$\boxed{1}$ $A_1, A_2, \ldots \in \mathcal{M}$ *e* $A_n \uparrow A \implies A \in \mathcal{M}$

$\boxed{2}$ $A_1, A_2, \ldots \in \mathcal{M}$ *e* $A_n \downarrow A \implies A \in \mathcal{M}$.

As classes monótonas estão intimamente relacionadas às σ-álgebras. De fato, para nós, seu único uso será ajudar a verificar se uma determinada coleção de subconjuntos é uma σ-álgebra. Toda σ-álgebra é uma classe monótona, porque σ-álgebras são fechadas sob uniões enumeráveis.

Temos a seguinte relação entre as classes de conjuntos

1.22 Teorema ($\sigma = \lambda + \pi = \mathcal{M} + f$) *São equivalentes as seguintes afirmações:*

$\boxed{1}$ *$\mathcal{F}$ é uma σ-álgebra*

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

$\boxed{3}$ *$\mathcal{F}$ é um λ-sistema e um π-sistema.*

Demonstração. *Provaremos essa equivalência provando as implicações*

$$\begin{aligned} \mathcal{F} \text{ é uma } \sigma\text{-álgebra} &\iff \mathcal{F} \text{ é uma álgebra e classe monótona} \\ &\iff \mathcal{F} \text{ é um } \lambda\text{-sistema e um } \pi\text{-sistema.} \end{aligned}$$

Começamos observando que as implicações ($\Longrightarrow$) são triviais. Vamos demonstrar as implicações inversas.

$\boxed{3} \Longrightarrow \boxed{1}$

Considere que $\mathcal{F}$ é um λ-sistema e um π-sistema. Vamos demonstrar que $\mathcal{F}$ é uma σ-álgebra. Observe que $\Omega \in \mathcal{F}$ é óbvio pelas propriedades dos λ-sistemas. Além disso $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$

é direto das propriedades de λ-sistema. Para demonstrar que $A_1, A_2, \ldots \in \mathcal{F} \Rightarrow \bigcup_{k=1}^{\infty} A_k \in \mathcal{F}$ usaremos um pequeno truque para converter união em união disjunta.

$$\bigcup_{k=1}^{\infty} A_k = \bigcup_{k=1}^{\infty} \underbrace{A_k - (A_1 \cup \cdots \cup A_{k-1})}_{\text{disjuntos}}$$
$$= \bigcup_{k=1}^{\infty} A_k \cap A_1^{\mathsf{c}} \cap \cdots \cap A_{k-1}^{\mathsf{c}}$$

Agora $A_k^{\mathsf{c}} \in \mathcal{F}$ pelas propriedades dos λ-sistemas e logo $A_k \cap A_1^{\mathsf{c}} \cap \cdots \cap A_{k-1}^{\mathsf{c}} \in \mathcal{F}$ pelas propriedades dos π-sistemas. Logo $\bigcup_{k=1}^{\infty} A_k$ pode ser escrito como união disjunta de eventos em $\mathcal{F}$. Logo $\bigcup_{k=1}^{\infty} A_k \in \mathcal{F}$ por propriedades dos λ-sistemas.

$\boxed{2} \Longrightarrow \boxed{1}$ *Vamos provar que se $\mathcal{F}$ é uma álgebra e classe monótona então $\mathcal{F}$ é fechado sobre união enumerável. Se $A_1, A_2, \ldots \in \mathcal{F}$ então*

$$\bigcup_{k=1}^{\infty} A_k = \lim_{n}{\uparrow} \bigcup_{k=1}^{n} A_k$$

com $\bigcup_{k=1}^{n} A_k \in \mathcal{F}$ pelas propriedades de álgebra e logo $\bigcup_{k=1}^{\infty} A_k \in \mathcal{F}$ pelas propriedades de classe monótona. □

Classes Geradas

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

Demonstração. *Basta verificar que $\bigcap_{\lambda \in \Lambda} \mathcal{F}_\lambda$ satisfaz os axiomas listados na Definição 1.8. Por exemplo,*

$$A_i \in \bigcap_{\lambda \in \Lambda} \mathcal{F}_\lambda \Rightarrow \forall \lambda \in \Lambda, \bigcup A_i \in \mathcal{F}_\lambda$$
$$\Rightarrow \bigcup A_i \in \bigcap_{\lambda \in \Lambda} \mathcal{F}_\lambda.$$

Do mesmo modo podemos provar que:

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

Essas observações nos permitem fazer as seguintes definições

1.24 Definição *Seja C uma família arbitrária de subconjuntos de Ω.*

1. *Definimos a **σ-álgebra gerada** por C como*

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

2. *Definimos a **álgebra gerada** por C como*

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

3. *Definimos a **classe monótona gerada** por C como*

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

4. *Definimos o **λ-sistema gerado** por C como*

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

A σ-álgebra $\sigma\langle C\rangle$ será assim a menor σ-álgebra que contém C. Observamos que a definição da σ-álgebra gerada é implícita, e as σ-álgebras mais importantes que são construídas como σ-álgebras geradas não podem ser descritas explicitamente. A σ-álgebra gerada contém em todos os subconjuntos de Ω que podem ser obtidos a partir de elementos de C por um número enumerável de operações de complemento, união e interseção, e contém todos subconjuntos de Ω que podem ser obtidos dos anteriores por um número enumerável de operações de complemento, união e interseção e assim por diante.

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

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

a. *conjuntos finitos*

b. *conjuntos enumeráveis*

c. *conjuntos cujo complementar é enumerável ou finito.*

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

Há um tipo de raciocínio que ocorre com tanta frequência em problemas que envolvem classes de conjuntos que merece ser enunciado explicitamente.

1.27 **Teorema (Princípio dos Bons Conjuntos)** *Dado $\mathcal{C}$ e $\mathcal{G}$ duas família de subconjuntos de Ω. Se*

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

b *$\mathcal{G}$ é uma σ-álgebra*

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

Demonstração. *A demonstração é imediata do fato que:*

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

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

A operação de obter σ-álgebras se porta bem em relação a uma restrição a subespaços de Ω.

1.28 **Teorema (Geradores Restringidos)** *Dados Ω um espaço amostral e $\mathcal{C}$ uma classe de subconjuntos de Ω. Se $\Omega_0 \subset \Omega$ então*

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

Demonstração.

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

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

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

Observe que essa inclusão é equivalente à afirmação de que, para cada $A \in \sigma\langle\mathcal{C}\rangle$, $A \cap \Omega_0 \in \sigma\langle\mathcal{C} \cap \Omega_0\rangle$. Para demonstrar esse fato seja

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

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

□ $\mathcal{C} \subset \mathcal{G}$

$A \in \mathcal{C} \Longrightarrow A \cap \Omega_0 \in \mathcal{C} \cap \Omega_0 \subset \sigma\langle\mathcal{C} \cap \Omega_0\rangle$.

□ $\Omega \in \mathcal{G}$

$$\begin{aligned}\Omega_0 \subset \Omega \Longrightarrow \Omega \cap \Omega_0 &= \Omega_0 \in \sigma\langle\mathcal{C} \cap \Omega_0\rangle \\ &\quad \textit{pois uma } \sigma\textit{-álgebra em } \Omega_0 \textit{ deve conter } \Omega_0 \\ &\Longrightarrow \Omega \in \mathcal{G}.\end{aligned}$$

□ $A \in \mathcal{G} \Longrightarrow A^c \in \mathcal{G}$

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

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

□ $A_1, A_2, \ldots \in \mathcal{G} \Longrightarrow \bigcup_k A_k \in \mathcal{G}$

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

1.29 Teorema (Teorema da Classe Monótona de Halmos) *Se $\mathcal{F}_0$ é uma álgebra então $\mathcal{M}\langle\mathcal{F}_0\rangle = \sigma\langle\mathcal{F}_0\rangle$.*

Demonstração. *Seja $\mathcal{M} = \mathcal{M}\langle\mathcal{F}_0\rangle$. É suficiente provar que $\sigma\langle F_0\rangle \subset \mathcal{M}$. Faremos isso demonstrando que $\mathcal{M}$ é uma álgebra.*

Considere a classe $G = \{A : A^C \in \mathcal{M}\}$. Como $\mathcal{M}$ é classe monótona, o mesmo acontece com G. Como F_0 é uma álgebra e $F_0 \subset G$ temos que $\mathcal{M} \subset G$. Portanto, $\mathcal{M}$ é fechada sob complementação.

Defina G_1, como a classe de todos os conjuntos A tais que $A \cup B \in \mathcal{M}$ para todo $B \in F_0$.

$$G_1 = \{A \subset \Omega : A \cup B \in \mathcal{M} \textit{ para todo } B \in F_0\}$$

Então, novamente, G_1 *é uma classe monótona e* $F_0 \subset G_1$ *e logo* $\mathcal{M} \subset G_1$*.*

Defina G_2*, como a classe de todos os conjuntos* B *tais que* $A \cup B \in \mathcal{M}$ *para todos os* $A \in \mathcal{M}$*.*

$$G_2 = \{B \subset \Omega : A \cup B \in \mathcal{M} \text{ para todo } A \in \mathcal{M}\}$$

Então, G_2 *é uma classe monótona. Agora, como* $\mathcal{M} \subset G_1$ *segue que* $A \in \mathcal{M}$ *e* $B \subset F_0$ *e esse fatos implicam que* $A \cup B \subset \mathcal{M}$*. em outras palavras,* $B \subset F_0$ *implica que* $B \in G_2$*. Assim* $F_0 \subset G_2$*; por minimalidade,* $\mathcal{M} \subset G_2$ *e, portanto,* $A, B \in \mathcal{M}$ *implica que* $A \cup B \in \mathcal{M}$*.* □

1.30 Teorema ($\pi - \lambda$ de Dynkin) *Se* $\mathcal{P}$ *é um* π*-sistema então* $\lambda\langle\mathcal{P}\rangle = \sigma\langle\mathcal{P}\rangle$*.*

Demonstração.

$\boxed{\Rightarrow}$ *Como toda* σ*-álgebra é um* λ*-sistema,* $\sigma\langle\mathcal{P}\rangle$ *é um* λ*-sistema que contém* $\mathcal{P}$*, assim* $\lambda\langle\mathcal{P}\rangle \subset \sigma\langle\mathcal{P}\rangle$*.*

$\boxed{\Leftarrow}$ *Para demonstrar a inclusão contrária, comecemos demonstrando que* $\lambda\langle\mathcal{P}\rangle$ *é fechado em relação a intersecções finitas. Para cada* $A \in \Omega$ *defina a família* $\mathcal{G}_A$ *de subconjuntos de* Ω*,*

$$\mathcal{G}_A = \{B \subset \Omega : A \cap B \in \lambda\langle\mathcal{P}\rangle\}$$

A demonstração é razoavelmente direta e faz uso inúmeras vezes do princípio dos Bons Conjuntos em contextos razoavelmente similares aumentando a generalidade, o que pode gerar confusão. Por isso estruturamos a demonstração em um conjunto de afirmativas preliminares.

a) *Se* $A \in \mathcal{P}$ *então* $\mathcal{G}_A$ *é um* λ*-sistema que contém* $\mathcal{P}$*.*

Como $\mathcal{P}$ *é fechado em relação à intersecções finitas e como* $\mathcal{P} \subset \lambda\langle\mathcal{P}\rangle$ *temos que* $\mathcal{P} \subset \mathcal{G}_A$*. O fato que* $\mathcal{P} \subset \mathcal{G}_A$ *implica que* $\Omega \in \mathcal{G}_A$*, e as igualdades de conjunto*

$$(B \setminus C) \cap A = (B \cap A) \setminus (C \cap A) \text{ e}$$
$$\Big(\bigcup_n B_n\Big) \cap A = \bigcup_n (B_n \cap A)$$

implicam que $\mathcal{G}_A$ *é fechado para diferenças e uniões de sequências crescentes de subconjuntos.*

Desta forma $\mathcal{G}_A$ *é um* λ*-sistema. Assim,* $\mathcal{G}_A$ *é um* λ*-sistema que inclui* $\mathcal{P}$

b) *Se* $A \in \lambda\langle\mathcal{P}\rangle$ *então* $\mathcal{G}_A$ *é um* λ*-sistema.*

Esse item é análogo ao anterior.

c) *Se* $A \in \mathcal{P}$ *então* $\lambda\langle\mathcal{P}\rangle \subset \mathcal{G}_A$

Se $A \in \mathcal{P}$ *então* $\mathcal{G}_A$ *é um* λ*-sistema contendo* $\mathcal{P}$*, logo contém o* λ*-sistema minimal que contém* $\mathcal{P}$*.*

d) *Se* $D \in \mathcal{P}$ *e* $E \in \lambda\langle\mathcal{P}\rangle$*, então* $D \in \mathcal{G}_E$

Como $D \in \mathcal{P}$ temos que $\lambda\langle\mathcal{P}\rangle \subset \mathcal{G}_D$.

Como $E \cap D \in \lambda\langle\mathcal{P}\rangle$ por definição $D \in \mathcal{G}_E$.

e) *Se $E \in \lambda\langle\mathcal{P}\rangle$ então $\lambda\langle\mathcal{P}\rangle \subset \mathcal{G}_E$.*

Seja $D \in \mathcal{P}$ e $E \in \lambda\langle\mathcal{P}\rangle$. Pela parte d) temos que $D \in \mathcal{G}_E$. Logo, como D é arbitrário $\mathcal{P} \in \mathcal{G}_E$.

Como $E \in \lambda\langle\mathcal{P}\rangle$ a parte b) nos fornece que $\mathcal{G}_E$ é um λ-sistema.

f) *Para terminar a demonstração considere $A,B \in \lambda\langle\mathcal{P}\rangle$. Como $A \in \lambda\langle\mathcal{P}\rangle$ o item* e) *nos permite concluir que*

$\lambda\langle\mathcal{P}\rangle \subset \mathcal{G}_A$. Logo $B \in \mathcal{G}_A$ e logo por definição $A \cap B \in \mathcal{G}_A$. E logo $\lambda\langle\mathcal{P}\rangle$ é um π-sistema.

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

1.31 Teorema (Princípio dos Bons Conjuntos 2) *Dados $\mathcal{P}$ e $\mathcal{G}$ duas sub-coleções de subconjuntos de Ω. Se*

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

b *$\mathcal{P}$ é uma π-sistema;*

c *$\mathcal{G}$ é uma λ-sistema*

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

1.6 σ-álgebra de Borel

As σ-álgebras de Borel permeiam toda a Teoria da Medida e Probabilidade. Nesta seção, faremos um tratamento mais detalhado dessas álgebras. O ponto principal é que as σ-álgebras de Borel admitem muitos geradores equivalentes. Geradores diferentes são úteis para demonstrar coisas diferentes. Por exemplo, a σ-álgebra de Borel em $(0,1]^d$ como gerado pela álgebra de uniões finitas disjuntas de retângulos é útil para construir a medida Lebesgue.

1.32 Definição (Espaço Métrico) *O par (Ω,d), no qual Ω é um conjunto e d é uma função*

$$d : X \times X \to R$$

*é dito **espaço métrico** se*

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

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

$\boxed{3}$ $d(x,y) = d(y,x)$

$\boxed{4}$ $d(x,y) \leq d(x,z) + d(z,y)$

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

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

A definição acima imediatamente nos permite definir a σ-álgebra de Borel $\mathcal{B}(\mathbb{R}^d)$ e $\mathcal{B}((0,1]^d)$.

1.35 Teorema (Restrição de Borel) *Dado Ω um espaço métrico $\Omega_o \subset \Omega$. Então a σ-álgebra de Borel $\mathcal{B}(\Omega_o)$, satisfaz*

$$\mathcal{B}(\Omega_o) = \mathcal{B}(\Omega) \cap \Omega_o$$

Se, $\Omega_o \in \mathcal{B}(\Omega)$ então $\mathcal{B}(\Omega_o) = \{B \in \mathcal{B}(\Omega) : B \subset \Omega_o\}$.

Demonstração. *Começaremos mostrando que $\mathcal{B}(\Omega_o) = \mathcal{B}(\Omega) \cap \Omega_o$*

Considere

$$\mathcal{G} := \text{subconjuntos abertos de } \Omega$$
$$\mathcal{G}_o := \text{subconjuntos abertos de } \Omega_o$$

Do curso de topologia você deve se lembrar que:

$$\mathcal{G}_o = \mathcal{G} \cap \Omega_o.$$

Logo

$$\begin{aligned}\mathcal{B}(\Omega_o) = \sigma\langle \mathcal{G}_o \rangle &= \sigma\langle \mathcal{G} \cap \Omega_o \rangle \\ &= \sigma\langle \mathcal{G} \rangle \cap \Omega_o, \\ &= \mathcal{B}(\Omega) \cap \Omega_o.\end{aligned}$$

Agora mostraremos que $\mathcal{B}(\Omega) \cap \Omega_o = \{B \in \mathcal{B}(\Omega) : B \subset \Omega_o\}$ quando $\Omega_o \in \mathcal{B}(\Omega)$)

Para demonstrar a contenção $\supset$ suponha que $B \subset \Omega_o$ e $B \in \mathcal{B}(\Omega)$. Então $B = B \cap \Omega_o \in \mathcal{B}(\Omega) \cap \Omega_o$. Para demonstrar a contenção $\subset$ seja $B \in \mathcal{B}(\Omega) \cap \Omega_o$ e assim $B = \widetilde{B} \cap \Omega_o$ onde $\widetilde{B} \in \mathcal{B}(\Omega)$. Como $\Omega_o \subset \Omega$ temos que $B \in \mathcal{B}(\Omega)$ e $B \subset \Omega_o$. □

1.36 Teorema (Geradores da σ-álgebra de Borel)

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

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

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

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

A parte mais importante do teorema acima é que $\mathcal{B}((0,1]^d) = \sigma\left\langle \mathcal{B}_0((0,1]^d) \right\rangle$. Esta caracterização permite construir probabilidades em $\mathcal{B}_0((0,1]^d)$, então utilizar o Teorema da Extensão de Carathéodory para estender essas probabilidades para um modelo de probabilidade total em $\mathcal{B}((0,1]^d)$.

Demonstração. *Mostraremos apenas uma igualdade:*

$$\sigma\left\langle (a,b] : 0 < a < b < 1 \right\rangle = \sigma\left\langle (a,b) : 0 < a < b < 1 \right\rangle$$

□ *Mostraremos que* $(a_0,b_0] \in \sigma\langle(a,b) : 0 < a < b < 1\rangle$*}*

$$(a_0,b_0] = \bigcap_{n=1}^{\infty}(a_0,b_0 + n^{-1}).$$

□ *Mostraremos que* $(a_0,b_0) \in \sigma\langle(a,b] : 0 < a < b < 1\rangle$*) Isso segue da identidade.*

$$(a_0,b_0) = \bigcup_{n=1}^{\infty}(a_0,b_0 - n^{-1}].$$

Os conjuntos na σ-álgebra de Borel são extremamente ricos. Na verdade, é difícil demonstrar que existem conjuntos que não estão em $\mathcal{B}(\mathbb{R})$. A maneira mais fácil de encontrar esse conjunto é utilizar as propriedades da medida de Lebesgue.

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

A construção da σ álgebra de Borel poderia ser feita de modo análogo em um espaço topológico.

Nesse contexto é interessante o seguinte teorema:

Teorema. Seja Ω um espaço polonês, ou seja, um espaço topológico de tal forma que existe uma métrica d em Ω que define a topologia de Ω e que faz com que Ω um espaço métrico separável completo. Então, Ω como um espaço de Borel é isomorfo a um dos conjuntos abaixo:

a $\mathbb{R}$

b $\mathbb{Z}$

c espaço finito

Exercícios

Ex. 1.1 — Mostre que se $\mathcal{A}_1$ e $\mathcal{A}_2$ são álgebras de subconjuntos de Ω, então $\mathcal{A}_1 \cap \mathcal{A}_2$ é álgebra.

Ex. 1.2 — Considere $\Omega = (0,1]$. Prove que

$$\mathcal{A} = \{A \subset \Omega : A = \cup_{i=1}^{n}(a_i, b_i],$$

$$(a_i, b_i] \subset (0,1] \text{ e a união é disjunta é álgebra. } \}$$

Ex. 1.3 — Prove que a coleção

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

é álgebra sobre Ω.

Ex. 1.4 — Mostre que a coleção

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

não é uma σ-álgebra sobre Ω.

Ex. 1.5 — Considere Ω infinito, não enumerável. Mostre que a coleção

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

é uma σ-álgebra sobre Ω.

Ex. 1.6 — Prove que

1. Se $A_n \uparrow$, então A_n converge, e

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

2. Se $A_n \downarrow$, então A_n converge, e

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

Ex. 1.7 — (Conjunto de Cantor) Mostre que o conjunto de Cantor é não enumerável em $\mathcal{B}((0,1])$ (uma boa maneira de ver isso é trabalhar com uma caracterização do conjunto de Cantor na base 3).

Ex. 1.8 — Prove o Princípio dos Bons Conjuntos:
Dado $\mathcal{C}$ e $\mathcal{G}$ duas coleção de subconjuntos de Ω. Se

- $\mathcal{C} \subset \mathcal{G}$;
- $\mathcal{G}$ é uma σ-álgebra

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

Ex. 1.9 — (Conjunto de Vitali II) Suponha uma medida μ não nula, sigma aditiva e invariante por translações nos plano. Prove detalhadamente que existe um subconjunto do quadrado [0,1]2 que não é μ mensurável.

Ex. 1.10 — Prove o Teorema da Classe Monótona de Halmos

Ex. 1.11 — Considere as seguintes propriedades
(λ_1) $\Omega \in \mathcal{L}$,
(λ_2) $A \in \mathcal{L}$ implica que $A^c \in \mathcal{L}$, e
(λ_3) Para toda sequência de conjuntos disjuntos $\{A_n\}_{n=1}^{\infty}$ em $\mathcal{L}$, temos $\bigcup_{n=1}^{\infty} A_n \in \mathcal{L}$.
(λ_2') Para todo $A, B \in \mathcal{L}, A \subseteq B$ então $B - A \in \mathcal{L}$;
(λ_4) Para todo A, $B \in \mathcal{L}, A \cap B = \emptyset$ então $A \cup B \in \mathcal{L}$;
(λ_5) Para toda sequência não decrescente $\{A_n\}_{n=1}^{\infty}$ em $\mathcal{L}$, $\bigcup_{n=1}^{\infty} A_n \in \mathcal{L}$;
(λ_6) Para toda sequência não crescente$\{A_n\}_{n=1}^{\infty}$ em $\mathcal{L}$, $\bigcap_{n=1}^{\infty} A_n \in \mathcal{L}$.
Mostre que

1. $\mathcal{L}$ é λ-sistema see $\mathcal{L}$ satisfaz $(\lambda_1), (\lambda_2')$, e (λ_3) .
2. Todo λ-sistema satisfaz também $(\lambda_4), (\lambda_5)$, e (λ_6) .
3. $\mathcal{L}$ é um λ-sistema see satisfaz $(\lambda_1), (\lambda_2')$, e (λ_6) .
4. Se a família $\mathcal{L}$ é não vazia e satisfaz (λ_2) e (λ_3) , então $\mathcal{L}$ é um λ-sistema.

Ex. 1.12 — Para qualquer família não vazia $\mathcal{A} \subset 2^{\Omega}$, se

$$\begin{aligned} C &:= \textit{a coleção de conjuntos de } \mathcal{A} \textit{ e seus complementos} \\ \mathcal{I} &:= \textit{a coleção de intersecções finitas de } C \\ \mathcal{U} &:= \textit{a coleção de uniões finitas de } \mathcal{I}. \end{aligned}$$

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

Ex. 1.13 — Usando o resultado anterior mostre que um subconjunto da álgebra de Borel de [0,1] pode ser escrito como união finita de intervalos.

CAPÍTULO 2

Medidas de Probabilidade

Nesse capítulo vamos definir as medidas de probabilidade, seguindo a motivações apresentadas no início do Capítulo 1. Começaremos por dois conceitos relacionados, porém mais simples: o conceito de probabilidade finitamente aditiva e de pré-medida de probabilidade. Esses conceitos são insuficientes para desenvolver uma teoria rica de probabilidade, mas serão o ponto de partida para a construção das medidas de probabilidade.

Uma observação relevante a ser feita é que descrever um experimento probabilístico, um modelo estocástico envolve fundamentalmente construir a medida de probabilidade desse modelo. Uma estratégia usual nesse caso é construir uma probabilidade finitamente aditiva, demonstrar que na verdade temos uma pré-medida de probabilidade e finalmente concluir que temos uma medida de probabilidade. Assim apresentaremos uma série de ferramentas e caracterizações de quando uma probabilidade finitamente aditiva é de fato uma pré-medida de probabilidade. Posteriormente apresentaremos o Teorema de Extensão de Carathéodory que nos diz que sempre podemos construir uma medida de probabilidade a partir de uma pré-medida de probabilidade.

Finalmente, usaremos esse ferramental para construir diversos modelos probabilísticos, e em especial a medida de Lebesgue.

2.1 Definição (Probabilidade Finitamente Aditiva) *Se $\mathcal{F}_0$ é uma álgebra em Ω, então $\mathbf{P} : \mathcal{F}_0 \to [0,1]$ é dita uma* ***probabilidade finitamente aditiva*** *ou* ***atribuição de probabilidade*** *em $\mathcal{F}_0$ se*

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

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

Podemos na definição anterior trocar a hipótese de aditividade finita por aditividade enumerável. Podemos fazer isso para eventos numa álgebra ou numa σ-álgebra.

2.2 Definição (Pré-Medida de Probabilidade) *Se* $\mathcal{F}_0$ *é uma álgebra em* Ω*, então* $\mathbf{P} : \mathcal{F}_0 \longrightarrow [0,1]$ *é dita uma* ***pré-medida de probabilidade*** *em* $\mathcal{F}_0$ *se*

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

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

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

1 $\mu(\varnothing) = 0$

2 *para toda sequência de eventos disjuntos* $A_1, A_2, \ldots \in \mathcal{F}$*, temos que*

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

Essa propriedade é denominada de $\sigma -$ ***aditividade***.

Se nas Definições 2.1 e 2.2 removermos a hipótese $\mathbf{P}(\Omega) = 1$ temos uma medida finitamente aditiva e uma pré-medida respectivamente.

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

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

2.5 Definição (Medidas Finitas e σ-finitas) *Dado* $(\Omega, \mathcal{F}, \mu)$ *um espaço de medida.*

a *Se* $\mu(\Omega) < \infty$ *então* μ *é dita* ***medida finita***;

b *Se* $\mu(\Omega) = \infty$ *então* μ *é dita* ***medida infinita***;

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

Note que toda medida de probabilidade é σ-finita.

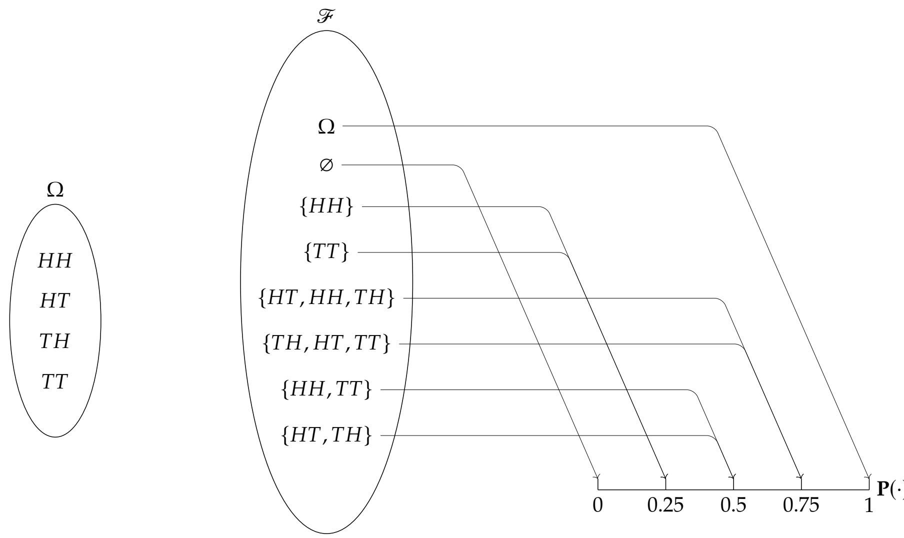


Figura 2.1: Relação entre o espaço amostral, σ-álgebra e a medida de probabilidade.

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

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

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

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

Alguns dos principais exemplos de probabilidades vistos em cursos básicos caem na categoria descrita no exemplo acima.

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

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

Como

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

a medida

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

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

2.8 Exemplo (Lançamento de Moedas) *Vamos considerar o experimento de lançar n moedas. Representando* coroa *por 0 e* cara *por 1, podemos definir o espaço amostral por*

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

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

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

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

Fica como exercício mostrar que

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

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

Provaremos agora algumas propriedades elementares da medida de probabilidades.

2.9 Proposição (Propriedades da Probabilidade)

1. $\mathbf{P}(A^c) = 1 - \mathbf{P}(A)$.
2. $A \subseteq B$ *implica* $\mathbf{P}(A) \leq \mathbf{P}(B)$. *(**monotonicidade**)*
3. *Dados eventos A e B então* $\mathbf{P}(A \cup B) \leq \mathbf{P}(A) + \mathbf{P}(B)$ *(**sub-aditividade**)*
4. $\mathbf{P}\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \mathbf{P}(A_n)$ *(*σ*-**sub-aditividade**)*
5. $\mathbf{P}[A \cup B] = \mathbf{P}[A] + \mathbf{P}[B] - \mathbf{P}[A \cap B]$; *(**Inclusão-exclusão**)*
6. $\mathbf{P}(A \triangle B) \geq \max\{\mathbf{P}(A - B), \mathbf{P}(B - A)\}$

Demonstração.

1. $\Omega = A \cup A^c$. *Logo,* $\mathbf{P}(\Omega) = \mathbf{P}(A) + \mathbf{P}(A^c)$. *Agora basta utilizar que* $\mathbf{P}(\Omega) = 1$.
2. *Use a decomposição* $B = A \cup (B \setminus A)$.
3. *Note que* $A \cup B$ *pode ser escrito como* $A \cup (B \setminus A)$. *Logo, pela monotonicidade da probabilidade, temos que* $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B \setminus A) \leq \mathbf{P}(A) + \mathbf{P}(B)$.
4. *Exercício*
5. *Exercício* □

2.1 Continuidade das Medidas de Probabilidade

Um dos conceito fundamentais para medidas de probabilidade é o conceito de continuidade:

2.10 Teorema (Continuidade da Probabilidade) *Dado uma sequência de eventos $A_1, A_2, \ldots \in \mathcal{F}$. Então*

1. *Se $A_n \uparrow A$, então $\mathbf{P}(A_n) \uparrow \mathbf{P}(A)$ quando $n \to \infty$.*

2. *Se $A_n \downarrow A$, então $\mathbf{P}(A_n) \downarrow \mathbf{P}(A)$ quando $n \to \infty$.*

Demonstração. *Provaremos apenas a primeira propriedade.*

Dado uma sequência de eventos $A_1, A_2, \ldots$. Vamos começar transformando a união em união disjunta. Para isso definiremos uma sequência auxiliar de eventos: defina $B_1 = A_1$ e para $k \geq 2$, defina $B_k = A_k \setminus A_{k-1}$.

Como $A_1 \subset A_2 \subset \cdots$ então os conjuntos B_n são disjuntos e $\bigcup_{n=1}^{j} B_n = \bigcup_{n=1}^{j} A_n = A_j$, para $1 \leq j \leq \infty$. Logo

$$A = \bigcup_k B_k$$

E assim $\mathbf{P}(A_n) = \mathbf{P}(\bigcup_{k=1}^n B_k)$, e como os conjuntos B_k são disjuntos temos que $\mathbf{P}(\bigcup_{k=1}^n B_k) = \sum_{k=1}^n \mathbf{P}(B_k)$, por aditividade. Todas as somas parciais são limitadas superiormente por 1. Além disso, a sequência de somas parciais é não decrescente. Portanto, converge. Assim

$$\sum_{k=1}^{\infty} \mathbf{P}(B_k) = \mathbf{P}\left(\bigcup_{k=1}^{\infty} B_k\right) = \mathbf{P}(A).$$

2.11 Teorema *Dado $A_1, A_2, \ldots$ uma sequência de eventos. Então*

1. $\mathbf{P}(\liminf A_n) \leq \liminf \mathbf{P}(A_n) \leq \limsup \mathbf{P}(A_n) \leq \mathbf{P}(\limsup A_n)$.

2. *Se $A_n \to A$ então $\mathbf{P}(A_n) \to \mathbf{P}(A)$ quando $n \to \infty$.*

Demonstração. *Exercício.*

Dica do 1. *Observe as inclusões*

$$\liminf A_n \uparrow \bigcap_{m=n}^{\infty} A_m \subset A_n \subset \bigcup_{m=n}^{\infty} A_m \downarrow \limsup A_n$$

2.12 Teorema (Caracterizações da σ-aditividade) *Dado $\mathbf{P} : \mathcal{F}_0 \to [0,1]$ uma probabilidade finitamente aditiva numa álgebra $\mathcal{F}_0$. Então as seguintes afirmações são equivalentes*

1. *$\mathbf{P}$ é uma pré-medida de probabilidade;*

2 **P** *é contínua por baixo. Isto é dados* $A_1, A_2, \ldots \in \mathcal{F}_0$ *e* $A_n \uparrow A \in \mathcal{F}_0$ *então* $\mathbf{P}(A_n) \uparrow \mathbf{P}(A)$*;*

3 **P** *é contínua por cima. Isto é dados* $A_1, A_2, \ldots \in \mathcal{F}_0$ *e* $A_n \downarrow A \in \mathcal{F}_0$ *então* $\mathbf{P}(A_n) \downarrow \mathbf{P}(A)$*;*

4 *Contínua por cima em* $\varnothing$*. Isto é dados* $A_1, A_2, \ldots \in \mathcal{F}_0$ *e* $A_n \downarrow \varnothing$ *então* $\mathbf{P}(A_n) \downarrow 0$.

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

(4. $\Longrightarrow$ 3.*) Suponha que* **P** *é contínua por cima em* $\varnothing$ *e que* $A_1, A_2, \ldots \in \mathcal{F}_0$ *e* $A_n \downarrow A \in \mathcal{F}_0$.

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

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

(2. $\Longrightarrow$ 1.*) Basta provar a* σ*-aditividade. Suponha que* $A_1, A_2, \ldots$ *são conjuntos disjuntos em* $\mathcal{F}_0$ *tais que* $\bigcup_{k=1}^{\infty} A_k \in \mathcal{F}_0$*. Então*

$$P\left[\bigcup_{k=1}^{\infty} A_k\right] = P\left[\lim_n\uparrow \bigcup_{k=1}^{n} A_k\right] = \lim_n\uparrow P\left[\bigcup_{k=1}^{n} A_k\right]$$

Na última igualdade usamos o item 2. □

2.13 Teorema (Unicidade das Medidas de Probabilidade) *Dado* $\mathcal{P}$ *uma família de subconjuntos de* Ω*. Se* **P** *e* **Q** *são duas medidas de probabilidade em* $(\Omega, \sigma\langle\mathcal{P}\rangle)$ *tais que*

1 **P** *e* **Q** *concordam em* $\mathcal{P}$*;*

2 $\mathcal{P}$ *é um* π*-sistema,*

então **P** *e* **Q** *concordam em* $\sigma\langle\mathcal{P}\rangle$.

Demonstração. *Para demonstrar esse fato usaremos o Teorema* $\pi - \lambda$ *de Dynkin.*

Definiremos os bons conjuntos como:

$$\mathcal{G} := \{A \subset \Omega \colon \mathbf{Q}[A] = \mathbf{P}[A]\}.$$

O Teorema $\pi - \lambda$ *de Dynkin afirma que* $\sigma\langle\mathcal{P}\rangle = \lambda\langle\mathcal{P}\rangle$ *quando* $\mathcal{P}$ *é um* π*-sistema. Logo para demonstrar que* $\sigma\langle\mathcal{P}\rangle = \lambda\langle\mathcal{P}\rangle \subset \mathcal{G}$ *basta demonstrar que* $\mathcal{G}$ *é um* λ*-sistema e utilizar o* Princípio dos Bons Conjuntos.

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

Isto é imediato pois $\mathbf{Q}[\Omega] = 1$ *e* $\mathbf{P}[\Omega] = 1$;

□ $A \in \mathcal{G} \Longrightarrow A^c \in \mathcal{G}$

$$\begin{aligned} A \in \mathcal{G} &\Longrightarrow \mathbf{Q}[A] = \mathbf{P}[A] \\ &\Longrightarrow 1 - \mathbf{Q}[A^c] = 1 - \mathbf{P}[A^c] \\ &\Longrightarrow \mathbf{Q}[A^c] = \mathbf{P}[A^c] \\ &\Longrightarrow A^c \in \mathcal{G}. \end{aligned}$$

□ *Dado uma família de conjuntos disjuntos* $A_1, A_2, \dots \in \mathcal{G} \Longrightarrow \bigcup_{k=1}^{\infty} A_k \in \mathcal{G}$

$$\begin{aligned} \underbrace{A_1, A_2, \dots}_{\text{disjuntos}} \in \mathcal{G} &\Longrightarrow \mathbf{Q}[A_k] = \mathbf{P}[A_k], \forall k \\ &\Longrightarrow \underbrace{\sum_{k=1}^{\infty} \mathbf{Q}[A_k]}_{=\mathbf{Q}[\bigcup_{k=1}^{\infty} A_k]} = \underbrace{\sum_{k=1}^{\infty} \mathbf{P}[A_k]}_{=\mathbf{P}[\bigcup_{k=1}^{\infty} A_k]} \\ &\Longrightarrow \bigcup_{k=1}^{\infty} A_k \in \mathcal{G}. \end{aligned}$$

Observe que esta última afirmação não poderia ser provada se os conjuntos A_k *não fossem disjuntos. Isso ilustra a necessidade do Teorema* $\pi - \lambda$ *de Dynkin.* □

2.14 Definição (Medida nula e μ-negligenciável) *Dado* $(\Omega, \mathcal{F}, \mu)$ *um espaço de medida. Então*

a *Um conjunto* $A \in \mathcal{F}$ *é dito de medida nula se* $\mu(A) = 0$.

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

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

Todo espaço de medida pode ser completado.

2.16 Teorema (O completamento $(\Omega, \bar{\mathcal{F}}, \bar{\mu})$) *Sejam* $(\Omega, \mathcal{F}, \mu)$ *um espaço de medida e* $\mathcal{N}_\mu$ *ser a família dos conjuntos* μ-*negligenciáveis.*

Então

a $\bar{\mathcal{F}} := \sigma\langle \mathcal{F}, \mathcal{N}_\mu \rangle = \{F \cup N : F \in \mathcal{F}, N \in \mathcal{N}_\mu\}$;

b *A função* $\bar{\mu}$ *em* $\bar{\mathcal{F}}$ *definida por* $\bar{\mu}(F \cup N) = \mu(F)$ *para* $F \in \mathcal{F}$ *e* $N \in \mathcal{N}_\mu$ *é a única extensão de* μ *para uma medida em* $(\Omega, \bar{\mathcal{F}})$;

c *O espaço de medida* $(\Omega, \bar{\mathcal{F}}, \bar{\mu})$ *é completo.*

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

2.2 Teorema de Extensão de Carathéodory

O Teorema de Extensão de Carathéodory afirma que qualquer medida na álgebra $\mathcal{F}_0$ de subconjuntos de um dado conjunto Ω pode ser estendida à σ-álgebra gerada por $\mathcal{F}_0$, e essa extensão é única se a medida for σ-finita. Consequentemente, qualquer medida em um espaço contendo todos os intervalos de números reais pode ser estendida para a álgebra Borel do conjunto de números reais. Este é um resultado extremamente poderoso da Teoria de Medida e através dele provaremos, por exemplo, a existência da medida de Lebesgue.

Vamos apresentar uma técnica para a construção de medidas. Faremos a descrição para medidas de probabilidade mas a construção funciona para medidas finitas. E com um pouco mais de cuidado para medidas σ-finitas.

Na sessão seguinte estudaremos como determinar, e de certa forma construir, medidas em espaços gerais. Nos preocuparemos em especial com medidas de probabilidade, mas os argumentos para medidas quaisquer são similares, com alguns pequenos ajustes para incluir com cuidado os conjuntos de medida infinita.

Medidas Exteriores

A ideia fundamental dessa seção é estender uma função de conjuntos $\mathbf{P}$, que foi apenas parcialmente definida, para uma medida exterior $\mathbf{P}^*$.

A extensão é notavelmente simples e intuitiva. Ele também "funciona"com praticamente qualquer função não negativa $\mathbf{P}$. Só mais tarde precisamos impor condições mais fortes em $\mathbf{P}$, como a aditividade.

2.17 **Definição (Medida Exterior Induzida)** *Considere* $\mathcal{A}$ *uma família qualquer de subconjuntos de um espaço* Ω*, e* $\mathbf{P}\colon \mathcal{A} \to [0, \infty]$ *ser uma função de conjunto não-negativa. Suponha* $\varnothing \in \mathcal{A}$ *e* $\mathbf{P}(\varnothing) = 0$.

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

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

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

Primeiramente mostraremos que $\mathbf{P}^*$ satisfaz as seguintes propriedades básicas, que transformamos em uma definição por conveniência.

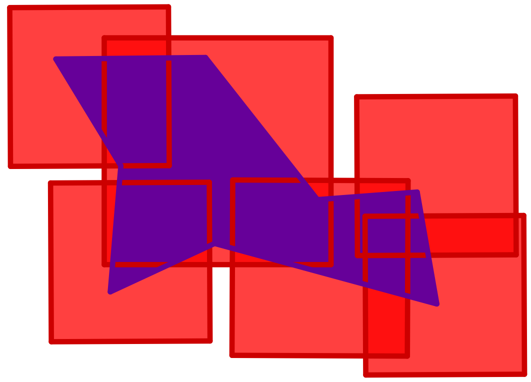
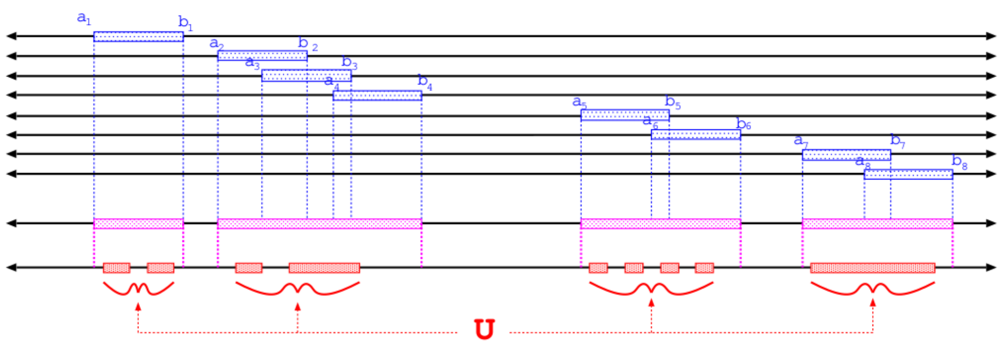


Figura 2.2: Medida externa

Figura 2.3: Os intervalos $(a_1, b_1], (a_2, b_2], \ldots, (a_8, b_8]$ cobrem U - [26]



2.18 Definição (Medida Exterior) *Uma função* $\mathbf{Q}\colon \mathcal{P}(\Omega) \to [0, \infty]$ *é uma* ***medida exterior*** *se:*

$\boxed{1}$ $\mathbf{Q}(\varnothing) = 0$.

$\boxed{2}$ *Monotonicidade* $\mathbf{Q}(E) \leq \mathbf{Q}(F)$ *quando* $E \subseteq F \subseteq \Omega$.

$\boxed{3}$ *Subaditividade enumerável Se* $E_1, E_2, \ldots \subseteq \Omega$*, então* $\mathbf{Q}(\bigcup_n E_n) \leq \sum_n \mathbf{Q}(E_n)$.

2.19 Teorema (Propriedades da Medida Exterior Induzida) *A medida exterior induzida* $\mathbf{P}^*$ *é uma medida exterior satisfazendo* $\mathbf{P}^*(A) \leq \mathbf{P}(A)$ *para todo* $A \in \mathcal{A}$.

Demonstração. *As propriedades 1. e 2. para uma medida exterior são obviamente satisfeitas por* $\mathbf{P}^*$*. E o fato que* $\mathbf{P}^*(A) \leq \mathbf{P}(A)$ *para* $A \in \mathcal{A}$ *também é óbvio, bastando observar que* A *é uma cobertura de* A.

A propriedade 3. é demonstrada por argumentos de aproximação. Seja $\varepsilon > 0$. Para cada E_n, pela definição de $\mathbf{P}^*$, existem conjuntos $\{A_{n,m}\}_m \subseteq \mathcal{A}$ cobrindo E_n, com

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

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

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

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

Nosso primeiro objetivo é procurar casos para os quais $\mathbf{Q} = \mathbf{P}^*$ seja σ-aditiva, de modo que se torne uma medida. Suspeitamos que possamos ter que restringir o domínio de $\mathbf{Q}$, e ainda assim este domínio deve ser suficientemente grande e ainda ser uma σ-álgebra.

Felizmente, há uma caracterização abstrata do "melhor"domínio para $\mathbf{Q}$, devido a Constantin Carathéodory:

2.20 Definição (Conjuntos Mensuráveis para a Medida Exterior) *Seja $\mathbf{Q}\colon \mathcal{P}(\Omega) \to [0,\infty]$ uma medida exterior. Então*

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

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

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

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

2.21 Heurística ([1]) *Suponha que $\mathbf{Q}^*$ seja uma medida externa finita em Ω*

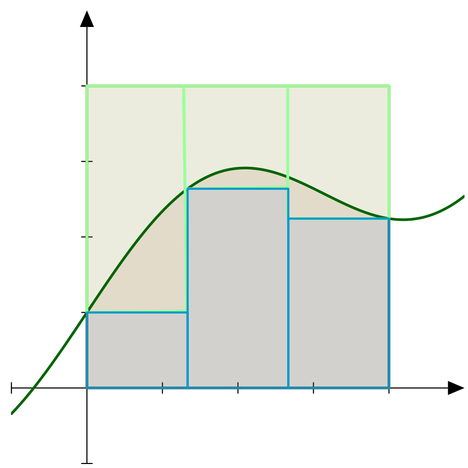
Podemos definir uma "medida interna"$\mathbf{Q}_*$ em Ω por

$$\mathbf{Q}_*(A) = \mathbf{Q}^*(\Omega) - \mathbf{Q}^*(A^c)$$

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

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

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



Claro, a construção de Carathéodory não exige que $\mathbf{Q}^*$ *seja finita, mas o argumento acima ainda nos fornece uma intuição decente para o caso geral*

O denominação "conjunto mensurável"é justificada pelo seguinte resultado sobre $\mathcal{M}$

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

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

$$\mathbf{Q}\big((A \cap B) \cap E\big) + \mathbf{Q}\big((A \cap B)^{\complement} \cap E\big) \tag{2.1}$$
$$= \mathbf{Q}(A \cap B \cap E) + \mathbf{Q}\big((A^{\complement} \cap B \cap E) \cup (A \cap B^{\complement} \cap E) \cup (A^{\complement} \cap B^{\complement} \cap E)\big) \tag{2.2}$$
$$\leq \mathbf{Q}(A \cap B \cap E) + \mathbf{Q}(A^{\complement} \cap B \cap E) + \mathbf{Q}(A \cap B^{\complement} \cap E) + \mathbf{Q}(A^{\complement} \cap B^{\complement} \cap E) \tag{2.3}$$
$$= \mathbf{Q}(B \cap E) + \mathbf{Q}(B^{\complement} \cap E)\,, \quad \text{pela definição de } A \in \mathcal{M} \tag{2.4}$$
$$= \mathbf{Q}(E)\,, \quad \text{pela definição de } B \in \mathcal{M}\,. \tag{2.5}$$

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

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

Para conveniência de notação, seja:

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

Precisamos também do seguinte fato

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

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

$$\mathbf{Q}(D_N \cap E) = \mathbf{Q}\big(D_{N-1} \cap (D_N \cap E)\big) + \mathbf{Q}\big(D_{N-1}^{\complement} \cap (D_N \cap E)\big) \tag{2.6}$$

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

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

Assim

$$\mathbf{Q}(E) = \mathbf{Q}(D_N \cap E) + \mathbf{Q}(D_N^{\complement} \cap E) \tag{2.9}$$

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

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

Tomando $N \longrightarrow \infty$, obtemos

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

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

2.23 Teorema (Aditividade Enumerável da Medida Exterior) *Dado uma medida exterior $\mathbf{Q}$, então $\mathbf{Q}$ é σ-aditiva em $\mathcal{M}$. Isso é, se $B_1, B_2, \ldots \in \mathcal{M}$ forem disjuntos, então $\mathbf{Q}(\bigcup_n B_n) = \sum_n \mathbf{Q}(B_n)$.*

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

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

A desigualdade na outra direção está implícita na subaditividade. □

Resumimos o que obtivemos até agora

2.24 Corolário (Teorema de Carathéodory) *Se $\mathbf{Q}$ for uma medida exterior (2.18), então a restrição de $\mathbf{Q}$ para os conjuntos mensuráveis $\mathcal{M}$ (2.20) produz uma medida em $\mathcal{M}$.*

Definindo uma Medida por Extensão

2.25 Teorema (Teorema de Extensão de Carathéodory) *Dada uma pré-medida de probabilidade $\mathbf{P}_0$ na álgebra $\mathcal{F}_0$ então esta possui uma única extensão para a medida de probabilidade $\mathbf{P}$ em $\sigma\langle\mathcal{F}_0\rangle =: \mathcal{F}$.*

Demonstração. *A unicidade segue diretamente do Teorema 2.13 pois $\mathcal{F}_0$ é um π-sistema.*

Para provar a existência da extensão considere $A \in \mathcal{F}_0$. Para qualquer $E \subseteq \Omega$ e $\varepsilon > 0$, por definição de $\mathbf{P}^*$ podemos encontrar $B_1, B_2, \ldots \in \mathcal{F}_0$ cobrindo E tal que $\sum_n \mathbf{P}(B_n) \leq \mathbf{P}^*(E) + \varepsilon$. Usando as propriedades das medidas exteriores induzidas temos

$$\mathbf{P}^*(A \cap E) + \mathbf{P}^*(A^\complement \cap E) \leq \mathbf{P}^*\Big(A \cap \bigcup_n B_n\Big) + \mathbf{P}^*\Big(A^\complement \cap \bigcup_n B_n\Big) \tag{2.12}$$

$$\leq \sum_n \mathbf{P}^*(A \cap B_n) + \sum_n \mathbf{P}^*(A^\complement \cap B_n) \tag{2.13}$$

$$\leq \sum_n \mathbf{P}(A \cap B_n) + \sum_n \mathbf{P}(A^\complement \cap B_n) \tag{2.14}$$

$$= \sum_n \mathbf{P}(B_n)\,, \quad \textit{da aditividade finita de } \mathbf{P}, \tag{2.15}$$

$$\leq \mathbf{P}^*(E) + \varepsilon\,. \tag{2.16}$$

Tomando $\varepsilon \downarrow 0$, temos que A é mensurável para $\mathbf{P}^*$.

Agora mostraremos que $\mathbf{P}^*(A) = \mathbf{P}(A)$. Considere qualquer cobertura de A por $B_1, B_2, \ldots \in \mathcal{F}_0$. Pela subaditividade enumerável e monotonicidade de $\mathbf{P}$,

$$\mathbf{P}(A) = \mathbf{P}\Big(\bigcup_n A \cap B_n\Big) \leq \sum_n \mathbf{P}(A \cap B_n) \leq \sum_n \mathbf{P}(B_n)\,.$$

Tomando o ínfimo sobre todas as coberturas temos que $\mathbf{P}(A) \leq \mathbf{P}^*(A)$. A outra desigualdade $\mathbf{P}(A) \geq \mathbf{P}^*(A)$ é automática. □

2.3 Classes Compactas

Agora, damos condições suficientes para que uma medida finitamente aditiva finita seja uma pré-medida e consequentemente uma medida. As condições são baseadas em uma propriedade relacionada à propriedade topológica de compacidade.

2.26 Definição (Classe Compacta) *Dizemos que $\mathcal{K}$ é uma **classe compacta** se para toda sequência $C_n \subset \mathcal{K}$ tal que $\bigcap_{n=1}^\infty C_n = \varnothing$ tivermos que existe $m \in \mathbb{N}$ tal que $C_1 \cap \cdots \cap C_m = \varnothing$.*

2.27 Exemplo *A família de conjuntos compactos num espaço métrico é uma classe compacta.*

Provaremos esse fato por contradição. Suponha que nenhuma intersecção finita seja vazia. Defina o conjunto $B_n = \bigcap_{m=1}^{n} C_m$. *Então cada* B_n *não é vazio e compacto. Escolha uma sequência* $x_1, x_2, \ldots,$ *tal que* $x_k \in B_k$. *Observe que esta sequência tem uma subsequência convergente porque está dentro do conjunto compacto* C_1. *Podemos mostrar que o limite dessa sequência está contido em cada* C_n, *porque a cauda da sequência está contida em* C_n *e* C_n *é fechado. Isso é uma contradição porque provamos que* $\bigcap_{n=1}^{\infty} C_n$ *contém um ponto.*

O mesmo resultado é verdadeiro para um espaço topológico de Hausdorff. Para demonstrar isso considere uma sequência de conjuntos $C_n \subset \mathcal{K}$. *Se* $\bigcap_{n=1}^{\infty} C_n = \varnothing$, *então*

$$C_0 = C_0 \setminus \emptyset = C_0 \setminus \bigcap_{n=1}^{\infty} C_n = \bigcup_{n=1}^{\infty} (C_0 \setminus C_n)$$

Mas $C_0 \setminus C_n$ *é aberto para todo* C_n, *então pela compacidade de* C_0 *temos que existem um* $\{\alpha_1,\ldots,\alpha_m\}$ *tais que*

$$C_0 = \bigcup_{i=1}^{m} (C_0 \setminus C_{\alpha_i}) = C_0 \setminus \bigcap_{i=1}^{m} C_{\alpha_i} \Longrightarrow \bigcap_{i=1}^{n} C_{\alpha_i} = \emptyset.$$

Exercício: Se $\mathcal{K}$ é uma classe compacta, então, a menor família, incluindo $\mathcal{K}$ e fechada sob uniões finitas e intersecções enumeráveis também uma classe compacta. □

2.28 Teorema (Teorema de Extensão Compacta) *Sejam* μ *uma medida finitamente aditiva definida numa álgebra* $\mathcal{F}_0$ *de subconjuntos de* Ω *e* $\mathcal{K}$ *uma subclasse compacta de* $\mathcal{F}_0$, *e suponha que, para cada* $A \in \mathcal{F}_0$, *tenhamos*

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

Então, μ *é pré medida em* $\mathcal{F}_0$.

Demonstração. *Suponha que os conjuntos* $A_n \in \mathcal{F}_0$ *são decrescentes e que sua intersecção seja vazia. Mostraremos que* $\mu(A_n) \to 0$. *Para isso considere* $\varepsilon > 0$. *Para cada* n *escolha* $C_n \in \mathcal{K}$ *com* $C_n \subset A_n$ *e* $\mu(A_n) \le \mu(C_n) + \frac{\varepsilon}{2^n}$. *Temos então que*

$$\left(\bigcap_{i=1}^{n} A_i\right) \setminus \left(\bigcap_{i=1}^{n} C_i\right) \subset \left(\bigcup_{i=1}^{n} A_i \setminus C_i\right) \tag{2.17}$$

Observe que $K_n = \bigcap_{i=1}^{n} C_i \subset \bigcap_{i=1}^{n} A_i \downarrow \varnothing$ *logo* $K_n \downarrow \varnothing$. *Como os conjuntos formam uma classe compacta existe* m *tal que* $K_m = \varnothing$.

Como os conjuntos A_n *são encaixantes temos que para* $n > m$ *a Equação (2.17) se reduz a*

$$A_n \subset \left(\bigcup_{i=1}^{n} A_i \setminus C_i\right).$$

e assim

$$\mu(A_n) \leq \sum_{i=1}^{n} \mu(A_i \setminus C_i) < \varepsilon$$

e logo $\lim_{n\to\infty} \mu(A_n) = 0$.

2.4 Aplicações

Teorema dos Números Normais de Borel

Nessa seção construiremos o primeiro modelo de espaço de probabilidade não trivial. Esse modelo fundamental representará a escolha de um número uniformemente no intervalo $(0,1]$, bem como representará o lançamento de uma moeda infinitas vezes. Ou seja, construiremos a medida de Lebesgue em $(0,1]$.

A construção desse modelo é paradigmática. Os passos que utilizaremos nessa construção se repetem em diversas outras construções.

Vamos começar colocando em destaque quais são esses passos:

Construção de Probabilidades

a Começaremos construindo uma probabilidade finitamente aditiva. Essa construção em geral é simples e intuitiva.

b Provaremos a σ-aditividade. Em alguns casos pode ser mais fácil provar a condição equivalente de continuidade no vazio. Em outros provar a condição do Teorema da Extensão Compacta.

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

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

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

2.30 Definição *Seja $\mathbf{P}$ uma atribuição de probabilidade em $\mathcal{B}_0((0,1])$ tal que:*

1 $\mathbf{P}[(a,b]] = b - a$ *para todo* $0 \leq a \leq b \leq 1$

2 *e estenda para todos os conjuntos de* $\mathcal{B}_0((0,1])$ *usando a identidade* $\mathbf{P}[A \cup B] = \mathbf{P}[A] + \mathbf{P}[B]$ *se* A,B *são conjuntos disjuntos em* $\mathcal{B}_0((0,1])$.

2.31 Teorema *A função de probabilidade finita* $\mathbf{P}$ *está bem definida.*

Demonstração. *Exercício* □

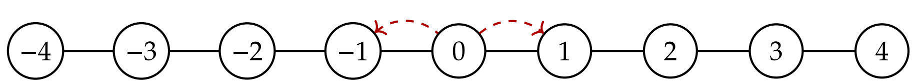
Podemos ver um número real $\omega \in (0,1]$ como a realização do lançamento de uma moeda infinitas vezes. Para isso considere a expressão binária desse número. Dessa forma $\omega \in (0,1]$ pode ser visto como uma sequência de dígitos em $\{0,1\}^{\mathbb{N}}$. Nessa representação estamos descartando o dígito 0 inicial de todas as sequência.

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

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

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

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



◁

Provaremos que

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

Veja que na expressão acima o limite está fora da probabilidade e como veremos o evento que está dentro da probabilidade é um evento na álgebra de Borel. Assim expressão anterior faz sentido com o que desenvolvemos até esse instante.

2.34 Teorema (Lei Fraca dos Grandes Números - Teorema de Bernoulli) *Para todo* $\varepsilon > 0$,

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

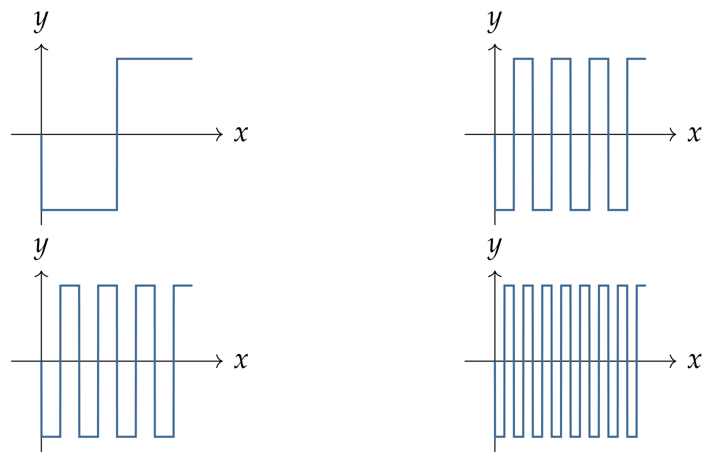


Figura 2.4: Funções de Rademacher z_k.

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

Cada intervalo diádico de nível $i-1$ se divide em dois no nível i. Logo a função de Rademacher toma valor $+1$ em metade do intervalo e valor -1 na outra metade. De modo mais geral suponha que $i < j$. Em um intervalo diádico de grau $j-1$, $z_i(w)$ é constante e $z_i(w)$ tem valor -1 na metade esquerda e $+1$ na direita. O produto $z_i(w)z_j(w)$, portanto, integra 0 sobre cada um dos intervalos diádicos de nível $j-1$ e logo

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

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

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

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

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

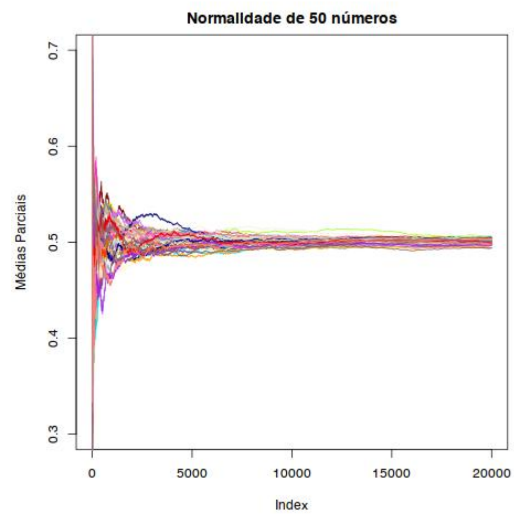


Figura 2.5: Dado uma base, um número normal é um número real cujos algarismos aparecem todos com a mesma frequência.

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

2.35 Definição *O conjunto dos* ***números normais***[1] *em* $(0,1]$ *é definido como*

$$N := \{\omega \in (0,1] : \lim_{n\to\infty} s_n(\omega)/n = 0\}$$
$$= \{\omega \in (0,1] : \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} d_k(\omega) = \frac{1}{2}\}.$$

O conjunto dos ***números anormais*** *é definido como* $A := (0,1] - N$.

Provaremos que os números anormais formam um conjunto negligenciável.

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

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

[1] na base 2

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

2.37 Teorema (Teorema dos Números Normais de Borel) *O conjunto dos números anormais, A, é negligenciável.*

Demonstração.

$\boxed{1}$ *Começaremos notando as seguintes inclusões:*

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

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

$\boxed{2}$ *Agora por (1.) temos que*

$$\begin{aligned}A = N^{\mathsf{c}} \subset \{\omega\colon \left|\frac{s_{k^2}(\omega)}{k^2}\right| \geq \varepsilon_k \text{ para infinitos } k\}\\ \subset \bigcup_{k=j}^{\infty} \underbrace{\{\omega\colon \left|\frac{s_{k^2}(\omega)}{k^2}\right| \geq \varepsilon_k\}}_{=:B_k}, \quad \text{para todo } j\end{aligned}$$

onde $B_k \in B_0^{(0,1])}$. Da demonstração da Lei Fraca

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

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

Esse é o máximo que podemos ir só com probabilidade finitamente aditiva. Gostaríamos de poder "passar o limite para dentro" e dizer que os conjuntos anormais tem medida nula.

Para isso precisamos estender esse modelo para um modelo de probabilidade. Começaremos provando a continuidade.

2.38 Teorema *A aplicação $\mathbf{P} : \mathcal{B}_0((0,1]) \to [0,1]$ definida pela Definição 2.30 é uma pré-medida de probabilidade.*

Demonstração. *Usaremos o Teorema 2.25 e demonstraremos que $\mathbf{P}$ é contínua por cima em $\emptyset$. Dado $A_n \downarrow \emptyset$ onde $A_n \in \mathcal{B}_0((0,1])$ (em particular temos que $\bigcap_{k=1}^{\infty} A_k = \emptyset$). Provaremos que $\mathbf{P}[A_n] \downarrow 0$.*

Observe que para todo $n \geq N$ temos

$$\mathbf{P}[A_n] \leq \mathbf{P}[A_N] = \mathbf{P}\Big[\bigcap_{k=1}^{N} A_k\Big]$$

como os conjuntos A_k são decrescentes. É suficiente demonstrar que para todo $\varepsilon > 0$ existe N_ε tal que

$$\mathbf{P}\Big[\bigcap_{k=1}^{N_\varepsilon} A_k\Big] \leq \varepsilon.$$

O argumento seguinte não funciona, mas ele vai motivar a solução. Dado $\varepsilon > 0$ e para todo A_k construa uma sequência de fechados F_k tal que $F_k \subset A_k$ e $\mathbf{P}[A_k - F_k] \leq \varepsilon/2^k$

(Se A_k tem a forma $\bigcup_i (a_i, b_i]$ tome $F_k := \bigcup_i [a_i + \tau, b_i]$ para τ suficientemente pequeno).

Como $\bigcap_{k=1}^{\infty} A_k = \emptyset$ temos que $\bigcap_{k=1}^{\infty} F_k = \emptyset$. Por compacidade existe N_ε tal que $\bigcap_{k=1}^{N_\varepsilon} F_k = \emptyset$. Logo

$$\mathbf{P}\Big[\bigcap_{k=1}^{N_\varepsilon} A_k\Big] - \underbrace{\mathbf{P}\Big[\underbrace{\bigcap_{k=1}^{N_\varepsilon} F_k}_{=\emptyset}\Big]}_{=0} \leq \sum_{k=1}^{N_\varepsilon} \mathbf{P}\big[A_k - F_k\big] \leq \sum_{k=1}^{N_\varepsilon} \frac{1}{2^k} \leq \varepsilon.$$

Isso estabeleceria 24 se não fosse pelo problema que $\mathbf{P}$ não está necessariamente definida nos F_k.

Agora é claro como remediar a situação. Para todo A_k encontre conjuntos fechados F_k e subconjuntos A_k^o de $\mathcal{B}_0((0,1])$ tais que $A_k \supset F_k \supset A_k^o$ e $\mathbf{P}[A_k - A_k^o] \leq \varepsilon/2^k$. Para todo $\varepsilon > 0$ temos que existe N_ε tal que $\bigcap_{k=1}^{N_\varepsilon} F_k = \emptyset$ o que implica $\bigcap_{k=1}^{N_\varepsilon} A_k^o = \emptyset$ e assim

$$\mathbf{P}\Big[\bigcap_{k=1}^{N_\varepsilon} A_k\Big] - \underbrace{\mathbf{P}\Big[\underbrace{\bigcap_{k=1}^{N_\varepsilon} A_k^o}_{=\emptyset}\Big]}_{=0} \leq \sum_{k=1}^{N_\varepsilon} \mathbf{P}\big[A_k - A_k^o\big] \leq \sum_{k=1}^{N_\varepsilon} \frac{1}{2^k} \leq \varepsilon.$$

□

2.39 Teorema (Medida de Lebesgue) *Dado $\mathbf{P} : \mathcal{B}_0((0,1]) \to [0,1]$ como na Definição 2.30. Então*

$\boxed{1}$ $\mathbf{P}$ *admite uma única extensão para uma medida de probabilidade em $\mathcal{B}((0,1])$ (também denotada $\mathbf{P}$);*

2 $\mathbf{P}$ *é a única medida em* $\mathcal{B}((0,1])$ *que satisfaz* $\mathbf{P}[(0,x]] = x$ *para todo* $x \in (0,1]$;

Demonstração.

□ *Demonstração do item 2.39.1*

O Teorema de Extensão de Carathéodory com o Teorema 2.38 mostra que existe uma única extensão $\mathbf{P} : \mathcal{B}_0((0,1]) \longrightarrow [0,1]$ *para* $\mathbf{P} : \mathcal{B}((0,1]) \longrightarrow [0,1]$ *como* $\mathcal{B}((0,1]) = \sigma\left\langle \mathcal{B}_0((0,1])\right\rangle$.

□ *Demonstração do item 2.39.2*

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

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

2.40 Corolário *Dado* $\mathbf{P} : \mathcal{B}_0((0,1]) \longrightarrow [0,1]$ *como na Definição 2.30. Então*

1 $N \in \mathcal{B}((0,1])$ *e* $\mathbf{P}[N] = 1$ *onde* N *é o conjunto dos números normais em* $(0,1]$;

2 $\mathcal{B}_0((0,1]) \subsetneq \mathcal{B}((0,1]) \subsetneq \overline{\mathcal{B}((0,1])} \subsetneq \mathcal{P}(\Omega)$. *Conjuntos em* $\mathcal{B}((0,1])$ *são denominados* ***Borel mensuráveis***. *Conjunto em* $\overline{\mathcal{B}((0,1])}$ *são denominados* ***Lebesgue mensuráveis***.

Demonstração.

□ *Demonstração do item 2.40.1*

Observamos inicialmente que N *e* N^{c} *estão ambas em* $\mathcal{B}((0,1])$. *Agora pelo Teorema 2.37 temos que* N^{c} *é negligenciável. Ou seja dado* $\varepsilon > 0$, *existe* $B_n \in \mathcal{B}_0((0,1])$ *tal que* $N^{\mathsf{c}} \subset \bigcup_{n=1}^{\infty} B_n$ *onde* $\sum_{n=1}^{\infty} \mathbf{P}[B_n] \leq \varepsilon$. *Como* $\bigcup_{n=1}^{\infty} B_n \in \mathcal{B}((0,1])$ *temos*

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

Logo

$$\mathbf{P}(N^{\mathsf{c}}) = \inf\{\mathbf{P}(B) \colon N^{\mathsf{c}} \subset B \in \mathcal{F}\} \leq \varepsilon$$

para todo ε. *Logo* $\mathbf{P}(N^{\mathsf{c}}) = 0$ *e* $\mathbf{P}(N) = 1$.

□

Nós já mencionamos que $N \in \mathcal{B}((0,1))$ *mas* $N \notin \mathcal{B}_0((0,1))$

A Proposição 2.50 mostra que $\mathcal{B}((0,1)) \subsetneq \overline{\mathcal{B}((0,1)}$.

O conjunto de Vitali mostra que é impossível estender P para uma medida de probabilidade em $\mathcal{P}(\Omega)$ *que estabelece que* $\overline{\mathcal{B}((0,1)} \subsetneq \mathcal{P}(\Omega)$.

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

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

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

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

Agora apresentaremos uma construção alternativa para o modelo de lançamento de infinitas moedas honestas.

Considere uma sequência infinita de lançamentos de moedas honestas. Desejamos construir um modelo probabilístico deste experimento de modo que cada sequência de resultados possível dos primeiros lançamentos tem a mesma probabilidade, $\frac{1}{2}^n$.

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

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

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

Álgebra e σ-álgebra de $\{0,1\}^n$

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

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

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

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

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

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

O evento A pertence a $\mathcal{F}_n$, e é fácil verificar que todos os elementos de $\mathcal{F}_n$ são desta forma.

Também é fácil verificar que $\mathcal{F}_n$ é uma σ-álgebra.

Exercício: Prove que $\mathcal{F}_n$ é uma σ-álgebra. □

As σ-álgebras $\mathcal{F}_n$, para qualquer n fixo, são muito pequenas. Elas só modelam os primeiros n lançamentos de moedas. Estamos interessados, em vez disso, em conjuntos que pertencem a $\mathcal{F}_n$, para n arbitrário, e isso nos leva a definir

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

a coleção de conjuntos que pertencem ao $\mathcal{F}_n$ para algum n. Intuitivamente, $A \in \mathcal{F}_0$ se a ocorrência ou não ocorrência de A pode ser decidido após um número fixo de jogadas de moedas.

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

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

Exercício: Prove que $\mathcal{F}_0$ é um álgebra. □

Gostaríamos de ter um modelo de probabilidade que atribua probabilidades a todos os eventos em $\mathcal{F}_n$, para cada n. Isso significa que precisamos de uma σ-álgebra que inclua $\mathcal{F}_0$. Por outro lado, gostaríamos que nossa σ-álgebra fosse tão pequena quanto possível, ou seja, contenha como poucos subconjuntos de $\{0,1\}^n$ quanto possível, para minimizar a possibilidade de incluir conjuntos patológicos aos quais as probabilidades não podem ser atribuídas . Isso nos leva a considerarmos F como a σ-álgebra gerada por $\mathcal{F}_0$.

Definindo a medida de Probabilidade

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

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

Antes de prosseguir, precisamos verificar se a definição acima é consistente. Observe que o mesmo conjunto A pode pertencer a $\mathcal{F}_n$ para vários valores de n. Portanto, precisamos verificar se, quando aplicamos a definição de $\mathbf{P}_0(A)$ para diferentes opções de n, obtemos o mesmo valor. Na verdade, suponha que $A \in \mathcal{F}_m$, que implica que $A \in \mathcal{F}_n$, para $n > m$. Neste caso,

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

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

É fácil verificar que $\mathbf{P}_0(\Omega) = 1$, e que $\mathbf{P}_0$ é finitamente aditiva. Também podemos provar que $(\mathcal{F}_0, \mathbf{P}_0)$ satisfaz

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

pois todo conjunto em $\mathcal{F}_0$ possui base finita e logo compacto na topologia produto.

Logo podemos invocar o Teorema de Extensão Compacta e concluir que existe uma única medida de probabilidade em $\mathcal{F}$, a σ-álgebra gerado por $\mathcal{F}_0$, que concorda com $\mathbf{P}_0$ em $\mathcal{F}_0$. Esta medida de probabilidade atribui a mesma probabilidade, $1/2^n$, a cada sequência possível de comprimento n, conforme desejado.

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

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

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

$$(\boldsymbol{i}, \boldsymbol{i}+1] \equiv (i_1, i_1+1] \times \cdots \times (i_d, i_d+1].$$

Esses conjuntos particionam o espaço,

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

e assim $\mathbb{R}^d$ pode ser decomposto como união enumerável de cubos unitários disjuntos. Seja $\mathcal{B}_0^{(\boldsymbol{i},\boldsymbol{i}+1]}$ a álgebra da união de finitos retângulos disjuntos em $(\boldsymbol{i},\boldsymbol{i}+1]$ e seja $\mathcal{B}((\boldsymbol{i},\boldsymbol{i}+1]) \equiv \sigma\langle\mathcal{B}_0^{(\boldsymbol{i},\boldsymbol{i}+1]}\rangle$ a σ-álgebra de Borel de $(\boldsymbol{i},\boldsymbol{i}+1]$. Finalmente denote por $\mathbf{P}_{\boldsymbol{i}}$ a única medida de probabilidade uniforme em $\mathcal{B}((\boldsymbol{i},\boldsymbol{i}+1])$ que atribui o volume Euclidiano aos retângulos em $(\boldsymbol{i},\boldsymbol{i}+1]$, i.e.

$$\mathbf{P}_{\boldsymbol{i}}\big((a_1,b_1]\times\cdots\times(a_d,b_d]\big) = \prod_{k=1}^{d}(b_k - a_k)$$

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

A construção de $\mathbf{P}_{\boldsymbol{i}}$ é feita exatamente da mesma forma que a medida de probabilidade uniforme foi construída em (0,1].

- Dado $A \in B_0^{(\boldsymbol{i},\boldsymbol{i}+1]}$ definimos $\mathbf{P}_{\boldsymbol{i}}(A)$ como sendo a soma dos volumes disjuntos retângulo que compõem A (este passo não é totalmente trivial uma vez que existem diferentes de composições de A em retângulos disjuntos, mas pode-se provar que P_{bsi} é está bem definido).

- Depois, mostra-se que $\mathbf{P}_{\boldsymbol{i}}$ é uma medida de probabilidade em $((\boldsymbol{i},\boldsymbol{i}+1],\mathcal{B}_0^{(\boldsymbol{i},\boldsymbol{i}+1]})$.A parte mais difícil deste passo é demonstrar a σ aditividade. No (0,1] nos usamos o fato equivalente que $\mathbf{P}_{\boldsymbol{i}}$ é contínua por cima no $\varnothing$. O argumento também funciona em $((\boldsymbol{i},\boldsymbol{i}+1],\mathcal{B}_0^{(\boldsymbol{i},\boldsymbol{i}+1]},\mathbf{P}_{\boldsymbol{i}})$.

- Finalmente utilizamos o Teorema de extensão de Carathéodory temos $((\boldsymbol{i},\boldsymbol{i}+1],\mathcal{B}((\boldsymbol{i},\boldsymbol{i}+1]),\mathbf{P}_{\boldsymbol{i}})$ (a unicidade segue do fato que os retângulos formam um π-sistema).

- A partir da medida de probabilidades $((\boldsymbol{i},\boldsymbol{i}+1],\mathcal{B}((\boldsymbol{i},\boldsymbol{i}+1]),\mathbf{P}_{\boldsymbol{i}})$ podemos definir a medida de Lebesgue $\mu_{\mathcal{L}}^d$ nos conjuntos $A \in \mathcal{B}((\boldsymbol{i},\boldsymbol{i}+1])$ como:

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

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

2.42 Teorema (Medida de Lebesgue) *$\mu_{\mathcal{L}}^d$ é uma medida em $(\mathbb{R}^d,\mathcal{B}(\mathbb{R}^d))$.*

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

- ***(i)*** $\mu_{\mathcal{L}}^d(A) \in [0,\infty]$: *Trivial.*

- ***(ii)*** $\mu_{\mathcal{L}}^d(\varnothing) = 0$: *Imediato pois* $\mathbf{P}_{\boldsymbol{i}}\big((\boldsymbol{i},\boldsymbol{i}+1]\cap\varnothing\big) = 0$.

□ ***(iii)*** *σ-aditividade: Suponha $A_1, A_2, \ldots \in \mathcal{B}(\mathbb{R}^d)$ são disjuntos. Então*

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

2.43 Teorema *$\mu_{\mathcal{L}}^d$ é a única medida em $(\mathbb{R}^d, \mathcal{B}((0,1]^d))$ que atribui o volume euclidiano padrão para os retângulos finitos, isto é:*

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

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

Demonstração. *Defina $\mathcal{P}$ como o π-sistema composto de retângulos finitos de $\{(a_1,b_1]\times\cdots\times(a_d,b_d] : -\infty < a_k < b_k < \infty\}$ e o conjunto vazio $\varnothing$. Então $\mathcal{B}(\mathbb{R}^d) = \sigma\langle\mathcal{P}\rangle$. Também observamos que $\mu_{\mathcal{L}}^d$ é σ-finita em $\mathcal{P}$ pois $\mu_{\mathcal{L}}^d((\boldsymbol{i},\boldsymbol{i}+1]) = 1$, $\mathbb{R}^d = \bigcup_{\boldsymbol{i}\in\mathbb{Z}^d}(\boldsymbol{i},\boldsymbol{i}+1]$ e cada $(\boldsymbol{i},\boldsymbol{i}+1]\in\mathcal{P}$. Logo o resultado decorre do Teorema 2.13.* □

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

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

Demonstração. *Para demonstrar que $A + x \in \mathcal{B}(\mathbb{R}^d)$ usaremos o Princípio dos Bons Conjuntos. Dado $x \in \mathbb{R}^d$ e o conjunto $\mathcal{G}_x := \{A \in \mathcal{B}(\mathbb{R}^d) : A + x \in \mathcal{B}(\mathbb{R}^d)\}$. É fácil ver que $\mathcal{G}_x$ é uma σ-álgebra. Por exemplo,*

$$\begin{aligned}
A \in \mathcal{G}_x &\Rightarrow A \in \mathcal{B}(\mathbb{R}^d) \text{ e } A + x \in \mathcal{B}(\mathbb{R}^d)\\
&\Rightarrow A^{\mathsf{c}} \in \mathcal{B}(\mathbb{R}^d) \text{ e } (A+x)^{\mathsf{c}} \in \mathcal{B}(\mathbb{R}^d)\\
&\Rightarrow A^{\mathsf{c}} \in \mathcal{B}(\mathbb{R}^d) \text{ e } A^{\mathsf{c}} + x \in \mathcal{B}(\mathbb{R}^d)\\
&\Rightarrow A^{\mathsf{c}} \in \mathcal{G}_x.
\end{aligned}$$

Para concluir a demonstração do teorema 2.44 usaremos a mesma ideia de 2.4327 sobre a unicidade de $\mu^d_{\mathcal{L}}$.

Para isso dado x *defina* $\mu_x(A) := \mu^d_{\mathcal{L}}(A+x)$. *Então* μ_x *é uma medida em* $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. *As medidas* μ_x *e* $\mu^d_{\mathcal{L}}$ *concordam no* π*-sistema dos retângulos finitos em* $\mathbb{R}^d$. *Logo pelo Teorema 2.43,* $\mu^d_{\mathcal{L}}(A) = \mu_x(A) := \mu^d_{\mathcal{L}}(A+x)$ *para todo* $A \in \mathcal{B}(\mathbb{R}^d)$. □

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

$$\mu^d_{\mathcal{L}}(TA) := |\det T| \mu^d_{\mathcal{L}}(A).$$

Demonstração. *Exercício*

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

Demonstração. *Dado* $\{B_i : i \in \mathcal{I}\}$ *uma família de conjuntos disjuntos tais que* $\mu(B_i) > 0$ *para todo* $i \in \mathcal{I}$. *Mostraremos que* $\mathcal{I}$ *deve ser enumerável.*

Como μ *é* σ*-finita existe* $A_1, A_2, \ldots \in \mathcal{F}$ *tal que* $\mu(A_k) < \infty$ *e* $\Omega = \bigcup_k A_k$. *Mostraremos os seguintes fatos:*

□ $\{i \in \mathcal{I} : \mu(A_k \cap B_i) > \varepsilon\}$ *é finita para todo* k : *Dado* $\varepsilon > 0$ *e suponha por contradição que existisse uma família enumerável de conjuntos* $\mathcal{I}_c \subset \mathcal{I}$ *tal que* $\mu(A_k \cap B_i) > \varepsilon$ *para todo* $i \in \mathcal{I}_c$ *e que*

$$\mu(A_k) \geq \mu(A_k \cap (\bigcup_{i \in \mathcal{I}_c} B_i)) = \sum_{i \in \mathcal{I}_c} \mu(A_k \cap B_i) > \sum_{i \in \mathcal{I}_c} \varepsilon = \infty$$

o que leva a uma contradição

□ $\{i \in \mathcal{I} : \mu(A_k \cap B_i) > 0\}$ *é enumerável para todo* k : *Esse fato segue de que*

$$\{i \in \mathcal{I} : \mu(A_k \cap B_i) > 0\} = \bigcup_{\varepsilon \text{ racional}} \underbrace{\{i \in \mathcal{I} : \mu(A \cap B_i) > \varepsilon\}}_{\text{finito por (i)}}.$$

□ $\mathcal{I} = \bigcup_k \{i \in \mathcal{I} : \mu(A_k \cap B_i) > 0\}$: *Para demonstrar que* $\mathcal{I} \cup \bigcup_k \{i \in \mathcal{I} : \mu(A_k \cap B_i) > 0\}$ *observe que* $i \in \mathcal{I}$ *então* $\mu(B_i) > 0$. *Como* $\Omega = \bigcup_k A_k$ *existe* k *tal que* $\mu(A_k \cap B_i) > 0$. *Logo* $i \in \bigcup_k \{i \in \mathcal{I} : \mu(A_k \cap B_i) > 0\}$. *A outra inclusão é óbvia.*

Para terminar a prova, basta notar que os últimos itens implicam $\mathcal{I}$ *é enumerável.* □

Se $k < d$ então $\mu^d_{\mathcal{L}}(A) = 0$ para todo hiperplano k-dimensional $A \subset \mathbb{R}^d$ com $k < d$.

Demonstração. *Dado* A *um hiperplano* k*-dimensional com* $k < d$. *Dado* x *um ponto em* $\mathbb{R}^d$ *que não está contido em* A. *então* $\{A + xt : t \in \mathbb{R}\}$ *é uma família não enumerável de subconjuntos de* $\mathcal{B}(\mathbb{R}^d)$. *Como* $\mu^d_{\mathcal{L}}$ *é invariante por translação* $\mu^d_{\mathcal{L}}(A) = \mu^d_{\mathcal{L}}(A + xt)$ *para todo* $t \in \mathbb{R}$. *Pelo Teorema 2.46,* $\mu^d_{\mathcal{L}}(A) = 0$. □

2.47 Definição (Conjuntos mensuráveis de Borel versus Lebesque) *Dado* $\left(\Omega, \overline{\mathcal{B}(\mathbb{R}^d)}, \overline{\mu^d_{\mathcal{L}}}\right)$ *o completamento de* $(\Omega, \mathcal{B}(\mathbb{R}^d), \mu^d_{\mathcal{L}})$*. Se* $A \in \mathcal{B}(\mathbb{R}^d)$ *então* A *é dito Borel mensurável. Se* $A \in \overline{\mathcal{B}(\mathbb{R}^d)}$ *então* A *é dito Lebesque mensurável.*

2.5 ↯Como são, como vivem e e quantos são os Borelianos?

Nesta seção apresentaremos alguns resultados sobre a cardinalidade dos Borelianos e dos conjuntos mensuráveis. Esses resultados não serão utilizados no restante do texto.

Começaremos com uma proposição, cuja demonstração é simples e deixaremos como exercício.

2.48 Proposição

- *Todo subconjunto aberto de* $\mathbb{R}$ *é uma união enumerável de intervalos abertos e logo é um boreliano*

- *Todo subconjunto fechado de* $\mathbb{R}$ *é um Boreliano*

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

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

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

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

Procedendo desta maneira, obtemos uma sequência encaixante $C_0 \supset C_1 \supset C_2 \supset \cdots$ de conjuntos fechados, onde C_n é a união de 2^n intervalos fechados, como ilustrado na Figura 2.5. Então a interseção

$$C = \bigcap_n C_n$$

é o conjunto de Cantor.

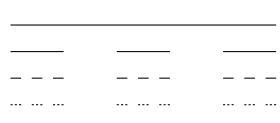


Figura 2.6: O conjunto de Cantor

2.49 Proposição *O conjunto de Cantor C tem as seguintes propriedades:*

1. *C é fechado, e logo é um Boreliano.*

2. *C é não enumerável. De fato, $|C| = |\mathbb{R}|$.*

3. $\mu(C) = 0$.

2.50 Proposição (Cardinalidade dos Conjuntos Lebesgue mensuráveis) *Seja $\mathcal{L}$ a coleção de todos os subconjuntos Lebesgue mensuráveis de $\mathbb{R}$. Então*

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

Demonstração. *Claramente $|\mathcal{L}| \leq |\mathcal{P}(\mathbb{R})|$. Mas como o conjunto de Cantor C tem medida de Lebesgue nula, temos que todo subconjunto do conjunto de Cantor é Lebesgue mensurável, ou seja, $\mathcal{P}(C) \subset L$. Mas como $|C| = |\mathbb{R}|$, segue que $\mathcal{P}(C) = \mathcal{P}(\mathbb{R})$ e, portanto, $\mathcal{P}(\mathbb{R}) \leq |\mathcal{L}|$.* □

2.51 Definição (Hierarquia Finita de Borel) *A hierarquia finita do Borel consiste em duas sequências Σ_n e Π_n de subconjuntos de $\mathcal{B}$ definidos da seguinte forma:*

- *Σ_1 é a coleção de todos os conjuntos abertos em $\mathbb{R}$ e Π_1 é a coleção de todos os conjuntos fechados em $\mathbb{R}$.*

- *Para cada $n \geq 1$, a coleção Σ_{n+1} consiste de todas as uniões enumeráveis de conjuntos de Π_n, e a coleção Π_{n+1} consiste de todas as interseções enumeráveis de conjuntos de Σ_n.*

2.52 Exemplo *Assim, cada conjunto em Σ_3 pode ser escrito como*

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

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

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

para conjuntos fechados $F_{m,n}$. ◁

Assim nós temos $\Sigma_1 \subseteq \Sigma_2, \Sigma_1 \subseteq \Pi_2, \Pi_1 \subseteq \Sigma_2$ e $\Pi_1 \subseteq \Pi_2$. E além disso, todas as quatro inclusões são próprias.

2.53 Teorema (Propriedades de Σ_n e Π_n) *Para cada $n \in \mathbb{N}$, as seguintes afirmações são válidas.*

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

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

Além disso, todas as inclusões são próprias.

Se $B \subseteq \mathbb{R}$ for um conjunto Borel, a **classificação de Borel** de B é o mínimo n tal que B se encontra em $\Sigma_n \subset \Pi_n$. Assim, conjuntos que são abertos ou fechados têm classificação de Borel 1.

Surpreendentemente, não é verdade que todo conjunto de Borel tenha classificação finita. Por exemplo, se $\{S_n\}$ for uma sequência de conjuntos do Borel, tal que cada S_n e tenha classificação n, então a união

$$S = S_1 \subset S_2 \subset S_3 \subset \dots$$

não pode ter classificação finita. Tal conjunto S é dito de classificação ω, e a coleção de todos esses conjuntos é conhecida como Σ_ω. O complemento de qualquer conjunto em Σ_ω também é dita de classificação ω, e a coleção de todos esses conjuntos é conhecida como Π_ω.

A hierarquia do Borel continua até além de ω. Por exemplo, $\Sigma_{\omega+1}$ consiste em todas as uniões enumeráveis de conjuntos de Π_ω e $\Pi_{\omega+1}$ consiste em todas as intersecções enumeráveis de conjuntos Σ_ω De fato, nós temos uma sequência de conjuntos

$$\subset \Sigma_2 \subset \Sigma_3 \subset \cdots \subset \Sigma_\omega \subset \Sigma_{\omega+1} \subset \Sigma_{2\omega} \subset \cdots \subset \Sigma_{2\omega} \subset \Sigma_{2\omega+1} \subset \cdots$$

e da mesma forma para o Π's. O resultado é que os conjuntos Σ_α e Π_α podem ser definidos para cada ordinal contável α (ou seja, para cada elemento de um conjunto bem ordenado e não enumerável e minimal, S_Ω). As famílias resultantes $\Sigma_\alpha : \alpha \in S_\Omega$ e $\Pi_\alpha : \alpha \in S_\Omega$ constituem a hierarquia de Borel completa, e a álgebra de Borel $\mathcal{B}$ é a união destas:

$$\mathcal{B} = \bigcup \Sigma_\alpha = \bigcup \Pi_\alpha.$$

Não é muito difícil provar que cada um dos conjuntos Σ_α e Π_α tem a mesma cardinalidade de $\mathbb{R}$. Como $|S_\Omega| \leq |\mathbb{R}|$, logo temos que a álgebra de Borel $\mathcal{B}$ tem cardinalidade de $|\mathbb{R}|$ também.

2.54 Teorema (Cardinalidade da Álgebra de Borel) *Seja $\mathcal{B}$ a σ-álgebra de Borel em $\mathbb{R}$. Então $|\mathcal{B}| = |\mathbb{R}|$.*

2.55 Proposição *A cardinalidade da coleção de conjuntos mensuráveis é igual a $|\mathcal{P}(\mathbb{R})|$,*

Como essa cardinalidade é maior que a cardinalidade da álgebra de Borel. Isso produz o seguinte corolário.

2.56 Corolário *Existe um conjunto mensurável de Lebesgue que não é um conjunto de Borel.*

2.6 Medidas em Espaços Métricos

Escrever: espaços poloneses, regularidade e tightness.

Exercícios

Ex. 2.1 — Dados $A_1, A_2, \ldots$. Mostre que $\mathbf{P}(A_n \quad i.v._n) = 1$ se e somente se $\sum_{n=1}^{\infty} \mathbf{P}(A_n|A)$ diverge para todo A de probabilidade não nula.

Ex. 2.2 — Suponha que temos uma pré-medida μ em um espaço X tal que $\mu(X) < \infty$. Prove que $E \subset X$ é mensurável Carathéodory se e somente se $\mu^*(E) + \mu^*(E^C) = \mu(X)$.

Ex. 2.3 — [Integral de Lebesgue-Stieltjes] Dado $\Omega = \mathbb{R}$ e $\mathcal{B}_0(\mathbb{R}) := f\langle(-\infty, a] : -\infty < a < \infty\rangle$ a álgebra de Borel de $\mathbb{R}$. Seja P uma probabilidade finitamente aditiva em $\mathcal{B}_0(\mathbb{R})$. Mostre que P é uma medida de probabilidade em $\mathcal{B}_0(\mathbb{R})$ se e somente se a função definida por $F(x) := P((-\infty, x])$ é não decrescente, continua a direita e satisfaz $\lim_{x\to-\infty} F(x) = 0$ e $\lim_{x\to\infty} F(x) = 1$.

Ex. 2.4 — (sub-additividade enumerável) Mostre que

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

Ex. 2.5 — Prove que a continuidade da medida de probabilidade. Dados $A_1, A_2, \ldots$ eventos, então

1. Se $A_n \uparrow A$, então $\mathbf{P}(A_n) \uparrow \mathbf{P}(A)$ quando $n \to \infty$.
2. Se $A_n \downarrow A$, então $\mathbf{P}(A_n) \downarrow \mathbf{P}(A)$ quando $n \to \infty$.
3. $\mathbf{P}(\liminf A_n) \leq \liminf \mathbf{P}(A_n) \leq \limsup \mathbf{P}(A_n) \leq \mathbf{P}(\limsup A_n)$.
4. $A_n \to A$ implica que $\mathbf{P}(A_n) \to \mathbf{P}(A)$ as $n \to \infty$.

Ex. 2.6 — Considere um modelo probabilístico cujo espaço amostral é a reta real. Mostre que

1. $\mathbf{P}([0, \infty)) = \lim_{n\to\infty} \mathbf{P}([0, n])$
2. $\lim_{n\to\infty} \mathbf{P}([n, \infty)) = 0$.

Ex. 2.7 — (Uma condição insuficiente) Dados $\Omega = \{1, 2, 3, 4\}$, e $A = \{\{1, 2\}, \{1, 3\}\}$.

1. Mostre que a σ-algebra $\mathcal{F}$ gerada por A é o conjunto das partes $2^(\Omega)$.
2. Mostre que existem duas medidas de probabilidade P_1 e P_2 em $(\Omega, \mathcal{F})$, tal que $P_1(E) = P_2(E)$ para qualquer $E \in A$, mas $P_1 \neq P_2$. Este problema mostra que o fato de que duas medidas de probabilidade coincidem com uma família geradora

de uma σ-álgebra não garante que elas sejam iguais. No entanto, isso é verdade se adicionarmos a suposição de que a família geradora A é fechada sob interseções finitas.

Ex. 2.8 — Desigualdade de Bonferroni
Definindo
$$S_1 := \sum_{i=1}^{n} \mathbb{P}(A_i),$$
e
$$S_2 := \sum_{1 \le i < j \le n} \mathbb{P}(A_i \cap A_j),$$
bem como
$$S_k := \sum_{1 \le i_1 < \cdots < i_k \le n} \mathbb{P}(A_{i_1} \cap \cdots \cap A_{i_k})$$
para todos os números inteiros de k em $\{3, \ldots, n\}$.
Então, para k ímpar
$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) \le \sum_{j=1}^{k} (-1)^{j-1} S_j$$
e para $k > 2$ par
$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) \ge \sum_{j=1}^{k} (-1)^{j-1} S_j.$$
(Se tiver com muita preguiça prove apenas para $k = 2,3$.)

Ex. 2.9 — Problema do Pareamento
No final de um dia agitado, os pais chegam ao jardim de infância para pegar seus filhos. Cada pai escolhe uma criança para levar a casa uniformemente ao acaso. Use o fórmula de inclusão-exclusão para mostrar que a probabilidade de pelo menos um pai escolhe seu próprio filho é igual a
$$1 - \frac{1}{2!} + \frac{1}{3!} \cdots + (-1)^{n-1} \frac{1}{n!}$$
E que essa probabilidade converge a $1 - 1/e$ quando $n \to \infty$

Ex. 2.10 — Qual é a probabilidade de que um ponto escolhido aleatoriamente no interior de um triângulo equilátero esteja mais perto do centro do que de suas bordas?

Ex. 2.11 — Agulhas de Buffon
Considere um plano horizontal dividido em faixas por uma série de linhas paralelas a distância fixa d, como as tábuas do chão.
E considere uma agulha, cujo comprimento seja igual à distância entre as linhas paralelas, que é jogada no plano aleatoriamente.

1. Defina precisamente o espaço de Probabilidade.
2. Prove que a probabilidade da agulha não interceptar uma das linhas paralelas é $\frac{2}{\pi}$.

Ex. 2.12 — O conjunto de números normais e anormais está em $\mathcal{B}((0,1])$.

Ex. 2.13 — Todos os subconjuntos contáveis, co-contáveis (i.e. complementos de conjuntos contáveis) e perfeitos de (0,1] estão em $\mathcal{B}((0,1])$. Em particular, o conjunto dos números irracionais em (0,1] é um conjunto de Borel.

Ex. 2.14 — Para todo número $\omega \in (0,1]$, denotamos por $d_k(\omega)$ o k-ésimo digito da representação de ω. Seja $z_k(\omega) := 2d_k(\omega) - 1$ e

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

Mostre que

$$M(t) := \int_0^1 e^{ts_n(\omega)}\, d\omega = \left(\frac{e^t + e^{-t}}{2}\right)^n \tag{2.18}$$

para todo $t \in \mathbb{R}$. Diferenciando com respeito à t, mostre que $\int_0^1 s_n(\omega)\, d\omega = M'(0) = 0$ e $\int_0^1 s_n^2(\omega)\, d\omega = M''(0) = n$.

A ideia é dividir a integral $\int_0^1$ nos intervalos diádicos da forma $(\frac{k-1}{2^n}, \frac{k}{2^n}]$

Ex. 2.15 — Seja Ω um espaço métrico com distância d. Ω é dito **separável** se existe um conjunto enumerável $\Omega_0 \subset \Omega$ que é denso em Ω (i.e., todo ponto em Ω é limite de uma sequência de pontos em Ω_0).

1. Mostre que $\sigma\langle\text{bolas abertas em } \Omega\rangle \subset \mathcal{B}(\Omega)$.
2. Mostre que $\sigma\langle\text{bolas abertas em } \Omega\rangle = \mathcal{B}(\Omega)$ se Ω é separável.
3. Mostre que $\Omega = \mathbb{R}$ é separável com a métrica usual
4. Conclua que $\sigma\langle\text{bolas abertas em } \mathbb{R}\rangle = \mathcal{B}(\mathbb{R})$.

Ex. 2.16 — Suponha que μ_1 e μ_2 são medidas em $\sigma\langle\mathcal{F}_0\rangle$ geradas pela classe $\mathcal{F}_0$. Suponha também que a desigualdade

$$\mu_1(A) \le \mu_2(A) \tag{2.19}$$

é válida para todo A em $\mathcal{F}_0$.

1. Mostre que se $\mathcal{F}_0$ é uma álgebra e μ_1 e μ_2 são σ-finitas em $\mathcal{F}_0$, então (2.19) vale para todo $A \in \sigma\langle\mathcal{F}_0\rangle$.
(Dica: primeiramente trate o caso em que μ_2 é finita.)

2. Mostre através de um exemplo que (2.19) pode não ser verdade para algum $A \in \sigma\langle \mathcal{F}_0 \rangle$ se:
   a) $\mathcal{F}_0$ é uma álgebra mas μ_1 e μ_2 não são σ-finita em $\mathcal{F}_0$;
   b) Se μ_1 e μ_2 são σ-finitas em $\mathcal{F}_0$, mas $\mathcal{F}_0$ é somente um π-sistema.

Ex. 2.17 — Príncipio de Littlewood Suponha que $\mathcal{F}_0$ é uma álgebra, μ é uma medida em $\sigma\langle \mathcal{F}_0 \rangle$ e μ é σ-finita em $\mathcal{F}_0$.

1. Suponha $B \in \sigma\langle \mathcal{F}_0 \rangle$ e $\epsilon > 0$. Mostre que existe uma sequência disjunta de $\mathcal{F}_0$-conjuntos $A_1, A_2, \ldots$ tais que $B \subset \bigcup_{n=1}^{\infty} A_n$ e $\mu\big(\bigcup_{n=1}^{\infty} A_n - B\big) \leq \epsilon$.
2. Suponha $B \in \sigma\langle \mathcal{F}_0 \rangle$, $\mu(B) < \infty$ e $\epsilon > 0$. Mostre que existe um $\mathcal{F}_0$-conjunto A tal que $\mu(A \triangle B) \leq \epsilon$.
3. Mostre através de um exemplo que a conclusão anterior pode ser falsa se B tiver uma medida infinita.

Ex. 2.18 — O seguinte exercício mostra que não importa em qual ordem somamos um número infinito de termos positivos. Isso é útil para mostrar rigorosamente os axiomas de medida para a medida de contagem em qualquer Ω. Observe que, se houver termos negativos, a ordem é importante.
Dado I um conjunto infinito e seja f uma função de I para $[0,\infty]$. A soma de f sobre I é definida como

$$\sum_{i \in I} f(i) := \sup\Big\{\sum_{i \in H} f(i) : \ H \subset I, H \text{ é finito}\Big\}.$$

Mostre que

1. para toda partição $I = \bigcup_{k \in K} I_k$ de I em subconjuntos disjuntos não-vazios I_k,

$$\sum_{i \in I} f(i) = \sum_{k \in K}\Big(\sum_{i \in I_k} f(i)\Big)$$

2. Se I é enumerável, então

$$\sum_{i \in I} f(i) = \lim_{n \to \infty} \sum_{m=1}^{n} f(i_m)$$

   para toda enumeração $i_1, i_2, \ldots$ dos pontos em I.

3. Prove que

$$\mu(A) := \sum_{a \in A} f(a), \forall A \subseteq X$$

   definida acima é uma medida

Ex. 2.19 — Mostre que

$$P\Big[|s_n/n| \geq \epsilon\Big] \leq 2e^{-n\epsilon^2/2}$$

para todo $\epsilon > 0$.
Dica : Use (2.18) e a desigualdade $(e^x + e^{-x})/2 \leq \exp(x^2/2)$ que é verdadeira para todo $x \in \mathbb{R}$.

CAPÍTULO 3

Independência

Intuitivamente, dois eventos A e B são independentes se a ocorrência do evento A não modifica a probabilidade de ocorrência do evento B. Como veremos, a definição formal é que A e B são independentes se

$$\mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B)$$

Para motivar essa definição, começaremos introduzindo a noção de probabilidade condicional.

Suponha que realizamos um experimento aleatório e nos é dada a informação parcial de que um evento B foi observado, com $\mathbf{P}(B) > 0$. Como essa informação altera nosso modelo? O primeiro ponto importante a ser considerado é que qualquer "resultado" fora de B deve ter probabilidade 0 nesse novo modelo. Em outras palavras, ao medirmos a nova probabilidade de um evento A, devemos considerar apenas o evento $A \cap B$, uma vez que $A \cap B^c$ deve ter probabilidade nula. Além disso, observamos que o conjunto B deve ter probabilidade 1 nessa nova medida. Outro ponto importante é que, as probabilidades de eventos dentro de B devem manter a mesma razão na nova medida que possuíam na medida anterior. Ou seja, se um evento $A \cap B$ era 3 vezes mais provável que o evento $C \cap B$ na medida original, ele deve ainda ser 3 vezes mais provável na nova medida.

Essas propriedades motivam a seguinte definição.

3.1 Definição (Probabilidade Condicional) *Dado um espaço de probabilidade* $(\Omega, \mathcal{F}, \mathbf{P})$ *e* $A, B \in \mathcal{F}$ *eventos com* $P(B) > 0$. *A* ***probabilidade condicional*** *de que* A *ocorre dado que* B *ocorre,* *indicado por $P(A|B)$, é definida como*

$$\mathbf{P}(A|B) = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)}$$

Observe que as propriedades que comentamos acima são respeitadas por essa definição.

- $\mathbf{P}(A|B)$ é proporcional a $\mathbf{P}(A \cap B)$;
- $\mathbf{P}(B|B) = 1$;
- Dados quaisquer $A, C \in \mathcal{F}$ com $\mathbf{P}(C \cap B) > 0$, então

$$\frac{\mathbf{P}(A|B)}{\mathbf{P}(A|C)} = \frac{\mathbf{P}(A \cap B}{\mathbf{P}(A \cap C)}.$$

Frequentemente pensamos na probabilidade condicional como a probabilidade quando restringimos ao evento B, tomando este como o novo espaço amostral.

3.2 Teorema *Dado $(\Omega, \mathcal{F}, \mathbf{P})$ um espaço de probabilidade e dado B com $P(B) > 0$, então $(B, \mathcal{F} \cap B, \mathbf{P}(|B))$ é um espaço de probabilidade.*

Demonstração. *Exercício.* □

3.3 Exemplo *Um estudante realiza um teste cuja duração máxima é de uma hora. Suponha que a probabilidade de que o estudante termine o teste em menos que x horas seja igual a $\frac{x}{2}$, para todo $0 \leq x \leq 1$. Então, dado que o estudante continua a realizar o teste após $0{,}75$ horas, qual é a probabilidade condicional de que o estudante utilize a hora completa?* ◁

Seja T_x o evento em que o estudante termina o teste em menos que x horas, $0 \leq x \leq 1$, e C o evento em que o estudante utiliza a hora completa ou seja, C é o evento em que o estudante não finalizou o teste em menos que 1 hora. Temos que o evento em que o estudante ainda está trabalhando após 0,75horas é o complemento do evento $T_{0,75}^{\complement}$, então a probabilidade desejada é

$$\mathbf{P}(C|T_{0,75}^{\complement}) = 0{,}8$$

3.4 Teorema (Lei da Probabilidade Total) *Dado $(\Omega, \mathcal{F}, \mathbf{P})$ um espaço de probabilidade. Seja $\{B_1, B_2, \ldots\}$ uma partição enumerável de Ω. Então*

$$\forall A \in \mathcal{F} : \mathbf{P}(A) = \sum_{i=1}^{\infty} \mathbf{P}(A|B_i)\,\mathbf{P}(B_i)$$

Demonstração.

$$\mathbf{P}(A) = \mathbf{P}\left(A \cap \left(\bigcup_{i=1}^{\infty} B_i\right)\right) \tag{3.1}$$

$$= \mathbf{P}\left(\bigcup_{i=1}^{\infty} (A \cap B_i)\right) \tag{3.2}$$

$$= \sum_{i=1}^{\infty} \mathbf{P}(A \cap B_i) \tag{3.3}$$

$$= \sum_{i=1}^{\infty} \mathbf{P}(A \mid B_i)\,\mathbf{P}(B_i) \tag{3.4}$$

3.5 Teorema (de Bayes) *Dado* $\mathbf{P}$ *uma medida de probabilidade num espaço de probabilidade* $(\Omega, \mathcal{F}, \mathbf{P})$. *Suponha que* $\mathbf{P}(A) > 0$ *e que* $\mathbf{P}(B) > 0$

Então

$$\mathbf{P}(B \mid A) = \frac{\mathbf{P}(A \mid B)\,\mathbf{P}(B)}{\mathbf{P}(A)}$$

Demonstração. *Da definição de probabilidade condicional temos:*

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

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

Manipulando temos:

$$\mathbf{P}(A \mid B)\,\mathbf{P}(B) = \mathbf{P}(A \cap B) = \mathbf{P}(B \mid A)\,\mathbf{P}(A)$$

Finalmente, dividindo por $\mathbf{P}(A)$ *temos:*

$$\mathbf{P}(B \mid A) = \frac{\mathbf{P}(A \mid B)\,\mathbf{P}(B)}{\mathbf{P}(A)}$$

3.6 Teorema *Sejam* $A_1, A_2, \ldots, A_n$ *eventos que formam uma partição do espaço amostral, e assuma que* $\mathbf{P}(A_i) > 0$ *para todo* i. *Então, para qualquer evento* B *tal que* $\mathbf{P}(B) > 0$, *temos que*

$$\mathbf{P}(A_i|B) = \frac{\mathbf{P}(A_i)\mathbf{P}(B|A_i)}{\mathbf{P}(B)} = \frac{\mathbf{P}(A_i)\mathbf{P}(B|A_i)}{\mathbf{P}(A_1)\mathbf{P}(B|A_1) + \ldots + \mathbf{P}(A_n)\mathbf{P}(B|A_n)}.$$

3.1 Independência

Um dos conceito mais significativo na teoria da probabilidade é o conceito de independência. Intuitivamente, dois eventos A e B são independentes se a ocorrência do evento B não trás nenhuma informação sobre a ocorrência de A. Em outras palavras, se a ocorrência de B não altera a probabilidade de A. Expressando isso de maneira mais forma chegamos a seguinte definição provisória.

3.7 Definição (Eventos Independentes - Definição Provisória) *Dado um espaço de probabilidade* $(\Omega,\mathcal{F},\mathbf{P})$*, dois eventos* A *e* B *com* $\mathbf{P}(A) > 0$ *e* $\mathbf{P}(B) > 0$ *são ditos* ***independentes*** *se* $\mathbf{P}(A|B) = \mathbf{P}(A)$.

A definição acima, além de limitada a eventos de probabilidade positiva, possui uma aparente inconsistência na maneira como é enunciado. A definição $\mathbf{P}(A|B) = \mathbf{P}(A)$ não trás a mesma simetria presente na expressão "A e B são independentes". De fato, baseado apenas nesta expressão parece mais razoável dizer que *A é independente de B*, mas não necessariamente que *B é independente de A*.

Mas note que se $\mathbf{P}(A|B) = \mathbf{P}(A) > 0$, então

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

justificando a simetria da definição.

Mais do que isso, para eventos de probabilidade positiva, $\mathbf{P}(A|B) = \mathbf{P}(A)$ ocorre se, e só se, $\mathbf{P}(A\cap B)/\mathbf{P}(B) = \mathbf{P}(A)$. Ou seja,

$$\mathbf{P}(A|B) = \mathbf{P}(A) \Leftrightarrow \mathbf{P}(A\cap B) = \mathbf{P}(A)\mathbf{P}(B).$$

Como a igualdade da direita não depende de que $\mathbf{P}(A) > 0$ e $\mathbf{P}(B) > 0$, temos assim uma definição mais geral e mais elegante de independência.

3.8 Definição (Eventos Independentes) *Dado um espaço de probabilidade* $(\Omega,\mathcal{F},\mathbf{P})$*, dois eventos* A *e* B *são ditos* ***independentes*** *se*

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

Resumindo: as duas definições são equivalentes no caso em que $\mathbf{P}(A),\mathbf{P}(B) > 0$, mas a segunda definição é preferível pois inclui eventos tais que $\mathbf{P}(A) = 0$ ou $\mathbf{P}(B) = 0$. Além disso, a segunda definição deixa claro por simetria que, quando A é independente de B, B também é independente de A.

3.9 Proposição *Um evento A é independente dele mesmo se, e somente se,* $\mathbf{P}(A) = 0$ *ou* $\mathbf{P}(A) = 1$.

Demonstração.

$$\mathbf{P}(A) = \mathbf{P}(A\cap A) = \mathbf{P}(A)^2$$

Resolvendo a equação do segundo grau em $\mathbf{P}(A)$ *temos que* $\mathbf{P}(A) = 0$ *ou* $\mathbf{P}(A) = 1$ □

3.10 Exemplo *Considere o exemplo de rolarmos dois dados não viciados. Considere o espaço amostral* $\Omega = \{(i,j) : i,j \in \{1,2,3,4,5,6\}\}$ *e* $\mathbf{P}$ *a medida de probabilidade equiprovável em* $(\Omega, 2^{\Omega})$ *tal que*

$$\mathbf{P}(\{(i,j)\}) = \frac{1}{36},$$

para todo $(i,j) \in \Omega$.

Defina agora os exemplos

$$\begin{aligned} A &= \{\textit{observamos 2 no primeiro dado}\} = \{(2,j) : j \in \{1,2,3,4,5,6\}\}, \\ B &= \{\textit{observamos 4 no segundo dado}\} = \{(i,4) : j \in \{1,2,3,4,5,6\}\}, \quad e \\ C &= \{\textit{a soma dos valores observados é 7}\} = \{(i,j) \in \Omega : i + j = 7\}. \end{aligned}$$

Observe que $\mathbf{P}(A) = \mathbf{P}(B) = \mathbf{P}(C) = 1/6$. *Além disso*

$$\mathbf{P}(A \cap B) = \frac{1}{36} = \frac{1}{6} \cdot \frac{1}{6} = \mathbf{P}(A)\mathbf{P}(B),$$

e analogamente $\mathbf{P}(A \cap C) = \mathbf{P}(A)\mathbf{P}(C)$ *e* $\mathbf{P}(B \cap C) = \mathbf{P}(B)\mathbf{P}(C)$.

Isso mostra que os eventos A *e* B *são independentes, assim como os eventos* A *e* C *e os eventos* B *e* C. ◁

A definição de independência para mais que dois eventos é um pouco mais sensível. No exemplo anterior descrevemos 3 eventos tais que cada par destes é independente, mas não parece razoável dizer que o trio A, B, C é independente. De fato, se sabemos a informação de que o valor do primeiro dado é 2 e do segundo é 4 nos dá que a soma dos valores é 6, e portanto

$$\mathbf{P}(C|A \cap B) = 0 \neq \mathbf{P}(C).$$

Isso nos mostra que a independência dois a dois, apesar de necessária, não funciona por si só como uma definição consistente de independência de múltiplos eventos.

Deixamos como exercício, enunciado a seguir, demostrar que satisfazer $\mathbf{P}(A \cap B \cap C) = \mathbf{P}(A)\mathbf{P}(B)\mathbf{P}(C)$ também não é suficiente para independência de três eventos.

Exercício: Construa um exemplo de espaço de probabilidade e 3 eventos A, B e C tais que $\mathbf{P}(A \cap B \cap C) = \mathbf{P}(A)\mathbf{P}(B)\mathbf{P}(C)$, mais os pares de eventos não são todos independentes. □

A solução para tal questão é dada da seguinte definição.

3.11 Definição (Eventos Independentes) *Uma coleção de subconjuntos* $\{A_k\}_{k \in \mathcal{K}}$ *da* σ*-álgebra* $\mathcal{F}$ *é dita* ***independente*** *se, para cada conjunto finito de índices* $\mathcal{H} \subset \mathcal{K}$ *a seguinte identidade vale:*

$$\mathbf{P}\Big(\bigcap_{h \in \mathcal{H}} A_h\Big) = \prod_{h \in \mathcal{H}} \mathbf{P}(A_h).$$

3.12 Exemplo (Lançamento uma moeda) *Este é **o** exemplo canônico de independência. Suponha que você jogue uma moeda duas vezes. Seja A ser o evento "na primeira jogada saiu cara" e B ser o evento que "na segunda jogada saiu coroa". Calculando $\mathbf{P}(A) = \frac{1}{2}$ e $\mathbf{P}(B) = \frac{1}{2}$. Para calcular a probabilidade da ocorrência de ambos eventos, listamos os casos e observamos que $\mathbf{P}(A \cap B) = \frac{1}{4}$. Assim, A e B são independentes.* ◁

3.13 Exemplo (Amostragem) *Suponha que uma urna contenha 5 bolas brancas e 5 bolas pretas. Agora escolhemos 2 bolas sem substituição.*

Considere os eventos:

$$A = \textit{primeira bola é branca}$$
$$B = \textit{segunda bola é branca.}$$

Os eventos A e B são independentes? Não. Depois de retirar uma bola branca, podemos dizer que a urna conterá apenas 4 bolas brancas e 5 bolas pretas. Então, $\mathbf{P}(B|A) = \frac{4}{4+5}$, e a probabilidade de B em si (quando não sabe o que acontece com a primeira bola) é $\mathbf{P}(B) = \frac{1}{2}$ (porque $\mathbf{P}(B) = \mathbf{P}(B^{\mathsf{c}})$, e assim o papel das cores é simétrico.).

Mas, por outro lado, se nós considerarmos o caso de retirada com substituição, então os eventos A e B serão independente. ◁

É frequente a independência aparecer como hipótese na construção de medidas. Isso acontece frequentemente quando dizemos que "5 moedas são lançadas de maneira independente" ou que "repetimos um experimento seguidas vezes de modo independente". O exemplo a seguir mostra como a hipótese de independência pode ser usada na construção da medida de probabilidade associada ao lançamento de infinitas moedas.

3.14 Exemplo *Jogamos uma moeda não viciada infinitas vezes de maneira independente. Queremos construir o espaço de probabilidade para esse problema. Vamos tomar $\Omega = \{0,1\}^{\infty}$. Vamos também considerar as σ-álgebra $\mathcal{F}_n$, $n \geq 1$ de eventos cilíndricos, dadas por*

$$\mathcal{F}_n := \{\{\omega \in \Omega : (\omega_1, \ldots, \omega_n) \in B\}; B \subset \{0,1\}^n\}.$$

A σ-álgebra que usaremos será dada então por $\mathcal{F} = \sigma\langle\mathcal{F}_0\rangle$, onde $\mathcal{F}_0$ é a álgebra dada por

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

Dado $n \geq 1$ defina agora o evento $H_n = \{\omega \in \Omega : \omega_n = 1\}$. Ou seja, H_n é o evento onde observamos cara no n-ésimo lançamento.

As hipóteses do problema nos dizem duas coisas importantes:

a *como a moeda é não viciada,* $\mathbf{P}(H_n) = 1/2$ *para todo* $n \geq 1$;

b $H_1, H_2, \ldots$ *são eventos independentes.*

Tome $\epsilon = (\epsilon_1, \ldots, \epsilon_n) \in \{0,1\}^n$ *uma sequência de* n *0's e 1's representando um resultado possível dos* n *primeiros lançamentos da coroa. Faça agora* $F_k(\epsilon) = H_k$ *se* $\epsilon_k = 1$ *e* $F_k(\epsilon) = H_k^c$ *se* $\epsilon_k = 0$, *e note que*

$$A(\epsilon) := \{\omega \in \Omega : \omega_k = \epsilon_k, k = 1, \cdots, n\} = \bigcap_{k=1}^{n} F_k(\epsilon) \in \mathcal{F}_n.$$

Vale assim que

$$\mathbf{P}(A(\epsilon)) = \mathbf{P}\left[\bigcap_{k=1}^{n} F_k(\epsilon)\right] = \prod_{k=1}^{n} \mathbf{P}(F_n(\epsilon)) = \frac{1}{2^n}.$$

Deste ponto em diante seguimos exatamente como na seção 5.2 da parte 2, e construímos exatamente a mesma medida (por que podemos afirmar isso?) ◁

A seguir apresentamos alguns resultados importantes sobre independência.

3.15 Proposição *Dado* $\mathbf{P}$ *uma medida de probabilidade e* $\{E_i \mid i \geq 1\}$ *uma família enumerável de eventos independentes. Então:*

$$\mathbf{P}\left(\bigcap_{i\geq 1} E_i\right) = \prod_{i\geq 1} \mathbf{P}(E_i).$$

Demonstração. $\left(\bigcap_{i=1}^{n} E_i\right)$ *é uma sequência de eventos decrescentes, logo*

$$\mathbf{P}\left(\bigcap_{i\geq 1} E_i\right) = \lim_{n\to+\infty} \mathbf{P}\left(\bigcap_{i=1}^{n} E_i\right) = \lim_{n\to+\infty} \prod_{i=1}^{n} \mathbf{P}(E_i) = \prod_{i\geq 1} \mathbf{P}(E_i).$$

3.16 Proposição *Se* $\{E_i \mid i \in I\}$ *é uma família de eventos independentes, então a família de eventos* $\{F_i \mid i \in I\}$, *onde* $F_i = E_i$ *ou* E_i^c, *também é independente.*

Demonstração. *Exercício. Dica: use que* $P(A \cap B) + \mathbf{P}(A \cap B^c) = \mathbf{P}(A)$ *para substituir* E_{i_k} *por* $E_{i_k}^c$ *em* $\mathbf{P}(E_{i_1} \cap \cdots \cap E_{i_n})$ *e siga por indução.* □

A seguir, como exemplo de aplicação da noção de independência, vamos resolver um problema sem nenhuma relação aparente com a teoria das probabilidades.

3.17 Teorema (Fórmula dos Números Primos de Euler) *Dado* $s > 1$. *Então*

$$\zeta(s) := \sum_{n\geq 1} \frac{1}{n^s} = \prod_{p\in\mathbf{P}} \frac{1}{1 - p^{-s}}.$$

Demonstração. *Sejam* $\Omega = \mathbb{N}$, $\mathbf{P}(n) = \dfrac{\zeta(s)^{-1}}{n^s}$ *e* P *o conjunto dos números primos.*

Considere E_k *o evento "números divisíveis por* k*". Observe que:*

$$\mathbf{P}(E_k) = \sum_{i \geq 1} \mathbf{P}(i \cdot k) = \sum_{i \geq 1} \frac{\zeta(s)^{-1}}{i^s k^s} = k^{-s}.$$

Logo, quaisquer primos $k_1, \ldots, k_r$,

$$\mathbf{P}\left(\bigcap_{j=1}^{r} E_{k_j}\right) = \mathbf{P}\left(E_{\prod_{j=1}^{r} k_j}\right) = \prod_{j=1}^{r} k_j^{-s} = \prod_{j=1}^{r} \mathbf{P}(E_{k_j}).$$

Mostramos que $\{E_k \mid k \text{ primo}\}$ *é uma família de eventos independentes. Como* $n = 1$ *se e somente se não é divisível por nenhum primo, temos:*

$$\frac{1}{\zeta(s)} = \mathbf{P}(1) = \mathbf{P}\left(\bigcap_{p \in \mathrm{P}} E_p^{\mathsf{c}}\right) = \prod_{p \in \mathrm{P}} \big(1 - \mathbf{P}(E_p)\big) = \prod_{p \in \mathrm{P}} (1 - p^{-s})$$

Terminamos a seção generalizando o conceito de independência para coleções de famílias de eventos, seguido de um resultado de caracterização desta forma de independência. Como não usaremos muito este conceito ao longo desse livro, não nos estenderemos sobre o assunto e deixaremos o resultado sem demonstração.

3.18 Definição (Famílias Independentes) *Dado* $\mathscr{A}_k$ *uma família de conjuntos em* $\mathcal{F}$ *para cada* $k \in \mathcal{K}$. *Então* $\{\mathscr{A}_k\}_{k \in \mathcal{K}}$ *são* ***famílias independentes*** *se para cada escolha de* $A_k \in \mathscr{A}_k$ *os eventos* $\{A_k\}_{k \in \mathcal{K}}$ *são independentes.*

3.19 Teorema *Dado* $(\Omega, \mathcal{F}, P)$ *um espaço de probabilidade e* $\mathcal{K}$ *um conjunto de índices*

1. ***Subclasses*** *Se* $\mathscr{A}_k \subset \mathscr{B}_k \subset \mathcal{F}$ *para todo* $k \in \mathcal{K}$ *e* $\{\mathscr{B}_k\}_{k \in \mathcal{K}}$ *são famílias independentes então* $\{\mathscr{A}_k\}_{k \in \mathcal{K}}$ *são famílias independentes.*

2. ***Acréscimo*** $\{\mathscr{A}_k\}_{k \in \mathcal{K}}$ *são famílias independentes se e somente se* $\{\mathscr{A}_k \cup \{\Omega\}\}_{k \in \mathcal{K}}$ *são famílias independentes.*

3. ***Produto Simplificado*** *se* $\mathscr{A}_1, \ldots, \mathscr{A}_n$ *são famílias de conjuntos em* $\mathcal{F}$ *e* $\Omega \in \mathscr{A}_k$ *para cada* k, *então* $\mathscr{A}_1, \ldots, \mathscr{A}_n$ *são famílias independentes se e somente se*

$$P\left(\bigcap_{k=1}^{n} A_k\right) = \prod_{k=1}^{n} \mathbf{P}(A_k).$$

para cada escolha $A_k \in \mathscr{A}_k$.

As definições acima são ambas generalizadas pela seguinte definição de independência para σ-álgebras.

3.20 **Definição (σ-Álgebras Independentes)** *Dado $(\Omega, \mathcal{F}, \mathbf{P})$ um espaço de probabilidade e seja $\mathcal{A}$ e $\mathcal{B}$ duas sub-σ-álgebras de $\mathcal{F}$. Então $\mathcal{A}$ e $\mathcal{B}$ são ditas independentes se, dados $A \in \mathcal{A}$ e $B \in \mathcal{B}$,*

$$\mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B)$$

3.21 **Teorema (Famílias Independentes)** *Suponha que $\{\mathscr{A}_k\}_{k\in\mathcal{K}}$ são famílias independentes de conjuntos em $\mathcal{F}$ tais que cada $\mathscr{A}_k$ é um π-sistema. Então $\{\sigma\langle\mathscr{A}_k\rangle\}_{k\in\mathcal{K}}$ são famílias independentes.*

3.22 **Proposição** *Considere uma família de eventos independentes $\{A_{k\in\mathcal{K}}\}$. Se $I \cap J = \emptyset$, então $\sigma\langle A_k : k \in I\rangle$ e $\sigma\langle A_k : k \in J\rangle$ são independentes.*

3.2 Lei 0-1 de Kolmogorov e Lemas de Borel-Cantelli

Dado uma sequência de eventos arbitrários $A_1, A_2, \ldots$. Queremos definir o que significa um evento A depender somente dos eventos $A_n, A_{n+1}, \ldots$, para n suficientemente grande, ou dito de outra forma um evento A não depender dos eventos iniciais $A_1, \ldots, A_n$ para todo n.

3.23 **Definição** *A $\sigma -$ **álgebras caudal** é*

$$\mathcal{T} := \bigcap_{n=1}^{\infty} \sigma\langle A_n, A_{n+1}, A_{n+2}, \ldots\rangle.$$

*Os elementos de $\mathcal{T}$ são denominados **eventos caudais**.*

Se pensarmos no índice n como o tempo, então $\sigma\langle A_n, A_{n+1}, A_{n+2}, \ldots\rangle$ contém a informação após o tempo n e $\mathcal{T}$ contém a informação "após o tempo n para todo n", ou seja, vagamente falando, a informação no infinito. Outro nome para a σ-álgebra caudal é σ-álgebra de eventos remotos.

Poderia-se pensar que poucos eventos seriam caudais, mas em algumas situações a σ-álgebra caudal pode conter muitos eventos como mostra o próximo exemplo.

3.24 **Exemplo** *Por exemplo, considerando o lançamento infinito de moedas, e denotando por H_n o evento no qual no n-ésimo lançamento aparece cara, então $\mathcal{T}$ inclui o evento $\limsup H_n$ no qual obtemos infinitas caras; o evento $\liminf H_n$ no qual obtemos apenas um número finito de coroas; o evento $\limsup H_{2n}$ que obtemos infinitamente muitas caras nos lances 2, 4, 8, e o evento "uma sequência de 100 caras consecutivas ocorre infinitas vezes".* $\triangleleft$

3.25 Teorema (Lei 0-1 de Kolmogorov) *Dados $A_1, A_2, \ldots$ eventos independentes. Então todo evento caudal A tem probabilidade 0 ou 1.*

Demonstração. *Considere $A_1, A_2, A_3, \ldots, A_{n-1}, A_n, A_{n+1}, \ldots, \ldots$ eventos. Então, pelo Corolário 3.22, $\sigma\langle A_1, \ldots, A_{n-1}\rangle$ e $\sigma\langle A_n, \ldots A_{n+1}, \ldots\rangle$ são independentes. Note que , $A \in \sigma\langle A_n, \ldots A_{n+1}, \ldots\rangle$. Logo, os eventos $A_1, A_2, \ldots, A_{n-1}, A$ são independentes para todo n.*

Consequentemente os eventos $A_1, A_2, \ldots, A$ são independentes. Então, pela Proposição 3.22, $\sigma\langle A_1, A_2, \ldots\rangle$ e $\sigma\langle A\rangle$ são independentes. Mas $A \in \sigma\langle A_1, A_2, \ldots\rangle$ e $A \in \sigma\langle A\rangle$, que são independentes. Logo, A é independente de si próprio. Logo

$$\mathbf{P}(A) = \mathbf{P}(A \cap A) = \mathbf{P}(A)^2.$$

e assim $\mathbf{P}(A)$ é ou 0 ou 1. □

3.26 Definição (Infinitas vezes e quase sempre) *Dados $A_1, A_2, \ldots$ subconjuntos de Ω. Então dizemos que os eventos ocorrem **infinitas vezes** se*

$$\{A_n \quad i.v.\} := \bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} A_n =: \limsup_{n\to\infty} A_n$$

*Dizemos que os eventos ocorrem **quase sempre** se*

$$\{A_n \quad q.s.\} := \bigcup_{m=1}^{\infty} \bigcap_{n=m}^{\infty} A_n =: \liminf_{n\to\infty} A_n$$

3.27 Teorema (Primeiro Lema de Borel-Cantelli) *Dados $E_1, E_2, \ldots$ subconjuntos em $\mathcal{F}$. Então*

$$\sum_{n=1}^{\infty} \mathbf{P}(E_n) < \infty \Longrightarrow \mathbf{P}(E_n \quad i.v.) = 0.$$

Demonstração. *Suponha que:*

$$\sum_{n=1}^{\infty} \mathbf{P}(E_n) < \infty.$$

Como a série converge devemos ter que

$$\sum_{n=N}^{\infty} \mathbf{P}(E_n) \to 0, \qquad \text{quando } N \to \infty.$$

Logo

$$\inf_{N\geqslant 1} \sum_{n=N}^{\infty} \mathbf{P}(E_n) = 0.$$

Assim

$$\begin{aligned}
\mathbf{P}(E_n \quad i.v.) &= \mathbf{P}\left(\limsup_{n\to\infty} E_n\right) \\
&= \mathbf{P}\left(\bigcap_{N=1}^{\infty}\bigcup_{n=N}^{\infty} E_n\right) \\
&\leqslant \inf_{N\geqslant 1} \mathbf{P}\left(\bigcup_{n=N}^{\infty} E_n\right) \\
&\leqslant \inf_{N\geqslant 1} \sum_{n=N}^{\infty} \mathbf{P}(E_n) \\
&= 0
\end{aligned}$$

3.28 Teorema (Lema de Fatou para Conjuntos) *Dados $E_1, E_2, \ldots$ subconjuntos de $\mathcal{F}$. Então*

$$\begin{aligned}
\mathbf{P}(E_n \quad q.s.) &\leq \liminf_n \mathbf{P}(E_n) \\
&\leq \limsup_n \mathbf{P}(E_n) \leq \mathbf{P}(E_n \quad i.v.).
\end{aligned}$$

Demonstração. *Exercício* □

Se pensarmos em n como tempo ou algo análogo, o que o resultado acima nos diz é que se uma sequência de eventos tem probabilidade decaindo muito rápido, então eventualmente deixaremos de observar essa sequência de eventos. De fato, note que

$$\left(\limsup_n A_n\right)^c = \left(\bigcap_{N=1}^{\infty}\bigcup_{n=N}^{\infty} A_n\right)^c = \left(\bigcup_{N=1}^{\infty}\bigcap_{n=N}^{\infty} A_n^c\right) = \liminf_n A_n^c,$$

de modo que se $\mathbf{P}(A_n \quad \text{i.v.}) = 0$, então $\mathbf{P}(A_n^c \quad q.s.) = 1$.

3.29 Exemplo (Passeio Aleatório Assimétrico) *Considere um processo aleatório em $\mathbb{Z}$ onde uma partícula inicia em $x = 0$ e a cada instante n salta uma unidade para a direita com probabilidade $p \in (0,1)$ ou uma unidade para a esquerda com probabilidade $1 - p$. Denote por $S_n \in \mathbb{Z}$ a posição da partícula no instante $n \geq 0$.*

O processo S_n é conhecido na literatura como passeio aleatório simples, e no caso em que $p \neq 1/2$, dizemos que o passeio é assimétrico.

Para construir tal processo tomaremos como espaço amostral o espaço

$$\Omega = \{-1,1\}^\infty = \{\omega = (\omega_1, \omega_2, \ldots) : \omega_k \in \{-1,1\}, k \geq 1\},$$

e seguindo os mesmos argumentos do exemplo das infinitas moedas, construímos uma medida tal que os eventos $A_n = \{\omega \in \Omega : \omega_n = 1\}, n \geq 1$ são independentes com $\mathbf{P}(A_n) = p$.

Agora definimos $S_0(\omega) = 0$ *e* $S_n(\omega) = \omega_1 + \omega_2 + \cdots + \omega_n$.

Assim para cada $n \geq 1$

$$\mathbf{P}(S_n = S_{n-1} + 1) := \mathbf{P}(\{\omega : S_N(\omega) = S_{n-1}(\omega) + 1\}) = \mathbf{P}(A_n) = p,$$

enquanto

$$\mathbf{P}(S_n = S_{n-1} - 1) := \mathbf{P}(\{\omega : S_N(\omega) = S_{n-1}(\omega) - 1\}) = \mathbf{P}(A_n) = 1 - p,$$

como gostaríamos.

Estamos interessados agora em estudar o retorno de tal processo à origem do espaço. Mais especificamente, queremos saber com que frequência o processo retorna para o ponto onde começou o passeio. Em outras palavras, estamos interessados na sequência de eventos $\{S_n = 0\} := \{\omega : S_n(\omega) = 0\}$.

Para que tenhamos $S_n = 0$ *é necessário que o total de passos para a direita até o instante* n *seja igual ao total de passos para a esquerda. Com isso concluímos, antes de mais nada, que* n *deve ser par.*

Fica como exercício para o leitor justificar agora que

$$\mathbf{P}(S_{2n} = 0) = \binom{2n}{n} p^n (1-p)^n.$$

Usando a fórmula de Stirling, encontramos que

$$\mathbf{P}(S_{2n} = 0) \approx \frac{\sqrt{4\pi n}\, 4^n\, n^{2n}\, e^{-2n}}{(\sqrt{2\pi n}\, n^n\, e^{-n})^2} p^n (1-p)^n = \frac{[4p(1-p)]^n}{\sqrt{\pi n}}.$$

Note agora que se $p \neq 1/2$ *então* $r := 4p(1-p) < 1$ *e portanto*

$$\sum_{n=0}^{\infty} \mathbf{P}(S_{2n} = 0) \approx \sum_{n=0}^{\infty} \frac{r^n}{\sqrt{\pi n}} < \infty.$$

E portanto $\mathbf{P}(S_{2n} = 0 \text{ i.v.}) = 0$*, mostrando que eventualmente o processo visita a origem do espaço uma última vez, para nunca mais retornar.* $\triangleleft$

No exemplo anterior, caso considerássemos o caso simétrico, onde $p = 1/2$ teríamos $r = 1$ e assim

$$\sum_{n=0}^{\infty} \mathbf{P}(S_{2n} = 0) \approx \sum_{n=0}^{\infty} \frac{1}{\sqrt{\pi n}} = \infty.$$

Infelizmente, a soma divergir não é, por si só, suficiente para garantir que $\mathbf{P}(S_{2n} = 0 \text{ i.v.}) = 1$. Para isso precisaríamos de alguma hipótese adicional, como mostramos no próximo resultado.

O segundo Lema de Borel fornece uma recíproca parcial do primeiro lema de Borel-Cantelli.

3.30 Teorema (Segundo Lema de Borel-Cantelli) *Dados $E_1, E_2, \ldots$ conjuntos independentes em $\mathcal{F}$. Então*

$$\sum_{n=1}^{\infty} \mathbf{P}(E_n) = \infty \Longrightarrow \mathbf{P}(E_n \quad i.v.) = 1.$$

Demonstração. *Suponha que $\sum_{n=1}^{\infty} \mathbf{P}(E_n) = \infty$ e que os eventos $(E_n)_{n=1}^{\infty}$ são independentes. É suficiente mostrar que os eventos E_n que não ocorrem para infinitos valores de n tem probabilidade 0.*

Começamos observando que

$$\mathbf{P}\left(\{E_n \quad i.v.\}^{\mathsf{c}}\right) = \mathbf{P}\left(\liminf_{n\to\infty} E_n^c\right) = \mathbf{P}\left(\left(\bigcap_{N=1}^{\infty}\bigcup_{n=N}^{\infty} E_n\right)^{\mathsf{c}}\right) =$$

$$= \mathbf{P}\left(\bigcup_{N=1}^{\infty}\bigcap_{n=N}^{\infty} E_n^{\mathsf{c}}\right) = \lim_{N\to\infty} \mathbf{P}\left(\bigcap_{n=N}^{\infty} E_n^{\mathsf{c}}\right)$$

e desta forma temos que é suficiente demonstrar que: $\mathbf{P}\left(\bigcap_{n=N}^{\infty} E_n^c\right) = 0$. *Como os eventos $(E_n)_{n=1}^{\infty}$ são independentes temos que:*

$$\begin{aligned}
\mathbf{P}\left(\bigcap_{n=N}^{\infty} E_n^{\mathsf{c}}\right) &= \prod_{n=N}^{\infty} \mathbf{P}(E_n^{\mathsf{c}}) \\
&= \prod_{n=N}^{\infty} (1 - \mathbf{P}(E_n)) \\
&\leq \prod_{n=N}^{\infty} \exp(-\mathbf{P}(E_n)) \text{ pois } 1 - x \leq e^{-x} \\
&= \exp\left(-\sum_{n=N}^{\infty} \mathbf{P}(E_n)\right) \\
&= 0.
\end{aligned}$$

Resumindo, dados $A_1, A_2, \ldots$ eventos independentes. Então

$$\mathbf{P}(A_n \quad \text{i.v.}) = \begin{cases} 0 & \text{se } \sum_n \mathbf{P}(A_n) < \infty \\ 1 & \text{se } \sum_n \mathbf{P}(A_n) = \infty. \end{cases}$$

3.31 Exemplo *Considere o exemplo de infinitos lançamentos independentes de uma moeda não-viciada.*

Intuitivamente, esperamos que o total de caras observados, assim como o total de coroas, sejam infinitos. Para mostrar isso, tome o evento

$$A_n = \{\text{o } n\text{-ésimo lançamento foi cara}\} = \{\omega \in \Omega : \omega_n = 1\}.$$

Já sabemos que os eventos A_n, $n \geq 1$ são independentes e que $\mathbf{P}(A_n) = 1/2$. Segue então que

$$\sum_{n=1}^{\infty} \mathbf{P}(A_n) = \infty,$$

e pelo segundo lema de Borel-Cantelli, $\mathbf{P}(A_n \;\; i.v.) = 1$. Ou seja, observaremos infinitas caras com probabilidade 1.

De modo análogo podemos mostrar que $\mathbf{P}(A_n^c \;\; i.v.) = 1$.

Antes de seguir para os exemplos na próxima sessão, tente resolver o exercício abaixo.

Exercício: Calcule a probabilidade de saírem 100 caras consecutivas infinitas vezes.

□

3.3 Aplicações

Sequências de Cara

Se jogarmos uma moeda infinitas vezes, mais cedo ou mais tarde, veremos sair duas caras seguidas. Se tivermos muita paciência, veremos saírem 100 coras seguidas. Para formalizar essa ideia, seja $\ell(n)$ = a função que retorna o número de caras consecutivas após o n-ésimo lançamento. Estamos adotando a convenção que $\ell(n) = 0$ se sair coroa no n-ésimo lançamento.

Dado um inteiro fixo k é fácil ver que

$$\mathbf{P}(\ell(n) \geq k \quad \text{i.v.}\ ) = 1.$$

Vamos considerar uma situação um pouco mais geral. O que ocorreria se trocássemos o inteiro fixo por uma função $k(n)$? O quão rápido essa função pode crescer de modo que a probabilidade permaneça 1.? Ou seja, para quais funções $k(n)$ temos que

$$\mathbf{P}(\ell(n) \geq k(n) \quad \text{i.v.}\ ) = 1.$$

O próximo teorema nos diz que essa função pode crescer no máximo como $\log_2 n$ como provamos no próximo teorema:

3.32 Teorema

$\boxed{1}$ *Para todo $\epsilon > 0$, $\mathbf{P}(\ell(n) \geq (1+\epsilon)\log_2 n \;\; i.v.) = 0$.*

$\boxed{2}$ $\mathbf{P}(\ell(n) \geq \log_2 n \quad i.v.) = 1.$

Demonstração.

$\boxed{1}$ *Começamos pela observação que* $\mathbf{P}(\ell(n) \geq r) = \frac{1}{2^r}$. *E aplicando Borel-Cantelli após a substituição* $(1+\epsilon)\log_2 n$ *obtemos:*

$$\mathbf{P}(\ell(n) \geq (1+\epsilon)\log_2 n) = \frac{1}{n^{1+\epsilon}}.$$

Esta série converge e logo podemos utilizar o Primeiro Lema de Borel-Cantelli. Isso termina a demonstração.

$\boxed{2}$ *A segunda parte é mais interessante. Ela nos fornece uma técnica comum na teoria da probabilidade. Por que a segunda parte é mais elaborada? O problema é que os eventos* $\{\ell(n) \geq \log_2 n\}$ *não são independentes, então não podemos simplesmente aplicar o lema Borel-Cantelli.*

Mas podemos "podar" esses eventos de modo a tratarmos apenas de eventos que não se sobrepõem. Vamos denotar $r(n) := \log_2 n$.

Agora olhamos para $n_1 = 2$. *Mas só voltaremos a olhar a sequência após* $\log_2 n_1$, *ou seja* $n_2 = n_1 + r(1)$. *De modo geral,*

$$\begin{aligned} n_1 &= 2 \\ n_{k+1} &= n_k + r(n_k). \end{aligned}$$

Ou seja, os eventos $A_k = \{\ell(n_k) \geq r(n_k)\}$ *são independentes.*

Começamos em 2, *pois* $\log_2 1 = 0$. *Retornando a demonstração do teorema.*

Os eventos A_k *são independentes, porque os* A_k^c *envolvem índices não sobrepostos. Então temos*

$$\mathbf{P}(A_k) = \frac{1}{2^{r(n_k)}}.$$

Agora, só queremos ver que a soma desta série diverge. Mas a soma

$$\sum_{k=1}^{\infty} \frac{1}{2^{r(n_k)}}$$

está definida recursivamente. Então, preenchemos as lacunas na soma fazendo

$$\sum_{k=1}^{\infty} \frac{1}{2^{r(n_k)}} = \sum_{k=1}^{\infty} \frac{1}{2^{r(n_k)}} \cdot \frac{n_{k+1} - n_k}{r(n_k)}.$$

E agora convertemos os saltos $n_{k+1} - n_k$ como soma de 1s:

$$= \sum_{k=1}^{\infty} \sum_{n_k \le n < n_{k+1}} \frac{1}{2^{r(n_k)} r(n_k)},$$

Como $r(n_k) \le r(n)$, a última soma é maior ou igual à

$$\le \sum_{n=1}^{\infty} \frac{1}{2^{r(n)} r(n)}$$

e assim, usando a definição de $r(n)$, temos

$$\sum_{n=1}^{\infty} \frac{1}{n \log_2 n},$$

que diverge.

Para que a demonstração anterior fique correta precisaríamos tomar a parte inteira de $r(n_k)$ e fazer pequenas alterações. Deixamos essas alterações como exercício ao leitor $\square$

Teorema do Macaco Infinito

> "I heard someone tried the monkeys-on-typewriters bit trying for the plays of W. Shakespeare, but all they got was the collected works of Francis Bacon."
> - Bill Hirst

Vamos considerar uma pequena adaptação do Teorema 3.32. Vamos considerar que estamos sorteando caracteres uniformemente no alfabeto $\{a, b, c \ldots, x, y, z\}$. Então obtemos uma sequência infinita de caracteres, onde cada caractere é escolhido uniformemente ao acaso. E uma das consequências do Teorema é que qualquer texto finito ocorre quase-certamente nessa sequência.

Isso nos leva ao seguinte "teorema"

3.33 Teorema (Teorema do Macaco Infinito) *Um macaco que pressiona as teclas aleatoriamente de uma máquina de escrever por uma quantidade infinita de tempo irá quase certamente digitar um determinado texto, como as obras completas de William Shakespeare. Na verdade, o macaco quase certamente digitaria todos os textos finitos possíveis um número infinito de vezes.*

Teoria de Percolação

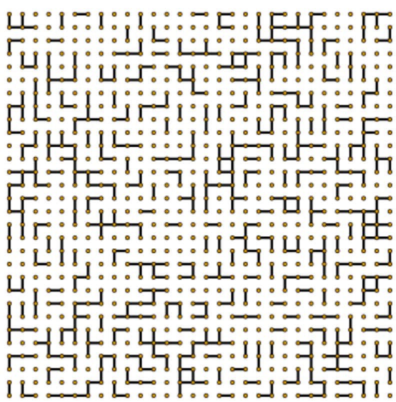
Suponha que, em dado material poroso, algum líquido seja derramado no topo. Este líquido chegará ao fundo? Quão rápido?

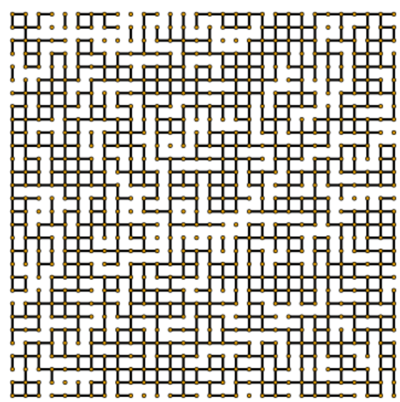
Em matemática, podemos pensar um material poroso como um grafo, digamos por exemplo que os vértices são $\mathbb{Z}^2$. Conectamos vértices vizinhos desta rede de forma independente com alguma probabilidade p. O resultado é um grafo aleatório. De modo alternativo, podemos fazer uma construção análoga com vértices em vez de arestas, e considere o subgrafo como aquele determinado pelos vértices escolhidos.

Alguma terminologia:

- As arestas são denominadas elos e os vértices são denominados sítios.
- As aresta/vértices escolhidos são ditos abertos.
- O gráfico escolhido é o subgrafo aberto.
- As componentes conexas do subgrafo aberto são denominados clusters abertos.

De acordo com a terminologia acima, os dois modelos descritos anteriormente são **percolação de elos** e **percolação do sítio**. No que se segue vamos restringir nossa atenção a percolação de elos.





Como dissemos, a motivação física para este modelo é um líquido que percola através de uma rede sólida, onde as ligações abertos representam os locais pelos quais o líquido pode passar. A principal questão sugerida por esta interpretação é se o líquido é capaz de passar completamente pela rede, o que corresponde matematicamente à existência de um cluster aberto infinito.

O Teorema de Harris-Kesten responde esta pergunta para a percolação de elos em $\mathbb{Z}^2$. Esse teorema afirma que para $p > \frac{1}{2}$, existe um cluster aberto infinito com probabilidade 1 e para $p < \frac{1}{2}$ existe um cluster aberto infinito com probabilidade 0.

Vamos fazer a construção probabilística desse modelo e para isso definimos:

- $\Omega = \{0,1\}^{|E|} = \prod_{e \in E} \{0,1\}$ como o conjunto de todos as configurações de arestas abertas ou fechadas. Dada uma configuração $\omega \in \Omega$ e uma aresta $e \in E$, dizemos que e está aberta se $\omega_b = 1$ e fechada $\omega_b = 0$
- $\mathcal{F}_0$ como a álgebra em Ω gerada pelos cilindros finito dimensionais

$$C(F,\sigma) = \{\omega \in \Omega \mid \omega_f = \sigma_f \text{ para } f \in F\}, \quad ;$$

  Sendo $F \subset E$ um conjunto finito e $\sigma \in \{0,1\}^{|F|}$ uma configuração em F.
- $\mathcal{F}$ como a σ-álgebra em Ω gerada pelos cilindros finito dimensionais
- No modelo mais simples de percolação cada aresta tem uma probabilidade p de estar aberta e essa probabilidade independe das outras arestas. Dessa forma escolhemos $\mathbf{P}$ como a pré-medida de probabilidade em $\mathcal{F}_0$ induzida por

$$\mathbf{P}(C(F,\sigma)) = \left(\prod_{\substack{f \in F \\ \sigma_f = 1}} p \right) \left(\prod_{\substack{f \in F \\ \sigma_f = 0}} 1 - p \right).$$

Por argumentos de classe compacta, podemos provar que $\mathbf{P}$ é uma pré-medida e pelo Teorema de Extensão de Carathéodory podemos definir uma probabilidade $\mathbf{P}$ em $\mathcal{F}$.

A tripla $(\Omega, \mathcal{F}, \mathbf{P})$ definida acima é o modelo de percolação em $\mathbb{Z}^2$. Não é difícil alterar a construção acima para construir o modelo de percolação em $\mathbb{Z}^d$ ou em um grafo G conexo e com um número enumerável de vértices.

Probabilidades Críticas

Para um sítio x, definimos C_x como o cluster aberto contendo x e $\theta_x(p)$ a probabilidade de que C_x seja infinito.

3.34 Proposição *Existe uma* ***probabilidade crítica*** p_c *tal que* $\theta_x(p) = 0$ *se* $p < p_c$ *e* $\theta_x(p) > 0$ *se* $p < p_c$*, para todo* x.

Demonstração. *Considere dois sítios* x, y *tais que* $|x - y| = d$.

Como o grafo é conexo, temos um caminho ligando x e y. Esse caminho nos permitirá estabelecer uma relação entre $\theta_x(p)$ e $\theta_y(p)$. Vejamos como: $\theta_x(p) \geq p^d \theta_y(p)$, pois se houver um cluster aberto infinito contendo y e um caminho de y para x, existirá um cluster aberto infinito contendo x também. Assim, se $\theta_x(p) = 0$ para um sítio, então $\theta_x(p) = 0$ para todos os sítios, e se $\theta_x(p) > 0$ para um sítio, então $\theta_x(p) > 0$ para todos os sítios.

Se mostramos que $\theta_x(p)$ é uma função crescente de p, a existência de p_c segue. Para ver o comportamento crescente, utilizaremos a técnica de acoplamento. Atribua a cada uma das arestas e uma variável aleatória independente $X_e \sim \text{Uni}[0,1]$ e, em seguida, decida torna-lá aberta se $X_e < p$. Se $p_1 < p_2$, o subgrafo obtido com probabilidade p_1 será sempre contido no subgrafo obtido com probabilidade p_2, e dessa forma $\theta_x(p)$ deve ser crescente.

Também observamos que para $p = 0$ temos que $\theta(0) = 0$ e para $p = 1$ temos que $\theta(1) = 1$. Gostaríamos de provar que essa transição é não trivial, ou seja, ou seja que a probabilidade crítica é estritamente maior que 0 e menor que 1. Mas não faremos isso agora.

Discutiremos mais sobre a técnica de acoplamento no capítulo 12. □

Agora, conecte este p_c à nossa pergunta original sobre a existência de um cluster aberto infinito.

3.35 **Teorema** *Seja I ser o evento que existe um cluster aberto infinito. Então $\mathbf{P}(I) = 1$ se $p > p_c$ e $\mathbf{P}(I) = 0$ se $p < p_c$.*

Demonstração. *A demonstração será feita mostrando que I é um evento caudal. Se I for um evento caudal, pela lei Kolmogorov 0-1, sua probabilidade só pode tomar os valores 0 ou 1. Assim, se $p < p_c$, então $\mathbf{P}(I) \leq \sum_x \theta_x(p) = 0$ então $\mathbf{P}(I) = 0$. Se $p > p_c$ então $\mathbf{P}(I) \geq \theta_x(p) > 0$ para pelo menos um sítio, então $\mathbf{P}(I) = 1$.*

Mostraremos agora que I é um evento caudal. Para isso, enumere os elos como $E = \{e_o, e_1, \ldots\}$, e para $k \geq 0$ defina

$$A_k = \{\omega \in \Omega : \omega_{e_k} = 1\}.$$

Segue que os eventos $A_0, A_1, \ldots$ são independentes, e além disso temos que $\mathcal{F} = \sigma\langle A_0, A_1, \ldots\rangle$.

Dado $\omega \in \Omega$ seja $G_0(\omega)$ o subgrafo obtido a partir da remoção de e_0 de E

$$G_0(\omega) = \{e \in E : \omega_e = 1\} \backslash \{e_0\}.$$

Deste modo o evento

$$I_0 = \{\textit{existe um cluster infinito em } G_0\},$$

é tal que $I_0 \subseteq I$ e $I_0 \in \sigma\langle A_1, A_2, \ldots\rangle$.

Além disso, se $\omega \in I$, então o cluster $C_x(\omega)$ é infinito. Se a componente C_x não contem e_0, então C_x também é um cluster infinito em $G_0(\omega)$, e segue que $\omega \in I_0$.

Se C_x contem e_0, a retirada de e_0 dividirá C_x em, no máximo, dois clusters (duas componentes conexas) disjuntos, e ao menos um deles ainda será infinito, e como não contem e_0, será infinito em $G_0(\omega)$. Deste modo, $\omega \in I_0$. Concluímos assim que $I = I_0 \in \sigma\langle A_1, A_2, \ldots\rangle$.

De modo análogo, podemos demonstrar que $I \in \sigma\langle A_n, A_{n+1}, \ldots\rangle$ para todo $n \geq 0$, e portanto I e caudal.

Exercícios

Ex. 3.1 — Deixe Ω ser os inteiros $\{1, 2, \ldots, 9\}$ com probabilidade de $1/9$ cada. Mostre que os eventos $\{1, 2, 3\}, \{1, 4, 5\}, \{2, 4, 6\}$ são dois a dois independentes, mas a família não é independente.

Ex. 3.2 — Construa um exemplo de três eventos A, B, C que não são independentes mas que satisfazem $\mathbf{P}(A \cap B \cap C) = \mathbf{P}(A)\mathbf{P}(B)\mathbf{P}(C)$.

Ex. 3.3 — Dado uma família finita de eventos $\{A_i\}$ que são independentes mostre que o conjunto $\{A_i^C\}$ que é o conjunto de complementos dos eventos originais, também é independente.

Ex. 3.4 — Construa um espaço de probabilidade $(\Omega, \mathbf{P})$ e k eventos,cada um com probabilidade $1/2$, que são $k-1$ independentes entre si, mas não são independentes. Escolha o espaço amostral tão pequeno quanto possível.

Ex. 3.5 — Uma bolsa contém uma bola preta e m bolas brancas. Uma bola é retirada aleatoriamente. Se ela for preta, ele é retornada para o saco. Se for branca, ela e uma bola branca adicional são devolvidas à bolsa. Deixe A_n indicar o evento em que a bola negra não é retirada nas primeiras n tentativas . Discuta a reciproca do Lema de Borel-Cantelli com referência aos eventos A_n.

Ex. 3.6 — Dados A_n, $n \geq 1$, conjuntos de Borel no espaço de Lebesgue $([0, 1], \mathcal{F}(0, 1), \mu)$. Mostre que se existe $\Upsilon > 0$, tal que $\mu(A_n) \geq \Upsilon$ para todo n, então existe pelo menos um ponto que pertence a infinitos conjuntos A_n

Ex. 3.7 —

1. Para todo $k = 1, \ldots, n$ deixe $\mathcal{P}_k$ ser uma partição de Ω em conjuntos enumeráveis de $\mathcal{F}$. Mostre que as σ-álgebras $\sigma\langle\mathcal{P}_1\rangle, \ldots, \sigma\langle\mathcal{P}_n\rangle$ são independentes se e somente se (3.5) vale para cada escolha de A_k em $\mathcal{P}_k$ $k = 1, \ldots, n$.
2. Mostre que os conjuntos $A_1, \ldots, A_n$ são independentes se e somente se

$$\mathbf{P}\Big(\bigcap_{k=1}^{n} B_k\Big) = \prod_{k=1}^{n} \mathbf{P}(B_k)$$

   para toda escolha de B_k como A_k ou A_k^c para $k = 1, \ldots, n$.

Ex. 3.8 — Deixe $\mathscr{A}_1, \ldots, \mathscr{A}_n$ serem π-sistemas de conjuntos em $\mathcal{F}$ tais que

$$\mathbf{P}\Big(\bigcap_{k=1}^{n} A_k\Big) = \prod_{k=1}^{n} \mathbf{P}(A_k) \tag{3.5}$$

para toda escolha de $A_k \in \mathscr{A}_k$ para $k = 1, \ldots, n$.

1. Mostre usando um exemplo simples que $\mathscr{A}_k$'s não precisam ser independentes.
2. Mostre que $\mathscr{A}_k$'s serão independentes se para todo k, Ω é união enumerável de conjuntos em $\mathscr{A}_k$.
   Dica: Inclusão- Exclusão

Ex. 3.9 — Dados $A_1, A_2, \ldots$. Mostre que $\mathbf{P}(A_n \text{ i.v.}_n) = 1$ se e somente se $\displaystyle\sum_{n=1}^{\infty} \mathbf{P}(A_n|A)$ diverge para todo A de probabilidade não nula

Dica: Mostre que $\mathbf{P}(A_n \text{ i.v.}_n) < 1 \iff \displaystyle\sum_{n=1}^{\infty} \mathbf{P}(A_n|A) < \infty$ para algum conjunto A com $\mathbf{P}(A) > 0$.

Ex. 3.10 — Medida de Lebesgue
Faça a construção da Medida de Lebesgue na reta usando o teorema da Classe Compacta.

Ex. 3.11 — Infinitas Moedas I
Dado $\Omega = \{0, 1\}^{\infty}$ ser o espaço das sequências $0 - 1$, e deixe $\mathcal{F}_0$ denotar a álgebra da união finita de conjuntos da forma
$A_n(\epsilon_1, \ldots, \epsilon_n) = \{\omega = (\omega_1, \omega_2, \ldots) \in \Omega : \omega_1 = \epsilon_1, \ldots \omega_n = \epsilon_n\}$.
Fixe $p \in [0, 1]$ e defina

$$\mathbf{P}_p(A_n(\epsilon_1, ..., \epsilon_n)) = p^{\sum \epsilon_i}(1 - p)^{n - \sum \epsilon_i}$$

1. Mostre que a extensão naturalmente aditiva de $\mathbf{P}_p$ para $\mathcal{F}_0$ define uma medida na álgebra $\mathcal{F}_0$. [Dica: pelo Teorema de Tychonov da topologia, o conjunto Ω é compacto para a topologia do produto, verifique que os conjuntos $C \subset \mathcal{F}_0$ são abertos e fechados na topologia produto, de modo que, por compacidade, qualquer união disjunta enumerável pertencente a $\mathcal{F}_0$ deve ser uma união finita.]
2. Mostre que $\mathbf{P}_p$ tem uma única extensão para $\sigma\langle\mathcal{F}_0\rangle$. Esta probabilidade $\mathbf{P}_p$ define a probabilidade do produto infinito. [Dica: aplique o Teorema de extensão de Carathéodory.]
3. Mostre que as funções projeções de coordenadas

   $$X_n(\omega) = \omega_n, \quad \omega = (\omega_1, \omega_2,)$$

   para $n = 1, \ldots$ definem uma sequência de variáveis aleatórias i.i.d. de Bernoulli.

Ex. 3.12 — Percolação de Elos
Para um gráfico (infinito) $G = (V, E)$, tomamos:

- $\Omega = \{0,1\}^{|E|}$ como o conjunto de todos os arranjos de arestas abertas, onde 0 representa uma aresta fechada e 1 representa uma aresta aberta;

- Σ como a σ -algebra em Ω gerado pelos cilindros
$$C(F, \sigma) = \{\omega \in \Omega \mid \omega_f = \sigma_f \text{ para } f \in F\},$$
$$F \subset E, |F| < \infty \text{ e } \sigma = \{0,1\}^{|F|};$$
- **P** como a medida de probabilidade em Σ induzida por
$$\mathbf{P}_p(C(F, \sigma)) = \left(\prod_{\substack{f \in F \\ \sigma_f = 1}} p \right) \left(\prod_{\substack{f \in F \\ \sigma_f = 0}} 1 - p \right).$$

1. Mostre que a extensão naturalmente aditiva de $\mathbf{P}_p$ para $\mathcal{F}_0$ define uma medida na álgebra $\mathcal{F}_0$.
2. Mostre que $\mathbf{P}_p$ tem uma única extensão para $\sigma\langle\mathcal{F}_0\rangle$.

Capítulo 4

Variáveis Aleatórias

4.1 Funções Mensuráveis

Como vimos na parte anterior, espaços mensuráveis são o primeiro ingrediente na modelagem de uma probabilidade (ou de uma medida). São eles que carregam a estrutura que dará suporte ao nosso modelo. Assim, ao mapearmos um espaço $(\Omega_1, \mathcal{F}_1)$ em outro espaço $(\Omega_2, \mathcal{F}_2)$ é importante garantirmos que tais estruturas estejam preservadas. E quem faz esse papel na teoria da medida são as **funções mensuráveis**.

4.1 Definição *Dados $(\Omega_1, \mathcal{F}_1)$ e $(\Omega_2, \mathcal{F}_2)$ espaços mensuráveis, uma função $f : \Omega_1 \to \Omega_2$ é dita $\mathcal{F}_1 - \mathcal{F}_2$-**mensurável** ou simplesmente **mensurável** se a pré-imagem de todo conjunto mensurável é mensurável. Ou seja, se*

$$A \in \mathcal{F}_2 \implies f^{-1}(A) \in \mathcal{F}_1.$$

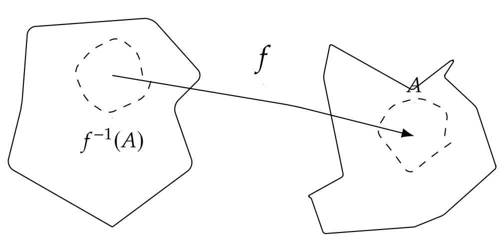


Figura 4.1: A função f é mensurável se $f^{-1}(A)$ é mensurável para todo A mensurável

4.2 Observação

- *Dado uma função $f : \Omega_1 \to \Omega_2$ que é $\mathcal{F}_1 - \mathcal{F}_2$–**mensurável** então se $\mathcal{F}_1 \subset \mathcal{F}_1'$ claramente f é $\mathcal{F}_1' - \mathcal{F}_2$ – **mensurável**.*
- *O mesmo não ocorre se aumentarmos a σ-álgebra da imagem! Pode-se provar, por exemplo, que existem funções contínuas que não são Lebesgue mensuráveis, i.e, não são $\bar{\mathcal{B}}(\mathbb{R}) - \mathcal{B}(\mathbb{R})$-mensuráveis.* ◁

A família de funções mensuráveis é, em geral, bastante rica. No contexto das funções reais, por exemplo, continuidade é condição suficiente para mensurabilidade. Mas ainda assim, verificar mensurabilidade baseado apenas na sua definição pode ser um trabalho hercúleo e ingrato.

Por isso, antes de seguir com os exemplos, vamos primeiro colocar um resultado que facilita esse processo.

4.3 Teorema *Considere uma função $f : (\Omega_1, \mathcal{F}_1) \to (\Omega_2, \mathcal{F}_2)$, e $\mathcal{A} \subset \mathcal{F}_2$ tal que $\mathcal{F}_2 = \sigma\langle\mathcal{A}\rangle$. Nestas condições, se $f^{-1}(A) \in \mathcal{F}_1$ para todo $A \in \mathcal{A}$, então f é mensurável. Ou seja, se*

$$A \in \mathcal{A} \Rightarrow f^{-1}(A) \in \mathcal{F}_1.$$

então

$$A \in \sigma\langle\mathcal{A}\rangle \Rightarrow f^{-1}(A) \in \mathcal{F}_1,$$

Demonstração. *Considere $\mathcal{F} := \{A \in \sigma\langle\mathcal{A}\rangle : f^{-1}(A) \in \mathcal{F}_1\}$. Iremos demonstrar que $\mathcal{F} = \sigma\langle\mathcal{A}\rangle$.*

Sabemos que $\mathcal{A} \subseteq \mathcal{F} \subseteq \sigma\langle\mathcal{A}\rangle$. Além disso temos que $\sigma\langle\mathcal{A}\rangle$ é a σ-álgebra minimal contendo $\mathcal{A}$. Assim, basta demonstrar que $\mathcal{F}$ é uma σ-álgebra.

1 *Demonstraremos que se $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$.*

Para isso observe que $f^{-1}(A^c) = (f^{-1}(A))^c$.

2 *Demonstraremos que se $A_1, A_2, \ldots \in \mathcal{F}$, então $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$. Mas*

$$f^{-1}\left(\bigcup_n A_n\right) = \bigcup_{n=1}^{\infty} f^{-1}(A_n)$$

Logo, $\mathcal{F}$ é uma σ-álgebra e $\mathcal{A} \subseteq \mathcal{F} \subseteq \sigma\langle\mathcal{A}\rangle$ o que implica que $\mathcal{A} \subseteq \mathcal{F} = \sigma\langle\mathcal{A}\rangle$. □

4.4 Corolário *Sejam $(\Omega, \mathcal{F})$ um espaço mensurável e $f : \Omega \to \mathbb{R}$ uma função real. Nestas condições são equivalentes as seguintes afirmativas:*

a f *é* $\mathcal{F} - \mathcal{B}(\mathbb{R})$ *mensurável (mensurável a Borel);*

b $\{\omega \in \Omega : f(\omega) \leq x\} = f^{-1}((-\infty, x]) \in \mathcal{F}$ *para todo* $x \in \mathbb{R}$*;*

c $\{\omega \in \Omega : f(\omega) < x\} = f^{-1}((-\infty, x)) \in \mathcal{F}$ *para todo* $x \in \mathbb{R}$*;*

d $\{\omega \in \Omega : f(\omega) \geq x\} = f^{-1}([x, +\infty)) \in \mathcal{F}$ *para todo* $x \in \mathbb{R}$*;*

e $\{\omega \in \Omega : f(\omega) > x\} = f^{-1}((x, +\infty)) \in \mathcal{F}$ *para todo* $x \in \mathbb{R}$*;*

Demonstração. *Basta lembrar que*

$$\begin{aligned}
\mathcal{B}(\mathbb{R}) &= \sigma\langle\{(-\infty, x] : x \in \mathbb{R}\}\rangle \\
&= \sigma\langle\{(-\infty, x) : x \in \mathbb{R}\}\rangle \\
&= \sigma\langle\{[x, +\infty) : x \in \mathbb{R}\}\rangle \\
&= \sigma\langle\{(x, +\infty) : x \in \mathbb{R}\}\rangle
\end{aligned}$$

Sigamos para alguns exemplos...

4.5 Exemplo (Função Indicadora de um Evento) *Dado um espaço mensurável* $(\Omega, \mathcal{F})$ *e* $A \in \mathcal{F}$*, a função indicadora de* A*, geralmente denotada por* $\mathbb{1}_A$ *ou por* $I\{A\}$ *é definida por:*

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

Agora, dado $B \in \mathcal{B}(\mathbb{R})$ *note que*

$$\mathbb{1}_A^{-1}(B) = \begin{cases} \emptyset & , \quad 0, 1 \notin B; \\ \Omega & , \quad 0, 1 \in B; \\ A & , \quad 0 \notin B \text{ e } 1 \in B; \\ A^c & , \quad 1 \notin B \text{ e } 0 \in B. \end{cases}$$

E portanto, $\mathbb{1}_A$ *é uma função mensurável.*

4.6 Exemplo (Funções Simples) *Podemos modificar o exemplo anterior de modo a criar uma função* X *cuja imagem seja um conjunto finito* $\{a_1, \ldots, a_n\}$.

Para isso tomemos uma partição $A_1, A_2, \ldots, A_n \in \mathcal{F}$ *de* Ω *e defina*

$$\mathrm{X}(\omega) = \sum_{k=1}^{n} a_k \mathbb{1}_{A_k}(\omega).$$

Neste caso, tomando os a_ks distintos, teríamos que $A_k = \mathrm{X}^{-1}(\{a_k\})$ e a mensurabilidade de X pode ser verificada observando que

$$\mathrm{X}^{-1}(B) = \bigcup_{k : a_k \in B} A_k \in \mathcal{F},$$

onde a união acima é vazia se $\{k : a_k \in B\} = \emptyset$.

A próxima proposição é um resultado de central importância, e nos fornece uma gama enorme de funções mensuráveis.

4.7 Proposição (Mensurabilidade de aplicações contínuas) *Seja (Ω_1, d_1) e (Ω_2, d_2) espaços métricos e seja $f : \Omega_1 \to \Omega_2$ uma aplicação contínua. Então f é $\mathcal{B}(\Omega_1) - \mathcal{B}(\Omega_2)$-mensurável.*

Demonstração. *Da continuidade de f, sabemos que $f^{-1}(A)$ é um aberto de Ω_1 sempre que $A \in \mathcal{B}(\Omega_2)$ é aberto. E logo f é mensurável.* □

Os exemplos acima mostram que a família de funções mensuráveis é, em geral, bastante rica. Para terminar essa seção, um resultado bastante útil relacionado a funções mensuráveis.

4.8 Proposição *Dadas funções mensuráveis $g : \Omega_1 \to \Omega_2$ e $f : \Omega_2 \to \Omega_3$. Então a composição $f \circ g$ é mensurável.*

Demonstração. *Exercício* □

4.9 Exemplo *Se $f : (\Omega, \mathcal{F}) \to (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ é uma função mensurável, tal que*

$$f(\omega) = (f_1(\omega), \ldots, f_n(\omega)),$$

com $f_k : \Omega \to \mathbb{R}$, $k = 1, \ldots, n$, então f_k é uma função mensurável, para cada $k = 1, \ldots, n$.

Para ver isso tome a projeção canônica $\pi_k : \mathbb{R}^n \to \mathbb{R}$ dada por $\pi(x_1, \ldots, x_n) = x_k$ e note que

$$\pi_k^{-1}((-\infty, x]) = \mathbb{R}^{k-1} \times (0, x]\mathbb{R}^{n-k} \in \mathcal{B}(\mathbb{R}^n),$$

e, pelo corolário 4.4, π_k é mensurável.

Para ver que f_k é mensurável basta notar que $f_k = \pi_k \circ f$, e o resultado segue da proposição 4.8. ◁

4.2 Variáveis Aleatórias

Variáveis aleatórias são um elemento central na teoria das probabilidades. Informalmente, elas são usadas para representar quantidades associadas à um dado experimento aleatório. Ao rolar um dado duas vezes, podemos estar interessados no valor observado no primeiro dado, no segundo ou até mesmo na soma dos dois valores. Cada uma destas quantidades é uma variável aleatória.

Formalmente, definimos uma variável aleatória como uma função real mensurável quando colocamos em $\mathbb{R}$ a σ-álgebra de Borel $\mathcal{B}(\mathbb{R})$.

4.10 Definição *Uma **variável aleatória** é uma função mensurável que toma valores em* $\mathbb{R}$

$$\mathrm{X} : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R})).$$

4.11 Observação *Ao longo do curso veremos que o comportamento da imagem é a parte mais importante na caracterização de uma variável aleatória, e o domínio frequentemente assume um papel secundário. E é por essa razão que tais funções são conhecidas como* variáveis aleatórias, *e não simplesmente por funções mensuráveis.* ◁

No caso em que o espaço de chegada é $\mathbb{R}^n$, a função mensurável X é denominada **vetor aleatório**. Neste caso é comum denotarmos $X = (X_1, \ldots, X_n)$, onde cada $X_k, k = 1, \ldots, n$ é uma variável aleatória, como vimos no exemplo 4.9.

4.12 Exemplo *O primeiro exemplo que consideraremos é o lançamento de moeda. Nesse caso,temos que o espaço amostral é* $\Omega = \{H, T\}$ *e* $\mathcal{F} = \mathcal{P}(\Omega)$.

Então, temos a seguinte a variável aleatória

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

definida como:

$$\mathrm{X}(H) = 1,$$
$$\mathrm{X}(T) = 0.$$

O segundo exemplo, dado abaixo em forma de definição, é simplesmente uma adaptação de notação do exemplo sobre funções simples dada acima.

4.13 Definição (Variável Aleatória Simples) *Seja* $(\Omega, \mathcal{F})$ *um espaço mensurável. Uma aplicação* $X : \Omega \to \mathbb{R}$ *é dita variável aleatória simples se houver um* $n \in \mathbb{N}$ *e conjuntos mensuráveis mutuamente disjuntos* $A_1, \ldots, A_n \in \mathcal{F}$*, assim como um conjunto finito* $\{a_k, k = 1, \ldots, n\} \subset R$*, tal que*

$$X = \sum_{k=1}^{n} a_k \mathbb{1}_{A_k}.$$

É importante observar que uma variável aleatória simples pode ser expressa de várias maneiras diferentes, bastando para isso que escolhamos diferentes conjuntos mensuráveis mutuamente disjuntos. Além disso, para transformar $A_1, \dots, A_n \in \mathcal{F}$ em uma partição, como usamos no exemplo de função simples, basta fazer $A_0 = (\cup_{k=1}^n A_k)^c$ e $a_0 = 0$. Com isso teremos

$$\mathrm{X} = \sum_{k=1}^{n} a_k \mathbb{1}_{A_k} = \sum_{k=0}^{n} a_k \mathbb{1}_{A_k}.$$

O resultado abaixo já foi estudado na seção anterior, mas é interessante chamar a atenção e reescrevê-lo com a notação introduzida aqui.

4.14 Teorema *Considere a função* $\mathrm{X} : (\Omega, \mathcal{F}, \mathbf{P}) \to \mathbb{R}$. *Então* X *é a variável aleatória, se e somente se, o conjunto* $\{\omega : \mathrm{X}(\omega) \leq a\}$ *é um evento para todo* $a \in \mathbb{R}$.

Demonstração. *Usaremos o Teorema 4.3 com*

$$(\Omega_1, \mathcal{F}_1, \mu_1) = (\Omega, \mathcal{F}, \mathbf{P})$$
$$(\Omega_2, \mathcal{F}_2, \mu_2) = (\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu)$$

e lembrando que $\mathcal{B}(\mathbb{R}) = \sigma\big\langle \{(-\infty, a], a \in \mathbb{R}\}\big\rangle$. *Então* $\mathrm{X}^{-1}((-\infty, a]) = \{\omega : \mathrm{X}(\omega) \leq a\}$. □

Antes de dar sequência, vamos estabelecer uma notação que vamos usar bastante daqui para frente.

Para isso, dada uma variável aleatória X e um boreliano $A \in \mathcal{B}(\mathbb{R})$ vamos, a partir deste ponto, usar a seguinte notação:

$$\{\mathrm{X} \in A\} := \{\omega : \mathrm{X}(\omega) \in A\} = \mathrm{X}^{-1}(A).$$

Do mesmo modo, para $a, b \in \mathbb{R}$,

$$\{\mathrm{X} \leq a\} := \{\omega : \mathrm{X}(\omega) \leq a\} = \mathrm{X}^{-1}((-\infty, a]),$$
$$\{\mathrm{X} > b\} := \{\omega : \mathrm{X}(\omega) > b\} = \mathrm{X}^{-1}((b, +\infty)),$$
$$\{b \leq \mathrm{X} \leq a\} := \{\omega : b \leq \mathrm{X}(\omega) \leq a\} = \mathrm{X}^{-1}([b, a]),$$

e assim por diante.

Para variáveis aleatórias X, Y, faremos

$$\{X \in A, Y \in B\} = \{X \in A\} \cap \{Y \in B\}.$$

A notação acima, além de mais simples, coloca em evidência de forma mais clara a importância que a teoria das probabilidades dá ao comportamento da imagem de X, tirando o foco no domínio. Mais uma vez, o domínio ainda cumpre um papel importante, mas a imagem é, em geral, onde focaremos nossa atenção.

4.3 Funções de Variáveis Aleatórias

4.15 Proposição *Se* $X_1, \ldots, X_n$ *são variáveis aleatórias reais, então* $(X_1, \ldots, X_n)$ *é um vetor aleatório.*

Demonstração. *Utilizando o Teorema 4.14 basta verificar a mensurabilidade do mapa*

$$\omega \to (X_1(\omega), \ldots, X_n(\omega))$$

em uma classe de geradores da σ*-álgebra de Borel em* $\mathbb{R}^n$ *como por exemplo os geradores*

$$\{(-\infty,a_1] \times (-\infty, a_2] \times \cdots \times (-\infty, a_n] : a_1,a_2, \ldots ,a_n \in \mathbb{R}\}.$$

Para esse fim consideraremos as pré-imagens:

$$\begin{aligned}
&\{(X_1, \ldots, X_n) \in (-\infty,a_1] \times (-\infty, a_2] \times \cdots \times (-\infty, a_n]\} \\
&= \{X_1 \in (-\infty, a_1], \ldots, X_n \in (-\infty, a_n]\} \\
&= \bigcap_{k=1}^{n} \{X_k \in (-\infty, a_k]\} = \bigcap_{k=1}^{n} \{X_k \leq a_k\}.
\end{aligned}$$

Logo, cada conjunto $\{X_k \leq a\}$ *é mensurável (pois* X_k *é uma variável aleatória), e suas intersecções são mensuráveis pelas propriedades de* σ*-álgebra.* □

4.16 Teorema (Funções de Variáveis Aleatórias) *Dadas variáveis aleatórias reais* $X_1, \ldots, X_n$ *e* $f : \mathbb{R}^n \to \mathbb{R}$ *uma função mensurável. Então,* $f(X_1, \ldots, X_n)$ *é a variável aleatória real.*

Demonstração. *A função* $f(X_1, \ldots ,X_n)$ *é composição de duas funções mensuráveis:*

- *a função*
$$\omega \mapsto (X_1(\omega), \ldots, X_n(\omega))$$
é mensurável por 4.15.

- *a função*
$$(x_1, \ldots, x_n) \mapsto f(x_1, \ldots, x_n)$$
é mensurável por hipótese .

Como a composição de funções mensuráveis é mensurável temos o resultado desejado. □

4.17 Exemplo *Como corolário do Teorema anterior temos que se* X *é a variável aleatória, então:*

- X^2
- $\text{sen}(X)$
- e^X

são variáveis aleatórias.

Se $X_1, \dots, X_n$ *são variáveis aleatórias, então:*

- $X_1 + \dots + X_n$
- $X_1 \cdot X_2 \cdot \dots \cdot X_n$

são variáveis aleatórias. ◁

4.18 Teorema (Limites de Variáveis Aleatórias) *Se* $X_1, X_2, \dots$ *são variáveis aleatórias, então,* $\sup X_n$, $\inf X_n$, $\limsup X_n$, $\liminf X_n$, *e* $\lim X_n$ *são variáveis aleatórias quando existirem.*

Demonstração.

1 *Demonstração de que* $\sup X_n$ *é variável aleatória. Observamos que é suficiente demonstrar que* $\{\sup X_n \leq a\}$ *é uma variável aleatória* $\forall\, a \in \mathbb{R}$. *Mas esse conjunto é igual à*

$$\bigcap_{n=1}^{\infty} \{X_n \leq a\},$$

que é mensurável.

2 *Demonstração de* $\inf X_n$: *Exercício*

3 *Demonstração de* $\limsup X_n$: *Observe que este conjunto pode ser escrito como*

$$\inf_n \sup_{k \geq n} X_k$$

e assim o resultado segue dos itens 1 e 2. □

4.19 Exemplo *Se* $X_1, X_2, \dots$ *são variáveis aleatórias , então*

$$\sum_{n=1}^{\infty} X_n.$$

também o é.

Esse fato é verdadeiro pois o valor de uma série infinita é o supremo das somas parciais, e as somas parciais são variáveis aleatórias. ◁

4.4 Variáveis Aleatórias Geram σ-álgebras

Nesta seção entraremos um pouco mais a fundo na questão da mensurabilidade de uma variável aleatória, tentado entender um pouco mais afundo que tipo de informação sobre o nosso experimento aleatório a variável carrega.

Comecemos com um exemplo simples.

4.20 Exemplo *Considere o exemplo do lançamento de uma moeda n vezes, de maneira independente. Como de costume, tomamos o espaço amostral $\Omega = \{0,1\}^n$ ("0" representando coroa e "1" representando cara), e o munimos com a σ-álgebra 2^Ω, uma vez que o espaço é finito.*

Seja então a variável aleatória N = número de caras observadas. Explicitamente, $N : \Omega \to \mathbb{R}$ é dada por

$$N(\omega) = \sum_{k=1}^{n} \omega_k.$$

Esta variável só enxerga o número de caras, e portanto não carrega consigo nenhuma informação sobre quais lançamentos tais caras foram observadas ou se existem ou não lançamentos consecutivos de cara.

Assim, se soubermos que $N = 3$, por exemplo, sabemos apenas que o resultado está no evento

$$\{N = 3\} = \{\omega \in \Omega : N(\omega) = 3\} = \{\omega \in \Omega : \omega_1 + \cdots + \omega_n = 3\},$$

e nada mais.

Em geral, qualquer informação encapsulada por N pode ser representado por um evento da forma $\{N \in A\}$, com $A \in \mathcal{B}(\mathbb{R})$. Neste caso específico, temos que

$$\{N \in A\} = \bigcup_{k \in A \cap \{1,\ldots,n\}} \{N = k\},$$

e portanto toda informação dada por N pode ser vista como união de eventos $\{N = k\}$. $\triangleleft$

É interessante observar que a família $\{\{N \in A\} : A \in \mathcal{B}(\mathbb{R})\}$, que coleciona toda a *informação* contida na variável N, é de fato uma σ-álgebra. Deixamos essa verificação como exercício, apontando apenas que o resultado segue do bom comportamento de N^{-1} em relação às operações de união, interseção e complemento, e do fato de $\mathcal{B}(\mathbb{R})$ ser uma σ-álgebra.

Com isso estamos prontos para a próxima definição.

4.21 Definição *Dada uma variável aleatória* X *definida em um espaço mensurável* $(\Omega,\mathcal{F})$, *definimos a* σ*-álgebra gerada por* X, *e denotamos por* $\sigma\langle \mathrm{X}\rangle$, *como a família*

$$\begin{aligned}\sigma\langle \mathrm{X}\rangle &:= \{\mathrm{X}^{-1}(A) : A \in \mathcal{B}(\mathbb{R})\}\\ &= \{\{\mathrm{X} \in A\} : A \in \mathcal{B}(\mathbb{R})\}\end{aligned}$$

Analogamente, dado um vetor aleatório $X = (X_1, X_2, \dots, X_n)$, *a* σ*-álgebra gerada por* $X_1, X_2, \dots, X_n$ *é a família*

$$\begin{aligned}\sigma\langle X_1, \dots, X_n\rangle &:= \sigma\langle X\rangle\\ &= \{X^{-1}(A) : A \in \mathcal{B}(\mathbb{R}^n)\}\\ &= \{\{(X_1, \dots, X_n) \in A\} : A \in \mathcal{B}(\mathbb{R}^n)\}\end{aligned}$$

Como comentamos anteriormente, $\sigma\langle \mathrm{X}\rangle$ pode ser vista como a σ-álgebra que carrega toda a informação encapsulada pela variável aleatória X. Além disso, $\sigma\langle \mathrm{X}\rangle$ é a menor σ-álgebra de subconjuntos de Ω para a qual X é mensurável.

Antes de seguirmos para os exemplos, um rápido resultado.

4.22 Proposição *Seja* X *uma variável aleatória e* $\mathcal{A} \subset \mathcal{B}(\mathbb{R})$ *tal que* $\mathcal{B}(\mathbb{R}) = \sigma\langle\mathcal{A}\rangle$. *Nestas condições*

$$\sigma\langle \mathrm{X}\rangle = \sigma\langle\{\{\mathrm{X} \in A\} : A \in \mathcal{A}\}\rangle.$$

O mesmo vale se X *é um vetor aleatório e* $\mathcal{A} \subset \mathcal{B}(\mathbb{R}^n)$ *tal que* $\mathcal{B}(\mathbb{R}) = \sigma\langle\mathcal{A}\rangle$.

Demonstração. *Defina*

$$\begin{aligned}\mathcal{F} &:= \sigma\langle\{\{\mathrm{X} \in A\} : A \in \mathcal{A}\}\rangle, \quad e\\ \mathcal{G} &:= \{A \in \mathcal{B}(\mathbb{R}) : \{\mathrm{X} \in A\} \in \mathcal{F}\}.\end{aligned}$$

Observe que $\mathcal{A} \subset \mathcal{G} \subset \mathcal{B}(\mathbb{R})$, *e portanto basta mostrarmos que* $\mathcal{G}$ *é uma* σ*-álgebra que o resultado segue.*

A verificação de que $\mathcal{G}$ *é uma* σ*-álgebra deixaremos como exercício para o leitor.* ◻

4.23 Corolário *Se* $X = (X_1, X_2, \dots, X_n)$, *então*

$$\sigma\langle X_1, \dots, X_n\rangle = \sigma\langle\{X_1 \in A_1, \dots, X_n \in A_n\} : A_1, \dots, A_n \in \mathcal{B}(\mathbb{R})\}\rangle.$$

Demonstração. *Basta notar que* $\mathcal{B}(\mathbb{R}^n)$ *é gerada por* $\{A_1 \times \cdots \times A_n : A_k \in \mathcal{B}(\mathbb{R}), k = 1, \dots, n\}$ *e que*

$$\{X_1 \in A_1, \dots, X_n \in A_n\} = \{X \in A_1 \times \cdots \times A_n\}.$$

4.24 Exemplo *Dados* $\Omega = \{0,1\}^{\mathbb{N}}$ *e* $\omega = (\omega_1, \omega_2, \ldots)$ *um elemento de* Ω. *Considere a variável aleatória* $X_n : \Omega \to \mathbb{R}$ *definida por* $X_n(\omega) = \omega_n$ *e seja* $\mathcal{F}_n = \sigma\langle X_1, \ldots, X_n\rangle$.

Calcule explicitamente $\mathcal{F}_1$ *e* $\mathcal{F}_2$

Solução: *Como* X_1 *só toma dois valores* 0 *ou* 1, $\sigma\langle X_1\rangle$ *é gerado por*

$$X_1^{-1}(\{0\}) = A_1 = \{\omega : \omega_1 = 0\} \quad e$$

$$X_1^{-1}(\{1\}) = B_1 = \{\omega : \omega_1 = 1\} = A_1^c$$

Logo $\mathcal{F}_1 = \{\Omega, \emptyset, A_1, A_1^c\}$

De modo análogo, $\sigma\langle X_2\rangle$ *é gerado por*

$$X_2^{-1}(\{0\}) = A_2 = \{\omega : \omega_2 = 0\} \quad e$$

$$X_2^{-1}(\{1\}) = B_2 = \{\omega : \omega_2 = 1\} = A_2^c$$

Assim $\sigma\langle X_2\rangle = \{\Omega, \emptyset, A_2, A_2^c\}$

Logo $\mathcal{F}_2 = \sigma\langle X_1, X_2\rangle$ *é gerado por* $\sigma\langle X_1\rangle$ *e* $\sigma\langle X_2\rangle$ *e assim é igual a*

$\{\Omega, \emptyset, A_1, A_1^c, A_2, A_2^c, A_1 \cap A_2, A_1 \cap A_2^c, A_1^c \cap A_2, A_1^c \cap A_2^c, A_1 \cup A_2, A_1 \cup A_2^c, A_1^c \cup A_2, A_1^c \cup A_2^c\}$

4.25 Proposição *Dados* X *uma variável aleatória em* Ω, $\sigma\langle X\rangle$ *a* σ*-álgebra gerada por* X *e* Y *uma variável aleatória em* Ω. *Então* Y *é mensurável com respeito a* $\sigma\langle X\rangle$ *se e somente se existe uma função mensurável* $f : \mathbb{R} \to \mathbb{R}$ *tal que* $Y = f(X)$.

Então, é assim que as variáveis aleatórias geram σ-álgebras. Desta forma podemos esquecer o espaço inicial e sua sigma álgebra. A próxima coisa é esquecer da probabilidade.

4.5 Variáveis Aleatórias Induzem Medidas de Probabilidade em $\mathbb{R}$

Seja X uma variável aleatória real. Para todo $A \subseteq \mathbb{R}$, definimos a probabilidade $\mathbf{P}(A)$ de A como

$$\mathbf{P}(A) := \mathbf{P}(X \in A) = \mathbf{P}(X^{-1}(A))$$

Observamos que o primeiro $\mathbf{P}$ é diferente do segundo $\mathbf{P}$. Então $\mathbf{P}$ é a medida de probabilidade induzida em $\mathbb{R}$

A probabilidade $\mathbf{P}$ é denominada de **distribuição** de X.

Na teoria do medida, o objeto central é o estudo dos espaços de medida $(\Omega, \mathcal{F}, \mu)$. Na teoria da probabilidade, em muitas situações tentamos esquecer rapidamente os espaços da medida e discutir as distribuições.

Falaremos mais sobre distribuições no Capítulo 6.

4.6 Independência de Variáveis Aleatórias

4.26 Definição *As variáveis aleatórias $X_1, \dots, X_n$ são ditas **variáveis aleatórias independentes** se*

$$\mathbf{P}(X_1 \in B_1, \dots, X_n \in B_n) = \mathbf{P}(X_1 \in B_1) \cdot \dots \cdot \mathbf{P}(X_n \in B_n)$$

para todos conjuntos de Borel $B_1, \dots, B_n$ in $\mathbb{R}$.

4.27 Exemplo (Construindo duas moedas independentes) *Dado o espaço amostral de uma moeda $\Omega = \{1,0\}$. Considere $\Omega_1 = \Omega \times \Omega$. Considere a σ-álgebra das partes e a probabilidade uniforme em Ω_1. Ou seja, dado $\omega = (\omega_1, \omega_2) \in \Omega$, $\mathbf{P}(\omega) = 1/4$.*

Considere as variáveis aleatórias:

$$X(\omega_1, \omega_2) = \omega_1$$

$$Y(\omega_1, \omega_2) = \omega_2$$

As variáveis X e Y são independentes.

4.28 Teorema *As variáveis aleatórias $X_1, \dots, X_n$ são independentes se e somente se $\sigma\langle X_1\rangle, \dots \sigma\langle X_n\rangle$ são independentes.*

Demonstração. *Exercício.* □

Como consequência do Teorema anterior temos que:

4.29 Proposição *As variáveis aleatórias $X_1, \dots, X_n$ são independentes se e somente se*

$$\mathbf{P}(X_1 \leq x_1, \dots, X_n \leq x_n) = \mathbf{P}(X_1 \leq x_1) \cdot \dots \cdot \mathbf{P}(X_n \leq x_n)$$

para todo $x_1, \dots, x_n \in \mathbb{R}$.

4.30 Proposição *Dadas funções $\{f_k\}_{i=1}^n$ mensuráveis. Se $X_1, \dots, X_n$ são variáveis aleatórias independentes, então $f_1(X_1), \dots, f_n(X_n)$ são independentes.*

Demonstração. *Primeiro observamos que*

$$\mathbf{P}(f_1(X_1) \in B_1, \dots, f_n(X_n) \in B_k) = \mathbf{P}(f_1(X_1) \in B_1) \cdots \mathbf{P}(f_n(X_n) \in B_n). \tag{4.1}$$

e que $f_k(X_k) \in B_k$ é equivalente a $X_k \in f^{-1}(B_k)$.

Como consequência da última proposição temos que se X,Y são independentes, então X^2, Y^4 são independentes.

4.31 Teorema (Lei 0-1 de Kolmogorov para Variáveis Aleatórias) *Suponha que* $(X_n : n \in \mathbb{N})$ *é uma sequência de variáveis aleatórias independentes. Então a* σ*-álgebra caudal* T *de* $(\sigma\langle X_n\rangle : n \in \mathbb{N})$ *contém somente eventos de probabilidade* 0 *ou* 1.

Mais ainda, toda variável aleatória mensurável com respeito a T *é constante quase certamente.*

Demonstração. *A demonstração será deixada como exercício. Mas faremos alguns comentários que podem ajudar na demonstração.*

Observamos que a parte inicial é análoga a demonstração do Teorema 0-1 *de Kolmogorov.*

Para a parte final: se Y *variável aleatória mensurável com respeito a* T *então* $\mathbf{P}(Y \leq y)$ *toma valores em* $\{0, 1\}$*, logo* $\mathbf{P}(Y = c) = 1$*, onde* $c = \inf\{y : \mathbf{P}(Y \leq y) = 1\}$.

Exercícios

Ex. 4.1 — Mostre que a composta de funções mensuráveis é mensurável

Ex. 4.2 — Operações com Variáveis Aleatórias
Mostre que se X, Y são variáveis aleatórias em $(\Omega, F, \mathbf{P})$ então também são variáveis aleatórias:

1. $|X|$
2. $X + Y$
3. $X \cdot Y$
4. $\max(X, Y)$
5. $\min(X, Y)$

Ex. 4.3 — Dado X uma variável aleatória tal que $\mathbf{P}(X > 0) > 0$. Prove que existe $\delta > 0$ tal que $\mathbf{P}(X > \delta) > 0$

Ex. 4.4 — Amostragem de uma variável aleatória exponencial
Encontre um algoritmo para produzir uma variável aleatória com distribuição $\mathrm{Exp}(\lambda)$ usando um gerador de números aleatórios que produz números aleatórios uniformes em $(0, 1)$. Em outras palavras, se $U \sim U(0, 1)$, encontre uma função $g : (0, 1) \to \mathbb{R}$ tal que o aleatório variável $X = g(U)$ tem distribuição $\mathrm{Exp}(\lambda)$.

Ex. 4.5 — A Propriedade de Falta de Memória

1. Dizemos que uma variável não negativa $X \geq 0$ tem a propriedade de perda de memória se $\mathbf{P}(X \geq t | X \geq s) = \mathbf{P}(X \geq t \geq s)$ para todo $0 < s < t$.

2. Mostre que as variáveis aleatórias exponenciais têm a propriedade da memória de falta.
3. Prove que qualquer variável aleatória não-negativa que tenha a propriedade falta de memória tenha distribuição exponencial para algum parâmetro $\lambda > 0$. (Isso é mais fácil se você assumir que a função $G(x) = \mathbf{P}(X \geq x)$ é diferenciável em $[0, \infty)$, assim você pode fazer essa suposição se você não conseguir um argumento mais geral).

Ex. 4.6 — Dado uma variável aleatória X que é independente de si própria. Mostre que X é constante com probabilidade 1.

Ex. 4.7 — Dado $\epsilon > 0$, e deixe X_i ser uma sequência não negativa de variáveis aleatórias independentes tais que $\mathbf{P}(X_i > \delta) > \epsilon$ para todo i. Prove que com probabilidade 1 :

$$\sum_{i=1}^{\infty} X_i = \infty$$

Ex. 4.8 — Prove que se X e Y são independentes então as famílias $\{X^+, Y^+\}$, $\{X^+, Y^-\}$, $\{X^-, Y^+\}$ e $\{X^-, Y^-\}$ são independentes.

CAPÍTULO 5

Integral de Lebesgue

5.1 Integral de Lebesgue

A integral de Lebesgue é definida em um espaço de medida arbitrário $(\Omega, \mathcal{F}, \mu)$. Por simplicidade, assumiremos que μ é uma medida finita, isto é $\mu(\Omega) < \infty$. Esta suposição não é realmente necessária mas tampouco é um problema nas medidas de probabilidade.

Como não temos uma forma natural de simplesmente particionar o domínio Ω (porque é abstrato), particionamos a imagem $\mathbb{R}$. Ou seja, faremos aproximações de funções simples.

Definiremos a integral de Lebesgue através dos seguintes passos:

$\boxed{1}$ Para funções simples

$$f = \sum_{k=1}^{n} a_k \mathbb{1}_{A_k}$$

para alguma decomposição $\Omega = \bigcup_{k=1}^{n} A_k$, defina

$$\int f \, \mathrm{d}\mu = \sum_{k=1}^{n} a_k \mu(A_k).$$

$\boxed{2}$ Para funções mensuráveis não negativas $f \geq 0$. Defina

$$\int f \, \mathrm{d}\mu = \sup_{0 \leq \phi \leq f} \int \phi \, \mathrm{d}\mu$$

com ϕ simples.

Se esta integral for finita, diremos que a função não negativa f é **integrável**.

3 Para funções gerais f, definimos para a parte positiva e negativa separadamente. Assim decompomos

$$f = f^+ - f^-,$$

onde $f^+ = \max(f,0)$ e $f^- = \max(-f,0)$.

Vamos detalhar essa construção.

O primeiro passo é mostrar que a integral de funções simples independe da escolha da sua representação.

5.1 Proposição *Dadas duas representações da função simples f*

$$f = \sum_{i=1}^{n} a_i \mathbb{1}_{A_i} \text{ e } \sum_{j=1}^{m} b_j \mathbb{1}_{B_j}$$

então

$$\sum_{i=1}^{n} a_i \mu(A_i) = \sum_{j=1}^{m} b_j \mu(B_j)$$

Demonstração. *Se $\mu(A_i \cap B_j) > 0$ então $A_i \cap B_j \neq \varnothing$, e $f(x) = a_i = b_j$ para todo $x \in A_i \cap B_j$. Assim*

$$\sum_{i=1}^{n} a_i \mu(A_i) = \sum_{i=1}^{n} \sum_{j=1}^{m} a_i \mu(A_i \cap B_j) \tag{5.1}$$

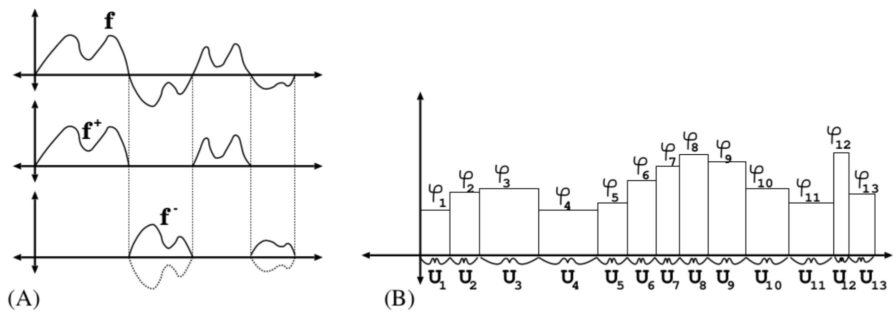
$$= \sum_{j=1}^{m} \sum_{i=1}^{n} b_j \mu(A_i \cap B_j) \tag{5.2}$$

$$= \sum_{j=1}^{m} b_j \mu(B_j) \tag{5.3}$$

5.2 Proposição (Aproximação por Funções Simples) *Toda função mensurável não-negativa $f : \Omega \to \mathbb{R}^+$ é o limite pontual de uma sequência monótona crescente de funções simples não negativas e se uma função f é o o limite pontual de uma sequência monótona crescente de funções simples então f é mensurável.*

Demonstração.

Figura 5.1: A- Parte Positiva e Negativa B- Função Simples - [26]



$\boxed{\Rightarrow}$ *Considere f uma função mensurável não-negativa definida sobre o espaço de medida $(\Omega, \mathcal{F}, \mu)$ como antes. Para cada $n \in \mathbb{N}$, subdivida o intervalo de f em $2^{2n}+1$ intervalos, 2^{2n} dos quais têm o comprimento 2^{-n}. Para cada n, considere*

$$I_{n,k} = \left[\frac{k-1}{2^n}, \frac{k}{2^n}\right)$$

para $k = 1, 2, \ldots, 2^{2n}$ e $I_{n,2^{2n}+1} = [2^n, \infty)$.

Observe que, para n fixo, os conjuntos $I_{n,k}$ são disjuntos e cobrem a reta real não negativa.

Agora, defina os conjuntos mensuráveis

$$A_{n,k} = f^{-1}(I_{n,k}) \text{ para } k = 1, 2, \ldots, 2^{2n}+1.$$

Então, a sequência crescente de funções simples

$$f_n = \sum_{k=1}^{2^{2n}+1} \frac{k-1}{2^n} \mathbf{1}_{A_{n,k}}$$

converge pontualmente para f quando $n \to \infty$. E se a função f for limitada, a convergência é uniforme.

A função anterior pode ser escrita também como

$$f_n := (2^{-n} \lfloor 2^n f \rfloor) \wedge n$$

onde $\wedge$ denota o mínimo entre os dois valores.

$\boxed{\Leftarrow}$ *Já provamos que o limite de funções mensuráveis é mensurável.* $\square$

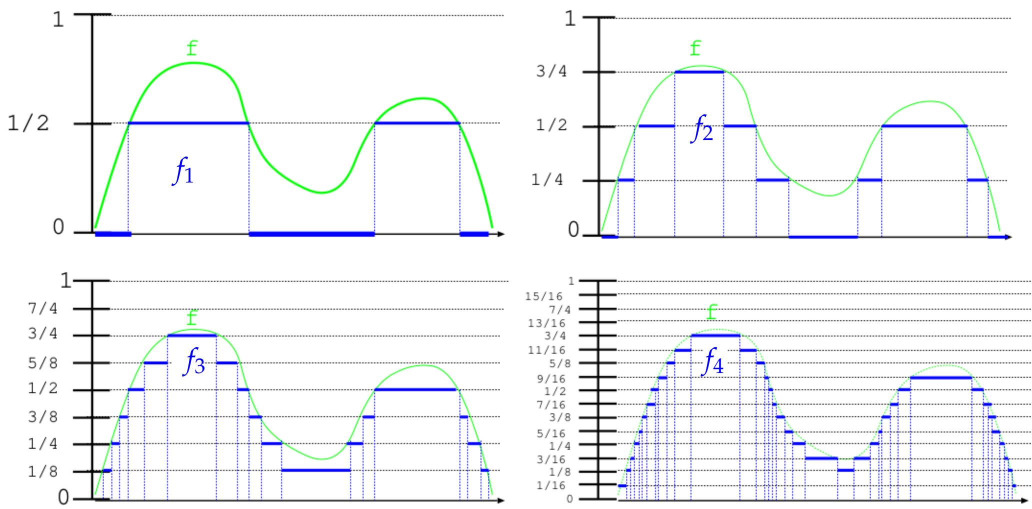
Ou seja, a classe das funções mensuráveis é o fecho por limites pontuais das funções simples.

Através de uma pequena adaptação dos argumentos anteriores podemos provar que:

5.3 Proposição (Aproximação por Funções Simples) *Dada função mensurável não-negativa $f : E \subset \Omega \to \mathbb{R}^+$. Se f é limitada em A então para todo $\varepsilon > 0$ existem funções simples ϕ_ε e ψ_ε em A tais que:*

1. *$\phi_\varepsilon(x) \leq f(x) \leq \psi_\varepsilon(x)$ para todo $x \in A$.*
2. *$0 \leq \psi_\varepsilon - \phi_\varepsilon < \varepsilon$ para todo $x \in A$.*

Figura 5.2: Aproximação por funções simples - [26]



5.4 Definição (Integrável)

1. *Dizemos que f é **integrável** se $|f|$ é integrável. Equivalentemente, se ambas as funções f^+ e f^- são integráveis. E nesse caso definimos*

$$\int f \ \mathrm{d}\mu = \int f^+ \ \mathrm{d}\mu - \int f^- \ \mathrm{d}\mu.$$

2. *Dizemos que f é **quase integrável** se pelo menos uma das funções f^+ e f^- é integrável. E nesse caso também definimos*

$$\int f \ \mathrm{d}\mu = \int f^+ \ \mathrm{d}\mu - \int f^- \ \mathrm{d}\mu.$$

2. *A integral num conjunto mensurável $A \subseteq \Omega$ é definida como*

$$\int_A f \ \mathrm{d}\mu = \int_\Omega f \cdot \mathbb{1}_A \ \mathrm{d}\mu.$$

5.2 Propriedades da Integral de Lebesgue

Uma das estratégias para demonstrar propriedades da Integral de Lebesgue é a denominada estratégia dos três passos:

Estratégia dos três passos

Para demonstrar uma propriedade da integral de Lebesgue a seguinte estratégia geral pode ser útil:

- Demonstramos que essa propriedade é satisfeita pela integral de funções simples.
- Utilizamos que uma função mensurável positiva pode ser aproximada monotonicamente por funções simples e de posse desse fato (e usando se necessário o Teorema da Convergência Dominada que demonstraremos a seguir) demonstramos essa propriedade para funções positivas.
- Para demonstrar para funções mensuráveis quaisquer usamos a decomposição em parte positiva e negativa.

5.5 Proposição

1. *(Monotonicidade). Se* $f \leq g$ *em quase todo ponto, então* $\displaystyle\int f \ \mathrm{d}\mu \leq \int g \ \mathrm{d}\mu$.
2. *Se* $f = g$ *em quase todo ponto, então as integrais são iguais.*
3. *(Linearidade)* $\displaystyle\int (af + bg) \ \mathrm{d}\mu = a \int f \ \mathrm{d}\mu + b \int g \ \mathrm{d}\mu$, *onde* $a, b \in \mathbb{R}$.
4. $\displaystyle\left|\int f \ \mathrm{d}\mu\right| \leq \int |f| \ \mathrm{d}\mu$.

Demonstração. *Como* $f \leq |f|$, *temos* $\displaystyle\int f \leq \int |f^+| \ \mathrm{d}\mu$. *De modo análogo,* $-f \leq |f|$, *logo* $\displaystyle-\int f \ \mathrm{d}\mu \leq \int |f^-| \ \mathrm{d}\mu$.

Para provar a Linearidade utilizaremos a estratégia dos três passos

A integral de Lebesgue para funções simples é linear. Se $\varphi = \sum_i a_i \mathbb{1}_{A_i}$, $\psi = \sum_j b_j \mathbb{1}_{B_j}$, *então*

$$\int_\Omega \varphi \, d\mu + \int_\Omega \psi \, d\mu = \sum_i a_i \mu(A_i) + \sum_j b_j \mu(B_j) \tag{5.4}$$

$$= \sum_i \sum_j a_i \mu(A_i \cap B_j) + \sum_j \sum_i b_j \mu(B_j \cap A_i) \tag{5.5}$$

$$= \sum_i \sum_j (a_i + b_j)\, \mu(A_i \cap B_j) \tag{5.6}$$

$$= \int_\Omega (\varphi + \psi) \, d\mu \,. \tag{5.7}$$

Para funções positivas $f, g > 0$ *e* a, b. *Observe que se* $f_n \uparrow f$ *e* $g_n \uparrow g$ *com* f_n *e* g_n *funções simples. Então*

$$a f_n + b g_n \uparrow a f + b g$$

e usamos que

$$\int a f + b g \, d\mu = \sup_\phi \{ \int \phi \, d\mu \} \geq \sup_{a f_n + b g_n} \{ \int a f_n + b g_n \, d\mu \} \tag{5.8}$$

$$= a \int f \, d\mu + b \int g \, d\mu \tag{5.9}$$

O outro lado da desigualdade segue da monotonicidade.

5.6 Teorema *Dado* $(\Omega, \mathcal{F}, \mu)$ *um espaço de medida. E funções* f, g *integráveis e não negativas, então*

1. $\int_\Omega f \, d\mu < \infty$ *então* $f < \infty$ μ*-q.c.*

2. $\int_\Omega f \, d\mu = 0 \iff f = 0$ μ*-q.c.*

3. *Se* $f = g$ μ*-q.c. então* $\int_\Omega f \, d\mu = \int_\Omega g \, d\mu$.

Observação. Uma consequência do Teorema 5.6 2 é que para qualquer função $f : \Omega \to \mathbb{R}$, tal que $\int f \, d\mu$ está definida, temos a liberdade de mudar os valores de $f(w)$ num conjunto μ-negligenciável sem alterar o valor da integral desde que a função resultante seja mensurável.

5.7 Definição (Integrável) *Seja $(\Omega, \mathcal{F}, \mu)$ um espaço de medida. Então $\mathcal{L}_1(\mu)$ denota o conjunto de funções $f : \Omega \to \mathbb{R}$ que são mensuráveis $\mathcal{F}$ com : $\int_\Omega f^+ \, \mathrm{d}\mu < \infty$ e $\int_\Omega f^- \, \mathrm{d}\mu < \infty$;*

Definimos a integral para funções simples e posteriormente para funções mensuráveis, que são limites de funções simples. Será que existe uma classe maior para a qual a integral definida como

$$\int f \, \mathrm{d}\mu = \sup_{0 \leq \phi \leq f} \int \phi \, \mathrm{d}\mu$$

converge? O próximo resultado nos tira essa esperança.

5.8 Teorema *Seja f uma função limitada em um conjunto de medida finita A. Então fé Lebesgue integrável sobre E se e somente se f for Lebesgue mensurável.*

Demonstração. *Faremos a demonstração assumindo sem perda de generalidade que $f \geq 0$. Como f é limitada e Lebesgue integrável pela Proposição 5.3 existem funções simples $\phi_n \leq f \leq \phi_n + \varepsilon$ com $\phi_n \uparrow$ tais que $\int \phi_n \uparrow \int f$.*

Como as funções ϕ_n são não decrescentes e limitadas temos a convergência pontual $\phi_n \to \phi$. Claramente teremos que $\phi \leq f$ e $\int_A \phi \, \mathrm{d}\mu = \int_A f \, \mathrm{d}\mu$

Então

$$0 \leq \int_A f \, \mathrm{d}\mu - \phi \leq \int_A \phi_n - \phi \, \mathrm{d}\mu \to 0$$

Logo $f = \phi$ quase certamente. Como ϕ é mensurável e a medida é completa terminamos.

5.3 Comparando a Integral de Lebesgue e de Riemann

Lembramos que

5.9 Teorema (Critério de Lebesgue para Riemann-Integrabilidade) *Uma função limitada num intervalo compacto $[a, b]$ é Riemann integrável se e somente se o conjunto de seus pontos de descontinuidade tem medida nula.*

5.10 Teorema *Dado uma função limitada $f : [a,b] \to \mathbb{R}$. Se f é Riemann integrável, então f é Lebesgue integrável, e ambas as integrais concordam*

Demonstração. *No que se segue denotaremos por $\int \ldots \, \mathrm{d}x$ as integrais de Riemann e por $\int \ldots \, \mathrm{d}\lambda(x)$ as integrais de Lebesgue.*

Como a função é Riemann integrável existem sequências φ_n e ψ_n de funções escada tais que $\varphi_n \leq f \leq \psi_n$ e

$$\int \varphi_n \, dx \to \int f \, dx \leftarrow \int \psi_n \, dx.$$

Fazendo a troca de φ_n *e* ψ_n *por* $\max\{\varphi_1, \ldots, \varphi_n\}$ *e* $\min\{\psi_1, \ldots, \psi_n\}$*, podemos assumir sem perda de generalidade que as sequências* φ_n *e* ψ_n *são crescente e decrescente respectivamente.*

Para funções escadas $\gamma : [a,b] \to \mathbb{R}$*, a integral de Lebesgue e a de Riemann coincidem. E assim,*

$$\int |\psi_n - \varphi_n| \, d\lambda(x) = \int \psi_n - \varphi_n \, dx \to \int f \, dx - \int f \, dx = 0. \tag{5.10}$$

Pela monotonicidade, $\varphi_n \to \varphi$ *e* $\psi_n \to \psi$ *pontualmente com* $\varphi_1 \leq \varphi \leq f \leq \psi \leq \psi_1$.

Pelo Teorema da Convergência Dominada

$$\int |\psi_n - \varphi_n| \, d\lambda(x) \to \int |\psi - \varphi| \, d\lambda(x).$$

Pela Equação (5.10), temos $\displaystyle\int |\psi - \varphi| \, d\lambda(x) = 0$ *e logo* $\psi = \varphi$ *quase certamente.*

Como $\varphi \leq f \leq \psi$*,temos* $f = \varphi = \psi$ *quase certamente, logo* f *é Lebesgue mensurável e*

$$\int f \, d\lambda(x) = \int \psi \, d\lambda(x) = \lim_n \int \psi_n \, d\lambda = \lim \int \psi_n \, dx = \int f \, dx.$$

5.11 Exemplo *Considere a função indicadora dos racionais* $f = \mathbb{1}_{\mathbb{Q}}$*. Como* $\mathbb{Q}$ *é enumerável,* $\mu(\mathbb{Q}) = 0$*. Logo* $\displaystyle\int f \, d\mu = 0$*. Portanto,* f *é Lebesgue integrável.*

Mas f *não é Riemann integrável pois o valor mais alto em cada intervalo de qualquer partição é* 1 *e o menor é* 0. ◁

5.12 Teorema (Teorema de Lusin) *Dado* $A \subset \mathbb{R}$ *um conjunto mensurável,* $\mu(A) < \infty$ *e* f *uma função mensurável com domínio* A*. Então, para todo* $\varepsilon > 0$ *existe um conjunto compacto* $K \subset A$ *com* $\mu(A \setminus K) < \varepsilon$*, tal que a restrição de* f *a* K *é contínua.*

Demonstração. *Dado* $\{V_n\}_{n\in\mathbb{N}}$ *uma enumeração dos intervalos abertos com extremos racionais. Fixe conjuntos compactos* $K_n \subset f^{-1}[V_n]$ *e* $K'_n \subset A \setminus f^{-1}[V_n]$ *para cada* n*, de modo que* $\mu\big(A \setminus (K_n \cup K'_n)\big) < \varepsilon/2^n$.

Seja

$$K = \bigcap_{n=1}^{\infty} \big(K_n \cup K'_n\big).$$

Claramente $\mu(A \setminus K) < \varepsilon$.

Dado $x \in K$ *e* n*, tal que* $f(x) \in V_n$*, podemos provar a continuidade de* f *quando restrita a* K*, escolhendo uma vizinhança compacta* $\tilde{K}_n$*, tal que* $x \in \tilde{K}_n$ *e* $f(\tilde{K}_n \cap K) \subset V_n$. □

5.4 Integração e Limites

5.13 Teorema (Teorema da Convergência Monótona) *Se $f_n \geq 0$ e $f_n \uparrow f$, então $\int f_n \, \mathrm{d}\mu \to \int f \, \mathrm{d}\mu$.*

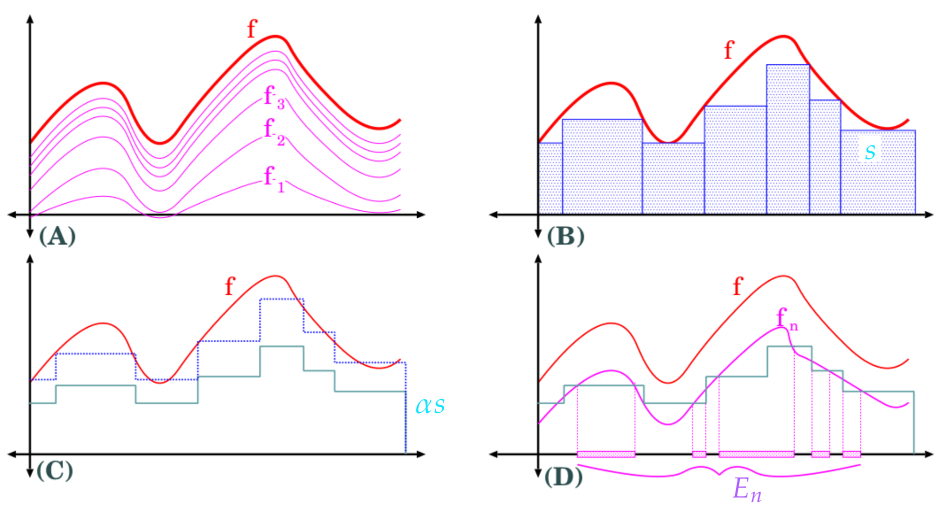
Demonstração. *Como $\int f_n \, \mathrm{d}\mu \leq \int f_{n+1} \, \mathrm{d}\mu$ temos que existe $\alpha \in [0, \infty]$ tal que*

$$\lim_{n\to\infty} \int f_n \, \mathrm{d}\mu = \alpha$$

Como $f_n \leq f$ para todo n temos que $\int f_n \, \mathrm{d}\mu \leq \int f \, \mathrm{d}\mu$ para todo n. E logo $\alpha \leq \int f \, \mathrm{d}\mu$. Seja s uma função simples mensurável tal que $0 \leq s \leq f$ e seja c uma constante tal que $0 < c < 1$ e defina

$$E_n = \{x | f_n(x) \geq c \cdot s(x)\}$$

Figura 5.3: Teorema da Convergência Monótona - [26]



Cada conjunto E_n é mensurável, a sequência é não decrescente, i.e, $E_1 \subset E_2 \subset \cdots$, e $\Omega = \bigcup_n E_n$. Para ver a última igualdade tome $x \in \Omega$. Se $f(x) = 0$ então $x \in E_1$, caso contrário $f(x) > 0$ e nesse caso $c \cdot s(x) < f(x)$ pois $c < 1$. Logo, temos que $x \in E_n$ para algum n. Além disso,

$$\int_\Omega f_n \, \mathrm{d}\mu \geq \int_{E_n} f_n \, \mathrm{d}\mu \geq c \int_{E_n} s \, \mathrm{d}\mu$$

para todo n. Fazemos $n \to \infty$ e assim obtemos que $\alpha \geq c \int_\Omega s \, \mathrm{d}\mu$. Como esse fato é verdadeiro para todo $c < 1$ temos que

$$\alpha \geq \int s \, \mathrm{d}\mu$$

para toda função simples $s \leq f$. *Logo, temos*

$$\alpha \geq \int_\Omega f \, \mathrm{d}\mu$$

Isso prova a desigualdade reversa. □

5.14 Proposição *Seja* $f : \Omega \to [0, \infty]$ *uma aplicação mensurável. Então:*

$\boxed{1}$ $f = 0$ *quase certamente se e somente se* $\int_\Omega f \, \mathrm{d}\mu = 0$.

$\boxed{2}$ *Se* $\int_\Omega f \, \mathrm{d}\mu < \infty$, *então* $f < \infty$ *quase certamente.*

Demonstração. *Provaremos para funções não negativas.*

$\boxed{1}$ $\boxed{\Rightarrow}$ *Suponha que* $f = 0$ *quase certamente e seja* $N : \{x : f(x) > 0\}$. *Então* $f \leq \infty \mathbb{1}_N$. *Também temos que* $n\mathbb{1}_N \uparrow \infty\mathbb{1}_N$. *Logo*

$$0 \leq \int f \, \mathrm{d}\mu \leq \int \infty \mathbb{1}_N \, \mathrm{d}\mu = \lim \int n\mathbb{1}_N \, \mathrm{d}\mu = 0$$

$\boxed{1}$ $\boxed{\Leftarrow}$ *Seja* $N_n : \{x : f(x) > \frac{1}{n}\}$. *Então* $N_n \uparrow N$ *e*

$$0 = \int f \, \mathrm{d}\mu \geq \int \frac{1}{n} \mathbb{1}_{N_n} \, \mathrm{d}\mu = \frac{\mu(N_n)}{n}$$

Logo $\mu(N) = 0$.

$\boxed{2}$ *Seja* $I : \{x : f(x) = \infty\}$. *Então para todo* n

$$\mu(I) = \int \mathbb{1}_I \, \mathrm{d}\mu \leq \int \mathbb{1}_{\{f \geq n\}} \, \mathrm{d}\mu \tag{5.11}$$

$$\leq \frac{1}{n} \int f \, \mathrm{d}\mu \leq \frac{1}{n} \int f \mathbb{1}_{\{f \geq n\}} \, \mathrm{d}\mu \to 0 \tag{5.12}$$

5.15 Lema (Lema de Fatou) *Se* $f_n \geq 0$, *então* $\displaystyle\int \liminf f_n \, \mathrm{d}\mu \leq \liminf \int f_n \, \mathrm{d}\mu.$ ◁

Demonstração. *Dado* $g_k(x) = \inf_{i \geq k} f_i(x)$. *Então* $g_k \leq f_k$,*e assim*

$$\int g_k \, \mathrm{d}\mu \leq \int f_k \, \mathrm{d}\mu$$

Mais ainda temos que $0 \leq g_1 \leq g_2 \leq \cdots$, *e cada* g_k *é mensurável com* $g_k(x) \to \liminf f_n(x)$ *quando* $k \to \infty$. *O Teorema da Convergência Monótona implica que o lado esquerdo tende ao lado direito quando* $k \to \infty$. □

5.16 Definição (Convergência quase todo ponto) *Dizemos que* $f_n \to f$ *em quase todo ponto ou quase certamente (q.c.) se* $\mu(\{\omega : f_n(\omega) \nrightarrow f(\omega)\}) = 0$.

5.17 Teorema (Teorema da Convergência Dominada) *Se $f_n \to f$ q.c. e $|f_n| \leq g$ para todo n, e g é integrável, então*

$$\int f_n \; \mathrm{d}\mu \to \int f \; \mathrm{d}\mu.$$

Demonstração. *Também é claro que $f \in L^1(\mu)$ pois $|f| \leq g$ e f é mensurável. Também temos que $|f_n - f| \leq 2g$ e assim podemos aplicar o Lema de Fatou para a função $2g - |f_n - f|$ e obter que*

$$\begin{aligned}\int_\Omega 2g \; \mathrm{d}\mu &\leq \liminf_{n\to\infty} \int_\Omega (2g - |f_n - f|) \, \mathrm{d}\mu \\ &= \int_\Omega 2g \; \mathrm{d}\mu + \liminf_{n\to\infty} \left(- \int_\Omega |f_n - f| \, \mathrm{d}\mu \right) \\ &= \int_\Omega 2g \; \mathrm{d}\mu - \limsup_{n\to\infty} \int_\Omega |f_n - f| \, \mathrm{d}\mu\end{aligned}$$

Como $\int_\Omega 2g \; \mathrm{d}\mu$ é finita podemos subtraí-la

$$\limsup_{n\to\infty} \int_\Omega |f_n - f| \, \mathrm{d}\mu \leq 0$$

E assim

$$\lim_{n\to\infty} \int_\Omega |f_n - f| \, \mathrm{d}\mu = 0$$

Se $f \in L^1$ então $|\int_x f| \leq \int_\Omega |f|$ e temos

$$\lim_{n\to\infty} \int_\Omega f_n \; \mathrm{d}\mu = \int_\Omega f \; \mathrm{d}\mu$$

5.5 Espaços $\mathcal{L}^p$

Nessa seção vamos estabelecer alguns resultados de análise funcional referente aos espaços das funções integráveis.

5.18 Definição

- *Dado um espaço de medida $(\Omega, \mathcal{F}, \mu)$ e $p \geq 1$ definimos $\mathcal{L}^p(\Omega, \mathcal{F}, \mu)$ como o espaço das classes de equivalência das funções mensuráveis tais que*

$$\int |f|^p \;\; \mathrm{d}(\mu) < \infty$$

sob a relação de equivalência de igualdade quase certa.

- *Para qualquer elemento $f \in \mathcal{L}^p(\Omega, \mathcal{F}, \mu)$ definimos a p-norma como*

$$\|f\|_p = \left(\int |f|^p \ \mathrm{d}(\mu) \right)^{1/p}$$

Nosso primeiro objetivo é demonstrar que $\mathcal{L}^p(\Omega, \mathcal{F}, \mu)$ é um espaço vetorial normado completo, ou seja, que é um espaço de Banach. Como primeiro passo nessa direção precisamos provar a desigualdade triangular.

5.19 Proposição (Desigualdade de Minkowski) *Dadas $f, g \in \mathcal{L}^p(\Omega, \mathcal{F}, \mu)$ então*

- $f + g \in \mathcal{L}^p(\Omega, \mathcal{F}, \mu)$
- $\|f + g\|_p \leq \|f\|_p + \|g\|_p$.

Demonstração. *Note que podemos assumir que $f \geq 0$ e $g \geq 0$ pois se provarmos a desigualdade para funções positivas, então o caso geral pode ser obtido aplicando a desigualdade triangular e usando o fato de que a função x^p é crescente*

Vamos provar então que

$$\|f + g\|_p \leq \|\,|f| + |g|\,\|_p \leq \|\,|f|\,\|_p + \|\,|g|\,\|_p = \|f\|_p + \|g\|_p$$

O caso $p = 1$ segue imediatamente da linearidade da integral.

Para $1 < p < \infty$, primeiro demonstraremos que $f + g \in \mathcal{L}^p(\Omega, \mathcal{F}, \mu)$:

$$(f + g)^p \leq (f \vee g + f \vee g)^p = 2^p (f^p \vee g^p) \leq 2^p (f^p + g^p)$$

e, portanto, $\|f + g\|_p^p \leq 2^p (\|f\|_p^p + \|g\|_p^p) < \infty$

Para provar a desigualdade triangular, note que podemos supor que $\|f + g\|_p > 0$ caso contrário, a desigualdade triangular segue da positividade da norma. Escrevemos

$$\|f + g\|_p^p = \int (f + g)^p \ \mathrm{d}(\mu) = \int f(f + g)^{p-1} \ \mathrm{d}(\mu) + \int g(f + g)^{p-1} \ \mathrm{d}(\mu)$$

Agora podemos aplicar a desigualdade Hölder para cada um dos termos do lado direito e usar o fato de que $\frac{1}{p} + \frac{1}{q} = 1$ é equivalente a $p = (p - 1)q$ para ver que

$$\int f(f + g)^{p-1} \ \mathrm{d}(\mu) \leq \left(\int f^p \ \mathrm{d}(\mu) \right)^{\frac{1}{p}} \left(\int (f + g)^{(p-1)q} \ \mathrm{d}(\mu) \right)^{\frac{1}{q}} = \|f\|_p \|f + g\|_p^{p/q}$$

Aplicando este argumento ao termo $\int g(f + g)^{p-1} \ \mathrm{d}(\mu)$ temos

$$\|f + g\|_p^p \leq (\|f\|_p + \|g\|_p) \cdot \|f + g\|_p^{p/q}$$

dividindo por $\|f + g\|_p^{p/q}$ e usando que $p - \frac{p}{q} = 1$ podemos concluir que $\|f + g\|_p \leq \|f\|_p + \|g\|_p$. □

5.20 Proposição (Completude de $\mathcal{L}^p$) *Para $p \geq 1$ o espaço vetorial normado $\mathcal{L}^p(\Omega, \mathcal{F}, \mu)$ é completo.*

Demonstração. *Seja f_n uma sequência de Cauchy em $\mathcal{L}^p(\Omega, \mathcal{F}, \mu)$. O primeiro passo consiste em mostrar que existe uma subsequência de f_n que converge quase certamente para um elemento $f \in \mathcal{L}^p(\Omega, \mathcal{F}, \mu)$.*

Pela propriedade de Cauchy, para cada $j \in \mathbb{N}$ podemos encontrar um $n_j > 0$ tal que $\|f_m - f_n\|_p \leq \frac{1}{2^j}$ para todo $m > n_j$. Desta forma, obtemos uma subsequência f_{n_j} tal que $\|f_{n_{j+1}} - f_{n_j}\|_p \leq \frac{1}{2^j}$ para todos os $j \in \mathbb{N}$. Agora aplicando o Teorema da Convergência Monótona e a desigualdade triangular obtemos

$$\begin{aligned}
\| \sum_{j=1}^{\infty} \left|f_{n_{j+1}} - f_{n_j}\right| \|_p &= \lim_{N\to\infty} \| \sum_{j=1}^{N} \left|f_{n_{j+1}} - f_{n_j}\right| \|_p \\
&\leq \lim_{N\to\infty} \sum_{j=1}^{N} \|f_{n_{j+1}} - f_{n_j}\|_p \\
&\leq \lim_{N\to\infty} \sum_{j=1}^{N} \frac{1}{2^j} < \infty
\end{aligned}$$

e, portanto, podemos concluir que a soma $\sum_{j=1}^{\infty} \left|f_{n_{j+1}} - f_{n_j}\right|$ é quase certamente finita.

No conjunto onde a soma é finita temos que f_{n_j} é uma sequência de Cauchy em $\mathbb{R}$. Para ver essa afirmação, suponha que $\epsilon > 0$ escolhemos $N > 0$ tal que $\sum_{j=N}^{\infty} \left|f_{n_{j+1}} - f_{n_j}\right| < \epsilon$, então para qualquer $k \geq j \geq N$ temos

$$\left|f_{n_k} - f_{n_j}\right| = \left|\sum_{m=j}^{k} (f_{n_{m+1}} - f_{n_m})\right| \leq \sum_{m=j}^{k} |f_{n_{m+1}} - f_{n_m}| < \epsilon$$

Sabemos que o conjunto onde f_{n_j} converge é mensurável então podemos definir f como o limite da sequência de Cauchy f_{n_j} onde válido e defini-la como zero no complementar, um conjunto de medida nula .

Agora vamos mostrar que $f \in \mathcal{L}^p(\Omega, \mathcal{F}, \mu)$ e que f_n converge para f. Suponha que $\epsilon > 0$ e escolha $N \in \mathbb{N}$ tal que para todos $m, n \geq N$ temos $\|f_m - f_n\|_p < \epsilon$. Agora podemos usar o lema de Fatou para provar que para qualquer $n \geq N$,

$$\int |f - f_n|^p \, d\mu \leq \liminf_{j\to\infty} \int \left|f_{n_j} - f_n\right|^p \, d\mu \leq \sup_{m \geq n} \int |f_m - f_n|^p \, d\mu < \epsilon^p$$

Portanto, pela Desigualdade Minkowski, temos que $f = f_n + (f - f_n)$ está em $\mathcal{L}^p(\Omega, \mathcal{F}, \mu)$ e $f_n \xrightarrow{p} f$. □

Exercícios

Ex. 5.1 — Calcule a integral

$$\int_C x^2 d\mu$$

onde C é o conjunto de Cantor.

Ex. 5.2 — Use o teorema da convergência dominada para provar que a função

$$U(t) = \int_{\mathbb{R}} u(x)\cos(xt)dx$$

é contínua para todo $t \in \mathbb{R}$

Ex. 5.3 — Dado $(X,\mathcal{F},\mu)$ um espaço de medida. Sejam $A_1, A_2, \ldots \in \mathcal{F}$. então

$$\mu\left(\bigcup_N \bigcap_{n=N}^{\infty} A_n\right) \leq \liminf \mu(A_n)$$

Ex. 5.4 — Toda função mensurável não-negativa $f : X \to \mathbb{R}^+$ é o limite pontual das funções simples

$$f_n := (2^{-n}\lfloor 2^n f \rfloor) \wedge n$$

onde $\wedge$ denota o mínimo entre os dois valores

CAPÍTULO 6

Distribuições

6.1 Distribuição de uma Variável Aleatória

Em diversos modelos probabilísticos o domínio no qual a variável aleatória está definida não é tão importante quanto os observáveis que queremos modelar. Se quisermos, por exemplo, criar um modelo para estudar a transmissão de uma certa doença, estaremos interessados em quantidades como o número de infectados em um certo intervalo de tempo e o total destes que se recuperam. Estas quantidades podem ser vistas como variáveis aleatórias em algum espaço amostral, mas as especificidades da natureza deste espaço amostral Ω podem não ser tão importantes quanto os valores das probabilidades com a qual as variáveis assumem certos valores.

Nesse capítulo introduziremos o conceito de distribuição que nos permitirá abstrair, quando possível, o espaço de probabilidade subjacente.

Para isso considere uma variável aleatória X definida em um espaço de probabilidade $(\Omega, \mathcal{F}, \mathbf{P})$ e observe que

1. $\mathbf{P}(X \in \mathbb{R}) = \mathbf{P}(\Omega) = 1$;

2. se $A_1, A_2, \ldots \in \mathcal{B}(\mathbb{R})$ são disjuntos então $X^{-1}(A_1), X^{-1}(A_2), \ldots \in \mathcal{F}$ são também disjuntos (verifique) e

$$\mathbf{P}\left[X^{-1}\left(\bigcup_{n=1}^{\infty} A_n\right)\right] = \mathbf{P}\left[\bigcup_{n=1}^{\infty}\{X^{-1}(A_n)\}\right] = \sum_{n=1}^{\infty} \mathbf{P}(X^{-1}(A_n)).$$

Isso nos permite seguir com a seguinte definição

6.1 Definição (Distribuição de uma Variável Aleatória) *Se* X *for uma variável aleatória definida em um espaço de probabilidade* $(\Omega, \mathcal{F}, \mathbf{P})$, *então* X *induz uma medida de probabilidade* $\mathbf{P}_X$ *em* $\mathbb{R}$ *definida por*

$$\mathbf{P}_X(A) = \mathbf{P}(X^{-1}(A)) = \mathbf{P}(X \in A), \tag{6.1}$$

para $A \in \mathcal{B}(\mathbb{R})$.

Esta medida de probabilidade $\mathbf{P}_X$ *é denominada* ***distribuição*** *de* X.

Observe que na equação (6.1), não há referência a nenhuma estrutura específica do espaço amostral Ω. Nessa equação estamos definindo uma nova probabilidade na reta real, e são essas probabilidades que nos interessam no estudo da variável X.

De maneira análoga podemos definir a distribuição de um vetor aleatório

6.2 Definição (Distribuição de Vetores Aleatórios) *Se* X *for uma vetor aleatório, então* X *induz uma medida de probabilidade* $\mathbf{P}_X$ *em* $\mathbb{R}^n$ *definida tomando a pré-imagem para cada conjunto de Borel* $A \subset \mathbb{R}^n$:

$$\mathbf{P}_X(A) = \mathbf{P}(X^{-1}(A)) = \mathbf{P}(X \in A). \tag{6.2}$$

Esta medida de probabilidade $\mathbf{P}_X$ *é denominada* ***distribuição*** *do vetor* X.

Se $X = (X_1, \ldots X_n)$ *é um vetor aleatório então as distribuições de cada uma das variáveis aleatórias* X_i *para* $i = 1, \ldots, n$ *são chamadas de* ***distribuições marginais*** *de* X.

Não é difícil ver que a distribuição não define a variável aleatória de maneira única. Se tivermos duas variáveis aleatórias, elas podem ser bastante diferentes e ainda sim terem a mesma distribuição.

6.3 Exemplo *Considere o experimento de lançar uma moeda não viciada, e tome como espaço amostral o conjunto* $\Omega = \{H, T\}$, *onde* H *representa cara e* T *representa coroa. Tomamos* $\mathbf{P}$ *como a medida de probabilidade equiprovável em* Ω.

Podemos agora definir a variável aleatória $X(H) = 1, X(T) = 0$.

Note que $\mathbf{P}(X = 1) = \mathbf{P}(X = 0) = \frac{1}{2}$, *e para* $A \in \mathcal{B}(\mathbb{R})$ *temos*

$$\mathbf{P}(X \in A) = P(X = 0)\mathbb{1}_A(0) + \mathbf{P}(X = 1)\mathbb{1}_A(1) = \begin{cases} \frac{1}{2} & ;\ \text{se } 0 \in A, 1 \in A^c \text{ ou } 0 \in A^c, 1 \in A; \\ 1 & ;\ \text{se } 0, 1 \in A; \\ 0 & ;\ \text{se } 0, 1 \in A^c \end{cases}.$$

Por outro lado poderíamos definir a variável aleatória $Y(H) = 0, Y(T) = 1$.

Estas variáveis são muito diferentes (de fato $X = 1 - Y$*), mas* X *e* Y *possuem a mesma distribuição.* $\triangleleft$

O exemplo anterior mostrou que variáveis aleatórias distintas podem ter a mesma distribuição. Antes de continuar, vamos definir alguns conceitos importante.

6.4 Definição (Equivalências de Variáveis Aleatórias)

1 *Dizemos que as variáveis aleatórias* X *e* Y *são iguais* ***quase certamente***, *fato que denotaremos por*

$$X \overset{q.c.}{=} Y,$$

se $\mathbf{P}(X \neq Y) = 0$.

2 *Dizemos que as variáveis aleatórias* X *e* Y *são iguais em* ***distribuição***, *fato que denotaremos por*

$$X \overset{d.}{=} Y,$$

se eles tiverem a mesma distribuição. Equivalentemente,

$$\mathbf{P}(X \in A) = \mathbf{P}(Y \in A) \text{ para todo } A \text{ de Borel.}$$

Observe que se $X \overset{q.c}{=} Y$, então $X \overset{d.}{=} Y$, mas a recíproca não é verdadeira. De fato, se $X \overset{q.c}{=} Y$, então

$$\mathbf{P}(X \in A) = \mathbf{P}(X \in A; X = Y) = \mathbf{P}(Y \in A; X = Y) = \mathbf{P}(Y \in A).$$

Por outro lado, podemos ter $X \overset{d.}{=} Y$ sem que X e Y estejam definidas em um mesmo espaço de probabilidade!!

6.5 Exemplo *Seja agora* $\mathbf{P}$ *a medida de Lebesgue em* $\Omega = [0,1]$, *e defina* $X(\omega) = \omega$ *e* $Y(\omega) = 1 - \omega$. *Note que para* $x \in \mathbb{R}$

$$\{X \leq x\} = \{\omega \in [0,1] : X(\omega) \leq x\} = \{\omega \in [0,1] : \omega \leq x\} = \begin{cases} \emptyset & ; \text{ se } x < 0; \\ [0, x] & ; \text{ se } 0 \leq x \leq 1; \\ [0,1] & ; \text{ se } x \geq 1 \end{cases},$$

e

$$\{Y \in A\} = \{\omega \in [0,1] : Y(\omega) \leq x\} = \{\omega \in [0,1] : 1 - \omega \leq x\} = \begin{cases} \emptyset & ; \text{ se } x < 0; \\ [1 - x, 1] & ; \text{ se } 0 \leq x \leq 1; \\ [0,1] & ; \text{ se } x \geq 1 \end{cases}.$$

Segue que para $x \in [0,1]$

$$\mathbf{P}_X((-\infty, x]) = \mathbf{P}(X \leq x) = \mathbf{P}([0, x]) = x = \mathbf{P}([1 - x, 1]) = \mathbf{P}(Y \leq x) = \mathbf{P}_Y((-\infty,x]),$$

e portanto

$$\mathbf{P}_X((-\infty, x]) = \mathbf{P}_Y((-\infty, x]),$$

para todo $x \in \mathbb{R}$.

Agora, como $\{(-\infty, x] : x \in \mathbb{R}\}$ *é um* π*-sistema que gera* $\mathcal{B}(\mathbb{R})$*, então* $\mathbf{P}_X = \mathbf{P}_Y$. ◁

No exemplo anterior comparamos a distribuição de duas variáveis aleatórias olhando apenas para as probabilidades em intervalos do tipo $(\infty, x]$, $x \in \mathbb{R}$. Para tanto argumentamos que

1. a classe $\mathcal{P} = \{(-\infty, x] : x \in \mathbb{R}\}$ é π-sistema;

2. $\mathcal{B}(\mathbb{R})$ é gerada por tais intervalos.

Com isso pudemos concluir que duas medidas de probabilidade em $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ que coincidam em $\mathcal{P}$ são de fato a mesma medida.

Isso demonstra o seguinte resultado.

6.6 Teorema *Sejam* X *uma variável aleatória definida em um espaço de probabilidade* $(\Omega_1, \mathcal{F}_1, \mathbf{P}_1)$ *e* Y *uma variável aleatória definida em um espaço de probabilidade* $(\Omega_2, \mathcal{F}_2, \mathbf{P}_2)$*. Nestas condições vale que* $X \overset{d.}{=} Y$ *se, e somente se,* $\mathbf{P}_1(X \le x) = \mathbf{P}_2(Y \le x)$ *para todo* $x \in \mathbb{R}$.

De maneira equivalente, a distribuição $\mathbf{P}_X$ *de* X *está completamente determinada pelos valores de* $\mathbf{P}_1(X \le x)$, $x \in \mathbb{R}$.

Este resultado motiva a definição do que chamamos de função distribuição de uma variável aleatória, que introduziremos na próxima seção.

6.2 A Função Distribuição

6.7 Definição (Função de Distribuição de uma Variável Aleatória) *A* ***função de distribuição*** $F : \mathbb{R} \to [0,1]$ *de uma variável aleatória* X *é definida como*

$$F(x) := \mathbf{P}(X \le x), \qquad x \in \mathbb{R}.$$

O teorema 6.6 nos garante que a função de distribuição determina unicamente a distribuição de X.

A função de distribuição é uma abstração poderosa do experimento aleatório. Ao falarmos da função de distribuição não precisamos mais de Ω, álgebras, assim por diante. Tudo é resumido por uma função $\mathbb{R} \to \mathbb{R}$.

6.8 Exemplo *Seja* X = *número de H em dois lançamentos independentes de uma moeda justa. Então temos*

$$X = \begin{cases} 0, & \textit{com probabilidade } \frac{1}{4} \\ 1, & \textit{com probabilidade } \frac{1}{2} \\ 2, & \textit{com probabilidade } \frac{1}{4} \end{cases}$$

Então, podemos criar a função de distribuição. Esta função apresentará descontinuidades do tipo salto nos pontos 0, 1 *e* 2.

$$F(X) = \mathbf{P}(X \le x) = \begin{cases} 0, & x < 0 \\ \frac{1}{4}, & x \in [0,1) \\ \frac{1}{4} + \frac{1}{2}, & x \in [1,2) \\ \frac{1}{4} + \frac{1}{2} + \frac{1}{4}, & x \in [2,+\infty) \end{cases}$$

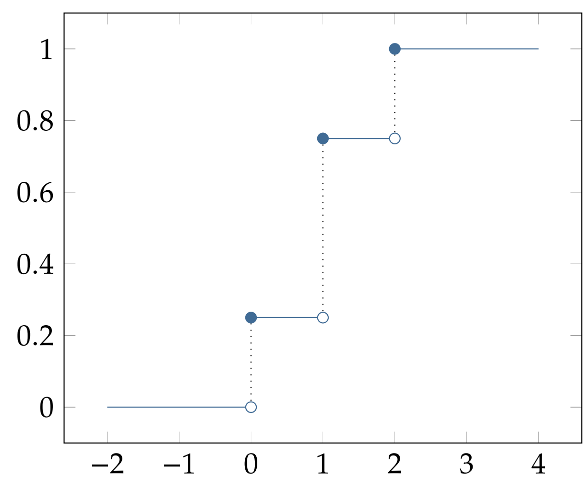


Figura 6.1: Distribuição do Exemplo 6.8

6.9 Exemplo *Considere o experimento de sortear ao acaso um ponto na bola unitária* $B = \{(a,b) : a^2 + b^2 \le 1\}$ *em* $\mathbb{R}^2$, *e defina* X *como a distância ao centro da bola.*

Para modelar este problema podemos considerar $\mathbf{P}(\cdot) = \mu(\cdot)/\pi$, *onde* μ *é a medida de Lebesgue em* B. *Ou seja, a probabilidade de sortearmos um ponto em uma região* $A \in \mathcal{B}(B)$ *é proporcional à área de* A. *Mais especificamente*

$$\mathbf{P}(A) = \frac{\mu(A)}{\pi}.$$

Agora, note que $X(a,b) = \sqrt{a^2 + b^2}$ *e portanto para* $x \in [0,1]$

$$\mathbf{P}(X \le x) = \mathbf{P}(\{(a,b) : a^2 + b^2 \le x^2\}) = \frac{\pi x^2}{\pi} = x^2,$$

e portanto

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

6.10 Teorema (Propriedades da Função de Distribuição) *A função de distribuição $F(x)$ de uma variável aleatória X possui as seguintes propriedades:*

1. *F é não decrescente e $0 \leq F(x) \leq 1$.*

2. *$F(x) \longrightarrow 0$ quando $x \longrightarrow -\infty$ e $F(x) \longrightarrow 1$ quando $x \longrightarrow +\infty$.*

3. *F é contínua à direita, ou seja*
$$\lim_{y \to x_+} F(y) = F(x).$$

4. *F possui limites à esquerda em todos os pontos. Além disso,*
$$F(x^-) := \lim_{y \to x^-} F(y) = \mathbf{P}(X < x).$$
Em particular,
$$\mathbf{P}(X = x) = F(x) - F(x^-).$$

5. *F possui no máximo número enumerável de descontinuidades.*

Uma função F que é contínua à direita e possui limites à esquerda em todos os pontos é dita **cadlag**, do francês "continu à droite, limites à gauche".

Demonstração. *Abaixo rascunhamos todas as ideias da demonstração, e deixamos para o leitor completar as provas.*

1. *Basta observar que os conjuntos $(-\infty, x]$ são crescentes em x. Assim se $x < y$, então $\{X \leq x\} \subseteq \{X \leq y\}$, e o resultado segue.*

2. *Os eventos $\{X \leq x_n\} \downarrow \emptyset$ quando $x_n \longrightarrow -\infty$. Pela continuidade, $\mathbf{P}(X \leq x_n) \longrightarrow \mathbf{P}(\emptyset) = 0$.*

   Da mesma forma, $\{X \leq x_n\} \uparrow \Omega$ quando $x_n \longrightarrow +\infty$. Então, $\mathbf{P}(X \leq x_n) \longrightarrow \mathbf{P}(\Omega) = 1$. Isso completa a prova de que $F(x) \longrightarrow 0$ como $x \longrightarrow -\infty$ e $F(x) \longrightarrow 1$ quando $x \longrightarrow \infty$.

3. *Para demonstrar que F é contínua à direita suponha que tenhamos uma sequência convergente para x pela direita. Queremos mostrar que esta sequência arbitrária $y_n = F(x_n)$ converge para x (até $y_n \downarrow x$). Nós temos*
$$\mathbf{P}(X \leq y_n) \to \mathbf{P}(X \leq x), \qquad n \to \infty.$$

A conclusão resulta da convergência dos conjuntos

$$\{X \le y_n\} \downarrow \{X \le x\}$$

e da continuidade.

4 *Agora vamos demonstrar que F possui limites à esquerda em todos os pontos, e*

$$F(x^-) := \lim_{y \to x^-} F(y) = \mathbf{P}(X < x).$$

Em particular, como consequência dos dois últimos itens podemos definir os saltos em x como:

$$\mathbf{P}(X = x) = F(x) - F(x^-).$$

Semelhante a (ii), nós fazemos $y_n \uparrow x$, e queremos mostrar

$$\mathbf{P}(X \le y_n) \to \mathbf{P}(X < x), \qquad n \to \infty.$$

Isso decorre da convergência dos conjuntos

$$\{X \le y_n\} \uparrow \{X < x\}$$

5 *F possui no máximo número enumerável de descontinuidades.*

Como $0 \le F(x) \le 1$,

- *pode haver no máximo 2 saltos (ou seja, descontinuidades) de altura $\ge \frac{1}{2}$. Aqui, "saltos"significa pontos x em que $F(x) - F(x^-) \ge \frac{1}{2}$.*
- *pode haver no máximo 3 saltos de tamanho $\in [\frac{1}{3}, \frac{1}{2})$.*
- *pode haver no máximo 4 saltos de tamanho $\in [\frac{1}{4}, \frac{1}{3})$.*
- $\cdots$

O número total de saltos é enumerável, e como cada salto está em $[\frac{1}{n}, \frac{1}{n-1})$ por algum n, então esta lista contém todos os saltos possíveis. □

6.11 Proposição $\mathbf{P}(X \in (a, b]) = F(b) - F(a)$.

6.12 Definição (Função de distribuição) *Uma função de monótona contínua à direita $F : \mathbb{R} \to [0, 1]$ com $\lim_{x \to -\infty} F(x) = 0$ e $\lim_{x \to \infty} F(x) = 1$ é chamada de candidata a função de distribuição de probabilidade.*

O primeiro passo nesta comparação é contribuir uma variável aleatória que possua uma função distribuição dada. Isso nos permitirá construir uma sequência de variáveis com funções de distribuição pré-determinadas, sempre no mesmo espaço de probabilidade, permitindo a análise mais precisa do comportamento da sequência.

Para isso, chamaremos de **função distribuição** qualquer função $F : \mathbb{R} \to [0,1]$ tal que

1. F é não-decrescente;

2. F é contínua pela direita;

3. $\lim_{x \to a^-} F(x)$ existe para todo $a \in \mathbb{R}$ (possui limite pela esquerda);

4. $\lim_{n \to +\infty} F(x) = 1$ e $\lim_{n \to -\infty} F(x) = 0$.

Nosso objetivo a partir de agora é, dada uma função F com as propriedades acima, encontrar uma variável aleatória X em algum espaço de probabilidade $(\Omega, \mathcal{F}, \mathbf{P})$ cuja função de distribuição acumulada seja F. Ou seja, uma variável X tal que $\mathbf{P}(X \leq x) = F(x)$, para todo $x \in \mathbb{R}$.

Uma maneira interessante seria definir a variável X como uma função de alguma variável aleatória cuja distribuição seja conhecida. Tomemos, por exemplo, uma variável $U \sim Uniforme(0,1)$, e tentemos escrever $X = F^*(U)$ para alguma função $F^* : (0,1) \to \mathbb{R}$.

Para escolher F^*, lembre que o que gostaríamos é que $\mathbf{P}(X \leq x) = \mathbf{P}(F^*(U) \leq x) = F(x)$.

Para isso note primeiro que, como $F(x) \in [0,1]$, então $\mathbf{P}(U \leq F(x)) = F(x)$. Ou seja, queremos F^* tal que

$$\mathbf{P}(F^*(U) \leq x) = \mathbf{P}(U \leq F(x)).$$

Ou ainda, que $F^*(u) \leq x \Leftrightarrow u \leq F(x)$.

Se F fosse inversível, poderíamos fazer $F^* = F^{-1}$ e, como F é crescente teríamos que

$$F*(u) \leq x \Leftrightarrow F(F*(u)) \leq F(x) \Leftrightarrow u \leq F(x).$$

Em geral F não é inversível, mas as considerações acima nos mostram que a chave para conseguir o que queremos ainda está em, de algum modo, "*inverter*" a função distribuição $F : \mathbb{R} \to [0,1]$.

Observe então que, como F é uma função não-decrescente, apesar de em geral não ser inversível, os problemas são facilmente identificado. O que impede a inversão de F são os seguintes problemas:

1. intervalos (a,b) onde a função é constante;

$\boxed{2}$ pontos de descontinuidade de F.

Mas não precisaremos de uma inversa no sentido clássico da palavra. O que precisamos é uma função $F^* : (0,1) \longrightarrow \mathbb{R}$ que reaja bem à desigualdades. Mais diretamente, precisamos que $F^*(\omega) \leq x$ se, e só se, $F(x) \geq \omega$.

Uma maneira de resolver esse problema é tentar definir F^* de modo que

$\boxed{1}$ se F é inversível em (a,b) então queremos que $F^*(\omega) = F^{-1}(\omega)$ em $F((a,b))$;

$\boxed{2}$ se $F(x) = \omega$ para todo $a < x < b$, então F^* é descontínua em ω;

$\boxed{3}$ se $F(x)$ é descontínua em $x = x_0$ com $\omega_1 = \lim\limits_{x \longrightarrow x_0^-} F(x) < F(x_0) = \omega_2$, então $F^*(\omega) = x_0$ para $\omega \in (\omega_1,\omega_2)$.

Essas ideias são melhor visualizadas na imagem abaixo.

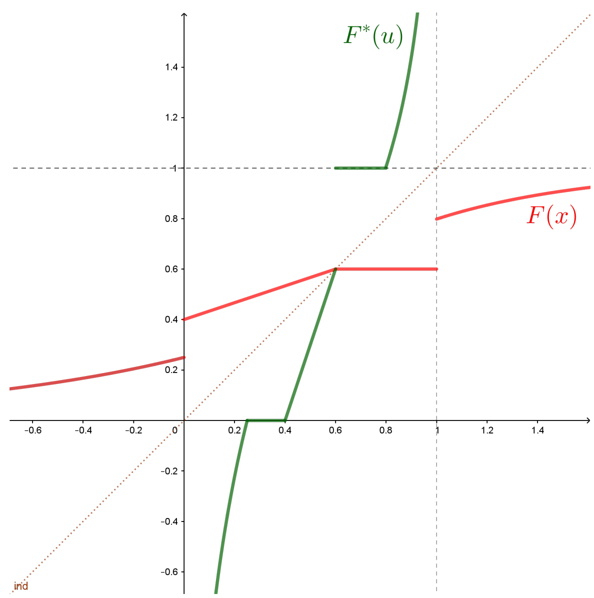


Figura 6.2: Função distribuição e sua inversa generalizada.

Existem diversas formas de definir uma função F^* com tais propriedades, e a seguir apresentaremos uma das mais comuns.

6.13 Definição *Dada uma função distribuição $F : \mathbb{R} \longrightarrow [0,1]$, definimos a **inversa generalizada** de F como a função $F^* : (0,1) \longrightarrow \mathbb{R}$ dada por*

$$F^*(\omega) = \inf\{x \in \mathbb{R} : F(x) \geq \omega\}.$$

Fica como exercício verificar que as propriedades listadas anteriormente são satisfeitas por F^* como definida acima.

Apesar de não ser a inversa de F, F^* ainda herda algumas propriedades de F.

6.14 Proposição (Propriedades da Inversa Generalizada) *Se F é não decrescente e contínua à direita, então definimos a função inversa generalizada como*

$$F^{-1}(x) := \inf\{y : F(y) \geq x\}$$

A inversa generalizada satisfaz:

1. $F^{-1}(y) = \sup\{y : F(y) < x\}$
2. $F^{-1}(y)$ *é não decrescente.*
3. $F^{-1}(F(x)) \leq x$.
4. $F(F^{-1}(y)) \geq y$.
5. $F^{-1}(y) \leq x$ *se e somente se* $y \leq F(x)$.

Demonstração.

1. *Seja* $F_1^{-1}(y) = \sup\{y : F(y) < x\}$. *Para todo* $x \in \mathbb{R}$, *se* $y_0 \in \{y|F(y) \geq x\}$, *então* $y_0 \notin \{y|F(y) < x\}$, *logo* $y_0 \geq F^{-1}(x)$, *o que implica na desigualdade* $F_1^{-1}(x) \geq F^{-1}(x)$.

   Para todo $\varepsilon > 0$, *da definição de* $F_1^{-1}(x)$, *temos* $F^{-1}\left(F_1^{-1}(x) - \varepsilon\right) < x$. *Logo* $F_1^{-1}(x) - \varepsilon \leq F^{-1}(x)$, *o que implica na desigualdade* $F_1^{-1}(x) \leq F^{-1}(x)$.

2. *Seja* $y_1 < y_2$. *Então* $F(x) \geq y_2$ *implica que* $F(x) \geq y_1$, *e assim* $\{x : F(x) \geq y_1\} \supset \{x : F(x) \geq y_2\}$. *Portanto, claramente,* $\inf\{x : F(x) \geq y_1\} \leq \inf\{x : F(x) \geq y_2\}$.

3. $F^{-1}(F(x)) = \inf\{z : F(z) \geq F(x)\}$ *e* $x \in \{z : F(z) \geq F(x)\}$, *então o resultado segue.*

4. *Seja* $x_n \in \{x : F(x) \geq y\}$ *tal que* $x_n \to x_0$. *Então* $\liminf_n F(x_n) \geq y$, *mas como F é monótona e não-decrescente e contínua à direita,* $\liminf_n F(x_n) \leq F(x_0)$. *Daí* $F(x_0) \geq y$, *ou seja,* $x_0 \in \{x : F(x) \geq y\}$. *Portanto,* $\{x : F(x) \geq y\}$ *é fechado e, portanto, contém seu ínfimo. Assim,* $F(F^{-1}(y)) \geq y$.

5. $F^{-1}(y) \leq x$ *implica em* $x \in \{z : F(z) \geq y\}$ *e assim* $y \leq F(x)$. *Por outro lado, se* $y \leq F(x)$, *em seguida,* $x \in \{z : F(z) \geq y\}$ *e, portanto,* $F^{-1}(y) \leq x$ *já que* $F^{-1}(y)$ *é o mínimo.* □

Com isso, dada uma variável aleatória U definida em um espaço de probabilidade $(\Omega, \mathcal{F}, \mathbf{P})$, e uniformemente distribuída no intervalo (0,1), podemos definir a variável aleatória

$$X = F^*(U),$$

e perceber que para $x \in \mathbb{R}$ temos $F(x) \in [0,1]$, e portanto

$$\mathbf{P}(X \le x) = \mathbf{P}(F^*(U) \le x) = \mathbf{P}(U \le F(x)) = F(x)$$

onde na última igualdade usamos que a probabilidade que a uniforme seja menor que y é y, i.e., $\mathbf{P}(U \le y) = y$

Mostramos assim o seguinte resultado.

6.15 Proposição (Existência de variável aleatória com distribuição prescrita) *Dada uma função distribuição $F : \mathbb{R} \longrightarrow [0,1]$, então existe uma variável aleatória X, definida em algum espaço de probabilidade $(\Omega, \mathcal{F}, \mathbf{P})$, tal que a função distribuição de X seja igual a F. Ou seja, $\mathbf{P}(X \le x) = F(x)$, para todo $x \in \mathbb{R}$.*

Antes de seguirmos vamos apresentar um resultado importante sobre os pontos de descontinuidades para função não-decrescentes, como F e F^*.

6.16 Proposição *Se $f : I \subset \mathbb{R} \longrightarrow \mathbb{R}$ é uma função não-decrescente definida em um intervalo $I \subset \mathbb{R}$, então o conjunto D_f formado pelos pontos de descontinuidade de f é no máximo enumerável.*

O mesmo resultado vale para f não-crescente.

Demonstração. *Se f é não-decrescente, então os limites laterais*

$$f(a-) = \lim_{x \to a^-} f(x) \quad e \quad f(a+) = \lim_{x \to a^+} f(x),$$

existem e se $a \in D_f$, temos

$$f(a-) < f(a+),$$

de modo que o intervalo $I_a := (f(a-), f(a+))$ é não vazio.

Note também que, se $a_1 < a_2$ então $f(a_1+) \le f(a_2-)$, e os intervalos I_{a_1} e I_{a_2} são disjuntos.

Escolhendo agora um racional $r(a) \in I_a$ para cada $a \in D_f$ criamos um mapa injetivo $a \mapsto r(a)$, que leva D_f em um subconjunto de $\mathbb{Q}$, concluindo o resultado □

6.17 Corolário *Dada F um função distribuição e F^* a sua inversa generalizada, temos que os pontos de continuidade de F e os pontos de continuidade de F^* são, no máximo, enumeráveis.*

Terminamos essa seção com a definição do significado de duas variáveis serem identicamente distribuídas

6.18 Definição (Variáveis Aleatórias Identicamente Distribuídas) *Duas variáveis aleatórias X e Y são ditas identicamente distribuídas se*

$$\mathbf{P}(X \leq x) = \mathbf{P}(Y \leq x) \quad \text{para todo } x \in \mathbb{R}.$$

6.3 Tipos de Distribuições

6.19 Definição *Uma função de distribuição F é dita*

1. ***discreta*** *se para algum conjunto enumerável de números reais* $\{x_j\}$ *e massas pontuais* $\{p_j\}$

$$F(x) = \sum_{x_j \leq x} p_j \quad \text{para todo } x \in \mathbb{R}$$

2. ***contínua*** *se for contínua para todo* x.

3. ***absolutamente contínua*** *se existe uma função* f *não negativa Lebesgue integrável tal que*

$$F(b) - F(a) = \int_a^b f(x)dx \text{ para todo } a < b$$

4. ***singular*** *se* $F \not\equiv 0$, F' *existe e igual a* 0 *q.c.*

Temos o seguinte Teorema de decomposição

6.20 Teorema (Decomposição de Distribuição) *Toda função de distribuição pode ser decomposta em uma combinação convexa de três tipos puros, discretas, absolutamente contínuas, e contínuas singulares. Assim, se F é uma função de distribuição, então*

$$F = \alpha F_{ac} + \beta F_d + \gamma F_s,$$

onde $\alpha, \beta, \gamma \geq 0$ *e* $\alpha + \beta + \gamma = 1$.

- $F_{ac} = \int_{-\infty}^{x} f(y)dy$ *com* $f(x) = F'_{ac}$ *q.c.*
- F_d *é uma função saltos com no máximo um número enumerável de saltos.*
- F_s *é singular.*

Começaremos provando que toda distribuição pode ser decomposta como soma de uma discreta e uma contínua.

6.21 Teorema *Toda função de distribuição pode ser decomposta em uma combinação convexa de uma discreta e uma contínua. Assim, se F é uma função de distribuição, então*

$$F = \alpha F_c + \beta F_d$$

onde $\alpha, \beta, \geq 0$ e $\alpha + \beta = 1$.

Demonstração. *Já provamos que F possui no máximo um número enumerável de saltos. Seja $\{x_j\}$ ser uma enumeração dos saltos. Defina*

$$p(x_j) = F(x_j) - F(x_j^-)$$

E defina também a função:

$$F_d^*(x) = \sum_{x_j \leq x} p(x_j), \quad x \in \mathbb{R}$$

A função $F_d^*(x)$ está bem definida pois a soma anterior é menor que 1 é discreta.

A função satisfaz $F_d^*(x)$ todas as propriedades de distribuição exceto possivelmente a última. Mas nesse caso faremos uma renormalização: se $\lim_{x \to \infty} F_d^* = \beta$ então consideramos $F_d = F_d^*/\beta$

Seja $F_c^*(x) = F(x) - F_d^*(x)$. Claramente $F_c^*(x)$ é crescente e não negativa.

Provaremos que $F_c^*(x)$ é contínua. Observamos que inicialmente que soma de funções contínuas à direita é contínua à direita e assim $F_c^*(x) = F_c^*(x^+)$. Agora provaremos a continuidade à esquerda

$$\begin{aligned} F_c^*(x) - F_c^*(x^-) &= F(x) - F_d(x) - (F(x^-) - F_d(x^-)) \\ &= F(x) - F(x^-) - (F_d(x) - F_d(x^-)) \\ &= \begin{cases} p_j - p_j = 0 \text{ se } x = x_j \\ 0 \text{ caso contrário} \end{cases} \end{aligned}$$

A função $F_c^*(x)$ satisfaz todas as propriedades de distribuição exceto possivelmente a última. Mas podemos renormalizá-la a uma distribuição. Se $\lim_{x \to \infty} F_c^*(x) = \alpha$ definimos $F_c = F_c^*/\alpha$.

Como

$$F(x) = F_c^*(x) + F_d^*(x) \tag{6.3}$$
$$= \alpha F_c + \beta F_d \tag{6.4}$$

Por último, o fato que $\alpha + \beta = 1$ decorre do fato que $F(x)$ é uma distribuição.

6.22 Teorema *Toda função de distribuição contínua pode ser decomposta em uma combinação convexa de uma absolutamente continua e uma contínua singular. Assim, se F é uma função de distribuição, então*

$$F = \alpha F_{ac} + \beta F_s$$

onde $\alpha, \beta, \geq 0$ e $\alpha + \beta = 1$.

Para a demonstração desse teorema usaremos os seguintes fatos

6.23 Lema

1. *Se F é uma função monótona em $[a,b]$ então F é diferenciável quase certamente em $[a,b]$. Nesse caso a derivada é não negativa.*

2. $\displaystyle\int_a^b F'(x)dx \leq F(b) - F(a)$ ◁

A demonstração desse lema será feita no Capítulo **??**

Demonstração. *Se definirmos $F^*_{ac}(x) = \displaystyle\int_{-\infty}^{b} F'(y)dy$ e $F^*_s = F(x) - F_{ac}$. Como $F(^*_{ac})' = F'(x)$ quase certamente temos que $(F^*_s)' = 0$ quase certamente.*

As funções F^*_s e F^*_{ac} satisfazem todas as propriedades de distribuição exceto possivelmente a última. Mas podemos renormalizá-las a distribuições F_s e F_{ac} de modo análogo ao que foi feito na demonstração do Teorema .

Distribuições Discretas

Seja X uma variável aleatória com a função de distribuição F (com probabilidade induzida **P**).

6.24 Definição *A distribuição de X (e X em si) é dita **discreta** se existir um conjunto enumerável S tal que*

$$\mathbf{P}(S^c) = 0.$$

Podemos enumerar o conjunto S, escrevendo $S = \{x_1, x_2, \ldots\}$, e definimos $p_k := \mathbf{P}(X = x_k)$. Então temos

$$F(x) = \sum_{k:\ x_k \leq x} p_k,$$

e esta é uma soma sobre um conjunto enumerável. Estes p_k às vezes são denominados **massas pontuais**.

1 Massa de um ponto em zero:

$$\mathbf{P}(X = 0) = 1.$$

Portanto, a função de distribuição F é:

$$F(x) = \begin{cases} 0 & , x < 0 \\ 1 & , x \geq 0 \end{cases}$$

Essa função de distribuição também denominada **medida de Dirac** no 0.

2 Qualquer variável aleatória simples é discreta. Uma variável aleatória simples é uma variável aleatória que toma valores finitos.

3 Finalmente, há exemplos "mais selvagens"do que isso. Por exemplo, onde o conjunto de S de valores é todos racionais: $S = \mathbb{Q}$, onde você possui massas pontuais arbitrárias que somam até 1, por exemplo, $p_k = \frac{1}{2^k}$.

Algumas Distribuições Discretas

Distribuição	Notação	Função de Probabilidade	Domínio
Um ponto	$\delta(a)$	$p(a) = 1$	$\times$
Uniforme Discreta	$p(x_i) = \frac{1}{n}$	$x_1, \ldots x_n$	
Bernoulli	$\mathrm{Be}(p)$	$p(0) = q, p(1) = p$	$\{0,1\}$
Binomial	$\mathrm{Bin}(n,p)$	$p(k) = p^k q^{n-k}$	$k = 0,1, \ldots ,n$
Geométrica	$\mathrm{Ge}(p)$	$p(k) = pq^k$	$k \in \mathbb{N}$
Primeiro Sucesso	$\mathrm{Po}(p)$	$p(k) = pq^{k-1}$	$k \in \mathbb{N}^*$
Poisson	$\mathrm{Po}(m)$	$p(k) = e^{-m} \frac{m^k}{k!}$	$k \in \mathbb{N}$

A distribuição uniforme discreta assume os valores $x_1, \ldots x_n$ e $p(x_i) = \dfrac{1}{n}$. A distribuição $\mathrm{Be}(p)$ descreve o resultado de um experimento de lançamento de moeda (não necessariamente honesta e a distribuição $\mathrm{Bin}(n, p)$ é o número de sucessos em n lançamentos. A distribuição $\mathrm{Ge}(p)$ descreve o número de fracassos antes do primeiro sucesso e a distribuição $\mathrm{Po}(p)$ o número de lançamentos necessários para ter sucesso uma vez.

A distribuição de Poisson surge como o limite da distribuição binomial quando $n \to \infty$ e $p \to 0$ e $np \to \lambda$.

6.25 Teorema (Teorema Limite de Poisson) *Se $n \to \infty$, $p \to 0$, de tal modo que $np \to \lambda$ então*

$$\frac{n!}{(n-k)!k!} p^k (1-p)^{n-k} \to e^{-\lambda} \frac{\lambda^k}{k!}.$$

Demonstração. *Exercício. Dica use a aproximação de Stirling.* □

Distribuições Absolutamente Contínuas

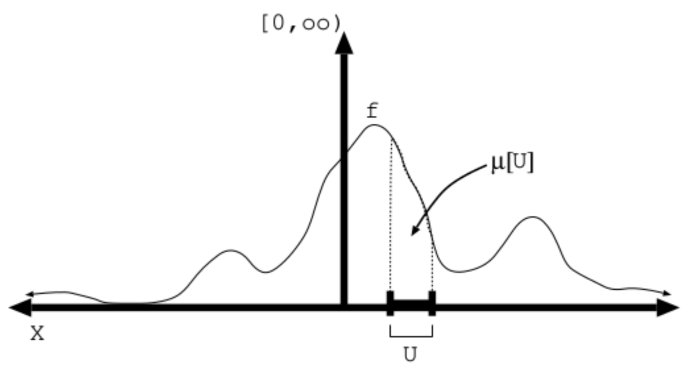
A distribuição de X (e X) é **absolutamente contínua** se existir uma função f, denominada **densidade de** X, tal que

$$F(x) = \int_{-\infty}^{x} f(y)\, dy \qquad \text{para cada} x.$$

Claramente, é necessário que $f \geq 0$ e que

$$\int_{-\infty}^{\infty} f(x)\, dx = 1.$$

Figura 6.3: Densidade - [26]



$$X = \sum_{k=1}^{n} a_k \mathbb{1}_{A_k}.$$

- Em particular, se f for contínua, então

$$F'(x) = f(x),$$

  ou seja, é a derivada de uma função de distribuição.

- Então, podemos dizer

$$\mathbf{P}(X \in (a, b]) = F(b) - F(a) = \int_a^b f(x)\, dx.$$

- Alternativamente, podemos começar com uma função f tal que

$$f \geq 0 \qquad \text{e} \qquad \int_{-\infty}^{\infty} f(x)\, dx = 1,$$

  e definir a função de distribuição pela fórmula

$$F(x) := \int_{-\infty}^{x} f(y)\, dy.$$

6.26 Exemplo (XKCD) *Alice secretamente escolhe dois números reais diferentes por um processo desconhecido e os coloca em dois envelopes. Bob escolhe um dos dois envelopes aleatoriamente (com um lançamento de uma moeda justa) e mostra o número desse envelope. Agora você deve adivinhar se o número no outro envelope fechado é maior ou menor que o que você viu.*

Existe uma estratégia que lhe dê uma chance melhor do que 50% de adivinhar corretamente, não importando o procedimento que Alice usasse para escolher seus números?

Solução:

Antes de abrir qualquer envelope, você escolhe um número r de forma aleatória usando qualquer distribuição de probabilidade absolutamente contínua que quiser. Vamos chamar de x o número do envelope aberto por Bob.

*A **estratégia** agora é: se $r < x$, então você prevê que o número oculto y é menor que x; Caso contrário, você adivinha que y é maior do que x.*

Por que esta é uma estratégia vencedora? Existem três casos a serem analisados:

- *r é inferior a x e y. Neste caso, você adivinha "menor"e ganha o jogo se $x > y$. Como as variáveis x e y foram atribuídas aos números ocultos uniformemente , $P(x > y) = 1/2$. Assim, neste caso você ganha com probabilidade meio.*
- *r é maior do que x e y. Por um argumento análogo ao primeiro caso , você adivinha "maior"e ganha com probabilidade meio.*
- *r está entre x e y. Neste caso, você adivinha "maior"se $x < y$ e "menor"se $x > y$ - isto é, você sempre ganha o jogo.*

O caso 3 ocorre com uma probabilidade não-zero finita ε, equivalente à integral da sua distribuição de probabilidade entre x e y. Com a média de todos os casos, sua chance de ganhar é $(1 + \varepsilon)/2$, que é estritamente maior do que meio.

Porque isso funciona? ◁

Distribuição Uniforme Contínua Uma variável aleatória contínua X é **uniforme** no intervalo $[a,b]$, para $a < b$, se sua f.d.p. é

$$f(x) = \begin{cases} \frac{1}{b-a}, & \text{se } x \in [a,b] \\ 0 & \text{caso contrário} \end{cases}$$

e denotamos esse fato por $X \sim \text{Uniforme}([a,b])$.

Nesse caso, a probabilidade de X estar num subintervalo de $[a,b]$ é proporcional ao comprimento de tal subintervalo; de fato, para $y < z$ reais

$$\mathbf{P}(y \leq X \leq z) = \int_y^z \frac{1}{b-a}\,\mathrm{d}x = \frac{z-y}{b-a}.$$

Distribuição Exponencial Uma variável aleatória contínua X é **exponencial** com parâmetro $\lambda > 0$, e denotamos esse fato por

$$X \sim \text{Exponencial}(\lambda)$$

se sua função de densidade é

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{se } x \geq 0 \\ 0 & \text{caso contrário.} \end{cases}$$

A função de distribuição é

$$F_X(t) = \int_0^t \lambda e^{-\lambda x}\,\mathrm{d}x = 1 - e^{-\lambda x}$$

portanto $\mathbf{P}(X > a) = e^{-\lambda a}$.

Distribuição Normal A variável aleatória X possui distribuição **normal** com parâmetros μ e σ^2, abreviado por $X \sim \mathcal{N}(\mu;\sigma^2)$, se sua função densidade de probabilidade é dada por

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

para todo $x \in \mathbb{R}$

Distribuição normal padrão $\mathcal{N}(0;1)$:

Distribuições Singulares

Distribuição uniforme no conjunto de Cantor Considere F a família de todos os subconjuntos do intervalo [0,1], que são a união de um número finito de intervalos compactos disjuntos de [0,1] e defina a aplicação $\Phi : F \to F$, por

$$\Phi\left(\sum_{i=1}^m [a_i,b_i]\right) := \sum_{i=1}^m \left(\left[a_i, \frac{2a_i+b_i}{3}\right] + \left[\frac{a_i+2b_i}{3}, b_i\right]\right)$$

Então para cada $A \in F$ temos que a sequência $\{\Phi^n(A)\}_{n\in\mathbb{N}}$ é decrescente e

$$\mu(\Phi^n(A)) = \left(\frac{2}{3}\right)^n \mu(A) \tag{6.5}$$

para todo $n \in \mathbb{N}$.

Para todo $n \in \mathbb{N}$ definimos a n-ésima etapa da construção do conjunto de Cantor como

$$C_n = \Phi^n([0,1])$$

Os conjuntos C_n são compactos e satisfazem $\mu(C_n) = \left(\frac{2}{3}\right)^n$.

Definimos o conjunto de cantor como

$$C = \bigcap_{n\in\mathbb{N}} C_n$$

6.27 Observação *De maneira mais geométrica o conjunto de Cantor C é definido recursivamente para isso começamos removendo* $(1/3, 2/3)$ *de* $[0,1]$ *e depois removendo o terço do meio de cada intervalo que permanece e assim por diante. Através dessa construção obtemos a sequência de conjuntos*

$$C_0 = [0,1]$$
$$C_1 = [0,1/3] \cup [2/3,1]$$
$$C_2 = [0,1/9] \cup [2/9,1/3] \cup [2/3,7/9] \cup [8/9,1]$$
$$C_3 = \cdots$$

Assim o conjunto Cantor contém todos os pontos no intervalo que não são excluídos em qualquer etapa neste processo infinito.

Existe ainda outra maneira de descrever o conjunto Cantor. Se representarmos os números no intervalo $[0,1]$ *na base 3 então podemos escrever os números no conjunto de Cantor como:*

$$C = \{x | x = \sum_{i=1}^{\infty} \frac{a_i}{3^i} \text{ com } a_i \in \{0,2\}\}.$$

Agora podemos definir uma variável aleatória de Cantor cuja função de distribuição aumenta no conjunto de Cantor e permanece constante fora desse conjunto. Definimos esta função da seguinte forma:

Começamos definindo uma sequência de funções $f_n : \mathbb{R} \to \mathbb{R}$ como

$$f_n(x) = \left(\frac{3}{2}\right)^n \mathbb{1}_{C_n}$$

e a funções

$$F_n(x) = \int_{-\infty}^{x} f_n(x) d\mu$$

e associada a função de conjuntos

$$F_n(A) = \int_A f_n(x) d\mu$$

Se denotarmos por $\mathcal{C}_n$ a coleção de 2^n intervalos que particionam C_n então se $J \in \mathcal{C}_n$

$$\begin{aligned}\int_J f_n d\mu &= \left(\frac{3}{2}\right)^n \mu(J) \\ &= \left(\frac{3}{2}\right)^{n+k} \mu(\Phi^k(J) \\ &= \int_J f_{n+k} d\mu\end{aligned}$$

E assim $F_n(J) = F_{n+k}(J)$ para todo $k \in \mathbb{N}$.

O sequência de funções $\{F_n\}_{n\in\mathbb{N}}$ é uniformemente de Cauchy. Para ver isso considere $n, k \in \mathbb{N}$

Se $x \in \mathbb{R} \setminus [0,1]$ então claramente

$$|F_{n+k}(x) - F_n(x)| = 0$$

Se $x \in \mathbb{R} \setminus C_n$ então

$$\begin{aligned}|F_{n+k}(x) - F_n(x)| &= \left| \sum_{J\in\mathcal{C}_n, J\subset(-\infty,x]} F_{n+k}(J) \right. \\ &\quad \left. - \sum_{J\in\mathcal{C}_n, J\subset(-\infty,x]} F_n(J) \right| \\ &= \left| \sum_{J\in\mathcal{C}_n, J\subset(-\infty,x]} F_{n+k}(J) - F_n(J) \right| \\ &= 0\end{aligned}$$

Se $x \in C_n$ isso significa que existe $J \in \mathcal{C}$ tal que $x \in J$ e assim

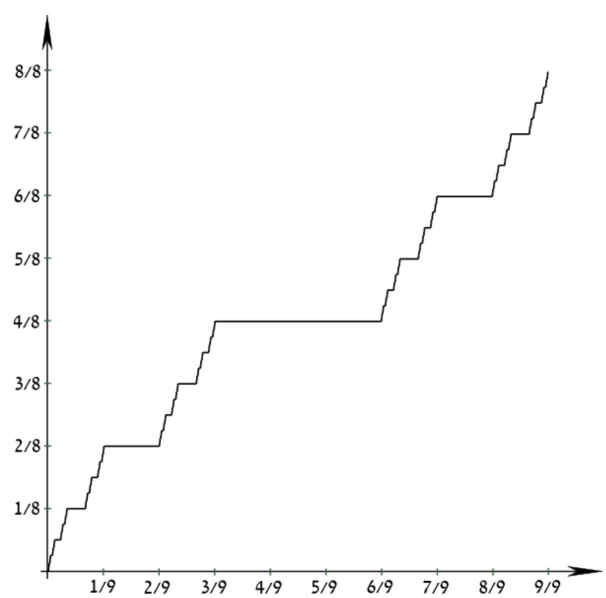
$$|F_{n+k}(x) - F_n(x)| = \left| \sum_{J \in C_n, J \subset (-\infty, x]} F_{n+k}(J) + F_{n+k}(J \cap (-\infty, x]) \right.$$
$$\left. - \sum_{J \in C_n, J \subset (-\infty, x]} F_n(J) - F_n(J \cap (-\infty, x]) \right|$$
$$= |F_{n+k}(J \cap (-\infty, x]) - -F_n(J \cap (-\infty, x])|$$
$$\leq 2F_n(J)$$
$$= 2\left(\frac{3}{2}\right)^n \mu(J)$$
$$= 2\frac{1}{2^n}$$

Logo

$$|F_{n+k}(x) - F_n(x)| \leq 2\frac{1}{2^n}$$

para todo $x \in \mathbb{R}$. Assim a sequência é uniformemente de Cauchy e logo converge para uma função contínua $F : \mathbb{R} \to [0,1]$ e assim é uma distribuição.

Este procedimento define uma função contínua, não decrescente de $\mathbb{R}$ em $[0,1]$



No entanto, uma vez que esta função é constante, exceto no conjunto Cantor, vemos que sua derivada fora do conjunto Cantor deve ser zero.

Observamos que não existe uma função de densidade para F, pois se existisse tal função seria igual a 0 em um conjunto de medidas 1.

Exercício: No k-ésimo lançamento de uma moeda justa, um jogador recebe a seguinte premiação: 0 se for cara e $2/3^k$ se for coroa. Seja X o ganho total do jogador após uma sequência infinita de lançamentos da moeda. Mostre que X possui a distribuição Cantor. □

Outros exemplos selvagens

6.28 Exemplo *Variável aleatória uniforme nos irracionais*

$$f(x) = \begin{cases} 1 & \textit{para } x \in [0,1 \backslash \mathbb{Q} \\ 0 & \textit{para } x \in [0,1 \cap \mathbb{Q} \end{cases}$$

6.29 Exemplo *Dado $\{r_k\}$ uma enumeração dos racionais e defina*

$$p(r_k) = \begin{cases} \frac{6}{\pi^2 k^2} & \textit{para } x \in \mathbb{Q} \\ 0 & \textit{caso contrário} \end{cases}$$

Como $\sum_{k=1}^{\infty} 1/k^2 = \pi^2/6$ a função acima é realmente uma distribuição.

Exercícios

Ex. 6.1 — Variáveis Aleatórias Discretas Independentes

1. Mostre que se X e Y são variáveis aleatórias tomando valores inteiros independentes,então
$$\mathbf{P}(X+Y=n) = \sum_k \mathbf{P}(X=k)\,\mathbf{P}(Y=n-k).$$
2. Dadas X e Y variáveis aleatórias independentes de Poisson com parâmetros e μ respectivamente. Prove que $X+Y$ é uma variável aleatória de Poisson com parâmetro $+\mu$.
3. Dadas X e Y variáveis aleatórias independentes de Bernoulli de parâmetros (n,p) e (m,p) respectivamente. Prove que $X+Y$ é uma variável aleatória de Bernoulli com parâmetro $(n+m,p)$.

Ex. 6.2 — Variáveis aleatórias induzem medidas de probabilidade em $\mathbb{R}$

Considere uma variável aleatória X definida no espaço de probabilidade $(\Omega, \mathcal{F}, \mathbf{P})$. Prove que X induz uma medida de probabilidade em $\mathbb{R}$ no seguinte sentido: Para todo conjunto de Borel A de $\mathbb{R}$, defina

$$\mathbf{P}(A) := \mathbf{P}(X \in A) = \mathbf{P}(\omega \in \Omega : \ X(\omega) \in A).$$

Prove que $(\mathbb{R}, \mathbb{R}, \mathbf{P})$ é um espaço de probabilidade.

Ex. 6.3 — Se uma função F satisfizer

- F é não decrescente e $0 \leq F(x) \leq 1$.
- $F(x) \to 0$ quando $x \to -\infty$ e $F(x) \to 1$ quando $x \to +\infty$.
- F é contínua a direita, ou seja
$$\lim_{y \to x_+} F(y) = F(x).$$

então F é a função de distribuição de alguma variável aleatória X.

Ex. 6.4 — Suponha que $Y_1, Y_2, \ldots$ é uma sequência infinita de variáveis aleatórias independentes, todas definidas no mesmo espaço de probabilidade $(\Omega, \mathcal{F}, \mathbf{P})$, tomando valores 0 e 1 com probabilidade 1/2 cada. Mostre que $U := \sum_{k=1}^{\infty} 2^{-k} Y_k$ é uniformemente em [0,1]. Dica: mostre que

$$\mathbf{P}[U \leq x] = \begin{cases} x & \text{quando } x \in [0,1]; \\ 1 & \text{quando } x > 1; \\ 0 & \text{quando } x < 0. \end{cases}$$

para todo $x \in \mathbb{R}$ analisando $\mathbf{P}[U_n \leq x]$ quando $n \to \infty$ onde $U_n := \sum_{k=1}^{n} 2^{-k} Y_k$.

Ex. 6.5 — Se

$$X = X^+ - X^-, \qquad Y = Y^+ - Y^-.$$

Prove que

$$X \stackrel{\text{d.}}{=} Y \text{ implica } X^+ \stackrel{\text{d.}}{=} Y^+ \text{ e } X^- \stackrel{\text{d.}}{=} Y^-.$$

* **Ex. 6.6 — Outra construção da Distribuição de Cantor**

Considere Y_n como no problema anterior. Defina

$$Y = \sum_{n=1}^{\infty} \frac{2Y_n}{3^n}$$

1. Prove que a função de distribuição F_Y é contínua.
2. Prove que F_Y é diferenciável q.c. com derivada igual a 0 q.c. Dica Prove que F_Y é constante no complemento do conjunto de Cantor.

3. Dado μ_Y a distribuição de Y. E m a medida de Lebesgue na reta real. Prove que μ_Y e m são mutualmente singulares. Ou seja que existe um conjunto de Borel A com $m(A) = 0$ e $\mu_Y(A^c) = 0$.

Definição: Suponha X e Y duas variáveis aleatórias, não necessariamente definidas no mesmo espaço de probabilidade. A variável aleatória Y é dita **estocasticamente maior** que X se $\mathbf{P}[X \leq x] \geq \mathbf{P}[Y \leq x]$ para todo $x \in \mathbb{R}$.

Ex. 6.7 — Suponha X e Y duas variáveis aleatórias e que Y é estocasticamente maior que X. Mostre que existem variáveis aleatórias X^* e Y^* definidas no mesmo espaço de probabilidade $(\Omega, \mathcal{F}, \mathbf{P})$ tais que $X^* \sim X$, $Y^* \sim Y$ e $X^*(\omega) \leq Y^*(\omega)$ para todo $\omega \in \Omega$.

Ex. 6.8 — Existem quatro maneiras significativas de definir uma inversa generalizada de uma função de distribuição F :

- $F_1^{-1}(x) := \sup\{x : F_X(x) < t\}, \quad t \in [0,1]$,
- $F_2^{-1}(x) := \inf\{x : F_X(x) \geq t\}, \quad t \in [0,1]$,
- $F_3^{-1}(x) := \inf\{x : F_X(x) > t\}, \quad t \in [0,1]$,
- $F_4^{-1}(x) := \sup\{x : F_X(x) \leq t\}, \quad t \in [0,1]$,

CAPÍTULO 7

Valor Esperado

Temos ainda um elemento central e fundamental para a teoria da probabilidade que precisamos abordar. Este é o conceito de *esperança* ou *valor esperado*.

Intuitivamente, o valor esperado de uma variável aleatória X é a generalização do conceito de média ponderada dos valores que X assume ou equivalentemente, de modo mais geométrico, o centro de massa dos valores que X assume.

Como motivação considere X é uma variável simples dada por

$$X = \sum_{k=1}^{n} a_k \cdot \mathbb{1}_{A_k},$$

com $A_1, A_2, \ldots, A_n$ disjuntos. Então a média ponderada $m(X)$ dos valores que X assume é

$$m(X) = \sum_{k=1}^{n} a_k \, \mathbf{P}(A_k) = \int X \, \mathrm{d}\mathbf{P}.$$

o que motiva a definição:

7.1 Definição (Valor Esperado ou Esperança) *Se X é uma variável aleatória definida em* $(\Omega, \mathcal{F}, \mathbf{P})$, *seu* ***valor esperado***, ***esperança*** *ou* ***média*** *é definido como*

$$\mathbf{E}[\mathrm{X}] := \int_\Omega \mathrm{X} d\mathbf{P}$$

Usaremos aqui a mesma nomenclatura usada em integrais de Lebesgue, e diremos que X é **integrável** se $\mathbf{E}|X| < \infty$ (o que é equivalente a dizer que $\mathbf{E}[X^+] < \infty$ e $\mathbf{E}[X^-] < \infty$ e de modo mais geral falaremos em esperança de X se X for quasi-integrável, isto é se $\mathbf{E}[X^+]$ ou $\mathbf{E}[X^-]$ for finita.

Agora podemos retomar o exemplo de variáveis aleatórias simples de posse da definição da esperança

7.2 Exemplo (Variáveis Aleatórias Simples) *Seja X uma variável aleatória X dada por*

$$X = \sum_{k=1}^{n} a_k \cdot \mathbb{1}_{A_k},$$

com $A_1, A_2, \ldots, A_n$ disjuntos. Então

$$\mathbf{E}[X] = \int X \, \mathrm{d}\mathbf{P} = \sum_{k=1}^{n} a_k \, \mathbf{P}(A_k).$$

Se além disso, os valores $a_1, \ldots, a_k \in \mathbb{R}$ são distintos, então $A_k = X^{-1}(\{a_k\}) = \{X = a_k\}$, e

$$\mathbf{E}[X] = \sum_{k=1}^{n} a_k \, \mathbf{P}(X = a_k).$$

Podemos restringir o exemplo anterior é obter um dos exemplos mais simples e também um dos mais úteis.

7.3 Exemplo (A Esperança da Função Indicadora) *Seja $\mathbb{1}_A$ a indicadora de um eventos $A \in \mathcal{F}$. $\mathbb{1}_A$ é uma variável aleatória simples, e pelo exemplo anterior, o seu valor esperado é dado por*

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

Ao ser definida como a integral de Lebesgue, a esperança herda uma série de propriedades importantes, que listaremos a seguir.

7.4 Teorema *Suponha que X, Y sejam variáveis aleatórias integráveis. Então*

1. *(Linearidade)* $\mathbf{E}[aX + bY] = a\mathbf{E}[X] + b\mathbf{E}[Y]$ *para quaisquer números reais a, b.*

2. *Se $X \equiv b \in \mathbb{R}$, constante, então $\mathbf{E}[X] = b$ (escrevemos também que $\mathbf{E}[b] = b$).*

3. *Se* $X \geq 0$ *então* $\mathbf{E}[X] \geq 0$.

4. *(Monotonicidade) Se* $X \geq Y$ *então* $\mathbf{E}[X] \geq \mathbf{E}[Y]$.

5. *(Desigualdade triangular)* $|\mathbf{E}[X]| \leq \mathbf{E}[|X|]$

6. *Se* $X_n \uparrow X$ *então* $\mathbf{E}[X] = \lim_{n\to\infty} \mathbf{E}[X_n]$

Demonstração. *Essas propriedades são herança direta das propriedades das integrais de Lebesgue, e não serão demonstradas.*

Mostraremos apenas a propriedade 2, que é típica de medidas de probabilidade.

Tome então uma variável aleatória X constante igual a b. Ou seja, $X(\omega) = b$ para todo ω. Essa variável é uma variável simples, dada por

$$X = b \cdot \mathbb{1}_\Omega,$$

e portanto

$$\mathbf{E}[b] = \mathbf{E}[X] = b \cdot \mathbf{P}(\Omega) = b.$$

7.5 Exemplo *Se uma esfera é colorida de modo que 90% de sua área seja vermelha e 10% seja azul, então existe um modo de inscrever um cubo de modo que os 8 vértices estejam tocando pontos vermelhos da esfera.*

Sejam $X_1, \dots, X_8$ as variáveis aleatórias indicadoras do fato de cada vértice ser vermelho ou não. Então, claramente $X_1 + \cdots + X_8$ é o número de vértices vermelhos. Por linearidade da esperança,

$$\mathbf{E}[X_1 + \cdots + X_8] = \mathbf{E}[X_1] + \cdots + \mathbf{E}[X_8] = 8 \cdot 0.9 = 7.2.$$

Para que o número médio de vértices seja 7.2, deve haver algum cubo com mais de 7 vértices vermelhos, e portanto deve ter 8 vértices vermelhos.

Esse é um exemplo da utilização de técnicas probabilísticas para provar fatos determinísticos.

Observe que não exigimos que as variáveis sejam independentes. Verifique que de fato isso não é necessário. ◁

Temos também a seguinte caracterização do operador esperança.

7.6 Teorema (Caracterização da Esperança) *A esperança é o único operador $\mathbf{E} : \mathcal{L}^1(\Omega) \to \mathbb{R}$ que satisfaz:*

- *Linearidade. Para $a, b \in \mathbb{R}$, $\mathbf{E}[aX + bY] = a\mathbf{E}[X] + b\mathbf{E}[Y]$.*
- *Continuidade. Se $X_n \to X$ então $\mathbf{E}[X_n] \to \mathbf{E}[X]$*
- *Relação com a probabilidade. Para cada evento A, $\mathbf{E}[\mathbb{1}_A] = \mathbf{P}(A)$.*

Demonstração. *Veja que*

$$\mathbf{E}_1[X] := \int_\Omega X d\mathbf{P}$$

tem as propriedades pedidas. Assim temos trivialmente a existência.

Vamos provar a unicidade mostrando a coincidência de um operador $\mathbf{E}$ *com as propriedades listadas e* $\mathbf{E}_1$.

O item 3. e a linearidade nos permite mostrar a coincidência para funções simples.

Se $X = \sum_i a_i \mathbb{1}_{A_i}$ *então* $\mathbf{E}_1[X] = \sum_i a_i \mathbf{P}(A_i) = \mathbf{E}[X]$.

Provaremos agora para funções positivas. Dado $X > 0$, *temos que existem funções simples* X_n *com* $X_n \uparrow X$. *Por continuidade*

$$\mathbf{E}_1[X] = \lim \mathbf{E}_1[X_n] = \lim \mathbf{E}[X_n] = E[X]$$

Para funções quaisquer. Consideramos a decomposição $X = X^+ + X^-$ *e usamos a linearidade*

ntes de darmos continuidade, vamos estabelecer uma nomenclatura que será central deste ponto em diante.

7.7 Definição *Dada um espaço de probabilidade* $(\Omega, \mathcal{F}, \mathbf{P})$, *diremos que uma afirmativa* $\mathbb{A}$ *ocorre* ***quase-certamente*** *se*

$$\mathbf{P}(\{\omega \in \Omega : \omega \text{ satisfaz a afirmativa } \mathbb{A}\}) = 1.$$

De notaremos isso por "$\mathbb{A}$ *ocorre q.c.".*

Em particular, dadas duas variáveis aleatórias X, Y *definidas em* Ω, *temos que* $X = Y$ *q.c., se* $\mathbf{P}(X = Y) = 1$. *Ou ainda que* $X \leq Y$ *q.c se* $\mathbf{P}(X \leq Y) = 1$.

Integrais de Lebesgue ignoram eventos de medida nula, e isso é mostrado no seguinte resultado.

7.8 Proposição *Dadas variáveis aleatórias* X *e* Y, *vale que*

1. *Se* $X = 0$ *q.c então* $\mathbf{E}[X] = 0$;

2. *Se* $X = Y$ *q.c e* X *é integrável ou* $\mathbf{E}[X] = \pm\infty$, *então* $\mathbf{E}[Y] = \mathbf{E}[X]$;

3. *Se* $X \leq Y$ *q.c. e* Y *é integrável, então* $\mathbf{E}[X] \leq \mathbf{E}[Y]$ *(*$\mathbf{E}[X]$ *pode ser* $-\infty$*)* .

4. *Se* $X \geq 0$ *e* $\mathbf{E}[X] = 0$, *então* $X = 0$ *q.c..*

Demonstração.

1. *Suponha que* X *é simples. Temos então que*

$$X = \sum_{k=1}^{n} a_k \mathbb{1} A_k,$$

e como $\mathbf{P}(X = 0) = 1$*, então* $\mathbf{P}(A_k) = 0$ *sempre que* $a_k \neq 0$*. Com isso, temos que* $\mathbf{E}[X] = 0$*.*

Para X *não negativa, tome* $(X_n)_{n\geq 1}$ *uma sequência não-decrescente de variáveis aleatórias simples e não-negativas tal que* $X_n \uparrow X$*. Como* $0 \leq X_n \leq X$ *e* $\mathbf{P}(X = 0) = 1$*, temos que* $X_n = 0$ *q.c e portanto* $\mathbf{E}[X_n] = 0$*. Assim, pelo Teorema da Convergência Monótona,* $\mathbf{E}[X] = \lim_{n\to\infty} \mathbf{E}[X_n] = 0$*.*

Para X *qualquer, basta notar que se* $X = 0$ *q.c, então* $X^+ = 0$ *q.c. e* $X^- = 0$ *q.c..*

2 *Exercício (aplique o item anterior em* $X - Y$*)*

3 *Tome* X, Y *variáveis aleatórias tais que* $X \leq Y$ *q.c. e* Y *integrável. Defina*

$$X^* = X \cdot \mathbb{1}_{X\leq Y} + Y \cdot \mathbb{1}_{X>Y}.$$

Como $\mathbf{P}(X > Y) = 0$*, então* $X^* = X$ *q.c., e além disso* $X^* \leq Y$ *pontualmente.*

Se X^* *é integrável, então a monotonicidade da esperaça nos dá que* $\mathbf{E}[X^*] \leq \mathbf{E}[Y]$*, e pelo item anterior* $\mathbf{E}[X] = \mathbf{E}[X^*] \leq \mathbf{E}[Y]$*.*

Caso contrário, temos $(X^*)^+ \leq Y^+$*, de modo que* $\mathbf{E}(X^*)^+ \leq \mathbf{E}[Y^+] < \infty$ *e* $\mathbf{E}[(X^*)^-] = \infty$*, de modo que* $\mathbf{E}[X^*] = -\infty$*.*

Agora basta usar o item anterior para X^+ *e* X^- *para concluir o resultado.*

4 *Dado* $\epsilon > 0$*, defina* $Y_\epsilon = \epsilon \mathbb{1}_{\{X>\epsilon\}}$ *e observe que* $Y_\epsilon \leq X$ *para todo* $\epsilon > 0$*. Segue que*

$$\epsilon \mathbf{P}(X > \epsilon) = \mathbf{E}[Y_\epsilon] \leq \mathbf{E}[X] = 0,$$

e portanto $\mathbf{P}(X > \epsilon) = 0$ *para todo* $\epsilon > 0$*.*

Fazendo $\epsilon \longrightarrow 0$*, vemos que* $\mathbf{P}(X > 0) = 0$*, e como* $\mathbf{P}(X \geq 0) = 1$*, o resultado segue.* □

7.1 Calculando a Esperança

Nesta seção veremos como a esperança de uma variável aleatória pode ser determinada a partir de sua distribuição. Para isso, vamos seguir a estratégia de três passos, mas de um ponto de vista bem específico.

Tome agora uma variável aleatória X qualquer, e $\phi : \mathbb{R} \longrightarrow \mathbb{R}$ uma função mensurável e simples e não-negativa, dada por

$$\phi(x) = \sum_{k=1}^{n} a_k \mathbb{1}_{A_k}(x),$$

$A_1, \ldots, A_n \in \mathcal{B}(\mathbb{R})$ disjuntos.

Temos assim que

$$\phi(X) = \sum_{k=1}^{n} a_k \mathbb{1}_{A_k}(X) = \sum_{k=1}^{n} a_k \mathbb{1}_{X \in A_k},$$

que é uma variável aleatória simples.

Segue que

$$\mathbf{E}[\phi(X)] = \sum_{k=1}^{n} a_k \mathbf{P}(X \in A_k) = \sum_{k=1}^{n} a_k \mathbf{P}_X(A_k) = \int_{\mathbb{R}} \phi(x)\, \mathrm{d}\mathbf{P}_X(x).$$

Dada agora uma função $\varphi : \mathbb{R} \to \mathbb{R}$ mensurável e não negativa, tome uma sequencia não-decrescente de funções simples e mensuráveis $\phi_n : \mathbb{R} \to \mathbb{R}, n \geq 1$, não-negativas e tais que $\phi_n(x) \uparrow \varphi(x)$, para todo $x \in \mathbb{R}$.

Pelos mesmos argumentos anteriores, $\phi_n(X), n \geq 1$ é uma sequência de variáveis aleatórias simples e não-negativas, tais que $\phi_n(X) \uparrow \varphi(X)$.

Segue do Teorema da Convergência Monótona, que

$$\mathbf{E}[\varphi(X)] = \lim_{n\to\infty} \mathbf{E}[\phi_n(X)] = \lim_{n\to\infty} \int_{\mathbb{R}} \phi_n(x)\, \mathrm{d}\mathbf{P}_X(x) = \int_{\mathbb{R}} \varphi(x)\, \mathrm{d}\mathbf{P}_X(x).$$

Dada agora uma função $\varphi : \mathbb{R} \to \mathbb{R}$ mensurável, podemos escrever $\varphi = \varphi^+ - \varphi^-$, e aplicando o que descobrimos até agora, vemos que

$$\mathbf{E}[\phi(X)] = \mathbf{E}[\phi^+(X)] - \mathbf{E}[\phi^-(X)] = \int_{\mathbb{R}} \varphi^+(x)\, \mathrm{d}\mathbf{P}_X(x) - \int_{\mathbb{R}} \varphi^-(x)\, \mathrm{d}\mathbf{P}_X(x) = \int_{\mathbb{R}} \varphi(x)\, \mathrm{d}\mathbf{P}_X(x).$$

Com isso mostramos que

7.9 Teorema (Teorema da Mudança de Variáveis) *Dada uma variável aleatória X definida em um espaço de probabilidade $(\Omega, \mathcal{F}, \mathbf{P})$ com distribuição $\mathbf{P}_X$, e uma função mensurável $\varphi : \mathbb{R} \to \mathbb{R}$, então*

$$\mathbf{E}[\varphi(X)] = \int_{\mathbb{R}} \varphi(x)\, \mathrm{d}\mathbf{P}_X(x).$$

Ou seja, se qualquer um dos lados da equação acima estiver bem definido, o outro também está e a igualdade vale.

Em particular

$$\mathbf{E}[X] = \int_{\mathbb{R}} x\, \mathrm{d}\mathbf{P}_X(x)$$

Uma consequência direta do resultado anterior é a seguinte.

7.10 Corolário (Distribuição Determina a Esperança) *Dado* X *integrável então se* $\mathrm{X} \stackrel{d.}{=} \mathrm{Y}$ *então* Y *é integrável e* $\mathbf{E}[\mathrm{X}] = \mathbf{E}[\mathrm{Y}]$.

Agora que sabemos que a esperança de uma variável aleatória é determinada unicamente pela sua distribuição, falta apenas entender como surgem aquelas expressões para variáveis discretas e absolutamente contínuas.

Variáveis Aleatórias Discretas

Vamos lembrar que uma variável aleatória discreta é aquela que assumi uma quantidade enumerável de valores. Ou seja, existe um conjunto $S = \{x_1, x_2, \ldots\}$ tal que $\mathbf{P}(X \in S) = 1$.

Com isso, temos que a distribuição $\mathbf{P}_X$ é totalmente definida pelos valores de $\mathbf{P}(\{x_k\}) = \mathbf{P}(X = x_k)$.

De fato, para qualquer $A \in \mathcal{B}(\mathbb{R})$

$$\mathbf{P}_X(A) = \sum_{k : x_k \in A} \mathbf{P}_X(\{x_k\}).$$

Perceba também que

$$X = \sum_{k=1}^{\infty} x_k \, \mathbb{1}_{A_k},$$

onde $A_k = \{X = x_k\} \in \mathcal{F}$.

Para simplificar nossa vida, suponha que $X \geq 0$. Ou seja $x_k \geq 0$ para todo $k \geq 1$. Nestas condições, as variáveis

$$X_n = \sum_{k=1}^{n} x_k \, \mathbb{1}_{A_k},$$

são simples e portanto

$$\mathbf{E}[X_n] = \sum_{k=1}^{n} x_k \, \mathbf{P}(A_k) = \sum_{k=1}^{n} x_k \, \mathbf{P}(X = x_k).$$

Além disso, temos que $X_n \uparrow X$ e o Teorema da Convergência Monótona nos diz que

$$\mathbf{E}[X] = \lim_{n \to \infty} \mathbf{E}[X_n] = \sum_{k=1}^{\infty} x_k \, \mathbf{P}(X = x_k).$$

Deixamos como exercício o caso onde X não assume apenas valores positivos.

Fica assim (quase) demonstrado então o seguinte resultado.

7.11 Proposição *Para uma variável aleatória discreta X positiva ou integrável, assumindo valores em $\{x_1, x_2, \ldots\}$, temos que*

$$\mathbf{E}[\mathrm{X}] = \sum_{k=1}^{\infty} x_k \, \mathbf{P}(x = x_k), \tag{7.1}$$

de modo que a série à direita é absolutamente convergente sempre que X for integrável.

Variáveis Absolutamente Contínuas

Vamos recordar que uma variável aleatória é absolutamente contínua se existe uma função mensurável $f : \mathbb{R} \to \mathbb{R}^+$, conhecida como **função densidade de probabilidade de X** (ou simplesmente densidade de X) tal que

$$\mathbf{P}_X(A) = \mathbf{P}(X \in A) = \int_A f(x)\, \mathrm{d}\mu,$$

onde a integral à direita é a integral de Lebesgue (feita em relação à medida de Lebesgue da reta).

Para entender quem é $\mathbf{E}[\varphi(X)]$ devemos primeiro entender como calcular

$$\int_{\mathbb{R}} \varphi(x)\, \mathrm{d}\mathbf{P}_X(x).$$

Para isso, tome primeiro uma função simples

$$\phi(x) = \sum_{k=1}^{n} a_k \mathbb{1}_{A_k},$$

e note que

$$\begin{aligned}
\int_{\mathbb{R}} \phi(x)\, \mathrm{d}\mathbf{P}_X(x) &= \sum_{k=1}^{n} a_k \mathbf{P}_X(A_k) \\
&= \sum_{k=1}^{n} a_k \int_{A_k} f(x)\, \mathrm{d}\mu \\
&= \sum_{k=1}^{n} a_k \int_{\mathbb{R}} \mathbb{1}_{A_k}(x) f(x)\, \mathrm{d}\mu \\
&= \int_{\mathbb{R}} \sum_{k=1}^{n} a_k \mathbb{1}_{A_k}(x) f(x)\, \mathrm{d}\mu \\
&= \int_{\mathbb{R}} \phi(x) f(x)\, \mathrm{d}x.
\end{aligned}$$

Tomando uma sequência de funções simples $\phi_n, n \geq 1$ não negativas, com $\phi_n \uparrow \varphi \geq 0$, concluímos pelo Teorema da Convergência Monótona que

$$\int_{\mathbb{R}} \varphi(x)\, \mathrm{d}\mathbf{P}_X(x) = \lim_n \int_{\mathbb{R}} \phi_n(x)\, \mathrm{d}\mathbf{P}_X(x) = \lim_n \int_{\mathbb{R}} \phi_n(x) f(x)\, \mathrm{d}x = \int_{\mathbb{R}} \varphi(x) f(x)\, \mathrm{d}x.$$

E separando uma função mensurável φ em parte positiva e negativa, concluímos o seguinte resultado.

7.12 **Proposição** *Para variável aleatória absolutamente contínua X com função densidade de probabilidade $f : \mathbb{R} \to \mathbb{R}^+$, e uma função mensurável $\phi : \mathbb{R} \to \mathbb{R}$ vale que*

$$\mathbf{E}[\varphi(X)] = \int_{\mathbb{R}} \varphi(x) f(x)\, \mathrm{d}x, \tag{7.2}$$

onde um lado da equação está bem definido se, e somente se, o outro também está.

7.2 Funções de Vetores Aleatórios

Nesta seção falaremos brevemente sobre o cálculo da esperança para funções de vetores aleatórios.

Mas especificamente, dado um vetor aleatório $X = (x_1, \ldots, X_n)$ definido em um espaço de probabilidade $(\Omega, \mathcal{F}, \mathbf{P})$ e uma função mensurável $\varphi : \mathbb{R}^n \to \mathbb{R}$, queremos entender como calcular $\mathbf{E}[\varphi(X_1, \ldots, X_n)]$.

Vamos olhar primeiro para a distribuição $\mathbf{P}_X$ do vetor aleatório X. Lembre-se que, de modo análogo a variáveis aleatórias, $\mathbf{P}_X$ é uma medida de probabilidade em $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ dada por

$$\mathbf{P}_X(A) = \mathbf{P}((X_1, \ldots, X_n) \in A),$$

para $A \in \mathcal{B}(\mathbb{R}^n)$.

De posse da distribuição $\mathbf{P}_X$ podemos recuperar as **distribuições marginais** das variáveis X_k, fazendo

$$\begin{aligned}
\mathbf{P}_{X_k}(B) &= \mathbf{P}(X_k \in B) \\
&= \mathbf{P}(X_k \in B; X_j \in \mathbb{R}; j \neq k) \\
&= \mathbf{P}((X_1, \ldots, X_n) \in \mathbb{R}^{k-1} \times B \times \mathbb{R}^{n-k}) \\
&= \mathbf{P}_X(\mathbb{R}^{k-1} \times B \times \mathbb{R}^{n-k}).
\end{aligned}$$

A relação entre $\mathbf{P}_X$ e as marginais $\mathbf{P}_{X_k}$ pode ser complicada, e em geral precisamos de informações adicionais sobre o vetor para descreve-la com mais precisão. Uma das situações onde esta relação está bem definida é no caso onde as variáveis $X_1, \ldots, X_n$ são independentes. Deixaremos os detalhes para a lista de exercícios.

Voltando ao nosso problema, não é difícil mostrar, usando os mesmos argumentos usados para mostrar o Teorema 7.9, que vale o seguinte resultado.

7.13 **Teorema** *Dado um vetor aleatório $X = (X_1, \ldots, X_n)$ definido em um espaço de probabilidade $(\Omega, \mathcal{F}, \mathbf{P})$ com distribuição $\mathbf{P}_X$, e uma função mensurável $\varphi : \mathbb{R}^n \to \mathbb{R}$, então*

$$\mathbf{E}[\varphi(X_1, \ldots, X_n)] = \int_{\mathbb{R}^n} \varphi(x_1, \ldots, x_n)\, \mathrm{d}\mathbf{P}_X(x_1, \ldots, x_n).$$

Ou seja, se qualquer um dos lados da equação acima estiver bem definido, o outro também está e a igualdade vale.

No caso em que as variáveis $X_1, \dots, X_n$ são discretas, assumindo valores em $S_1, \dots, S_n$ respectivamente, então o vetor é também simples, assumindo valores em $S_1 \times \cdots \times S_n$, e a variável $\phi(X_1, \dots, X_n)$ é discreta, assumindo valores $\varphi(x_1, \dots, x_n) \in \phi(S_1 \times \cdots \times S_n)$ com probabilidade $\mathbf{P}(X_1 = x_1, \dots, X_n = x_n)$.

Vale então o seguinte resultado.

7.14 **Proposição** *Dadas variáveis aleatórias $X_1, \dots, X_n$ discretas assumindo valores em $S_1, \dots, S_n$, respectivamente, e uma função mensurável $\varphi : \mathbb{R}^n \to \mathbb{R}$, então*

$$\mathbf{E}[\varphi(X_1, \dots, X_n)] = \sum_{x_1 \in S_1} \cdots \sum_{x_n \in S_n} \varphi(x_1, \dots, x_n)\, \mathbf{P}(X_1 = x_1, \dots, X_n = x_n),$$

quando os dois lados estiverem bem definidos.

O caso onde $X_1, \dots, X_n$ são absolutamente contínuas é um pouco mais complicado, pois não garante que o vetor X seja absolutamente contínuo. Ou seja, a existência de densidades $f_1, \dots, f_n : \mathbb{R} \to \mathbb{R}$ não garante a existência de uma função densidade $f : \mathbb{R}^n \to \mathbb{R}^+$ tal que

$$\begin{aligned} \mathbf{P}_X(A) &= \mathbf{P}((X_1, \dots, X_n) \in A) \\ &= \int_{\mathbb{R}^n} f(x_1, \dots, x_n)\, \mathrm{d}\lambda_n(x_1, \dots, x_n) \\ &= \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} f(x_1, \dots, x_n)\, \mathrm{d}x_1 \cdots \mathrm{d}x_n, \end{aligned}$$

para $A \in \mathcal{B}(\mathbb{R}^n)$, onde λ_n é a medida de Lebesgue em $\mathbb{R}^n$.

7.15 **Observação** *Para $X = (X_1, \dots, X_n)$ é um vetor aleatório absolutamente contínuo, também é usual dizer que as variáveis $X_1, \dots, X_n$ são conjuntamente absolutamente contínuas.* $\triangleleft$

Não vamos entrar em maiores detalhes sobre esse tipo de vetor, mas podemos mostrar que

7.16 **Proposição** *Para um vetor aleatório absolutamente contínuo X com função densidade de probabilidade $f : \mathbb{R}^n \to \mathbb{R}^+$, e uma função mensurável $\phi : \mathbb{R}^n \to \mathbb{R}$ vale que*

$$\mathbf{E}[\varphi(X)] = \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} \varphi(x_1, \dots, x_n) f(x_1, \dots, x_n)\, \mathrm{d}x_1 \cdots \mathrm{d}x_n, \tag{7.3}$$

onde um lado da equação está bem definido se, e somente se, o outro também está.

7.3 Variância, Covariância e Momentos de uma Variável Aleatória

A esperança mede, como já discutimos, o comportamento médio de uma variável aleatória. Uma média, ponderada pelas probabilidades, dos diversos valores assumidos pela variável. Porém, essa a média falha em captar a dispersão dos valores e suas probabilidades.

Para entender isso, vamos ver um exemplo simples.

7.17 Exemplo *Tome X e Y variáveis aleatórias discretas tais que*

$$\mathbf{P}(X = -1) = \mathbf{P}(X = 1) = \frac{1}{2},$$

e

$$\mathbf{P}(Y = -100) = \mathbf{P}(Y = -50) = \mathbf{P}(Y = 50) = \mathbf{P}(Y = 100) = \frac{1}{4}.$$

Calculando, encontramos que $\mathbf{E}[X] = \mathbf{E}[Y] = 0$, mas a variável aleatória Y assume valores muito mais dispersos em torno da média. ◁

Uma das alternativas que temos para medir essa dispersão é medir uma espécie de "distância média" (usado aqui de forma livre) entre a variável e sua média. É isso que motiva as próximas definições.

É isso que motiva as próximas definições.

7.18 Definição (Momentos, Variância e Covariância) *Se X é uma variável aleatória em $(\Omega, \mathcal{F}, \mathbf{P})$:*

- *Dado $n \in \mathbb{N}^*$ e X tal que X^n seja integrável então*

  $$m_k := \mathbf{E}[\mathrm{X}^k] \quad M_k := \mathbf{E}[|\mathrm{X}|^k] \quad \text{para } k = 1, \ldots, n$$

  são denominados o k-ésimo ***momento*** *e k-ésimo* ***momento absoluto*** *de X respectivamente.*

- *Se X é quadrado integrável, então*

  $$\mathbf{Var}[\mathrm{X}] := \mathbf{E}[(\mathrm{X} - \mathbf{E}[\mathrm{X}])^2]$$

  é a variância de X. O número $\sigma := \sqrt{\mathbf{Var}[\mathrm{X}]}$ é denominado ***desvio padrão*** *de X.*

- *Se X e Y são quadrado integráveis definimos a covariância de X e Y como*

  $$\mathbf{Cov}[\mathrm{X}, \mathrm{Y}] := \mathbf{E}\left[(\mathrm{X} - \mathbf{E}[\mathrm{X}])(\mathrm{Y} - \mathbf{E}[\mathrm{Y}])\right]$$

  As variáveis X e Y são ditas não correlacionadas se $\mathbf{Cov}[\mathrm{X}, \mathrm{Y}] = 0$.

- *o* ***coeficiente de correlação*** *de X e Y não constantes por*

  $$\rho(\mathrm{X}, \mathrm{Y}) = \frac{\mathbf{Cov}(X, Y)}{\sigma_X \sigma_Y};$$

7.19 Proposição (Propriedades da Variância e Covariância) *Dadas X e Y variáveis aleatórias integráveis. Então:*

1. $\mathbf{Var}[X] \geq 0$
2. $\mathbf{Var}[X] = \mathbf{E}[X^2] - \mathbf{E}[X]^2$.
3. $\mathbf{Var}[X] = 0 \Longleftrightarrow X = \mathbf{E}[X]$ *quase certamente.*
4. *A aplicação* $f : \mathbb{R} \to \mathbb{R}$, $x \longmapsto \mathbf{E}[(X - x)^2]$ *tem um mínimo em* $\mathbf{E}[X]$ *com* $f(\mathbf{E}[X]) = \mathbf{Var}[X]$.
5. $\mathbf{Var}[aX + b] = a^2\mathbf{Var}[X]$ *para* $a, b \in \mathbb{R}$.
6. $\mathbf{Cov}[X,Y] = \mathbf{E}[XY] - \mathbf{E}[X]\mathbf{E}[Y]$.
7. $\mathbf{Var}[X + Y] = \mathbf{Var}[X] + \mathbf{Var}[Y] + 2\mathbf{Cov}[X,Y]$
8. $\mathbf{Cov}[X,X] = \mathbf{Var}[X]$;

Demonstração. 1 *Segue direto da definição e das propriedades da esperança, uma vez que* $(X - \mathbf{E}[X])^2 \geq 0$;

2 *Segue da linearidade da esperança. De fato*

$$
\begin{aligned}
\mathbf{Var}[X] &= \mathbf{E}[(X - \mathbf{E}[X])^2] \\
&= \mathbf{E}[X^2 - 2X\mathbf{E}[X] + \mathbf{E}[X]^2] \\
&= \mathbf{E}[X^2] - 2\mathbf{E}[X]^2 + \mathbf{E}[X]^2 \\
&= \mathbf{E}[X^2] - \mathbf{E}[X]^2.
\end{aligned}
$$

3 *Se* $\mathbf{E}[(X - \mathbf{E}[X])^2] = \mathbf{Var}[X] = 0$, *então* $(X - \mathbf{E}[X])^2 = 0$ *q.c., e o resultado segue.*

4 *Imediato.*

5 $\mathbf{Var}[aX+b] = \mathbf{E}[(aX+b)^2]-\mathbf{E}[aX+b]^2 = a^2\mathbf{E}[X^2]+2ab\mathbf{E}[X]+b^2-(a^2\mathbf{E}[X]^2+2ab\mathbf{E}[X]+b^2) = a^2\mathbf{Var}[X]$.

6 *Segue da linearidade da esperança que*

$$
\begin{aligned}
\mathbf{Cov}[X,Y] &= \mathbf{E}[(X - \mathbf{E}[X])(Y - \mathbf{E}[Y])] \\
&= \mathbf{E}[XY - X\mathbf{E}[Y] - Y\mathbf{E}[X] + \mathbf{E}[X]\mathbf{E}[Y]] \\
&= \mathbf{E}[XY] - \mathbf{E}[X]\mathbf{E}[Y].
\end{aligned}
$$

$\boxed{7}$ *Segue da linearidade da esperança e dos itens anteriores que*

$$\begin{aligned}\mathbf{Var}(X+Y) &= \mathbf{E}[(X+Y)^2] - \mathbf{E}[X+Y]^2 \\ &= \mathbf{E}[X^2+Y^2+2XY] - (\mathbf{E}[X]^2+\mathbf{E}[Y^2]+2\mathbf{E}[X]\mathbf{E}[Y]) \\ &= \mathbf{Var}[X]+\mathbf{Var}[Y]+2\mathbf{Cov}[X,Y].\end{aligned}$$

$\boxed{8}$ *Segue direto da definição.* □

O item $\boxed{4}$ nos fornece uma descrição geométrica: imagine que queremos minimizar o erro quadrático médio de uma estimativa x da variável X. Nesse caso a estimativa que minimiza esse erro é a esperança e a variância é o menor valor do erro quadrático médio.

7.20 EXEMPLO *Seja* $X \in \mathcal{L}^2(\mathbf{P})$ *integrável e* $Y = aX + b$, *com* $a,b \in \mathbb{R}$ *não nulos. Temos então que* $\mathbf{Var}(X) < \infty$, $\mathbf{E}[Y] = a\mathbf{E}[X] + b$ *e que* $\mathbf{Var}[Y] = \mathbf{Var}[aX+b] = a^2\mathbf{Var}[X]$.

Da mesma forma temos que $\mathbf{Cov}[X,Y] = \mathbf{E}[(X-\mathbf{E}[X])(aX+b-(a\mathbf{E}[X]+b))] = a\mathbf{Var}[X]$.

Segue que

$$\rho(X,Y) = \frac{a\mathbf{Var}[X]}{\sqrt{a^2(\mathbf{Var}[X])^2}} = \text{sinal}(a) = \begin{cases} 1; & a > 0 \\ -1; & a < 0 \end{cases}.$$

A covariância a correlação servem para medir o "nível de correlação" *entre as variáveis. Nesse exemplo vimos que se* $Y = aX + b$ *então* $\mathbf{Cov}[X,Y]$ *tem o mesmo sinal de* a *e* $\rho(X,Y)$ *é* 1 *ou* -1, *de acordo com o sinal de* a.

7.21 TEOREMA (VARIÁVEIS ALEATÓRIAS INDEPENDENTES SÃO NÃO CORRELACIONADAS) *Sejam* X, Y *são variáveis aleatórias independentes integráveis. Então* (XY) *é integrável e*

$$\mathbf{E}[\mathrm{XY}] = \mathbf{E}[\mathrm{X}]\mathbf{E}[\mathrm{Y}]$$

Em particular, variáveis aleatórias independentes não estão correlacionadas.

Demonstração. *Assuma primeiro que as variáveis* X *e* Y *sejam simples, de modo que* X *e* Y *são discretas assumindo uma quantidade finita de valores cada uma.*

Nestas condições XY *também toma apenas um número finito de valores, de modo que* XY *é* *integrável e*

$$\begin{aligned}
\mathbf{E}[XY] &= \sum_{i=1}^{n}\sum_{j=1}^{m} x_i y_j \mathbf{P}[X = x_i, Y = y_j] \\
&= \sum_{i=1}^{n}\sum_{j=1}^{m} x_i y_j \mathbf{P}[X = x_i]\mathbf{P}[Y = y_j] \quad \textit{(pela independência)} \\
&= \left(\sum_{i=1}^{n} x_i \mathbf{P}[X = x_i]\right)\left(\sum_{j=1}^{m} y_i \mathbf{P}[Y = y_j]\right) \\
&= \mathbf{E}[X]\mathbf{E}[Y].
\end{aligned}$$

Agora suponha $X, Y \geq 0$, *e tome uma sequência crescente* $(\phi_n)_{n\geq 0}$ *de funções* $\phi_n : \mathbb{R}^+ \to \mathbb{R}^+$ *simples, tal que* $\phi_n(x) \uparrow x$, *para todo* $x \in \mathbb{R}^+$.

Para cada $n \geq 1$ *defina agora* $X_n = \phi_n(X)$ *e* $Y_n = \phi_n(Y)$.

Como X, Y *são independentes,* X_n, Y_n *são independentes e simples, para cada* $n \geq 1$. *Além disso,* $X_n \uparrow X$ *e* $Y_n \uparrow Y$. *Assim, pelo Teorema da Convergência Monótona, segue que*

$$\mathbf{E}[XY] = \lim_{n\to\infty} \mathbf{E}[X_n Y_n] = \lim_{n\to\infty} \mathbf{E}[X_n]\mathbf{E}[Y_n] = \mathbf{E}[X]\mathbf{E}[Y] < \infty.$$

Para variáveis quaisquer basta utilizar a decomposição em parte positiva e negativa e observar que as famílias $\{X^+, Y^+\}$, $\{X^+, Y^-\}$, $\{X^-, Y^+\}$ *e* $\{X^-, Y^-\}$ *são formadas por variáveis independentes.* □

Vimos assim que variáveis independentes tem realmente covariância nula. Mas a recíproca não é verdadeira. Vamos deixar esse fato como exercício.

Exercício: Dê um exemplo de variáveis X e Y dependentes, mas com $\mathbf{Cov}[X, Y] = 0$.

(Dica: tente pensar em uma variável X com $\mathbf{E}[X] = 0$ e uma variável Y tal que $Y = 0$ sempre que $X \neq 0$.) □

Uma outra consequência importante é sobre a variância da soma de variáveis independentes.

7.22 COROLÁRIO *Se* $X_1, \ldots, X_n$ *são variáveis aleatórias independentes, então*

$$\mathbf{Var}[X_1 + \cdots + X_n] = \mathbf{Var}[X_1] + \cdots + \mathbf{Var}[X_n].$$

Demonstração. *Para* $n = 2$ *temos*

$$\mathbf{Var}[X_1 + X_2] = \mathbf{Var}[X_1] + \mathbf{Var}[X_2] + 2\mathbf{Cov}[X_1, X_2] = \mathbf{Var}[X_1] + \mathbf{Var}[X_2].$$

Os demais casos segue por indução. □

7.23 Proposição *A aplicação* $\mathbf{Cov} : \mathcal{L}^2(\mathbf{P}) \times \mathcal{L}^2(\mathbf{P}) \to \mathbb{R}$ *é uma forma simétrica bilinear positiva semidefinida e* $\mathbf{Cov}[X, Y] = 0$ *se* Y *é constante quase certamente.*

Demonstração. *A demonstração é direta e será deixada como exercício.* □

7.24 Proposição (Desigualdade de Cauchy–Schwarz) *Se* $X, Y \in \mathcal{L}^2(\mathbf{P})$*, então*

$$\mathbf{Cov}[X, Y]^2 \leq \mathbf{Var}[X]\mathbf{Var}[Y].$$

A igualdade é válida se, e somente se, existirem $a, b, c \in \mathbb{R}$*, não todas nulas, e tais que* $aX + bY + c = 0$ *q.c.*

Demonstração. *A desigualdade de Cauchy-Schwarz é verdadeira para qualquer forma bilinear positiva semidefinita e, portanto, em particular para a covariância. Ainda assim, faremos uma demonstração aqui para deixar o texto mais completo e auto-contido.*

Faremos a demonstração supondo que $\mathbf{Var}[Y] > 0$*. O caso no qual* $\mathbf{Var}[Y] = 0$ *será deixada como exercício.*

Dada $\mathbf{Var}[Y] > 0$*, seja* $\theta := -\frac{\mathbf{Cov}[X,Y]}{\mathbf{Var}[Y]}$.

Temos assim $\theta\mathbf{Var}[Y] = -\mathbf{Cov}[X,Y]$*, e portanto*

$$\begin{aligned} 0 &\leq \mathbf{Var}[X + \theta Y]\mathbf{Var}[Y] \\ &= (\mathbf{Var}[X] + 2\theta\mathbf{Cov}[X, Y] + \theta^2\mathbf{Var}[Y])\mathbf{Var}[Y] \\ &= \mathbf{Var}[X]\mathbf{Var}[Y] - \mathbf{Cov}[X, Y]^2. \end{aligned}$$

Note que $\mathbf{Var}[X + \theta Y] = 0$ *se, e somente se,* $X + \theta Y$ *for constante quase certamente.*

Com isso, como supomos $\mathbf{Var}[Y] \neq 0$*, segue que a igualdade na primeira inequação acima vale se, e somente se,* $X + \theta Y$ *for constante quase certamente.* □

7.25 Corolário *Dadas variáveis* $X,Y \in \mathcal{L}^2(\mathbf{P})$ *então*

$$|\rho(X,Y)| \leq 1,$$

com igualdade apenas se existirem constantes $a,b,c \in \mathbb{R}$*, não todas nulas, tais que* $aX + bY + c = 0$.

7.4 Integração com respeito à Distribuição

Nesta seção, definimos integrais em relação às funções de distribuição. Essas integrais, conhecidas como **integrais de Lebesgue- Stieltjes**, podem ser definidas a partir do zero, mas no entanto, elas são simplesmente as integrais com respeito a probabilidades induzidas em $\mathbb{R}$. Nossa principal aplicação ao cálculo de esperanças.

7.26 Definição *Dado uma função de distribuição F em $\mathbb{R}$, existe uma única probabilidade $\mathbf{P}_F$ em $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ tal que $\mathbf{P}_F((a,b]) = F(b) - F(a)$ para todo intervalo $(a,b]$.*

A demonstração desse fato é análoga a construção que fizemos na seção 2.4 da Integral de Lebesgue e foi feita no exercício 2.3.

7.27 Definição (Integração com respeito à distribuição) *Dada F uma função de distribuição em $\mathbb{R}$. Para toda função $g \in \mathcal{L}^1(\mathbf{P}_F)$ definimos a integral de g com respeito a F*

$$\mathbf{E}_F[g] := \int_{\mathbb{R}} g(x) dF(x)$$

onde estamos considerando g como uma variável aleatória no espaço de probabilidade $(\mathbb{R},\mathcal{B}(\mathbb{R}),\mathbf{P}_F)$.

7.28 Proposição *Se $F(t) = \sum p_i \mathbb{1}_{(t_i \leq t)}$ então para toda $g \geq 0$*

$$\mathbf{E}_F[g] = \int_{\mathbb{R}} g dF = \sum_i p_i g(t_i).$$

Demonstração. *Para provar esse fato suponha que $g = \sum_{j=1}^m b_j \mathbb{1}_{B_j}$ é uma função aleatória simples.*

$$\mathbf{E}_F[g] := \sum_{j=1}^m b_j \mathbf{P}_X(B_j) = \sum_{j=1}^m b_j \sum_i p_i \mathbb{1}_{t_i \in B_j} \tag{7.4}$$

$$= \sum_i p_i \sum_{j=1}^m b_j \mathbb{1}_{B_j}(t_i) \tag{7.5}$$

$$= \sum_i p_i g(t_i) \tag{7.6}$$

Para provar para as funções mensuráveis $g \geq 0$ basta utilizar o Teorema da Convergência Monótona. Se $g \geq 0$ e $g_n \uparrow g$

$$\mathbf{E}_F[g] = \lim_{n\to\infty} \mathbf{E}_F[g_n] = \lim_{n\to\infty} \sum_i p_i g_n(t_i) = \sum_i p_i g(t_i)$$

7.29 Proposição *Suponha que F é absolutamente contínua com densidade contínua por partes f. Então se g é positiva e contínua por partes*

$$\mathbf{E}_F[g] = \int_{\mathbb{R}} g dF = \int_{-\infty}^{\infty} g(x) f(x) dx$$

Demonstração. *Para provar esse fato suponha que* $g = \sum_{j=1}^{m} b_j \mathbb{1}_{B_j}$ *é uma função aleatória do tipo escada. Então*

$$\mathbf{E}_F[g] := \sum_{j=1}^{m} b_j \mathbf{P}(B_j) = \sum_{j=1}^{m} b_j \int_{B_j} f(x)dx \tag{7.7}$$

$$= \sum_{j=1}^{m} b_j \int_{-\infty}^{\infty} f(x)\mathbb{1}_{B_j} dx \tag{7.8}$$

$$= \int_{-\infty}^{\infty} f(x) \left[\sum_{j=1}^{m} b_j \mathbb{1}_{B_j} \right] dx \tag{7.9}$$

$$= \int_{-\infty}^{\infty} f(x)g(x)dx \tag{7.10}$$

A demonstração para uma função geral $g > 0$ *segue, aproximando* g *por funções escadas, o que é possível porque* g *é contínua por partes e, em seguida, usando o Teorema de Convergência Monótona.* □

Utilizando as Proposições 7.28 e 7.29 podemos integrar com respeito a uma distribuição que seja combinação de uma discreta e uma absolutamente contínua.

7.30 Teorema (Mudança de Variáveis) *Dada uma variável aleatória* $X \in \mathcal{L}^1$ *e* F_X *a função de distribuição de* X *então*

$$\mathbf{E}[X] = \int_{\mathbb{R}} x \, dF_X(x)$$

e de modo mais geral,

$$\mathbf{E}[g(X)] = \int_{\mathbb{R}} g(x) \, dF_X(x).$$

Demonstração. *Provaremos primeiro para funções simples não negativas* $X = \sum_{i=1}^{n} a_i \mathbb{1}_{A_i}$. *Nesse caso* $\mathbf{E}[X] = \sum_{i=1}^{n} a_i \mathbf{P}(A_i)$. *Por outro lado* $F_X(t) = \sum_i \mathbf{P}(A_i)\mathbb{1}_{a_i \le t}$ *de modo que*

$$\int_{\mathbb{R}} x \, dF_X(x) = \sum_{i=1}^{n} a_i \mathbf{P}(A_i)$$

Deixaremos o restante da demonstração, que segue a estratégia usual de 3 passos, como exercício.

7.31 Exemplo *Se* $X \sim \mathcal{N}(0;1)$ *então* $\mathbf{E}(X) = 0$.

$$\mathbf{E}[X] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{\frac{-x^2}{2}} \mathrm{d}x = -\frac{1}{\sqrt{2\pi}} \int_{*}^{*} e^u du = -\frac{1}{\sqrt{2\pi}} e^{-x^2/2}\Big|_{-\infty}^{\infty} = 0.$$

7.32 Exemplo *Se* $X \sim \mathrm{Exp}(\lambda)$ *então* $\mathbf{E}[X^n] = \frac{n!}{\lambda^n}$. ◁

7.5 Desigualdades de Markov e Chebyshev

Comecemos pelas chamadas desigualdade de Markov e Chebyshev, que mostram como os diversos momentos informam o comportamento da cauda da distribuição de uma variável.

7.33 Teorema (Desigualdade de Markov) *Dada uma variável aleatória* $X \geq 0$*, vale que*

$$\mathbf{P}(X \geq t) \leq \frac{\mathbf{E}[X]}{t} \text{ para todo } x > 0. \tag{7.11}$$

Em particular, dada uma variável aleatória X *qualquer*

$$\mathbf{P}(|X| \geq t) \leq \frac{\mathbf{E}|X|}{t}.$$

Demonstração. *Começamos definindo* $Y := t\mathbb{1}_{\{X \geq t\}}$*, e observando que* $Y \leq X$ *pontualmente.*

Assim sendo temos

$$\mathbf{E}[X] \geq \mathbf{E}[Y] = \mathbf{E}[t\mathbb{1}_{\{X \geq t\}}] = t\mathbf{P}(X \geq t),$$

concluindo o resultado. □

Essa desigualdade é bastante simples, e de maneira alguma esperamos que seja ótima. Mas podemos usar a própria desigualdade de Markov para melhorá-la, como mostramos nos resultados abaixo.

7.34 Teorema (Desigualdade de Chebyshev) *Dadas uma variável aleatória* $X \geq 0$ *e uma constante* $p > 0$*, então vale que*

$$\mathbf{P}(X \geq t) \leq \frac{\mathbf{E}[X^p]}{t^p} \text{ para todo } t > 0. \tag{7.12}$$

Em particular

$$\mathbf{P}(|X| \geq t) \leq \frac{\mathbf{E}[|X|^p]}{t^p},$$

para uma variável aleatória X *qualquer.*

Demonstração. *A prova é uma redução fácil para a desigualdade de Markov:* $\mathbf{P}(X \geq t) = \mathbf{P}(X^p \geq t^p) \leq \frac{\mathbf{E}[X^p]}{t^p}$*, pela desigualdade de Markov.* □

É claro que para que essa desigualdade não seja trivial precisamos que $X \in \mathcal{L}^p(\mathbf{P})$, mas ainda assim ela ajuda a evidenciar algo interessante.

Se $X \in \mathcal{L}^p(\Omega)$ para todo $p > 0$ (podemos considerar p inteiro), então o decrescimento de $\mathbf{P}(|X| > t)$ deve ser mais rápido que t^{-p} para todo p.

Assim, na desigualdade de Chebyshev, fazemos exigência fortes sobre X mas somos recompensado com um lado direito integrável.

E podemos melhorar um pouco mais a recompensa. Bastando para isso aumentar as exigências sobre a variável. Isso fica evidente na generalização que enunciamos abaixo, e cuja demonstração fica de exercício.

7.35 **Teorema (Desigualdade de Markov Estendida)** *Se φ é uma função monótona crescente definida nos reais não negativos e tomando valores nos reais não negativos, e X é uma variável aleatória e $a \geq 0$, e $\varphi(a) > 0$, então*

$$\mathbb{P}(|X| \geq t) \leq \frac{\mathbb{E}(\varphi(|X|))}{\varphi(t)}$$

Outra aplicação interessante da desigualdade de Markov é no estudo de como os valores de uma variável aleatória se dispersam ao redor da média, e como os diversos momentos centrais, e em particular a variância, informam esse comportamento.

O corolário abaixo é geralmente apresentado na literatura como sendo a própria desigualdade de Chebyshev, e trata exatamente deste assunto.

7.36 **Corolário (Variância)** *Suponha que X seja uma variável aleatória com média μ e variância $\sigma^2 < \infty$. Então*

$$\mathbf{P}(|X - \mu| \geq t\sigma) \leq \frac{1}{t^2} \text{ para qualquer } t > 0, \tag{7.13}$$

ou ainda

$$\mathbf{P}(|X - \mu| \geq t) \leq \frac{\sigma^2}{t^2} \text{ para qualquer } t > 0. \tag{7.14}$$

Demonstração. *Use a desigualdade de Chebyshev para a variável aleatória $|X - \mu| \geq 0$. Então, obviamente, vamos definir $x = t\sigma$ e $p = 2$. Então $\mathbf{E}|X - \mu|^2 = \sigma^2$, e nós temos da desigualdade de Chebyshev,*

$$\mathbf{P}(|X - \mu| \geq t\sigma) \leq \frac{\sigma^2}{(t\sigma)^2} = \frac{1}{t^2}.$$

7.37 **Exemplo** *Dada uma variável aleatória X com variância $\sigma^2 < \infty$ e média $\mu \in \mathbb{R}$, então*

$$\mathbf{P}(\mu - \sqrt{2} \cdot \sigma < X < \mu + \sqrt{2} \cdot \sigma) = 1 - \mathbf{P}(|X - \mu| \geq \sqrt{2} \cdot \sigma) \geq 1 - \frac{1}{2} = 12.$$

Ou seja, X está a uma distância menor que $\sqrt{2} \cdot \sigma$ de sua média, com probabilidade $\geq \frac{1}{2}$. ◁

7.38 **Exemplo** *Considere o experimento de lançar uma moeda de maneira sequencial e independente. Para cada $n \geq 1$, defina X_n como a indicadora de que observamos cara no n-ésimo lançamento. Suponha que em um lançamento da moeda, a probabilidade de observar cara é $p \in (0,1)$.*

Com isso temos que as variáveis $X_1, X_2, \ldots$ são independentes com $\mathbf{E}[X_n] = p$ e $\mathbf{Var}[X_n] = p(1-p)$.

Queremos medir a proporção de caras observadas em n lançamentos, e para isso definiremos $S_n = X_1 + \cdots + X_n$. Isto é, S_n é o total de caras observadas em n lançamentos.

Segue assim que

$$\mathbf{E}[S_n] = n\,p,$$

e, como $X_1, \ldots, X_n$ *são independentes,*

$$\mathbf{Var}(S_n) = n\,p(1-p).$$

Tome agora $\epsilon > 0$ *pequeno e note que*

$$\mathbf{P}\left(\left|\frac{S_n}{n} - p\right| > \epsilon\right) = \mathbf{P}(|S_n - n\,p| > n\epsilon) \leq \frac{\mathbf{Var}(S_n)}{n^2p^2} = \frac{(1-p)}{n\,p}.$$

Concluímos assim que

$$\mathbf{P}\left(\left|\frac{S_n}{n} - p\right| > \epsilon\right) \longrightarrow 0,$$

quando $n \longrightarrow \infty$, *para todo* $\epsilon > 0$. ◁

7.6 Desigualdade Maximal de Kolmogorov

Motivados pelo último exemplo, queremos estudar um pouco mais a distribuição de somas de variáveis independentes.

As desigualdades de Markov e Chebyshev dão uma primeira ideia do que esperar, mas como já vimos, são estimativas muito grosseiras, precisando de hipóteses fortes para ficarem mais robustas.

A próxima desigualdade trata justamente desta questão.

7.39 Teorema *Seja* $(X_n)_{n\geq 1}$ *uma sequência de variáveis aleatórias independentes, definidas no mesmo espaço de probabilidade, com esperança* 0 *e variância finita. Seja* $S_n = \sum_{k=1}^n X_k$. *Então para* $\lambda > 0$,

$$\mathbf{P}\left(\max_{1\leq k\leq n} |S_k| \geq \lambda\right) \leq \frac{\mathbf{Var}(S_n)}{\lambda^2}$$

Demonstração. *Comece observando que o evento* $\{\max_{1\leq k\leq n} |S_k| \geq \lambda\}$ *é a união disjunta dos eventos*

$$A_k = \{|S_k| \geq \lambda, |S_j| < \lambda \text{ para } j < k\},\ k = 1, \ldots, n.$$

Observe também que, para cada $k = 1, \ldots, n$, *temos* $S_n - S_k = X_{k+1} + \cdots + X_n$, *que é independente de* $S_k = X_1 + \cdots + X_k$.

Vale também que $\mathbf{E}[S_k] = \mathbf{E}[X_1] + \cdots + \mathbf{E}[X_k] = 0$ *e, analogamente,* $\mathbf{E}[S_n - S_k = 0]$, *e assim*

$$\mathbf{E}[S_k(S_n - S_k)] = \mathbf{E}[S_k]\mathbf{E}[S_n - S_k] = 0.$$

Além disso, dentro do evento A_k vale que $\lambda^{-2}S_k^2 \geq 1$, de modo que $\mathbb{1}_{A_k} \leq \lambda^{-2}S_k^2\mathbb{1}_{A_k}$, e

$$\begin{aligned}
\mathbf{P}(A_k) &\leq \lambda^{-2}\mathbf{E}[S_k^2\mathbb{1}_{A_k}] \\
&\leq \lambda^{-2}\mathbf{E}[\big(S_k^2 + (S_n - S_k)^2\big)\mathbb{1}_{A_k}] \\
&= \lambda^{-2}\mathbf{E}[\big(S_k^2 + 2S_k(S_n - S_k) + (S_n - S_k)^2\big)\mathbb{1}_{A_k}],
\end{aligned}$$

onde a última linha segue, pois as variáveis $S_k\mathbb{1}_{A_k}$ e $S_n - S_k$ são independentes e $\mathbf{E}[S_k(S_n - S_k)] = 0$.

Segue assim que

$$\mathbf{P}(A_k) \leq \lambda^{-2}\mathbf{E}[\big(S_k + (S_n - S_k)\big)^2\mathbb{1}_{A_k}] = \lambda^{-2}\mathbf{E}[S_n^2\mathbb{1}_{A_k}]$$

e somando encontramos

$$\mathbf{P}\left(\max_{1\leq k\leq n}|S_k| \geq \lambda\right) \leq \lambda^{-2}\mathbf{E}[S_n^2\mathbb{1}_{\{\max_{1\leq k\leq n}|S_k|\geq\lambda\}}] \leq \lambda^{-2}\mathbf{E}[S_n^2].$$

O resultado é concluído quando vemos que $\mathbf{Var}(S_n) = \mathbf{E}[S_n^2] - \mathbf{E}[S_n]^2 = \mathbf{E}[S_n^2]$.

7.7 Desigualdade de Jensen, Hölder e Minkowski

Nesta seção apresentaremos uma desigualdade que generaliza a desigualdade triangular, conhecida por desigualdade de Jensen. Mas antes precisamos lembrar alguns conceitos e resultados de análise real.

7.40 Definição *Uma função $\varphi : \mathbb{R} \to \mathbb{R}$ é **convexa** se para todo $a, b \in \mathbb{R}$ e $\alpha \in [0,1]$ vale que*

$$\varphi(\alpha\, a + (1-\alpha)b) \leq \alpha\varphi(a) + (1-\alpha)\varphi(b).$$

Geometricamente, isso se traduz como o gráfico da função, no intervalo $[a, b]$, estar abaixo da reta secante ao gráfico nos pontos $(a, \varphi(a))$ e $(b, \varphi(b))$.

Funções como $|x|$, x^2 e $e^{\lambda x}$ são exemplos clássicos de funções convexas.

Um resultado importante sobre funções convexas, e que utilizaremos na demonstração do resultado principal, é o seguinte.

7.41 Proposição *Se $\phi : \mathbb{R} \to \mathbb{R}$ é uma função convexa então, dado $t_0 \in \mathbb{R}$, o mapa*

$$t \mapsto \frac{\varphi(t) - \varphi(t_0)}{t - t_0},$$

é não-decrescente.

Demonstração. *Considere* $t > t_0$ *(o caso* $t < t_0$ *segue de modo análogo) e tome* $s \in \mathbb{R}$ *com* $t_0 < s < t$*. Faça então* $\alpha = \frac{s-t_0}{t-t_0}$ *e note que*

$$\alpha\, t + (1-\alpha)t_0 = \frac{s-t_0}{t-t_0}\, t + \left(1 - \frac{s-t_0}{t-t_0}\right) t_0 = s.$$

Segue então da convexidade de φ *que*

$$\varphi(s) \leq \frac{s-t_0}{t-t_0}\,\varphi(t) + \left(1 - \frac{s-t_0}{t-t_0}\right)\varphi(t_0),$$

e o resultado segue. □

7.42 Proposição (Desigualdade de Jensen) *Se* X *é uma variável aleatória integrável e* $\varphi : \mathbb{R} \to \mathbb{R}$ *uma função convexa, então*

$$\varphi(\mathbf{E}[X]) \leq \mathbf{E}[\varphi(X)]$$

Demonstração. *O fato da função* φ *ser convexa significa que para qualquer* t_0*,*

$$\frac{\varphi(t) - \varphi(t_0)}{t - t_0}$$

não decresce em $\mathbb{R} \setminus \{t_0\}$*.*

E desse modo podemos encontrar Φ *tal que*

$$\sup_{t<t_0} \frac{\varphi(t) - \varphi(t_0)}{t - t_0} \leq \Phi \leq \inf_{t>t_0} \frac{\varphi(t) - \varphi(t_0)}{t - t_0}$$

e, portanto, para todo t*, temos*

$$(t - t_0)\Phi \leq \varphi(t) - \varphi(t_0)$$

Agora, seja $t = X$ *e* $t_0 = \mathbf{E}[X]$ *e a equação 95 se torna*

$$(X - \mathbf{E}[X])\,\Phi \leq \varphi(X) - \varphi\left(\mathbf{E}[X]\right) \tag{7.15}$$

Tomando a esperança em ambos os lados de (7.15) *temos*

$$(\mathbf{E}[X] - \mathbf{E}[X])\,\Phi \leq \mathbf{E}[\varphi(X)] - \varphi\left(\mathbf{E}[X]\right)$$

que ao reorganizar, torna-se

$$\varphi\left(\mathbf{E}[X]\right) \leq \mathbf{E}[\varphi(X)].$$

Terminamos com uma aplicação bastante importante da desigualdade de Jensen.

7.43 Exemplo *Seja X uma variável aleatória e duas constantes $p,q \in \mathbb{R}$ tais que $0 < p < q$. Vamos mostrar que se $X \in \mathcal{L}^q(\mathbf{P})$, então $X \in \mathcal{L}^p(\mathbf{P})$.*

Para isso, note que como $q > p > 0$, emtão $q/p > 1$ e a função

$$\varphi(x) = |x|^{\frac{q}{p}}$$

é convexa.

Segue da desigualdade de Jensen

$$\mathbf{E}[|X|^p]^{\frac{q}{p}} = \varphi(\mathbf{E}[|X|^p]) \leq \mathbf{E}[\varphi(|X|^p)] = \mathbf{E}[|X|^q],$$

e portanto se $\mathbf{E}[|X|^q] < \infty$, então $\mathbf{E}[|X|^p] < \infty$, e o resultado segue. ◁

7.44 Proposição (Desigualdade de Hölder) *Dado um espaço de probabilidade $(\Omega,\mathcal{F},\mathbb{P})$, e X,Y variáveis aleatórias*

$$\mathbb{E}[|XY|] \leqslant \left(\mathbb{E}\left[|X|^p\right]\right)^{\frac{1}{p}} \left(\mathbb{E}\left[|Y|^q\right]\right)^{\frac{1}{q}}.$$

Demonstração. *Considere a medida de probabilidade $\nu := \dfrac{g^q}{\int_\Omega g^q d\mu} \cdot \mu$ e a função $h := \dfrac{f}{g^{q-1}}$. Então por Jensen*

$$\int_\Omega fg d\mu = \int_\Omega hg^q d\mu = \int_\Omega g^q d\mu \cdot \int_\Omega h d\nu \leqslant \int_\Omega g^q d\mu \left(\int_\Omega h^p d\nu\right)^{1/p} = \left(\int_\Omega g^q d\mu\right)^{1/q} \left(\int_\Omega f^p d\mu\right)^{1/p}.$$

7.45 Proposição (Desigualdade de Minkowski)

$$\|f+g\|_p \leq \|f\|_p + \|g\|_p \qquad (1 \leq p \leq \infty)$$

Demonstração. *Primeiro, provamos que $f+g$ tem p-norma finita se f e g tiverem*

$$|f+g|^p \leq 2^{p-1}(|f|^p + |g|^p).$$

Na verdade, aqui usamos o fato que $h(x) = x^p$ é uma função convexa e assim, pela definição de convexidade,

$$\left|\tfrac{1}{2}f + \tfrac{1}{2}g\right|^p \leq \left|\tfrac{1}{2}|f| + \tfrac{1}{2}|g|\right|^p \leq \tfrac{1}{2}|f|^p + \tfrac{1}{2}|g|^p.$$

Isso significa que

$$|f+g|^p \leq \tfrac{1}{2}|2f|^p + \tfrac{1}{2}|2g|^p = 2^{p-1}|f|^p + 2^{p-1}|g|^p.$$

Agora, podemos falar legitimamente sobre ($\|f+g\|_p$). Se for zero, então a desigualdade de Minkowski é válida. Agora assumimos que ($\|f+g\|_p$) não é zero. Usando a desigualdade triangular e depois a desigualdade de Hölder

$$\|f+g\|_p^p = \int_\Omega |f+g|^p \,\mathrm{d}\mu \tag{7.16}$$

$$= \int_\Omega |f+g| \cdot |f+g|^{p-1} \,\mathrm{d}\mu \tag{7.17}$$

$$\leq \int_\Omega (|f|+|g|)|f+g|^{p-1} \,\mathrm{d}\mu \tag{7.18}$$

$$= \int_\Omega |f||f+g|^{p-1} \,\mathrm{d}\mu + \int_\Omega |g||f+g|^{p-1} \,\mathrm{d}\mu \tag{7.19}$$

$$\leq \left(\left(\int_\Omega |f|^p \,\mathrm{d}\mu \right)^{\frac{1}{p}} + \left(\int_\Omega |g|^p \,\mathrm{d}\mu \right)^{\frac{1}{p}} \right) \left(\int_\Omega |f+g|^{(p-1)\left(\frac{p}{p-1}\right)} \,\mathrm{d}\mu \right)^{1-\frac{1}{p}} \tag{7.20}$$

$$\tag{7.21}$$

(Pela Desigualdade de Hölder)

$$\tag{7.22}$$

$$= \left(\|f\|_p + \|g\|_p\right) \frac{\|f+g\|_p^p}{\|f+g\|_p} \tag{7.23}$$

Obtemos a desigualdade de Minkowski multiplicando ambos os lados por

$$\frac{\|f+g\|_p}{\|f+g\|_p^p}.$$

7.8 ↯Funções Geradoras de Probabilidade

Considere uma variável aleatória X, tomando valores nos naturais. Seja $p_r = \mathbf{P}(X = r)$.

7.46 Definição (Função Geradora de Probabilidade) *A* ***função geradora de probabilidade*** *de X é definida como*

$$p(z) = \mathbf{E}[z^X] = \sum_{r=0}^{\infty} \mathbf{P}(X = r) z^r = p_0 + p_1 z + p_2 z^2 \cdots = \sum_{0}^{\infty} p_r z^r.$$

Observamos que função geradora de probabilidade é uma série de potência e converge absolutamente para $|z| \leq 1$, pois

$$|p(z)| \leq \sum_r p_r |z^r| \leq \sum_r p_r = 1.$$

Esta definição pode parecer um pouco inusitada. Mas como veremos ela resulta ser uma ferramenta algébrica útil que resume de forma concisa informações sobre a distribuição de probabilidade.

7.47 Exemplo *Considere um dado justo. Então $p_r = 1/6$ para $r = 1, \cdots, 6$. assim*

$$p(z) = \mathbf{E}[z^X] = \frac{1}{6}(z + z^2 + \cdots + z^6) = \frac{1}{6}z\left(\frac{1-z^6}{1-z}\right).$$

7.48 Exemplo *Neste exemplo, calculamos a função geradora de probabilidade para a distribuição de Poisson de parâmetro α. Com base na definição, temos:*

$$p(z) = \sum_{j=0}^{\infty} \frac{e^{-\alpha}\alpha^j}{j!} z^j \tag{7.24}$$

$$= \sum_{j=0}^{\infty} \frac{e^{-\alpha}(\alpha z)^j}{j!} \tag{7.25}$$

$$= \frac{e^{-\alpha}}{e^{-\alpha z}} \sum_{j=0}^{\infty} \frac{e^{-\alpha z}(\alpha z)^j}{j!} \tag{7.26}$$

$$= e^{\alpha(z-1)} \tag{7.27}$$

7.49 Teorema *A distribuição de X é determinada exclusivamente pela sua função geradora de probabilidade.*

Demonstração. *Diferenciando a série termo a termo temos*

$$\left.\frac{d^n}{dz^n}p(z)\right|_{z=0} = n!p_n.$$

Então podemos recuperar $(p_0, p_1, \cdots)$ de $p(z)$. □

7.50 Teorema (Lema de Abel)

$$\mathbf{E}[X] = \lim_{z\to 1} p'(z).$$

Se $p'(z)$ for contínua em 1, então $\mathbf{E}[X] = p'(1)$.

Demonstração. *Como $z < 1$, temos*

$$p'(z) = \sum_{1}^{\infty} r p_r z^{r-1} \leq \sum_{1}^{\infty} r p_r = \mathbf{E}[X].$$

Logo

$$\lim_{z\to 1} p'(z) \leq \mathbf{E}[X].$$

Por outro lado, para qualquer ε, escolhendo N suficientemente grande, temos

$$\sum_{1}^{N} rp_r \geq \mathbf{E}[\mathrm{X}] - \varepsilon.$$

assim

$$\mathbf{E}[\mathrm{X}] - \varepsilon \leq \sum_{1}^{N} rp_r = \lim_{z\to 1} \sum_{1}^{N} rp_r z^{r-1} \leq \lim_{z\to 1} \sum_{1}^{\infty} rp_r z^{r-1} = \lim_{z\to 1} p'(z).$$

Portanto, $\mathbf{E}[\mathrm{X}] \leq \lim_{z\to 1} p'(z)$. □

7.51 Teorema

$$\mathbf{E}[\mathrm{X}(\mathrm{X}-1)] = \lim_{z\to 1} p''(z).$$

Demonstração. *Exercício* □

7.52 Exemplo *Considere a distribuição de Poisson. Então*

$$p_r = \mathbf{P}(\mathrm{X} = r) = \frac{1}{r!}\lambda^r e^{-\lambda}.$$

Então

$$p(z) = \mathbf{E}[z^{\mathrm{X}}] = \sum_{0}^{\infty} z^r \frac{1}{r!}\lambda^r e^{-\lambda} = e^{\lambda z} e^{-\lambda} = e^{\lambda(z-1)}.$$

Temos

$$\mathbf{E}[\mathrm{X}] = \left.\frac{d}{dz} e^{\lambda(z-1)}\right|_{z=1} = \lambda,$$

e

$$\mathbf{E}[\mathrm{X}(\mathrm{X}-1)] = \left.\frac{d^2}{dx^2} e^{\lambda(z-1)}\right|_{z=1} = \lambda^2$$

assim

$$\mathrm{Var}(\mathrm{X}) = \mathbf{E}[\mathrm{X}^2] - \mathbf{E}[\mathrm{X}]^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

7.53 Teorema *Suponha que $\mathrm{X}_1, \mathrm{X}_2, \cdots, \mathrm{X}_n$ sejam variáveis aleatórias independentes com funções geradoras de probabilidade $p_1, p_2, \cdots, p_n$. Então, a função geradora de probabilidade de $\mathrm{X}_1 + \mathrm{X}_2 + \cdots + \mathrm{X}_n$ é $p_1(z)p_2(z)\cdots p_n(z)$.*

Demonstração.

$$\mathbf{E}[z^{\mathrm{X}_1+\cdots+\mathrm{X}_n}] = \mathbf{E}[z^{\mathrm{X}_1}\cdots z^{\mathrm{X}_n}] = \mathbf{E}[z^{\mathrm{X}_1}]\cdots\mathbf{E}[z^{\mathrm{X}_n}] = p_1(z)\cdots p_n(z).$$

7.54 Exemplo *Se* X *e* Y *forem variáveis aleatórias de Poisson independentes com parâmetros* λ, μ *respectivamente, então*

$$\mathbf{E}[t^{X+Y}] = \mathbf{E}[t^X]\mathbf{E}[t^Y] = e^{\lambda(t-1)}e^{\mu(t-1)} = e^{(\lambda+\mu)(t-1)}$$

Portanto, $X + Y \sim \mathbf{P}(\lambda + \mu)$.

Também podemos fazê-lo diretamente:

$$\mathbf{P}(X + Y = r) = \sum_{i=0}^{r} \mathbf{P}(X = i, Y = r - i) = \sum_{i=0}^{r} \mathbf{P}(X = i)\mathbf{P}(X = r - i),$$

mas é muito mais complicado. ◁

Somas de um número aleatório de termos

Uma aplicação útil das funções geradora de probabilidade é o estudo da soma de um número aleatório de termos aleatórios. Por exemplo, uma companhia de seguros pode receber um número aleatório de reivindicações, cada um exigindo uma quantia aleatória de dinheiro. Então, temos uma soma de um número aleatório de termos. Isso pode ser respondido usando funções geradoras de probabilidade.

7.55 Exemplo *Sejam* $X_1, X_2, \cdots, X_n$ *variáveis aleatórias independentes e identicamente distribuídas com função geradora de probabilidade* $p(z) = \mathbf{E}[z^X]$. *Deixe* N *ser uma variável aleatória independente de* X_i *com função geradora de probabilidade* $h(z)$. *Qual é o função geradora de probabilidade de* $S = X_1 + \cdots + X_N$*?*

$$\begin{aligned}
\mathbf{E}[z^S] &= \mathbf{E}[z^{X_1+\cdots+X_N}] \\
&= \mathbf{E}_N[\underbrace{\mathbf{E}_{X_i}[z^{X_1+\ldots+X_N} \mid N]}_{\text{assumindo fixo}N}] \\
&= \sum_{n=0}^{\infty} \mathbf{P}(N = n)\mathbf{E}[z^{X_1+X_2+\cdots+X_n}] \\
&= \sum_{n=0}^{\infty} \mathbf{P}(N = n)\mathbf{E}[z^{X_1}]\mathbf{E}[z^{X_2}]\cdots\mathbf{E}[z^{X_n}] \\
&= \sum_{n=0}^{\infty} \mathbf{P}(N = n)(\mathbf{E}[z^{X_1}])^n \\
&= \sum_{n=0}^{\infty} \mathbf{P}(N = n)p(z)^n \\
&= h(p(z))
\end{aligned}$$

como $h(x) = \sum_{n=0}^{\infty} \mathbf{P}(N = n)x^n$.

assim

$$\begin{aligned}\mathbf{E}[S] &= \frac{d}{dz}h(p(z))\Big|_{z=1} \\ &= h'(p(1))p'(1) \\ &= \mathbf{E}[N]\mathbf{E}[X_1]\end{aligned}$$

Para calcular a variância, use o fato de que

$$\mathbf{E}[S(S-1)] = \frac{d^2}{dz^2}h(p(z))\Big|_{z=1}.$$

Então podemos encontrar que

$$\mathrm{Var}(S) = \mathbf{E}[N]\mathrm{Var}(X_1) + \mathbf{E}[X_1^2]\mathrm{Var}(N).$$

7.9 ⚡Processos de Ramificação

Originalmente, processos de ramificações foram considerados por Galton e Watson nos anos de 1870 quando estes procuravam um modelo quantitativo para o fenômeno do desaparecimento de sobrenomes, mesmo no cenário de uma população crescente. O modelo é construído sob a suposição de que cada homem em uma dada família tem a probabilidade p_k de ter k filhos e que o sobrenome de família é passado aos filhos, então desejamos determinar a probabilidade que após n gerações um indivíduo não tenha descendentes homens.

Vamos fazer perguntas como o número esperado de indivíduos em uma geração específica e a probabilidade de extinção.

Considere $X_0, X_1, \cdots$, onde X_n é o número de indivíduos na geração n-ésima. Assumimos o seguinte:

1. $X_0 = 1$

2. Cada indivíduo vive por uma unidade de tempo e produz k descendentes com probabilidade p_k.

3. Suponha que todos os descendentes se comportem de forma independente. Então

$$X_{n+1} = Y_1^n + Y_2^n + \cdots + Y_{X_n}^n,$$

onde Y_i^n são variáveis aleatórias independentes e identicamente distribuídas, como X_1.

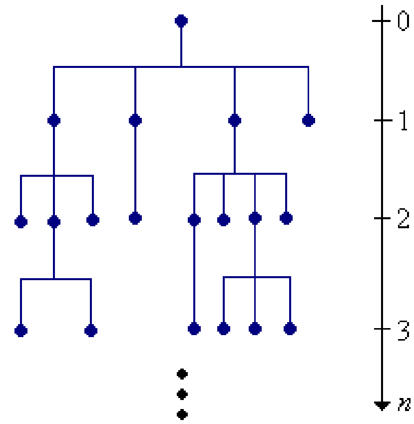


Figura 7.1: Processo de ramificação [14]

Como veremos é útil considerar o função geradora de probabilidade de um processo de ramificação. Seja $F(z)$ a função geradora de probabilidade de Y_i^n. Então

$$F(z) = \mathbf{E}[z^{Y_i^n}] = \mathbf{E}[z^{X_1}] = \sum_{k=0}^{\infty} p_k z^k.$$

Definimos

$$F_n(z) = \mathbf{E}[z^{X_n}].$$

O principal teorema dos processos de ramificação que provaremos nesta seção é

7.56 Teorema

$$F_{n+1}(z) = F_n(F(z)) = F(F(F(\cdots F(z)\cdots)))) = F(F_n(z)).$$

Demonstração.

$$\begin{aligned}
F_{n+1}(z) &= \mathbf{E}[z^{X_{n+1}}] \\
&= \mathbf{E}[\mathbf{E}[z^{X_{n+1}} \mid X_n]] \\
&= \sum_{k=0}^{\infty} \mathbf{P}(X_n = k)\mathbf{E}[z^{X_{n+1}} \mid X_n = k] \\
&= \sum_{k=0}^{\infty} \mathbf{P}(X_n = k)\mathbf{E}[z^{Y_1^n+\cdots+Y_k^n} \mid X_n = k] \\
&= \sum_{k=0}^{\infty} \mathbf{P}(X_n = k)\mathbf{E}[z^{Y_1}]\mathbf{E}[z^{Y_2}]\cdots\mathbf{E}[z^{Y_n}] \\
&= \sum_{k=0}^{\infty} \mathbf{P}(X_n = k)(\mathbf{E}[z^{X_1}])^k \\
&= \sum_{k=0} \mathbf{P}(X_n = k)F(z)^k \\
&= F_n(F(z))
\end{aligned}$$

7.57 Teorema *Suponha que*

$$\mathbf{E}[X_1] = \sum kp_k = \mu$$

e

$$\mathrm{Var}(X_1) = \mathbf{E}[(X-\mu)^2] = \sum (k-\mu)^2 p_k < \infty.$$

Então

$$\mathbf{E}[X_n] = \mu^n, \quad \mathrm{Var}X_n = \sigma^2\mu^{n-1}(1+\mu+\mu^2+\cdots+\mu^{n-1}).$$

Demonstração.

$$\begin{aligned}
\mathbf{E}[X_n] &= \mathbf{E}[\mathbf{E}[X_n \mid X_{n-1}]] \\
&= \mathbf{E}[\mu X_{n-1}] \\
&= \mu\mathbf{E}[X_{n-1}]
\end{aligned}$$

por indução, $\mathbf{E}[X_n] = \mu^n$ *(pois* $X_0 = 1$*).*

Para calcular a variância, observe que

$$\mathrm{Var}(X_n) = \mathbf{E}[X_n^2] - (\mathbf{E}[X_n])^2$$

e, portanto

$$\mathbf{E}[X_n^2] = \mathrm{Var}(X_n) + (\mathbf{E}[X])^2$$

Então, calculamos

$$\begin{aligned}
\mathbf{E}[X_n^2] &= \mathbf{E}[\mathbf{E}[X_n^2 \mid X_{n-1}]] \\
&= \mathbf{E}[\mathrm{Var}(X_n) + (\mathbf{E}[X_n])^2 \mid X_{n-1}] \\
&= \mathbf{E}[X_{n-1}\mathrm{Var}(X_1) + (\mu X_{n-1})^2] \\
&= \mathbf{E}[X_{n-1}\sigma^2 + (\mu X_{n-1})^2] \\
&= \sigma^2\mu^{n-1} + \mu^2\mathbf{E}[X_{n-1}^2].
\end{aligned}$$

assim

$$\begin{aligned}
\mathrm{Var}X_n &= \mathbf{E}[X_n^2] - (\mathbf{E}[X_n])^2 \\
&= \mu^2\mathbf{E}[X_{n-1}^2] + \sigma^2\mu^{n-1} - \mu^2(\mathbf{E}[X_{n-1}])^2 \\
&= \mu^2(\mathbf{E}[X_{n-1}^2] - \mathbf{E}[X_{n-1}]^2) + \sigma^2\mu^{n-1} \\
&= \mu^2\mathrm{Var}(X_{n-1}) + \sigma^2\mu^{n-1} \\
&= \mu^4\mathrm{Var}(X_{n-2}) + \sigma^2(\mu^{n-1} + \mu^n) \\
&= \cdots \\
&= \mu^{2(n-1)}\mathrm{Var}(X_1) + \sigma^2(\mu^{n-1} + \mu^n + \cdots + \mu^{2n-3}) \\
&= \sigma^2\mu^{n-1}(1 + \mu + \cdots + \mu^{n-1}).
\end{aligned}$$

Probabilidade de Extinção

Considere A_n o evento $X_n = 0$, ou seja, a extinção ocorreu na n-ésima geração. Sejam q ser a probabilidade da extinção eventualmente ocorrer e

$$A = \bigcup_{n=1}^{\infty} A_n = [\text{a extinção ocorre eventualmente}].$$

Como $A_1 \subseteq A_2 \subseteq A_3 \subseteq \cdots$, sabemos que

$$q = \mathbf{P}(A) = \lim_{n\to\infty} \mathbf{P}(A_n) = \lim_{n\to\infty} \mathbf{P}(X_n = 0).$$

Mas

$$\mathbf{P}(X_n = 0) = F_n(0),$$

Como $F_n(0) = \sum \mathbf{P}(X_n = k)z^k$. Logo

$$F(q) = F\left(\lim_{n\to\infty} F_n(0)\right) = \lim_{n\to\infty} F(F_n(0)) = \lim_{n\to\infty} F_{n+1}(0) = q.$$

assim

$$F(q) = q.$$

Alternativamente, usando a lei da probabilidade total

$$q = \sum_k \mathbf{P}(X_1 = k)\mathbf{P}(\text{extinção} \mid X_1 = k) = \sum_k p_k q^k = F(q),$$

onde a segunda igualdade vem do fato de que, para que toda a população se extingua, cada população individual deve se extinguir.

Isso significa que para encontrar a probabilidade de extinção, precisamos encontrar um ponto fixo de F.

7.58 Teorema *A probabilidade de extinção q é a menor raiz da equação $q = F(q)$. Escreva $\mu = \mathbf{E}[X_1]$. Então, se $\mu \leq 1$, então $q = 1$; se $\mu > 1$, então $q < 1$.*

Demonstração. *Para mostrar que é a menor raiz, seja α a menor raiz. Então note que $0 \leq \alpha \Rightarrow F(0) \leq F(\alpha) = \alpha$. Portanto, $F(F(0)) \leq \alpha$. Continuando indutivamente, $F_n(0) \leq \alpha$ para todo n. Assim*

$$q = \lim_{n\to\infty} F_n(0) \leq \alpha.$$

Então $q = \alpha$.

Para mostrar que $q = 1$ quando $\mu \leq 1$, mostramos que $q = 1$ é a única raiz. Sabemos que $F'(z), F''(z) \geq 0$ para $z \in (0, 1)$. Então F é crescente e convexa. Como $F'(1) = \mu \leq 1$, Então $z = 1$ é a única raiz. □

7.10 ↯Funções Geradoras de Momento

7.59 Definição *A **função geradora de momento** $\phi_X(t)$ é definida como*

$$\phi_X(t) = \mathbf{E}\left[e^{tX}\right]$$

desde que $\mathbb{E}(e^{tX})$ exista no intervalo $(-h,h)$ para $h > 0$.

Para variáveis discretas ou absolutamente contínuas temos que

$$\phi_X(t) = \mathbf{E}\left[e^{tX}\right] = \begin{cases} \sum_i e^{tx_i}p(x_i) & \text{se X é discreto} \\ \int_x e^{tx} f(x)\,dx & \text{se X for contínua} \end{cases}$$

7.60 Exemplo *A **distribuição degenerada de probabilidade** é a distribuição associada a uma variável aleatória discreta exibindo certeza do resultado. Se X for um variável aleatório degenerada, então $X = c$ com probabilidade 1. A função geradora de momento da variável aleatória degenerada é particularmente simples:*

$$\sum_{x_i = c} e^{x_i t} = e^{ct}.$$

7.61 TEOREMA *Seja* X *uma variável aleatória que possua função geradora de momento. Então*

$$M_X^{(n)}(t) = \frac{d^n}{dt^n}\mathbb{E}\left(e^{tX}\right) = \mathbb{E}\left(\frac{d^n}{dt^n}e^{tX}\right) = \mathbb{E}\left(X^n e^{tX}\right).$$

7.62 TEOREMA *Se* X *e* Y *são variáveis aleatórias independentes com função geradora de momento* $\phi_X(t)$ *e* $\phi_Y(t)$ *respectivamente, então* $\phi_{X+Y}(t)$*, a função geradora de momento* $X+Y$ *é dada por* $\phi_X(t)\phi_Y(t)$*.*

Demonstração.

$$\begin{aligned}\phi_{X+Y}(t) &= \mathbf{E}\left[e^{t(X+Y)}\right]\\ &= \mathbf{E}\left[e^{tX}e^{tY}\right]\\ &= \mathbf{E}\left[e^{tX}\right]\mathbf{E}\left[e^{tY}\right]\\ &= \phi_X(t)\phi_Y(t).\end{aligned}$$

7.63 TEOREMA *A função geradora de momento determina de forma única a distribuição de probabilidade.*

7.64 TEOREMA (A FUNÇÃO GERADORA DE MOMENTO DA VARIÁVEL ALEATÓRIA NORMAL) *Se* $Z \sim N(\mu, \sigma^2)$*, então* $\phi_Z(t) = \exp(\mu t + \sigma^2 t^2/2)$*.*

Demonstração.

$$\begin{aligned}\phi_Z(t) &= \mathbf{E}\left[e^{tX}\right]\\ &= \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty} e^{tx}e^{-(x-\mu)^2/(2\sigma^2)}\,dx\\ &= \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty}\exp\left(\frac{-(x^2-2\mu x+\mu^2-2\sigma^2 tx)}{2\sigma^2}\right)dx\end{aligned}$$

Completando o quadrado:

$$\begin{aligned}x^2-2\mu x+\mu^2-2\sigma^2 tx &= x^2-2(\mu+\sigma^2 t)x+\mu^2\\ &= (x-(\mu+\sigma^2 t))^2-(\mu+\sigma^2 t)^2+\mu^2\\ &= (x-(\mu+\sigma^2 t))^2-\sigma^4t^2-2\mu\sigma^2 t.\end{aligned}$$

Então voltando ao cálculo da função geradora de momento

$$\begin{aligned}\phi_Z(t) &= \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty}\exp\left(\frac{-\left((x-(\mu+\sigma^2 t))^2-\sigma^4t^2-2\mu\sigma^2 t\right)}{2\sigma^2}\right)dx\\ &= \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(\frac{\sigma^4t^2+2\mu\sigma^2 t}{2\sigma^2}\right)\int_{-\infty}^{\infty}\exp\left(\frac{-(x-(\mu+\sigma^2 t))^2}{2\sigma^2}\right)dx\\ &= \exp\left(\frac{\sigma^4t^2+2\mu\sigma^2 t}{2\sigma^2}\right)\\ &= \exp\left(\mu t+\sigma^2t^2/2\right)\end{aligned}$$

7.65 Teorema *Se $Z_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$, e $Z_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$ e Z_1 e Z_2 são independentes, então $Z_1 + Z_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$. Ou seja, a soma de variáveis aleatórias normais independentes é uma variável aleatória normal cuja média é a soma dos médias e cuja variância é a soma das variações.*

Demonstração. *Calculamos a função geradora de momento da soma usando o teorema sobre somas de variáveis aleatórias independentes.*

$$\begin{aligned}
\phi_{Z_1+Z_2}(t) &= \phi_{Z_1}(t)\phi_{Z_2}(t) \\
&= \exp(\mu_1 t + \sigma_1^2 t^2/2)\exp(\mu_2 t + \sigma_2^2 t^2/2) \\
&= \exp((\mu_1 + \mu_2)t + (\sigma_1^2 + \sigma_2^2)t^2/2)
\end{aligned}$$

7.11 ↯Entropia

Informação

Gostaríamos de desenvolver uma medida da informação que obtemos de observar a ocorrência de um evento de probabilidade p. Nossa primeira redução será ignorar quaisquer características específicas do evento e apenas observar se ou não aconteceu. Assim, pensamos no evento como o observação de um símbolo cuja probabilidade de ocorrência é p.

A abordagem que vamos adotar é axiomática: Queremos que a medida de informação $I(p)$ tenha as seguintes propriedades

7.66 Definição (Informação)

1. *A informação é uma quantidade não negativa: $I(p) \geq 0$.*

2. *Se um evento tiver probabilidade 1, não obtemos informações da ocorrência do evento: $I(1) = 0$.*

3. *Se ocorrerem dois eventos independentes (cuja probabilidade conjunta é o produto de suas probabilidades individuais), então a informação que obtemos ao observar os eventos são a soma das duas informações: $I(p_1 \cdot p_2) = I(p_1) + I(p_2)$. (Esta é a propriedade fundamental)*

4. *Queremos que nossa medida informação seja contínua função da probabilidade*

Podemos, portanto, mostrar o seguinte:

1. $I(p^2) = I(p \cdot p) = I(p) + I(p) = 2 \cdot I(p)$

$\boxed{2}$ Assim, além disso, $I(p^n) = n \cdot I(p)$
Epor indução

$\boxed{3}$ $I(p) = I((p^{1/m})^m) = m \cdot I(p^{1/m})$, então $I(p^{1/m}) = \frac{1}{m} \cdot I(P)$ e, portanto, em geral

$$I(p^{n/m}) = \frac{n}{m} \cdot I(p)$$

$\boxed{4}$ E, portanto, pela continuidade, obtemos, por $0 < p \leq 1$ e $a > 0$:

$$I(p^a) = a \cdot I(p)$$

Logo

$$I(p) = -\log_b(p) = \log_b(1/p)$$

por alguma base b.

Resumindo: a partir das quatro propriedades,

$\boxed{1}$ $I(p) \geq 0$

$\boxed{2}$ $I(p_1 \cdot p_2) = I(p_1) + I(p_2)$

$\boxed{3}$ $I(p)$ é monótona e contínua em p

$\boxed{4}$ $I(1) = 0$

podemos derivar que

$$I(p) = \log_b(1/p) = -\log_b(p),$$

para alguma constante positiva b. A base b determina o sistema de unidades que estão utilizando.

Normalmente, pensamos em termos de $\log_2(p)$.

7.67 Exemplo *Por exemplo, lançar uma moeda justa uma vez nos dará eventos H e T com probabilidade* $1/2$ *e, portanto, um único lançamento de uma moeda nos dá* $-\log_2(1/2) = 1$ *bit de informação.*

Lançando uma moeda justa n vezes temos $-\log_2((1/2)^n) = \log_2(2^n) = n \cdot \log_2(2) = n$ *bits de informação.*

Assim, obtemos que n lançamentos de uma moeda justa nos fornece n bits de informação, e leva n dígitos binários para especificar. Que estas duas medidas são as mesmas nos assegura que fizemos uma boa escolha de definição de informação. ◁

Entropia

Suponha agora que temos n símbolos $\{a_1, a_2, \ldots, a_n\}$ e que alguma fonte está nos fornecendo um fluxo desses símbolos. Suponha ainda que a fonte emite os símbolos com probabilidades $\{p_1, p_2, \ldots, p_n\}$, respectivamente. Por enquanto, também assumiremos que os símbolos são emitidos de forma independente

Qual é a quantidade média de informação que obtemos de cada símbolo que vemos no fluxo?

O que realmente queremos aqui é uma média ponderada. Se observarmos o símbolo a_i, nós obteremos $\log(1/p_i)$ informação dessa observação particular. A longo prazo, digamos N de observações, veremos aproximadamente $N \cdot p_i$ ocorrências do símbolo a_i. Assim, após N observações obteremos a informação total I de

$$I = \sum_{i=1}^{n} (N \cdot p_i) \cdot \log(1/p_i).$$

Mas, então, a média de informações que obtemos por símbolo observado será

$$\begin{aligned} I/N &= (1/N) \sum_{i=1}^{n} (N \cdot p_i) \cdot \log(1/p_i) \\ &= \sum_{i=1}^{n} p_i \cdot \log(1/p_i) \end{aligned}$$

Acima adotamos a convenção que defina $0 \cdot \log(1/0)$ é 0.

Isso nos leva a seguinte definição.

7.68 Definição *Seja uma distribuição de probabilidade $P = \{p_1, p_2, \ldots, p_n\}$. Definimos* ***entropia*** *da distribuição P por:*

$$H(P) = \sum_{i=1}^{n} p_i \cdot \log(1/p_i).$$

A definição anterior possui uma generalização óbvia, se tivermos uma distribuição contínua de probabilidade $P(x)$ definimos:

$$H(P) = \int P(x) \cdot \log(1/P(x))dx.$$

Em termos de valor esperado.

$$H(P) = \mathbf{E}[I(p)].$$

Em outras palavras, a entropia de uma distribuição de probabilidade é o valor esperado da informação da distribuição.

7.69 Teorema (Desigualdade de Gibbs) *Considere as distribuições de probabilidade,* $P = \{p_1, p_2, \ldots, p_n\}$ *e* $Q = \{q_1, q_2, \ldots, q_n\}$*, onde* $p_i, q_i \geq 0$ *e* $\sum_i p_i = \sum_i q_i = 1$*. Então*

$$-\sum p_i \log q_i \geq -\sum p_i \log p_i$$

com igualdade somente quando $p_i = q_i$ *para todos* i*.*

Demonstração. *A demonstração baseia-se na desigualdade* $\log x \leq x - 1$.

$$\begin{aligned}
\sum_{i=1}^{n} p_i \ln\left(\frac{q_i}{p_i}\right) &\leq \sum_{i=1}^{n} p_i \left(\frac{q_i}{p_i} - 1\right) = \sum_{i=1}^{n} (q_i - p_i) \\
&= \sum_{i=1}^{n} q_i - \sum_{i=1}^{n} p_i = 1 - 1 = 0,
\end{aligned}$$

Podemos usar a desigualdade de Gibbs para encontrar a distribuição de probabilidade que maximiza a função entropia. Suponha que $P = \{p_1, p_2, \ldots, p_n\}$ é um distribuição de probabilidade. Nós temos

$$\begin{aligned}
H(P) - \log(n) &= \sum_{i=1}^{n} p_i \log(1/p_i) - \log(n) \\
&= \sum_{i=1}^{n} p_i \log(1/p_i) - \log(n) \sum_{i=1}^{n} p_i \\
&= \sum_{i=1}^{n} p_i \log(1/p_i) - \sum_{i=1}^{n} p_i \log(n) \\
&= \sum_{i=1}^{n} p_i (\log(1/p_i) - \log(n)) \\
&= \sum_{i=1}^{n} p_i (\log(1/p_i) + \log(1/n)) \\
&= \sum_{i=1}^{n} p_i \log\left(\frac{1/n}{p_i}\right) \\
&\leq 0,
\end{aligned}$$

com igualdade somente quando $p_i = \frac{1}{n}$ para todos i.

Exercícios

Ex. 7.1 — Mostre que para variáveis aleatórias discretas,

$$\mathbf{E}X = \sum_{k=1}^{\infty} x_k \mathbf{P}(x = x_k), \tag{7.28}$$

se a série é absolutamente convergente

Ex. 7.2 — Dê exemplos de variáveis aleatórias X e Y definidas em [0,1] com a medida de Lebesque tal que $\mathbf{P}(X > Y) > 1/2$, mas $\mathbf{E}(X) < \mathbf{E}(Y)$.

Ex. 7.3 — Mostre que

$$\mathrm{X} \stackrel{\mathrm{d.}}{=} \mathrm{Y} \text{ implica } \mathrm{X}^+ \stackrel{\mathrm{d.}}{=} \mathrm{Y}^+ \text{ e } \mathrm{X}^- \stackrel{\mathrm{d.}}{=} \mathrm{Y}^-.$$

Ex. 7.4 — Suponha que $X_1, X_2, \ldots$ é uma sequência de variáveis independentes em (Ω, F, P). Mostre que as duas famílias $X_1, X_3, X_5, \ldots$ e $X_2, X_4, X_6, \ldots$ são independentes.

Ex. 7.5 — Deixe $X_1, X_2, \ldots$ serem v.a i.i.d. com média μ e variância σ^2 e deixe ser uma variável aleatória tomando valores nos inteiros com média m e variância v, com N independente de todas as X_i. Deixe $S = X_1 + ... + X_N = \sum_{i=1}^{\infty} X_i \mathbf{1}_{N>i}$.
Calcule $\mathbf{Var}(S)$.

Ex. 7.6 — Prove a Desigualdade de Cauchy-Schwarz: dadas duas variáveis aleatórias X e Y, temos

$$|\mathbf{E}XY| \leq \sqrt{\mathbf{E}[X^2]\mathbf{E}[Y^2]}, \tag{7.29}$$

e a igualdade ocorre se e somente se $X = \alpha Y$, para alguma constante $\alpha \in \mathbb{R}$.

Ex. 7.7 — Dados X e Z variáveis aleatórias independentes, cada uma seguindo a distribuição normal padrão, Deixe $a, b \in \mathbb{R}$ (não ambos nulos), e deixe

$$Y = aX + bZ$$

.

1. Calcule Corr(X,Y).
2. Mostre que $|\text{Corr}(X, Y)| \leq 1$ nesse caso.
3. Dê condições necessárias e suficientes sobre os valores de a e b para que $\text{Corr}(X, Y) = 1$.
4. Dê condições necessárias e suficientes sobre os valores de a e b tais que para que $\text{Corr}(X, Y) = -1$.

Ex. 7.8 — **Problema do Pareamento** Suponha que n cavalheiros saem para jantar e deixem seus chapéus no vestiário. Após o jantar (e vários copos de vinho) eles escolhem seus chapéus completamente aleatoriamente. Denote por X o número de senhores que tomam seus próprios chapéus. Encontre $\mathbf{E}[X]$ e $\mathbf{Var}[X]$. (esse problema já apareceu numa forma ligeiramente diferente na Lista 2)

Ex. 7.9 — Se $\mathbf{E}|X_1| = \infty$ então $\sum_{n=0}^{\infty} \mathbf{P}(|X_1| > n)$ diverge.

Ex. 7.10 — Dada uma variável aleatória X, então, para todo $\epsilon > 0$, existe uma variável aleatória limitada X_ϵ, tal que $\mathbf{P}(X \neq X_\epsilon) < \epsilon$

Ex. 7.11 — Coletor de Cupom
Cada vez que se compra um saco de salgadinho, obtém-se como um bônus uma figurinha (escondida dentro da embalagem) de um jogador de futebol. Suponha que existam n imagens diferentes que são igualmente susceptíveis de estar dentro de cada pacote.
Encontre o número esperado de pacotes para comprar para obter uma coleção completa de jogadores.

Ex. 7.12 — Se

$$P_X(s) := \mathbb{E}(s^X) = \sum_{i \geq 0} \mathbb{P}(X = i)s^i.$$

Mostre que

$$\mathrm{Var}(X) = P''(1) + P'(1) - P'(1)^2.$$

Desigualdades

Ex. 7.13 — Uma moeda honesta é lançada de forma independente n vezes. Seja S_n o número de caras obtidas nesses n lançamentos. Use a desigualdade de Chebyshev para provar que

$$\lim_{n \to \infty} \mathbb{P}(|\frac{S_n}{n} - \frac{1}{2}| < \epsilon) = 1$$

para todo $\epsilon > 0$.

Ex. 7.14 — Demonstração Probabilística do Teorema de Weierstrass
Utilize a desigualdade de Chebyshev para mostrar que para toda função continua $f : [0,1] \to \mathbb{R}$,

$$\sum_{k=0}^{n} f\left(\frac{k}{n}\right)\binom{n}{k}x^k(1-x)^k \to f(x)$$

uniformemente em $x \in [0,1]$ quando $n \to \infty$.
Dica: Sejam $X_1, X_2, \ldots, X_n$ variáveis aleatórias independentes cada uma assumindo os valores 0 e 1 com probabilidade p e $1-p$ respectivamente. Seja $S_n = X_1 + X_2 + \ldots + X_n$ o número de caras em n lançamentos. Defina o polinômio $r_n(p) = \mathbb{E}[f(\frac{S_n}{n})]$ e estude a expressão $|r_n(p) - f(p)|$.

Ex. 7.15 — Sharpness da desigualdade de Chebyshev

Para cada $t \geq 1$, construa uma variável aleatória X com média μ e variância σ^2, e tal que a desigualdade de Chebyshev se torna uma igualdade:

$$\mathbf{P}(|X-\mu| \geq t\sigma) = \frac{1}{t^2}.$$

Ex. 7.16 — Deixe X ser uma variável aleatória não-negativa, tal que $\mathbf{E}X$ existe.

1. Mostre através de um exemplo que $\mathbf{E}(X^2)$ pode não existir.
2. Considere o truncamento $X_n = \min(X, n)$. Prove que para todo $p > 2$,

$$\sum_{n=1}^{\infty} n^{-p}\, \mathbf{E}(X_n^2) < \infty.$$

3. Prove a propriedade anterior para $p = 2$.

Ex. 7.17 — Deixe $X_1, \ldots X_n$ serem variáveis aleatórias reais independentes e deixe $S_k = X_1 + \cdots + X_k$ para $k = 1, \ldots, n$. Mostre que para $t > 0$ a desigualdade de Etemadi é válida:

$$\mathbf{P}\left[\max_k |S_k| \geq \right] t \leq 3 \max_k \mathbf{P}\left[|Sk| \geq t/3\right].$$

Dica: Considere os conjuntos

$$A_j := \left\{\max_{1\leq k<j} |S_k| < 3r, |S_j| \geq 3r\right\}, \qquad j = 1, \ldots, n$$

e observe que

$$\left\{\max_{1\leq j\leq n} |S_j| \geq 3r\right\} = \bigcup_{j=1}^{n} A_j.$$

CAPÍTULO 8

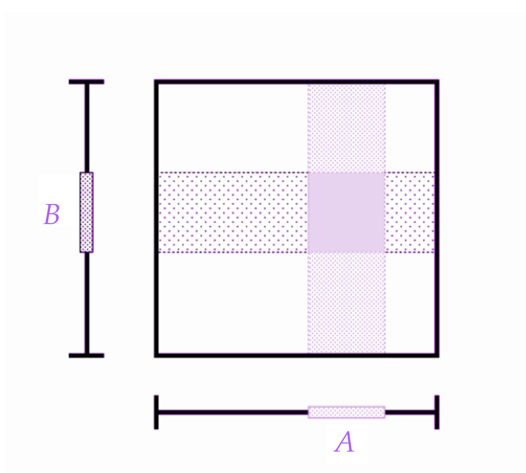
Medida Produto

Nesse capítulo generalizamos o conceito de medidas para o produto de espaços. A integral que obtemos como subproduto dessas medidas são os análogos abstratos das integrais de várias variáveis e como tais podem ser calculadas como integrais iteradas.

8.1 Definição *Dados $(\Omega_1, \mathcal{F}_1)$ e $(\Omega_2, \mathcal{F}_2,)$ espaços mensuráveis, definimos a σ-álgebra $\mathcal{F}_1 \otimes \mathcal{F}_2$ como a σ-álgebra gerada pela coleção dos retângulos mensuráveis*

$$\mathcal{F}_1 \otimes \mathcal{F}_2 = \sigma \langle \{A \times B : A \in \mathcal{F}_1, B \in \mathcal{F}_2\} \rangle$$

Figura 8.1: Retângulos mensuráveis em $\mathcal{F}_1 \otimes \mathcal{F}_2$



Nesse caso também diremos que o produto dos espaços mensuráveis $(\Omega_1, \mathcal{F}_1)$ e $(\Omega_2, \mathcal{F}_2)$ é o espaço mensurável $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2)$.

Antes de prosseguirmos, observamos que podemos realizar a construção da σ-álgebra produto de um modo ligeiramente diferente. Para isso note que o produto cartesiano $\Omega_1 \times \Omega_2$ de dois conjuntos Ω_1 e Ω_2 é caracterizado pelas aplicações de projeção associadas

$$\pi_{\Omega_1} : \Omega_1 \times \Omega_2 \to \Omega_1 \quad \pi_{\Omega_2} : \Omega_1 \times \Omega_2 \to \Omega_2.$$

Podemos fazer uso dessas aplicações para construir a σ-álgebra produto. Dados dois espaços mensuráveis $(\Omega_1, \mathcal{F}_1)$ e $(\Omega_2, \mathcal{F}_2)$, podemos formar uma coleção de subconjuntos em $\Omega_1 \times \Omega_2$ puxando de $\mathcal{F}_1$:

$$\pi^*_{\Omega_1}(\mathcal{F}_1) := \{\pi^{-1}_{\Omega_1}(E) : E \in \mathcal{F}_1\} = \{E \times \Omega_2 : E \in \mathcal{F}_1\}.$$

Agora, é fácil verificar que $\pi^*_{\Omega_1}(\mathcal{F}_1)$ é uma σ-álgebra.

Podemos fazer a mesma construção para a projeção no segundo espaço.

8.2 Teorema (Produto Finito de σ-Álgebras) *O produto das σ-álgebra $\mathcal{F}_1 \otimes \mathcal{F}_2$ é igual a σ-álgebra gerada pela união do pullback das σ-álgebras*

$$\mathcal{F}_1 \otimes \mathcal{F}_2 = \langle \pi^*_{\Omega_1}(\mathcal{F}_1) \cup \pi^*_{\Omega_2}(\mathcal{F}_2) \rangle.$$

Demonstração. *A demonstração é simples e será deixada como exercício ao leitor.* □

Suponha que $E \subset \Omega_1 \times \Omega_2$. Dado $x \in \Omega_1$ e $y \in \Omega_2$, definimos a x-seção $E^x \subset \Omega_2$ e a y-seção $E^y \subset \Omega_1$ de E como

$$E^y = \{x \in \Omega_1 : (x, y) \in E\}$$

$$E^x = \{y \in \Omega_2 : (x, y) \in E\}$$

Conforme indicado na próxima proposição, todas as seções de um conjunto mensurável são mensuráveis.

8.3 Proposição *Se $(\Omega_1, \mathcal{F}_1)$ e $(\Omega_2, \mathcal{F}_2)$ são espaços mensuráveis e $E \in \mathcal{F}_1 \otimes \mathcal{F}_2$, então $E^x \in \mathcal{F}_2$ para todo $x \in \Omega_1$ e $E^y \in \mathcal{F}_1$ para todo $y \in \Omega_2$.*

Demonstração. *Provaremos apenas para a seção E^y; a demonstração para a seção E^x é análoga.*

Considere a família $\mathcal{M} = \{E \in \mathcal{F}_1 \otimes \mathcal{F}_2 : E^y \in \mathcal{F}_1\}$. Mostraremos que $\mathcal{M}$ é uma σ-álgebra contendo os retângulos mensuráveis e logo deve ser $\mathcal{F}_1 \otimes \mathcal{F}_2$, e isso é suficiente para demonstrar a proposição.

- *Suponha que $E = A \times B$ é um retângulo mensurável. Então $E^y = A$ quando $y \in \mathcal{F}_2$, caso contrário $E^y = \emptyset$. Em ambos os casos $E^y \in \mathcal{F}_1$, logo $E \in \mathcal{M}$.*

- $\emptyset^y = \{x \in \Omega_1 : (x,y) \in \emptyset\} = \emptyset \in \mathcal{F}_1$, *assim* $\mathcal{M}$ *contém* $\emptyset$.

- *Se* $E_1, E_2, \ldots \in \mathcal{M}$, *então* $(\bigcup E_n)^y = \bigcup E_n^y \in \mathcal{F}_1$. *Logo* $\mathcal{M}$ *é fechado em relação a união enumerável.*

- *Se* $E \in \mathcal{M}$, *e* $F = E^{\complement}$, *então* $F^y = \{x \in \Omega_1 : (x,y) \notin E\} = \Omega_1 \setminus E^y \in \mathcal{F}_1$. *Logo* $\mathcal{M}$ *é fechada sob complementação.* □

8.4 Proposição *Suponha que* $\mathbb{R}^m$, $\mathbb{R}^n$ *estejam equipados com suas* σ*-álgebras Borel* $\mathcal{B}(\mathbb{R}^m), \mathcal{B}(\mathbb{R}^n)$ *e seja* $\mathbb{R}^{m+n} = \mathbb{R}^m \times \mathbb{R}^n$. *Então*

$$\mathcal{B}(\mathbb{R}^{m+n}) = \mathcal{B}(\mathbb{R}^m) \otimes \mathcal{B}(\mathbb{R}^n).$$

Suponha que $(\Omega_1, \mathcal{F}_1)$ e $(\Omega_2, \mathcal{F}_2)$ sejam espaços mensuráveis. A interseção de retângulos mensuráveis é um retângulo mensurável

$$(A \times B) \cap (C \times D) = (A \cap C) \times (B \cap D),$$

e o complemento de um retângulo mensurável é uma união finita de retângulos

$$(A \times B)^c = (A^c \times B) \cup (A \times B^c) \cup (A^c \times B^c).$$

Assim, a família de uniões finitas de retângulos mensuráveis em $\Omega_1 \times \Omega_2$ forma uma álgebra, que denotamos por $\mathcal{F}_0$. Esta álgebra não é, em geral, uma σ-álgebra, mas obviamente gera a mesma σ-álgebra produto que os retângulos mensuráveis.

8.5 Teorema (Existência da Medida Produto) *Dados* $(\Omega_1,\mathcal{F}_1,\mu_1)$ *e* $(\Omega_2,\mathcal{F}_2,\mu_2)$ *espaços de medidas. Então existe uma única medida* μ *em* $\mathcal{F} = \mathcal{F}_1 \otimes \mathcal{F}_2$ *satisfazendo*

$$\mu(A \times B) = \mu_1(A) \times \mu_2(B).$$

Diremos que μ *é a medida produto de* μ_1 *e* μ_2. *E será denotada por* $\mu_1 \otimes \mu_2$.

Demonstração. *Seja* $\mathcal{F}_0$ *a álgebra dos retângulos, i.e.,*

$$\mathcal{F}_0 = f\,\langle \{A \times B : A \in \mathcal{F}_1, B \in \mathcal{F}_2\} \rangle$$

Como $\mathcal{F}_0$ *é uma álgebra, e* $\mathcal{F} = \sigma\langle \mathcal{F}_0 \rangle$, *basta mostrar que se* $A \times B = \bigcup_{i=1}^{\infty}(A_i \times B_i)$ *é uma união disjunta, então*

$$\mu(A \times B) = \sum_{i=1}^{\infty} \mu(A_i \times B_i)$$

Observamos que a função indicadora tem a seguinte propriedade

$$\mathbb{1}_{A\times B}(x,y)=\sum_{i=1}^{\infty}\mathbb{1}_{A_i\times B_i}(x,y)=\sum_{i=1}^{\infty}\mathbb{1}_{A_i}(x)\mathbb{1}_{B_i}(y)$$

Integrando sobre y para um $x\in\Omega_1$ fixo e usando o Teorema da Convergência Dominada, temos

$$\mathbb{1}_A\mu_2(B)=\sum_{i=1}^{\infty}\mathbb{1}_{A_i}(x)\mu_2(B_i).$$

Integrando com respeito a x, temos

$$\mu_1(A)\mu_2(B)=\sum_{i=1}^{\infty}\mu_1(A_i)\mu_2(B_i).$$

Pelo Teorema de Extensão de Carathéodory temos que existe uma única medida definida na σ-álgebra $\mathcal{F}_1\otimes\mathcal{F}_2$. □

8.6 Teorema (Mensurabilidade de Funções com Variável Fixa) *Sejam $(\Omega_1\times\Omega_2,\mathcal{F}_1\otimes\mathcal{F}_2)$ e $(Z,\mathcal{M})$ espaços mensuráveis.*

Se $f:\Omega_1\times\Omega_2\to Z$ é mensurável, então as funções $f^y:\Omega_1\to Z$, $f^x:\Omega_2\to Z$ obtidas fixando uma variável também são mensuráveis.

Demonstração. *Provaremos apenas para f^y. Seja $I^y:\Omega_1\to\Omega_1\times\Omega_2$ a aplicação de inclusão $I^y(x)=(x,y)$. Dado $E\in\mathcal{F}_1\otimes\mathcal{F}_2$, por 8.3 $(I^y)^{-1}(E)=E^y\in\mathcal{F}_1$, logo I^y é uma função mensurável. Mas $f^y=f\circ I^y$.* □

8.1 Áreas de Seções Transversais

8.7 Teorema (Mensurabilidade da Função Área de Seção Transversal) *Sejam $(\Omega_1,\mathcal{F}_1,\mu)$, $(\Omega_2,\mathcal{F}_2,\nu)$ espaços de medida. Se $E\in\mathcal{F}_1\otimes\mathcal{F}_2$,*

- *$\nu(E^x)$ é uma função mensurável de $x\in\Omega_1$.*
- *$\mu(E^y)$ é uma função mensurável de $y\in\Omega_2$.*

Demonstração. *Provaremos para $\mu(E^y)$. Também assumiremos que $\mu(\Omega_1)<\infty$. Seja*

$$\mathcal{M}=\{E\in\mathcal{F}_1\otimes\mathcal{F}_2:\mu(E^y)\text{ é uma função mensurável de }y\}.$$

$\mathcal{M}$ é igual à $\mathcal{F}_1\otimes\mathcal{F}_2$, pois:

- *Se $E = A \times B$ é um retângulo mensurável, então $\mu(E^y) = \mu(A)\mathbb{1}_{y\in B}$ que é uma função mensurável de y. Se E é a união finita disjunta de retângulos mensuráveis E_n, então $\mu(E^y) = \sum_n \mu(E_n^y)$ que também mensurável.*

  Então $\mathcal{M}$ contém a álgebra das uniões finita disjuntas de retângulos mensuráveis.

- *Se E_n são conjuntos crescentes em $\mathcal{M}$ então $\mu\big((\bigcup_n E_n)^y\big) = \mu(\bigcup_n E_n^y) = \lim_{n\to\infty} \mu(E_n^y)$ é mensurável, e logo $\bigcup_n E_n \in \mathcal{M}$.*

  De modo análogo, se E_n são conjuntos decrescentes in $\mathcal{M}$, então $\bigcap_n E_n \in \mathcal{M}$. (Aqui é crucial que E_n^y tenham medida finita.)

- *Assim $\mathcal{M}$ é a classe monótona, contendo a álgebra das uniões finita de retângulos mensuráveis. Pelo Teorema da Classe Monótona (1.29), $\mathcal{M} = \mathcal{F}_1 \otimes \mathcal{F}_2$.*

Para demonstrar o caso de medida infinita seja $\Omega_n \uparrow \Omega$ com $\mu(\Omega_n) < \infty$, e reaplique o argumento trocando μ por uma medida finita $\mu_n(F) = \mu(F \cap \Omega_n)$. Então $\mu(E^y) = \lim_{n\to\infty} \mu_n(E^y)$ □

8.2 Integrais Iteradas

Até agora, construímos a medida produto que atribuí a um retângulo mensurável a medida que é o produto das medidas de cada um de seus lados. Essa medida foi obtida pelo processo de extensão de Carathéodory, agora apresentaremos uma fórmula integral explícita:

8.8 Teorema (Medida produto como integral de seções transversais) *Sejam $(\Omega_1, \mathcal{F}_1, \mu)$ e $(\Omega_2, \mathcal{F}_2, \nu)$ espaços de medida. Então existe uma única medida produto $\mu \otimes \nu \colon \mathcal{F}_1 \otimes \mathcal{F}_2 \to [0,\infty]$, tal que*

$$(\mu \otimes \nu)(E) = \int_{x\in\Omega_1} \underbrace{\int_{y\in\Omega_2} \mathbb{1}_E \, d\nu}_{\nu(E^x)} \; d\mu = \int_{y\in\Omega_2} \underbrace{\int_{x\in\Omega_1} \mathbb{1}_E \, d\mu}_{\mu(E^y)} \; d\nu \,.$$

Demonstração. *Denote por $\lambda_1(E)$ a integral dupla à esquerda, e por $\lambda_2(E)$ à integral a direita. Essa integrais existem pelo Teorema 8.7. As medidas λ_1 e λ_2 são σ-aditivas pelo Teorema da Convergência Monótona logo ambas são medidas em $\mathcal{F}_1 \otimes \mathcal{F}_2$. Mais ainda se $E = A \times B$, então expandindo as duas integrais temos $\lambda_1(E) = \mu(A)\nu(B) = \lambda_2(E)$.*

E pelo Teorema de Unicidade das Medidas de Probabilidade 2.13 temos que $\lambda_1 = \lambda_2 = \mu \otimes \nu$ para todo $\mathcal{F}_1 \otimes \mathcal{F}_2$. □

8.9 Teorema (Fubini) *Sejam $(\Omega_1, \mathcal{F}_1, \mu)$ e $(\Omega_2, \mathcal{F}_2, \nu)$ espaços de medida. Se $\mathrm{X} \colon \Omega_1 \times \Omega_2 \to \mathbb{R}$ é $\mu \otimes \nu$-integrável, então*

$$\int_{\Omega_1\times\Omega_2} \mathrm{X} \, d(\mu\otimes\nu) = \int_{x\in\Omega_1} \left[\int_{y\in\Omega_2} \mathrm{X}(x,y) \, d\nu \right] d\mu = \int_{y\in\Omega_2} \left[\int_{x\in\Omega_1} \mathrm{X}(x,y) \, d\mu \right] d\nu \,.$$

Demonstração. *Observamos primeiramente que se* $X = \mathbb{1}_E$ *então esse resultado é apenas 8.8.*

Uma vez que as três integrais são aditivas, e como elas são iguais para as funções simples não negativas X*, portanto, também o são para todas as funções* X *não negativas, por aproximação e convergência monótona.*

Para uma função X *não necessariamente não negativa, use a decomposição* $X = X^+ - X^-$ *como de costume e aplique linearidade. Agora pode acontecer que* $\int_{y\in\Omega_2} X^{\pm}(x,y)d\nu$ *possa ser* ∞ *para algum* $x \in \Omega_1$ *e, portanto,* $\int_{y\in\Omega_2} X(x,y)d\nu$ *não estaria definida. No entanto, se* X *é* $\mu\otimes\nu$*-integrável, isso pode acontecer apenas em um conjunto de medida nula em* Ω_1 *(veja o próximo teorema). A integral terá sentido desde que ignoremos este conjunto de medida nula.* □

8.10 Teorema (Tonelli) *Sejam* $(\Omega_1, \mathcal{F}_1, \mu)$ *e* $(\Omega_2, \mathcal{F}_2, \nu)$ *espaços de medida, e* $f : \Omega_1 \times \Omega_2 \to \mathbb{R}$ *uma função* $\mu\otimes\nu$*-mensurável. Então* f *é* $\mu\otimes\nu$*-integrável se e somente se*

$$\int_{x\in\Omega_1}\left[\int_{y\in\Omega_2} |f(x,y)|\ \mathrm{d}\nu\right]\mathrm{d}\mu < \infty$$

Demonstração. *Segue de imediato do Teorema de Fubini aplicado à função não negativa* $|f|$. □

Resumidamente, se uma integral dupla é absolutamente convergente, então é válido mudar a ordem de integração.

8.11 Exemplo *Se* X *não é não-negativa, a hipótese de convergência absoluta é crucial para mudar a ordem de integração.*

Um contra-exemplo elementar é a sequência duplamente indexada $a_{m,n} = (-m)^n/n!$ *Integrado em relação a medida de contagem temos* $\sum_n a_{m,n} = e^{-m}$*, então* $\sum_m \sum_n a_{m,n} = (1-e^{-1})^{-1}$*, mas* $\sum_m a_{m,n}$ *diverge para todo* n.

8.3 Distribuição Conjunta

Lembramos a definição de distribuição (conjunta) para vetores aleatórios:

8.12 Definição (Distribuição de Vetores Aleatórios) *Se* $X = (X_1, \ldots, X_n)$ *for uma vetor aleatório, então* X *induz uma medida de probabilidade* **P** *em* $\mathbb{R}^n$ *definida tomando a pré-imagem para cada conjunto de Borel* A :

$$\mathbf{P}(A) = \mathbf{P}(X^{-1}(A)) = \mathbf{P}(X \in A). \tag{8.1}$$

Esta medida de probabilidade **P** *é denominada* ***distribuição*** *de* X.

8.13 Definição (Função de Distribuição Conjunta) *Se* $X = (X_1, \ldots, X_n)$ *for uma vetor aleatório, então a função de distribuição conjunta de* $(X_1, \ldots, X_n)$

$$F_{(X_1,\ldots,X_n)}(x_1, \ldots, x_n) = \mathbf{P}(X \leq x_i, \ldots, X_n \leq x_n). \tag{8.2}$$

A partir da definição da independência vemos que a função de distribuição conjunta fatora como o produto de funções de distribuição.

$$F_{(X_1,\ldots,X_n)}(x_1, \ldots, x_n) = F_{X_1}(x_1) \times \cdots \times F_{X_n}(x_n),$$

8.14 Teorema *As variáveis aleatórias* $X_1, \ldots, X_n$ *são independentes se e somente se a distribuição conjunta* $\mathbf{P}$ *em* $\mathbb{R}^n$ *é o produto das distribuições* $\mathbf{P}_k$ *de* X_k *em* $\mathbb{R}$.

Demonstração. *Na direção* $\Longrightarrow$*, pelo Teorema de unicidade de medidas, basta verificar que* $\mathbf{P}(A) = (\mathbf{P}_1 \otimes \cdots \otimes \mathbf{P}_n)(A)$ *para* $A \in \mathcal{A}$*. Como sabemos, cada* $A \in \mathcal{A}$ *tem a forma* $A = A_1 \times \cdots A_n$*, onde* A_k *é Borel. Então, pela independência,*

$$\begin{aligned}
\mathbf{P}(A_1 \times \cdots \times A_n) &= \mathbf{P}\left((X_1, \ldots, X_n) \in A_1 \times \cdots \times A_n\right) && (8.3)\\
&= \mathbf{P}(X_1 \in A_1, \ldots, X_n \in A_n) && (8.4)\\
&= \mathbf{P}(X_1 \in A_1) \times \cdots \times \mathbf{P}(X_n \in A_n) && (8.5)\\
&= \mathbf{P}_1(A_1) \times \cdots \times \mathbf{P}_n(A_n). && (8.6)
\end{aligned}$$

Na outra direção $(\Longleftarrow)$*, para verificar a independência de* $X_1, \ldots, X_n$*, é suficiente considerarmos os semi-intervalos. Ou seja, se* $\mathbf{P}(X_1 \leq x_1, \ldots, X_n \leq x_n)$ *fatorar, o teorema estará demonstrado. Mas este último termo é (pela definição de distribuição conjunta)*

$$\begin{aligned}
\mathbf{P}((-\infty, x_1] \times \cdots \times (-\infty, x_n]) &= \mathbf{P}_1(-\infty, x_1] \times \cdots \times \mathbf{P}_n(-\infty, x_n] && (8.7)\\
&= \mathbf{P}(X_1 \leq x_1) \times \cdots \times \mathbf{P}(X_n \leq x_n). && (8.8)
\end{aligned}$$

E assim terminamos. □

Assim, o modo de variáveis aleatórias serem independentes é quando elas são definidas no espaço produto.

8.4 Produtos Infinitos: Teorema de Extensão de Kolmogorov I

Espaços produto são de importância central em diversas áreas da matemática, e na teoria da probabilidade não é diferente. Da construção de variáveis aleatórias independentes ao estudo formal de processos estocásticos, nos deparamos constantemente com necessidade de entender medidas de probabilidade em espaços produto.

Nessa seção nos concentraremos na construção de medidas no espaço $\mathbb{R}^{\mathbb{N}}$, onde todas as ideias centrais são apresentadas. Posteriormente, na Seção 8.8, o argumento que apresentaremos será generalizado para produtos mais gerais.

Para provar o Teorema de extensão de Kolmogorov, seguiremos os passos de Bochner[10, 2], e para isso, antes de entrarmos diretamente nos espaço produto, passaremos rapidamente por algumas definições e resultados que facilitarão nosso trabalho.

Considere então um conjunto Ω e uma sequência crescente $\{\mathcal{F}_n\}_{n=1}^{\infty}$ de σ-álgebras em Ω. E para cada n, seja μ_n uma medida de probabilidade na σ-álgebra $\mathcal{F}_n$.

8.15 Definição *Diremos que a sequência $\{(\mathcal{F}_n, \mu_n)\}_{n=1}^{\infty}$ é* ***Kolmogorov consistente*** *se sempre que $m \leq n$ temos $\mathcal{F}_m \subset \mathcal{F}_n$ e $\mu_n|_{\mathcal{F}_m} = \mu_m$.*

Dado $\mathcal{A} = \bigcup_{n=1}^{\infty} \mathcal{F}_n$, uma **extensão** da sequência de probabilidades $\{\mu_n\}_{n=1}^{\infty}$ é uma medida μ em $\sigma\langle\mathcal{A}\rangle$ satisfazendo $\mu_n = \mu|_{\mathcal{F}_n}$.

8.16 Teorema (Teorema de Extensão de Bochner) *Dado $\{(\mathcal{F}_n, \mu_n)\}_{n=1}^{\infty}$ uma sequência Kolmogorov consistente. Suponha que existe uma classe compacta $\mathcal{K} \subseteq \bigcup_{n=1}^{\infty} \mathcal{F}_n$ de Ω tal que*

$$\mu_n(A) = \sup\{\mu_n(C) : C \in \mathcal{K} \cap \mathcal{F}_n \text{ e } C \subset A\}.$$

Então existe uma única extensão de Kolmogorov para a σ-álgebra $\sigma\langle\bigcup_{n=1}^{\infty} \mathcal{F}_n\rangle$.

Demonstração.

Defina μ na álgebra $\mathcal{A} = \bigcup_{n=1}^{\infty} \mathcal{F}_n$ por $\mu(E) = \mu_n(E)$ para $E \in \mathcal{F}_n$. A condição de consistência de Kolmogorov garante que μ está bem definida. Também é simples ver que $\mu(\emptyset) = 0$ e $\mu(\Omega) = 1$.

A demonstração que μ é uma probabilidade finitamente aditiva é direta. Basta observar que se $A_i \in \mathcal{A}$ para $i = 1, \ldots, n$ então existe m tal que $A_i \in \mathcal{F}_m$ para todo $i = 1, \ldots, n$.

Para concluir a demonstração, observamos que se $A \in \mathcal{A}$, então existe n tal que $A \in \mathcal{F}_n$ e

$$\begin{aligned}
\mu(A) &= \mu_n(A) \\
&= \sup\{\mu_n(C) : C \in \mathcal{K} \cap \mathcal{F}_n \text{ e } C \subset A\} \\
&\leq \sup\{\mu_n(C) : C \in \mathcal{K} \text{ e } C \subset A\} \\
&\leq \mu(A).
\end{aligned}$$

Assim, pelo Teorema de Extensão Compacta (2.28) temos que μ é pré-medida e finalmente pelo Teorema de Extensão de Carathéodory podemos estender a uma probabilidade em $\sigma\langle\mathcal{A}\rangle$. □

Postas estas considerações iniciais, podemos nos concentrar no objeto central dessa seção, ou seja, medidas de probabilidade em

$$\mathbb{R}^{\mathbb{N}} = \{(\omega_1, \omega_2, \dots) : \omega_i \in \mathbb{R}, i \in \mathbb{N}\}.$$

O primeiro passo é eleger uma σ-álgebra onde definiremos nossa medida, e para isso

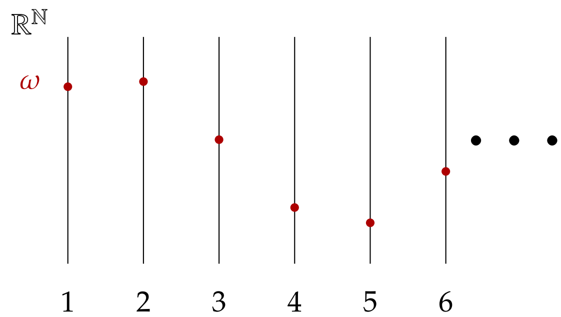


Figura 8.2: Representação visual de $\mathbb{R}^{\mathbb{N}}$.

temos um candidato natural. Nós equiparemos $\mathbb{R}^{\mathbb{N}}$ com a σ-álgebra produto gerada pelos retângulos finito dimensionais

8.17 Definição *A σ-álgebra produto gerada pelos retângulos finito dimensionais é dada por*

$$\mathcal{C} = \{A : A = \{\omega : \omega_i \in (a_i, b_i] \text{ para } i = 1, \dots, k \quad k \in \mathbb{N}\}\},$$

onde $-\infty \leq a_i < b_i \leq \infty$.

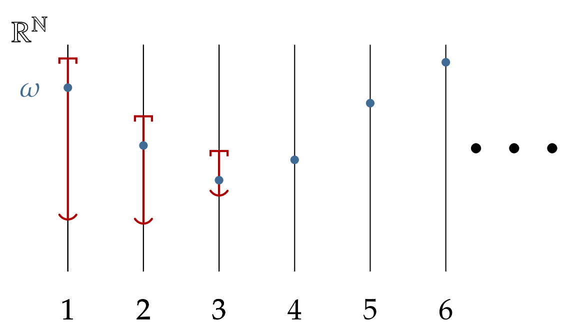


Figura 8.3: Representação visual de um cilindro em $\mathbb{R}^{\mathbb{N}}$. Em azul um ponto do cilindro.

8.18 Observação *Esta é, de fato, a σ-álgebra com a qual trabalhamos intuitivamente nos cursos mais básicos de probabilidade. Lembre, por exemplo, que quando queremos estudar o "lançamento de infinitas moedas", sempre descrevemos os eventos de interesse a partir de eventos que dependem do* *resultado de finitos lançamentos. Mesmo eventos do tipo $A = \{$observamos cara infinitas vezes$\}$ é escrito como*

$$A = \bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n,$$

com $A_n = \{$observamos cara na n-ésima jogada$\}$.

Dando sequência, vamos começar a preparar o caminho para usar o teorema 8.16 e introduziremos uma notação particularmente útil. Dado $B \subset \mathbb{R}^n$, denotaremos por $B \times \mathbb{R}^{\mathbb{N}-n}$ o conjunto:

$$B \times \mathbb{R}^{\mathbb{N}-n} := \{(\omega_1, \omega_2, \dots) \text{ tais que } (\omega_1, \dots, \omega_n) \in B\}$$

É fácil ver que se $A \in \mathcal{C}$ então existe $n \in \mathbb{N}$, e $B \in \mathcal{B}(\mathbb{R}^n)$ tal que $A = B \times \mathbb{R}^{\mathbb{N}-n}$.

Considere agora a sequência encaixante de σ-álgebras:

$$\mathcal{F}_n = \mathcal{B}(\mathbb{R}^n) \times \mathbb{R}^{\mathbb{N}-n} = \{A \times \mathbb{R}^{\mathbb{N}-n}, A \in \mathcal{B}(\mathbb{R}^n), \}$$

e note que dado $m < n$ e $B \in \mathcal{B}(\mathbb{R}^n)$, então

$$B \times \mathbb{R}^{\mathbb{N}-m} = (B \times \mathbb{R}^{n-m}) \times \mathbb{R}^{\mathbb{N}-n} \in \mathcal{F}_n,$$

de modo que $\mathcal{F}_m \subset \mathcal{F}_n$, e $\mathcal{A} = \bigcup_{n=1}^{\infty} \mathcal{F}_n$ é uma álgebra, com $\mathcal{C} \subseteq \mathcal{A}$.

8.19 Proposição $\sigma\langle\mathcal{A}\rangle = \sigma\langle\mathcal{C}\rangle$.

Demonstração. *Uma das inclusões é trivial. De fato, uma vez que $\mathcal{C} \subset \mathcal{A}$, segue trivialmente que $\sigma\langle\mathcal{C}\rangle \subset \sigma\langle\mathcal{A}\rangle$. Para ver a inclusão contrária, dado $n \geq 1$, faça $\mathcal{G}_n = \{A \in B(\mathbb{R}^n) : A \times \mathbb{R}^{\mathbb{N}-n} \in \sigma\langle\mathcal{C}\rangle\}$. Como $\mathcal{G}_n$ é uma σ-álgebra contendo os retângulos do tipo*

$$\Pi_{k=1}^{n}(a_k, b_k] \subset \mathbb{R}^n,$$

então $\mathcal{G}_n = \mathcal{B}(\mathbb{R}^n)$. Segue que $\mathcal{F}_n \subseteq \sigma\langle\mathcal{C}\rangle$ para todo $n \geq 1$ e

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

8.20 Teorema (Teorema de Extensão de Kolmogorov) *Dadas medidas de probabilidade $\mathbf{P}_n$ em $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ tais que*

$$\mathbf{P}_{n+1}((a_1, b_1] \times \cdots \times (a_n, b_n] \times \mathbb{R}) = \mathbf{P}_n((a_1, b_1] \times \cdots \times (a_n, b_n]),$$

para todo $n \geq 1$, então existe uma única medida de probabilidade $\mathbf{P}$ em $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^N))$ com

$$\mathbf{P}(\omega : \omega_i \in (a_i, b_i], 1 \leq i \leq n) = \mathbf{P}_n((a_1, b_1] \times \cdots \times (a_n, b_n])$$

Demonstração. *Para começar, dado $n \geq 1$ defina a medida de probabilidade Q_n em $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ por*

$$Q_n(A) = \mathbf{P}_{n+1}(A \times \mathbb{R}).$$

Como Q_n coincide com $\mathbf{P}_n$ no π-sistema dos retângulos do tipo $\Pi_{k=1}^n (a_k, b_k]$, então Q_n e $\mathbf{P}_n$ coincidem em todo $\mathcal{B}(\mathbb{R}^n)$. Segue que para todo $n \geq 1$ e $A \in \mathcal{B}(\mathbb{R}^n)$,

$$\mathbf{P}_{n+1}(A \times \mathbb{R}) = \mathbf{P}_n(A),$$

e por indução

$$\mathbf{P}_{n+m}(A \times \mathbb{R}^m) = \mathbf{P}_n(A).$$

Com a notação introduzida anteriormente, defina a medida μ_n em $(\mathbb{R}^{\mathbb{N}}, \mathcal{F}_n)$ por

$$\mu_n(B \times \mathbb{R}^{\mathbb{N}-n}) = \mathbf{P}_n(B).$$

Assim se $m < n$ então para $B \in \mathcal{F}_m$,

$$\mu_n(B \times \mathbb{R}^{\mathbb{N}-m}) = \mu_n(B \times \mathbb{R}^{n-m} \times \mathbb{R}^{\mathbb{N}-n}) = \mathbf{P}_n(B \times \mathbb{R}^{n-m}) = \mathbf{P}_m(B) = \mu_m(B \times \mathbb{R}^{\mathbb{N}-m}).$$

Concluímos assim que a sequência $\{(\mathcal{F}_n, \mu_n)\}_{n\geq 1}$ é Kolmogorov consistente.

Considere agora as classes de cilindros

$$\mathcal{K}_n = \{K \times \mathbb{R}^{\mathbb{N}-n}, K \in \mathcal{B}(\mathbb{R}^n) \text{ compacto }\} \quad e \quad \mathcal{K} = \bigcup_{n\geq 1} \mathcal{K}_n.$$

Observe que se $K \in \mathcal{K}_n$ e $L \in \mathcal{K}_m$, com $m < n$ então $K \cap L \in \mathcal{K}_n$.

Afirmamos que classe $\mathcal{K}$ é compacta, o que provaremos no lema a seguir.

Dado agora um conjunto $A \in \mathcal{C}$ então $A = D \times \mathbb{R}^{\mathbb{N}-n}$ com $D \in \mathcal{B}(\mathbb{R}^n)$, para algum $n \geq 1$, e logo existe um $K \subset D \in \mathcal{B}(\mathbb{R}^n)$ compacto, tal que $\mu_n(K \times \mathbb{R}^{\mathbb{N}-n}) = \mathbf{P}_n(K) < \mathbf{P}_n(D) + \epsilon = \mu_n(A) + \epsilon$. Consequentemente

$$\mu_n(A) = \sup\{\mu_n(C) : C \in \mathcal{K} \cap \mathcal{F}_n \text{ e } C \subset A\}.$$

e o teorema segue do Teorema 8.16.

Dado $A \in \mathcal{A}$ não vazio, definimos $\dim(A) := \min\{n \geq 1 : A \in \mathcal{F}_n\}$, e $\dim(\mathbb{R}^{\mathbb{N}}) = 0$.

Observe que se $A_1, A_2 \in \mathcal{A}$ são tais que $A_1 \cap A_2 \neq \emptyset$, então $\dim(A_1 \cap A_2) = \max\{\dim(A_1), \dim(A_2)\}$. De fato, se $m < n$

$$(B_1 \times \mathbb{R}^{\mathbb{N}-m}) \cap (B_2 \times \mathbb{R}^{\mathbb{N}-n}) = ((B_1 \times \mathbb{R}^{n-m}) \cap B_2) \times \mathbb{R}^{\mathbb{N}-n} \in \mathcal{F}_n.$$

8.21 Lema *A classe $\mathcal{K}$ definida na demonstração anterior como*

$$\mathcal{K}_n = \{K \times \mathbb{R}^{\mathbb{N}-n}, K \in \mathcal{B}(\mathbb{R}^n) \text{ compacto }\} \quad e \quad \mathcal{K} = \bigcup_{n\geq 1} \mathcal{K}_n.$$

é compacta. $\triangleleft$

Demonstração. *Para ver isso considere uma sequência de conjuntos* $\{C_i\}_{n=1}^{\infty} \in \mathcal{K}$ *e suponha que para todo* $m \geq 1$, $B_m = \bigcap_{i=1}^{m} C_i \neq \emptyset$. *Assim, para todo* $m \geq 1$ *e* $j \geq 1$, *temos que a projeção na* j*-ésima coordenada é não vazia, i.e.,* $\pi_j(B_m) \neq \emptyset$.

Temos agora dois casos: $\sup\{\dim(B_m); m \geq 1\} = d < \infty$ *ou* $\sup\{\dim(B_m); m \geq 1\} = \infty$.

No primeiro caso, como a sequência $(\dim(B_m))_{n\geq 1}$ *é não-decrescente, então existe* $m_0 \geq 1$ *tal que* $\dim(B_m) = d$ *para todo* $m \geq m_0$. *Assim, para todo* $m \geq m_0$ *existe* $K_m \subset \mathbb{R}^d$, *compacto de* $\mathbb{R}^d$ *tal que* $B_m = K_m \times R^{\mathbb{N}-d}$. *Vale assim que*

$$\bigcap_{m=m_0}^{\infty} K_m \neq \emptyset, \quad e \quad \bigcap_{i=1}^{\infty} C_i = \bigcap_{m=m_0}^{\infty} B_m = \left(\bigcap_{m=m_0}^{\infty} K_m\right) \times \mathbb{R}^{\mathbb{N}-d} \neq \emptyset.$$

Agora, se $\sup\{\dim(B_m); m \geq 1\} = \infty$, *então para todo* $j \geq 1$ *existe* $m_j \geq 1$ *tal que* $\dim(B_m) \geq j$, *para todo* $m \geq m_j$. *Em outras palavras, se* $m \geq m_j$, *então* $B_m = K_m \times \mathbb{R}^{\mathbb{N}-d}$ *para algum compacto* K_m *de* $\mathbb{R}^d$ *e algum* $d \geq j$. *Mas isso significa que* $\pi_j(B_m)$ *é um compacto de* $\mathbb{R}$ *para todo* $m \geq m_j$. *Além disso, como* $B_m \neq \emptyset$, *então* $\pi_j(B_m) \neq \emptyset$ *e*

$$\bigcap_{m=m_j}^{\infty} \pi_j(B_m) \neq \emptyset.$$

Como as projeções π_j *são contínuas e a sequência* $(B_m)_{m\geq 1}$ *é decrescente, então para todo* $j \geq 1$

$$\pi_j\left(\bigcap_{i=1}^{\infty} C_i\right) = \pi_j\left(\bigcap_{m=m_j}^{\infty} B_m\right) = \bigcap_{m=m_j}^{\infty} \pi_j(B_m) \neq \emptyset.$$

Segue, portanto que $\bigcap_{i=1}^{\infty} C_i \neq \emptyset$ *e* $\mathcal{K}$ *é uma classe compacta.*

8.5 Existência de Variáveis Aleatórias Independentes

8.22 Teorema *Dadas probabilidades* $\mathbf{P}_1, \ldots, \mathbf{P}_n$ *em* $\mathbb{R}$, *então existem variáveis aleatórias independentes* $X_1, \ldots, X_n$ *com distribuições* $\mathbf{P}_1, \ldots, \mathbf{P}_n$.

Demonstração. *Seja* $\mathbf{P} = \mathbf{P}_1 \otimes \cdots \otimes \mathbf{P}_n$ *a medida de probabilidade produto em* $\mathbb{R}^n$. *Como já provamos, toda medida de probabilidade em* $\mathbb{R}^n$ *é uma distribuição de alguma variável aleatória, veja Teorema 6.15.*

Vamos definir uma variável aleatória (ou melhor, um vetor aleatório). Começamos definindo o espaço de probabilidade $(\Omega, \mathcal{F}, \mathbf{P})$ *como* $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mathbf{P})$.

Então, agora, precisamos definir uma variável aleatória X *que é uma função desse conjunto. Seja* $X(x) = x$. *Como* X *é a função identidade, então* $X \in B$ *se e somente se* $x \in B$. *Então* $\mathbf{P}(X \in B) = \mathbf{P}(B)$, *então a distribuição de* X *é igual a* $\mathbf{P}$. *Considere* $X = (X_1, \ldots, X_n)$ *então as funções componentes* $X_1, \ldots, X_n$ *são independentes, pelo teorema anterior.* □

Podemos generalizar o resultado anterior para sequências de variáveis aleatórias i.i.d.

8.23 Teorema (Sequências de Variáveis i.i.d.) *Seja F uma distribuição em $\mathbb{R}$, então existe uma sequência de variáveis aleatórias independentes $X_1, \ldots, X_n, \ldots$ com distribuição F.*

Demonstração. *Exercício. Repita o argumento do teorema anterior usando o Teorema de Extensão de Kolmogorov.* □

Agora vamos demonstrar alguns fatos que mostram como a teoria desenvolvida até esse ponto reflete em densidades de variáveis aleatórias.

1 Uma função mensurável positiva $f \geq 0$ é a densidade de uma variável aleatória X se $\mathbf{P}(X \in A) = \int_A f(x) \quad dx$, para cada conjunto de Borel $A \subset \mathbb{R}$. Claro, a densidade é definida apenas q.c.

2 A definição é a mesma para um vetor aleatório: se X for um vetor aleatório e $\mathbf{f}$ está definido em $\mathbb{R}^n$ e A é um Boreliano de $\mathbb{R}^n$.

3 Densidade conjunta de n variáveis aleatórias $f_1, \ldots, f_n$ é a densidade do vetor aleatório $(X_1, \ldots, X_n)$.

8.24 Teorema *Sejam $X_1, \ldots, X_n$ variáveis aleatórias com densidades $f_1, \ldots, f_n$ e densidade conjunta $\mathbf{f}$. Então $X_1, \ldots, X_n$ são independentes se e somente se $\mathbf{f}(x_1, \ldots, x_n) = f_1(x_1) \cdot \cdots \cdot f_n(x_n)$ para quase todo $(x_1, \ldots, x_n)$ em $\mathbb{R}^n$.*

Existem algumas dificuldades técnicas. Para obter uma densidade de uma distribuição, precisamos diferenciar. Se temos funções que não são contínuas, então a diferenciação é um problema. Precisamos do seguinte resultado da teoria das medidas:

8.25 Teorema (Teorema de Diferenciação de Lebesgue) *Seja $f : \mathbb{R} \to \mathbb{R}$ uma função Lebesgue integrável. Então*

$$\frac{1}{\epsilon} \int_x^{x+\epsilon} X(y)\, dy$$

vai convergir para $X(x)$, para quase todo $x \in \mathbb{R}$.

Esta é uma versão unidimensional do Teorema de diferenciação de Lebesgue. Você também possui a versão n-dimensional.

8.26 Teorema *Se $f : \mathbb{R}^n \to \mathbb{R}$ é Lebesgue integrável, então*

$$\frac{1}{\mathrm{vol}(Q)} \int_Q f(y)\, dy \to X(x)$$

para quase todos os x, onde Q é um cubo contendo x.

Demonstração (Prova do Teorema 8.24). *Sem perda de generalidade, podemos assumir $n = 2$.*

X tem densidade f, Y tem densidade g e (X,Y) tem densidade φ. Isso significa que $P(X \in A) = \int_A f(x)\ \mathrm{d}x$, $P(Y \in B) = \int_B g(y)\, dy$ e $P((X,Y) \in A \times B) = \int_{A\times B} \varphi(x,y)\ \mathrm{d}x\ \mathrm{d}y$

Então, vamos começar a demonstrar a implicação. Então supomos que X e Y são independentes. Então, $\int_{A\times B} \varphi(x,y)\ \mathrm{d}x\ \mathrm{d}y$ deve ser igual a, por definição, o produto das probabilidades individuais $\int_A f(x)\ \mathrm{d}x \times \int_B g(y)\, dy$. Faça $A = [x_0, x_0 + \epsilon]$ e $B = [y_0, y_0 + \epsilon]$. Então $A \times B$ é exatamente um cubo Q que contém (x_0, y_0).

Fazendo $\epsilon \longrightarrow 0$, obtemos que o lado esquerdo será $\phi(x_0, y_0)$ e o lado direito será $f(x_0)g(y_0)$ para quase todos $(x_0 y_0)$.

Para demonstrar a recíproca. Assumiremos que a densidade conjunta é o produto das duas densidades para quase todo $(x,y) \in \mathbb{R}^2$. Agora, usaremos o Teorema de Fubini.

$$P(X \in A, Y \in B) = \int_{A\times B} \varphi(x,y)\ \mathrm{d}x\ \mathrm{d}y \tag{8.9}$$
$$= \int_{A\times B} f(x)g(y)\,\mathrm{d}x\,\mathrm{d}y \tag{8.10}$$
$$= \int_A f(x)\,\mathrm{d}x \int_B g(y)\,\mathrm{d}y, \text{ por Fubini} \tag{8.11}$$
$$= P(X \in A) \cdot P(Y \in B). \tag{8.12}$$

Então, por definição, X e Y são independentes. □

Vejamos alguns exemplos.

8.27 Exemplo (Distribuição Uniforme) *Sejam X e Y variáveis aleatórias independentes, distribuídas uniformemente em $[0,1]$. Assim, a densidade de X e Y é dada pela função de densidade*

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

Portanto, a densidade conjunta de X e Y é

$$\varphi(x,y) = \begin{cases} 1, & \text{se } (x,y) \in [0,1] \times [0,1] \\ 0, & \text{caso contrário.} \end{cases}$$

8.28 Exemplo (Distribuição Gaussiana em $\mathbb{R}^n$) *Também denominada de distribuição normal multivariada. A densidade de cada coordenada X_k é dada pela gaussiana padrão unidimensional*

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, x \in \mathbb{R}$$

e as coordenadas são independentes, então a densidade de $(X_1, \ldots, X_n)$ *é*

$$\varphi(x_1, \ldots, x_n) = \frac{1}{(2\pi)^{1/2}} e^{-\frac{1}{2}|x|^2},$$

onde $|x| = x_1^2 + x_2^2 + \cdots + x_n^2$.

Faremos agora uma demonstração alternativa da afirmação de que para variáveis independentes a esperança do produto é o produto das esperanças. Compare essa demonstração com a demonstração do Teorema 7.21.

8.29 Teorema *Sejam* $X, Y \in \mathcal{L}^1(\mathbf{P})$ *independentes. Então* $(XY) \in \mathcal{L}^1(\mathbf{P})$ *e*

$$\mathbf{E}[XY] = \mathbf{E}[X] \times \mathbf{E}[Y].$$

Demonstração. *Sejam* $\mathbf{P}_X$ *e* $\mathbf{P}_Y$ *as distribuições de* X *e* Y *respectivamente. Então* $\mathbf{P}_X$ *e* $\mathbf{P}_Y$ *são medidas em* $\mathbb{R}$. *Então*

$$\mathbf{E}[X] = \int x \, d\mathbf{P}_X$$

e

$$\mathbf{E}[Y] = \int y \, d\mathbf{P}_Y,$$

e, por outro lado

$$\mathbf{E}[XY] = \int xy \; \mathrm{d}(\mathbf{P}_X \times \mathbf{P}_Y)$$

uma vez que X *e* Y *são variáveis aleatórias independentes. Por Fubini, podemos escrever a última integral como uma integral iterada, que é exatamente* $\mathbf{E}[X] \times \mathbf{E}[Y]$.

Porquê podemos usar Fubini? ◻

8.6 Convolução e Soma de Variáveis Aleatórias Independentes

Vamos começar com um exemplo. Se X e Y são variáveis aleatórias com distribuições conhecidas, qual é a distribuição de X+Y é determinada? Certamente a expectativa é determinada (pela propriedade da esperança da soma). Mas a distribuição não é completamente determinada. Considere o seguinte exemplo: tome qualquer variável aleatória X, tal que $X \stackrel{d.}{=} -X$. Um exemplo de tal variável é o lançamento de uma moeda, onde estaremos denotando cara por 1 e coroa por −1.

Considere agora $X + X$ e $X + (-X)$. Apesar dos quatro somandos anteriores possuírem a mesma distribuição a primeira soma é igual a 2X e a segunda é igual a 0.

Ou seja, conhecer a distribuição dos somandos não é suficiente. Você realmente precisa da distribuição conjunta. Se X e Y forem independentes, então podemos calcular a distribuição da soma.

Se X e Y forem independentes podemos então calcular a distribuição de X + Y.

8.30 Teorema *Se X e Y são variáveis aleatórias independentes com funções de distribuição F e G, respectivamente. Então, a função de distribuição da soma X + Y é dada por*

$$H(z) = \int_{\mathbb{R}} F(z-y)dG(y) = \int_{\mathbb{R}} G(z-y)dF(y).$$

Demonstração. *A distribuição conjunta de X e Y é $\mathbf{P}_X \otimes \mathbf{P}_X$, uma vez que as variáveis aleatórias são independentes. Então $H(z) = \mathbf{P}(X + Y \leq z)$ é igual a $\mathbf{P}((X, Y) \in \{(x, y) \in \mathbb{R}^2 : x + y \leq z\})$.*

Então

$$H(z) = \int_{\{(x,y)\in\mathbb{R}^2 : x+y\leq z\}} d(\mathbf{P}_X \times \mathbf{P}_Y) \tag{8.13}$$

$$= \int \int_{\mathbb{R}} \mathbf{1}_{\{(x,y)\in\mathbb{R}^2 : x+y\leq z\}} d(\mathbf{P}_X \times \mathbf{P}_Y) \tag{8.14}$$

$$= \int_{\mathbb{R}} d\mathbf{P}_Y \cdot \int_{\mathbb{R}} \mathbf{1}_{\{(x,y)\in\mathbb{R}^2 : x+y\leq z\}} d\mathbf{P}_X \quad (\textit{por Fubini}) \tag{8.15}$$

$$= \int_{\mathbb{R}} d\mathbf{P}_Y \cdot \int_{\mathbb{R}} \mathbf{1}_{\{(x,y)\in\mathbb{R}^2 : x\leq z-y\}} d\mathbf{P}_X \tag{8.16}$$

$$= \int_{\mathbb{R}} d\mathbf{P}_Y \cdot \int_{\mathbb{R}} \mathbf{P}_X(X \in (-\infty, z-y]) d\mathbf{P}_X \tag{8.17}$$

$$= \int_{\mathbb{R}} F(z-y) d\mathbf{P}_Y(y) \tag{8.18}$$

Agora observamos que a integral em relação a uma medida é a mesma que a integral Stieltjes em relação a uma função de distribuição. e assim $= \int_{\mathbb{R}} F(z-y)\, \mathrm{d}G(y)$. □

8.31 Corolário *Se X e Y tiverem densidades f e g, respectivamente, então X + Y tem densidade*

$$Z(x) = \int_{\mathbb{R}} f(x-y)g(y)\, dy.$$

A integral anterior é denotada por $h = f \star g$ e é conhecida como **convolução** de f e g.

Demonstração (Prova do Corolário). *fazer.* □

8.32 Exemplo *Sejam X e Y variáveis aleatórias uniformes independentes em [0,1]. Calcule a soma das duas variáveis. A densidade de X + Y é dada por*

$$Z(x) = \int_{\mathbb{R}} \mathbf{1}_{[0,1]}(x-y)\mathbf{1}_{[0,1]}(y)$$

Então $0 \leq y \leq 1$, mas também $0 \leq x - y \leq 1$, então $y \leq x$ e $y \geq x - 1$.

A soma terá valores entre 0 e 2. Portanto, $x \in [0,2]$.

Se $x \in [0,1]$, então a única condição a ser satisfeita é $y \leq x$, então temos $\int_0^x dy = x$.

Se $x \in [1,2]$, então temos $\int_{x-1}^1 dy = 2 - x$. ◁

Nesta seção discutimos uma técnica para aproximar funções integráveis arbitrárias por funções suaves.

8.33 Definição (Funções Teste) *Uma função é dita* ***função teste*** *ou suave de suporte compacto, denotada por $f \in C_c^\infty(\mathbb{R})$ se f é de classe C^∞ e seu suporte*

$$\operatorname{supp}(f) := \overline{\{x \in X \mid f(x) \neq 0\}} = \overline{f^{-1}\left(\{0\}^c\right)}.$$

é um conjunto compacto.

Para começar, estabelecemos a existência de uma função teste em $[-1,1]$.

8.34 Lema *A função*

$$f(x) = \begin{cases} e^{\frac{-1}{1-x^2}} & |x| < 1 \\ 0 & |x| \geq 1 \end{cases}$$

é de suporte compacto em $[-1,1]$ e possui derivada contínua em todos os pontos. ◁

8.35 Lema *Seja $\rho(x)$ uma função positiva em $C_c^\infty(\mathbb{R})$ tal que $\rho(x)$ é tem suporte em $[-1,1]$ e*

$$\int_{-\infty}^{\infty} \rho(x)\, dx = \int_{-1}^{1} \rho(x)\, dx = 1.$$

Seja $f : \mathbb{R} \longrightarrow \mathbb{R}$ uma função contínua. Defina

$$f_n(x) = n \int_{-n}^{n} \rho(n(x-y)) f(y) dy$$

Então $f_n \in C_c^\infty(\mathbb{R})$, $f_n^{(m)}(x) = n \int_{-n}^{n} \rho^{(m)}(n(x-y)) f(y) dy$ e f_n convergem para f uniformemente em conjuntos compactos. Além disso, se f for limitada, então $\|f_n\|_\infty \leq \|f\|_\infty$. ◁

Demonstração. *Primeiro, note que, como $\rho(x)$ e todas as suas derivadas são compactas, elas também são limitadas. Em particular, existe um $M > 0$ tal que $|\rho'(x)| \leq M$. Para limpar a notação um pouco, defina $\rho_n(y) = n\rho(ny)$ e assim temos*

$$f_n(x) = \int_{-n}^{n} \rho_n(x-y) f(y) dy$$

Como o suporte de $\rho_n(x)$ está contido em $[-\frac{1}{n}, \frac{1}{n}]$, logo se consideramos $\rho_n(x-y)$ como uma função de y, seu suporte é contido em $[x-\frac{1}{n}, x+\frac{1}{n}]$. Assim, o suporte de $f_n(x)$ está contido em $[-n-\frac{1}{n}, n+\frac{1}{n}]$.

Para examinar a derivada de $f_n(x)$, escolha $h > 0$ e seja

$$\frac{f_n(x+h) - f_n(x)}{h} = \frac{1}{h}\int_{-n}^{n} (\rho_n(x+h-y) - \rho_n(x-y))\, f(y)dy$$

O Teorema de Taylor nos diz que $\frac{1}{h}(\rho_n(x+h-y) - \rho_n(x-y)) = \rho_n'(c)$ para algum $c \in [x+h-y, x-y]$. Assim sendo, $\left|\frac{1}{h}(\rho_n(x+h-y) - \rho_n(x-y))f(y)\right| \leq M\,|f(y)|$ e pela integrabilidade de $f(y)$ no intervalo $[-n, n]$ podemos usar a Convergência Dominada para concluir que

$$\begin{aligned}
f_n'(x) &= \lim_{h\to 0} \frac{f_n(x+h) - f_n(x)}{h} \\
&= \lim_{h\to 0} \frac{1}{h}\int_{-n}^{n} (\rho_n(x+h-y) - \rho_n(x-y))f(y)dy \\
&= \int_{-n}^{n} \lim_{h\to 0} \frac{1}{h}(\rho_n(x+h-y) - \rho_n(x-y))f(y)dy \\
&= \int_{-n}^{n} \rho_n'(x-y)f(y)dy
\end{aligned}$$

A continuidade de $f_n'(x)$ segue da continuidade de $f(y)$ e $\rho_n'(x-y)$ e o Teorema da Convergência Dominada como acima. Uma indução simples estende o resultado para derivadas de ordem arbitrária.

Em seguida, mostraremos a convergência. Escolha um conjunto compacto $K \subset \mathbb{R}$ e $\varepsilon > 0$, Como f é uniformemente contínua em K, existe um $\delta > 0$ tal que para qualqueis $x, y \in K$ temos que se $|x-y| \leq \delta$ então $|f(x) - f(y)| \leq \varepsilon$. Escolha $N_1 > 0$ tal que $\frac{1}{n} < \delta$ para todo $n \geq N_1$. A hipótese $\int_{-\infty}^{\infty} \rho(y)\,dy = \int_{-1}^{1} \rho(y)\,dy = 1$ e por uma mudança de variáveis temos $\int_{-\infty}^{\infty} \rho_n(x-y)\,dy = \int_{x-\frac{1}{n}}^{x+\frac{1}{n}} \rho_n(x-y)\,dy = 1$ para todos $x \in \mathbb{R}$ e $n > 0$. Escolha $N_2 > 0$ de modo que para todos $n > N_2$, temos $K \subset [-n+\frac{1}{n}, n-\frac{1}{n}]$. Portanto, podemos escrever $f(x) = \int_{-n}^{n} \rho_n(x-y)f(x)\,dy = 1$ para *qualquer $x \in K$ e $n > N_2$. Temos então que para qualquer $n \geq \max(N_1, N_2)$*

$$
\begin{aligned}
|f_n(x) - f(x)| &= \left| \int_{-n}^{n} (\rho_n(x-y) f(y) - \rho_n(x-y) f(x))\, dy \right| \\
&= \left| \int_{x-\frac{1}{n}}^{x+\frac{1}{n}} (\rho_n(x-y) f(y) - \rho_n(x-y) f(x))\, dy \right| \qquad \text{pois } n > N_2 \\
&\leq \int_{x-\frac{1}{n}}^{x+\frac{1}{n}} \rho_n(x-y)\, |f(y) - f(x)|\, dy \\
&\leq \varepsilon \int_{x-\frac{1}{n}}^{x+\frac{1}{n}} \rho_n(x-y)\, dy \qquad \text{pois } \frac{1}{n} < \delta \\
&\leq \varepsilon \qquad \text{pois } \rho_n \text{ é positivo e } \int_{-\infty}^{\infty} \rho_n(x)\, dx = 1
\end{aligned}
$$

A última coisa a provar é a desigualdade da norma caso f seja limitada.

$$
\begin{aligned}
|f_n(x)| &\leq n \int_{-n}^{n} \rho(n(x-y))\, |f(y)|\, dy \qquad \text{pois } \rho \text{ é positivo} \\
&\leq n \|f\|_\infty \int_{-\infty}^{\infty} \rho(n(x-y)) dy = \|f\|_\infty
\end{aligned}
$$

Testando Convergência com Funções Suaves

8.36 Lema *Seja $\{X_i\}_{i=1}^{\infty}$ uma sequência de variáveis aleatórias então $X_n \xrightarrow{d} X$ se e somente se $\lim_{n\to\infty} \mathbf{E}[f(X_n)] = \mathbf{E}[f(X)]$ para todos as funções $f \in C_c^{\infty}(\mathbb{R}, \mathbb{R})$.* $\lhd$

Demonstração. *Como qualquer $f \in C_c^{\infty}(\mathbb{R}, \mathbb{R})$ é temos que $X_n \xrightarrow{d} X$ implica $\lim_{n\to\infty} \mathbf{E}[f(X_n)] = \mathbf{E}[f(X)]$.*

Na outra direção, considere uma função limitada arbitrária f e escolha $\varepsilon > 0$. Então podemos encontrar $f_n \in C_c^{\infty}(\mathbb{R}, \mathbb{R})$ tal que f_n converge uniformemente em conjuntos compactos e $\|f_n\|_\infty \leq \|f\|_\infty$.

A ideia da prova é observar que, para qualquer $n, k \geq 0$, temos

$$
|\mathbf{E}[f(X_n) - f(X)]| \leq |\mathbf{E}[f(X_n) - f_k(X_n)]| + |\mathbf{E}[f_k(X_n) - f_k(X)]| + |\mathbf{E}[f_k(X) - f(X)]|
$$

e depois limitar cada termo no lado direito. O segundo termo será fácil de lidar devido à nossa hipótese e à suavidade de f_k. O primeiro e terceiros termos exigirão que examinemos a aproximação fornecida pela a convergência uniforme do f_k em todos os conjuntos compactos.

A primeira tarefa que temos é escolher o conjunto compacto. Para tanto basta considerar os intervalos fechados centrados na origem. Para qualquer $R \in \mathbb{R}$ com $R > 0$, existe um $\psi_R \in C_c^{\infty}(\mathbb{R}, \mathbb{R})$

com $\mathbb{1}|x| \leq \frac{R}{2} \leq \psi_R(x) \leq \mathbb{1}|x| \leq R$*, assim sendo*

$$\begin{aligned}
\lim_{n\to\infty} \mathbf{P}\left[|X_n| > R\right] &= 1 - \lim_{n\to\infty} \mathbf{E}\left[\mathbb{1}|X_n| \leq R\right] \\
&\leq 1 - \lim_{n\to\infty} \mathbf{E}\left[\psi_R(X_n)\right] \\
&= 1 - \mathbf{E}\left[\psi_R(X)\right] \\
&\leq 1 - \mathbf{E}\left[\mathbb{1}|X| \leq \frac{R}{2}\right] \\
&= \mathbf{P}\left[|X| > \frac{R}{2}\right]
\end{aligned}$$

Por outro lado, sabemos que $\lim_{R\to\infty} \mathbb{1}|X| \leq \frac{R}{2} = 0$ *q.c. e, portanto, pelo Teorema da Convergência Monótona,* $\lim_{R\to\infty} \mathbf{P}\left[|X| > \frac{R}{2}\right] = 0$*. Escolha* $R > 0$ *tal que*

$$\mathbf{P}\left[|X| > R\right] \leq \mathbf{P}\left[|X| > \frac{R}{2}\right] \leq \frac{\varepsilon}{4\|f\|_\infty}$$

Então podemos escolher $N_1 > 0$ *tal que* $\mathbf{P}\left[|X_n| > R\right] \leq \frac{\varepsilon}{2\|f\|_\infty}$ *para todo* $n > N_1$*.*

Tendo escolhido $R > 0$*, sabemos que* f_n *converge uniformemente para* f *em* $|x| \leq R$ *e, portanto, podemos encontrar um* $K > 0$ *tal que se* $k > K$ *e* $|x| \leq R$ *temos* $|f_k(x) - f(x)| < \varepsilon$*. Assim sendo,*

$$\begin{aligned}
\left|\mathbf{E}\left[f_k(X) - f(X)\right]\right| &\leq \mathbf{E}\left[|f_k(X) - f(X)|\,;|X| \leq R\right] + \mathbf{E}\left[|f_k(X) - f(X)|\,;|X| > R\right] \\
&\leq \varepsilon \mathbf{P}\left[|X| \leq R\right] + 2\|X\|_\infty \mathbf{P}\left[|X| > R\right] \\
&\leq \varepsilon + \frac{\varepsilon}{2} < 2\varepsilon
\end{aligned}$$

e pelo mesmo cálculo, por $n > N_1$

$$\left|\mathbf{E}\left[f_k(X_n) - f(X_n)\right]\right| \leq \varepsilon + 2\|X_n\|_\infty \mathbf{P}\left[|X_n| > R\right] \leq 2\varepsilon$$

Para terminar a prova, escolha $k > K$ *e então podemos encontrar* $N_2 > 0$ *tal que para todo* $n > N_2$*, temos* $\left|\mathbf{E}\left[f_k(X_n) - f_k(X)\right]\right| < \varepsilon$*. Juntando essas três estimativas, temos* $n > \max(N_1, N_2)$*,*

$$\left|\mathbf{E}\left[f(X_n) - f(X)\right]\right| \leq 5\varepsilon$$

8.7 Aplicações

Lei $0-1$ de Hewitt-Savage

8.37 Definição *Uma aplicação bijetiva* $\pi = (\pi_1, \pi_2, \dots)$ *do conjunto dos números naturais* $\mathbb{N}$ *em si mesmo é dita* ***permutação finita*** *se* $\pi_n = n$ *para todos os* $n > N_0$*.*

Se $X = (X_1, X_2, \dots) \in (\mathbb{R}^{\mathbb{N}}, \mathcal{B}^{\mathbb{N}})$ é uma sequência de variáveis aleatórias, $\pi(X)$ denota a sequência $(X_{\pi_1}, X_{\pi_2}, \dots)$. Se A é um evento $\{X \in B\}$ com $B \in \mathcal{B}^{\mathbb{N}}$ então $\pi(A)$ denotará o evento $\{\pi(X) \in B\}$.

Dado uma sequência $(X_n)_{n\in\mathbb{N}}$ de variáveis aleatórias definimos a σ-álgebra permutável ou σ-álgebra de eventos simétricos $\mathcal{E}$ como o conjunto de eventos na σ-álgebra das variáveis $\{X_n\}_{n=1}^{\infty}$ que são invariantes sob permutações finitas dos índices na sequência $\{X_n\}_{n=1}^{\infty}$. Isto é, o evento $A = \{X \in B\}$ é simétrico se $A = \pi(A)$.

8.38 Teorema (Lei 0 – 1 de Hewitt-Savage) *Dado uma sequência (X_n) de variáveis aleatórias independentes e identicamente distribuídas. Então a sigma álgebra dos eventos permutáveis $\mathcal{F}$ satisfaz*

$$\mathbf{P}(A) \in \{0,1\}, \forall A \in \mathcal{F}$$

Demonstração. *Seja $A = \{X \in B\}$ um evento simétrico, isto é*

$$A = \{X \in B\} = \pi_n(A) \equiv \{\pi_n(X) \in B\}.$$

Aproximaremos A por eventos cilíndricos. Para isso escolha conjuntos $B_n \in \mathcal{B}^{\mathbb{N}}$ tais que, para $A_n = \{\omega : (X_1, \dots X_n) \in B_n\}$,

$$\mathbf{P}(A \triangle A_n) \to 0, \ n \to \infty. \tag{8.19}$$

Como as variáveis aleatórias $\{X_i\}_{n\in\mathbb{N}}$ são independentes e identicamente distribuídas, temos que as probabilidades $\mathbf{P}(X \in B)$ e $\mathbf{P}(\pi_n(X) \in B)$ coincidem. Assim sendo

$$\mathbf{P}(A \triangle A_n) = \mathbf{P}(X \in B \triangle B_n) = \mathbf{P}(\pi_n(X) \in B \triangle B_n). \tag{8.20}$$

Consequentemente

$$\begin{aligned}\mathbf{P}(\pi_n(X) \in B \triangle B_n) &= \mathbf{P}\{(\pi_n(X) \in B) \triangle (\pi_n(X) \in B_n)\} \\ &= \mathbf{P}\{(X \in B) \triangle (\pi_n(X) \in B_n)\} = \mathbf{P}\{A \triangle \pi_n(A_n)\}.\end{aligned} \tag{8.21}$$

Por 8.20 e 8.21 temos que

$$\mathbf{P}(A \triangle A_n) = \mathbf{P}(A \triangle \pi_n(A_n)). \tag{8.22}$$

Logo por 8.19 temos que

$$\mathbf{P}(A \triangle (A_n \cap \pi_n(A_n))) \to 0, \ n \to \infty. \tag{8.23}$$

Assim, por 8.19, 8.22 e 8.23, obtemos

$$\mathbf{P}(A_n) \to \mathbf{P}(A) \ \ e \ \ \mathbf{P}(\pi_n(A)) \to \mathbf{P}(A)$$

$$\mathbf{P}(A_n \cap \pi_n(A_n)) \to \mathbf{P}(A).$$

Além disso, como $\{X_i\}$ *são independentes, escolhendo* π_n *como a permutação que leva* $(X_1, \ldots X_n)$ *para* $(X_{n+1}, \ldots X_{2n})$

temos

$$\begin{aligned}\mathbf{P}(A_n \cap \pi_n(A_n)) &= \mathbf{P}\{(X_1, \ldots X_n) \in B_n,\ (X_{n+1}, \ldots X_{2n}) \in B_n\} \\ &= \mathbf{P}\{\ (X_1, \ldots X_n) \in B_n\}.\ \mathbf{P}\{(X_{n+1}, \ldots X_{2n}) \in B_n\} \\ &= \mathbf{P}(A_n)\mathbf{P}(\pi_n(A_n)\end{aligned}$$

Logo $\mathbf{P}(A) = \mathbf{P}^2(A)$ *e, portanto,* $\mathbf{P}(A) = 0$ *ou 1.* □

Passeios Aleatórios

Dado uma sequência X_i de vetores aleatórios i.i.d. em $\mathbb{R}^n$ então

$$S_n = \sum_{i=1}^{n} X_i$$

é um passeio aleatório em $\mathbb{R}^n$

8.39 **Teorema** *Se* S_n *é um passeio aleatório em* $\mathbb{R}$ *então uma das seguintes alternativas ocorre*

- $S_n = 0$ *para todo* n
- $S_n \longrightarrow \infty$
- $S_n \longrightarrow -\infty$
- $-\infty = \liminf S_n < \limsup S_n = \infty$

Demonstração. *Começamos observando que* $\{\omega : \limsup S_n(\omega) = c\}$ *é um evento permutável para todo* $c \in \mathbb{R}$*. Assim pela lei 0-1 de Hewitt-Savage,*

$$\mathbf{P}(\{\omega : \limsup S_n(\omega) = c\}) \in \{0,1\}$$

e logo $\limsup Sn$ *é uma constante* $c \in [-\infty, \infty]$ *q.c.*

Agora considere $S'_n = S_n - X_1$*, como* S'_n *tem a mesma distribuição que* S_n *tomando limsup em ambos os lados temos*

$$c = c - X_1.$$

Assim, ou $X_1 = 0$ *ou* $c \in \{-\infty, \infty\}$*. De modo análogo,* $\liminf S_n \in \{-\infty, \infty\}$*.* □

8.8 ↯Teorema de Extensão de Kolmogorov II

Nessa seção generalizamos o Teorema de Extensão de Kolmogorov para produtos quaisquer de espaços de medida. Ou seja, nessa seção queremos construir uma medida no espaço produto

$$\Omega_A := \prod_{\alpha \in A} \Omega_\alpha$$

para famílias arbitrárias de conjunto $\{\Omega_\alpha\}_{\alpha \in A}$, onde os elementos nos espaços são $x_A = (x_\alpha)_{\alpha \in A}$.

De modo correspondente, temos as aplicações de projeção nas coordenadas $\pi_\beta : \Omega_A \to \Omega_\beta$, que leva x_A para x_β. Também podemos definir a composta de tais aplicações de projeção de coordenadas.

Por exemplo, dados $B \subset A$, podemos definir a projeção parcial $\pi_B : \Omega_A \to \Omega_B$ de modo natural. Mais geralmente se $C \subset B \subset A$, podemos definir $\pi_{C \leftarrow B} : \Omega_B \to \Omega_C$. Então temos as leis de composição

$$\pi_{D \leftarrow C} \circ \pi_{C \leftarrow B} = \pi_{D \leftarrow B}$$

se $D \subset C \subset B \subset A$.

Podemos definir a σ-algebras pullback para cada $\beta \in A$

$$\pi_\beta^*(\mathcal{B}_\Omega) := \{\pi_\beta^{-1}(E) : E \in \mathcal{B}_\Omega\}$$

e o produto infinito de σ-álgebras

$$\mathcal{B}_A = \prod_{\beta \in A} \mathcal{B}_\beta := \langle \bigcup_{\beta \in A} \pi_\beta^*(\mathcal{B}_\beta) \rangle.$$

Novamente, esta é a menor σ-álgebra tal que π_β e de modo mais geral π_B são mensuráveis. Os elementos em $\mathcal{B}_A$, por definição de geradores de σ-álgebra,são obtidos através de um número enumerável de operações de conjuntos.

Em particular, isso significa que se $E \in \mathcal{B}_A$, então existe um conjunto enumerável $B \subset A$ e um conjunto $E_B \in \mathcal{B}_B$ tal que $E_A = \pi_B^{-1}(E_B)$. Consequentemente, se $f : \Omega_A \to [0, +\infty]$ é mensurável, isso significa que existe um conjunto enumerável $B \subset A$ e uma função mensurável em $\mathcal{B}_B$ $f_B : \Omega_B \to [0, +\infty]$ tal que

$$f = f_B \circ \pi_B.$$

A formulação acima pode acomodar a situação na qual temos uma família infinita de variáveis aleatórias Ω_α, cada uma definida como uma função mensurável em um espaço amostral de Borel Hausdorff localmente compacto Ω_α. Então podemos expandir as variáveis aleatórias para espaço produto $\prod_{\alpha\in A}\Omega_A$ através das projeções

$$\tilde{\Omega}_\alpha = \Omega_\alpha \circ \pi_\alpha$$

e, portanto, todas as variáveis aleatórias da família podem ser consideradas definidas no mesmo espaço, se pudermos transformar esse espaço produto em um espaço de medida "compatível". Ou seja, se tivermos uma medida de probabilidade μ_A definida no produto σ -álgebra, então induzirá uma medida induzida μ_B em Ω_B para qualquer $B \subset A$, por

$$\mu_B(E_B) := (\pi_B)_* \mu_A = \mu_A(\pi_B^{-1}(E_B))$$

para todos $E_B \in \mathcal{B}_B$. Segue então

$$(\pi_{C\leftarrow B})_*\mu_B = \mu_C$$

se $C \subset B \subset A$.

8.40 **Teorema (Teorema de Extensão de Kolmogorov)** *Seja $((\Omega_\alpha,\mathcal{B}_\alpha),\mathcal{T}_\alpha)_{\alpha\in A}$ uma família de espaços mensuráveis $(\Omega_\alpha,\mathcal{B}_\alpha)$, munidos com uma topologia $\mathcal{T}_\alpha$. Para cada $B \subset A$, seja μ_B uma medida de probabilidade "inner regular" em $\mathcal{B}_B$, onde o espaço produto Ω_B é equipado com a topologia produto, e satisfaz a condição de compatibilidade*

$$(\pi_{C\leftarrow B})_*\mu_B = \mu_C$$

sempre que $C \subset B \subset A$, com C e B finitos. Então existe uma única medida de probabilidade μ_A em $\mathcal{B}_A$ que satisfaz $(\pi_B)_*\mu_A = \mu_B$ para todo $B \subset A$ finito.

Exercícios

Produto

Suponha que $E \subset \Omega_1 \times \Omega_2$. Dado $x \in \Omega_1$ definimos a x-seção $E^x \subset \Omega_2$ como

$$E^x = \{y \in \Omega_2 : (x,y) \in E\}$$

Ex. 8.1 — Se $E, F \subset \Omega_1 \times \Omega_2$ e $x \in \Omega_1$, mostre que

1. $(E \cap F)^x = E_x \cap F_x$,
2. $(E^c)_x = (E_x)^c$,
3. $(\cup E_n)_x = \cup (E_n)^x$, onde (E_n) é uma sequência de subconjuntos $\Omega_1 \times \Omega_2$.

Ex. 8.2 — Se μ_1 e μ_2 são medidas σ-finitas, mostre que a medida produto $m_1 \times m_2$ também é σ-finita.

Ex. 8.3 — Mostre que $\mathcal{B}^{\mathbb{R}^d} = \bigotimes_{i=1}^{d} \mathcal{B}^{\mathbb{R}}$.

Ex. 8.4 — Prove a existência da medida produto usando o Teorema da Extensão Compacta.

Ex. 8.5 — Seja F uma distribuição em $\mathbb{R}$, então existe uma sequência de variáveis aleatórias independentes $X_1, \ldots, X_n, \ldots$ com distribuição F.

Ex. 8.6 — Suponha que X seja uma variável exponencialmente distribuída com densidade

$$f(x) = \alpha e^{-\alpha x}$$

e Y uma variável aleatória independente que tenha distribuição uniforme $U(0,1)$. Encontre a densidade da soma $Z = X + Y$.

O seguinte exercício demonstra que a independência probabilística é análoga à independência linear:

Ex. 8.7 — Sejam V um espaço vetorial de dimensão finita sobre um corpo finito F e X uma variável aleatória uniforme em V. Seja $\langle \cdot, \cdot \rangle : V \times V \to F$ uma forma bilinear não degenerada em V, e sejam $v_1, \ldots, v_n$ vetores não-nulos em V. Mostre que as variáveis aleatórias$\langle X, v_1 \rangle, \ldots, \langle X, v_n \rangle$ são independentes se e somente se os vetores $v_1, \ldots, v_n$ são linearmente independentes.

CAPÍTULO 9

Modos de Convergência

Neste capítulo apresentaremos e compararemos diferentes noções de convergência de variáveis aleatórias. É de fundamental importância ressaltar que estamos tratando de **convergência em espaços de probabilidade**, e portanto alguns resultados desse capítulo não são válidos em espaços onde a medida não é finita.

9.1 Convergências Determinísticas

Para começar, falaremos muito brevemente dos modos de convergência mais naturais, e já estudados em outras disciplinas, que não levam em consideração o fato de estarmos em um espaço de medida.

Por serem modos estudados anteriormente, não entraremos em detalhes sobre eles, apresentando apenas as definições, alguns exemplos e breves comentários.

Convergência Pontual

Seja Ω um conjunto e $X_n : \Omega \to \mathbb{R}$ uma sequência de funções que compartilham do mesmo domínio Ω Dizemos que $X_n(x)$ **converge pontualmente** para uma função $X : \Omega \to \mathbb{R}$ se:

$$\lim_{n\to\infty} X_n(x) = X(x),$$

para todo $x \in \Omega$.

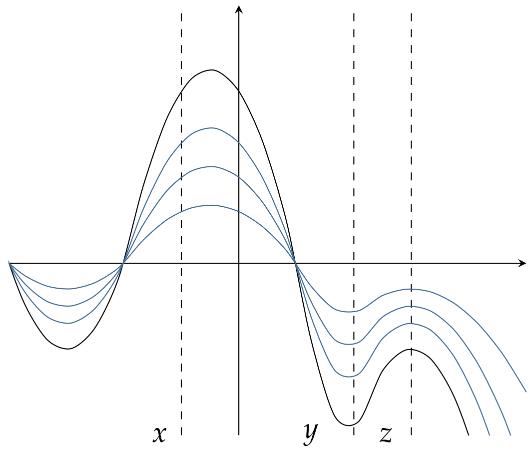
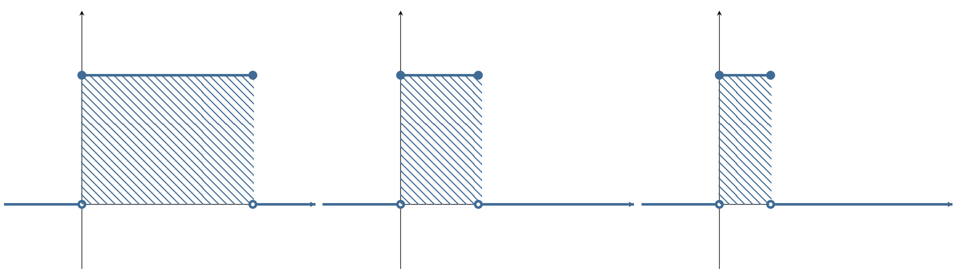


Figura 9.1: A sequência $\{X_n\}_{n=1}^{\infty}$ converge pontualmente para a função constante 0. Assim para todo ponto $x, y, z \in X$, temos $\lim X_n(x) = 0$, $\lim X_n(y) = 0$ e $\lim X_n(z) = 0$.



9.1 Exemplo *Para $n \in \mathbb{N}$ definimos* $X_n(x) = \begin{cases} 1 & \textit{se} \quad 0 < x < 1/n \\ 0 & \textit{caso contrário} \end{cases}$

Essa sequência converge pontualmente para a função constante.

Convergência Uniforme

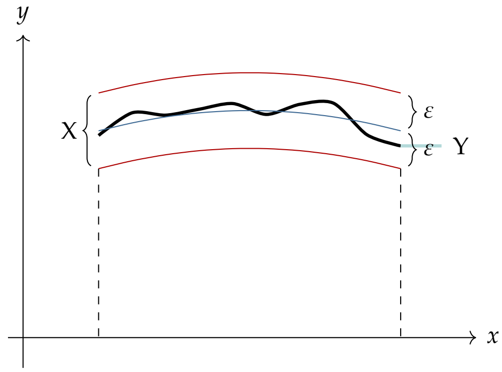
A **distância uniforme** entre duas funções X e Y é definida como

$$\|X - Y\|_u = \sup |X(x) - Y(x)|$$

Essa distância mede o afastamento máximo de X e Y.

Uma sequência de funções $\{X_n\}_{n=1}^{\infty}$ **converge uniformemente** para X se $\lim \|X_n - X\|_u = 0$. Isso significa não só que $\lim X_n(x) = X(x)$ para cada $x \in X$, mas além disso, que as funções X_n convergem para X em todos os lugares com uma "velocidade mínima comum".

9.2 Exemplo *Se $X_n(x) = 1/n$ para todo $x \in [0, 1]$, então a sequência $\{X_n\}_{n=1}^{\infty}$ converge para zero uniformemente em* $[0, 1]$. ◁



9.3 Exemplo *Para $n \in \mathbb{N}$ definimos* $X_n(x) = \begin{cases} 1 & \textit{se} \quad 0 < x < 1/n \\ 0 & \textit{caso contrário} \end{cases}$

Essa sequência converge pontualmente para a função constante, mas não uniformemente.

9.2 Convergência em Espaços de Probabilidade

Passaremos agora a definir e explorar modos que convergência típicos de espaços de medida. São convergências que levam em conta a estrutura mensurável imposta ao espaço, além da medida construída em cima desta estrutura.

Convergência em Distribuição

A convergência em distribuição é, em certo sentido, o tipo de convergência mais fraco. Tudo o que este tipo de convergência nos diz é que a distribuição de X_n converge para a distribuição de X, quando n vai para o infinito.

Assim, neste modo de convergência nos preocuparemos apenas com o comportamento das funções de distribuição das variáveis da sequência, e como estas se comparam com a função distribuição da variável limite.

Uma das principais diferenças desse modo de convergência é que, ao olharmos apenas as distribuições das variáveis, não estamos fazendo nenhuma referência ao espaço de probabilidade onde tais variáveis estão definidas. De fato, como ficará mais claro a seguir, as variáveis não precisam nem estar definidas em um mesmo espaço de probabilidade!

9.4 Definição (Convergência em Distribuição) *Uma sequência de variáveis aleatórias $\{X_n\}_{n=1}^{\infty}$, **converge em distribuição** para uma variável aleatória X, fato denotado por $X_n \xrightarrow{d} X$, se*

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x), \tag{9.1}$$

para todo x tal que $F_X(x)$ é contínua.

9.5 Exemplo *O exigência de que apenas os pontos de continuidade de F_X devem ser considerados na definição acima é, de fato, essencial.*

Para ver isso, tome uma sequência $\{X_n; n \geq 1\}$ de variáveis aleatórias, definidas em um mesmo espaço de probabilidade $(\Omega, \mathcal{F}, \mathbf{P})$, e tais que $X_n(\omega) = \frac{1}{n}$ para todo $\omega \in \Omega$. Tal sequência converge para 0 de todos os modos que vimos até agora: pontualmente e uniformemente, e em outros que descreveremos a seguir: quase-certamente, em $\mathcal{L}^p$ (para qualquer $p > 0$) e em probabilidade.

Além disso, temos

$$F_{X_n}(x) = \mathbf{P}(X_n \leq x) = \begin{cases} 0 & , \text{ se } x < \frac{1}{n}; \\ 1 & , \text{ se } x \geq \frac{1}{n} \end{cases},$$

e para a variável $X = 0$ temos

$$F_X(x) = \mathbf{P}(X \leq x) = \begin{cases} 0 & , \text{ se } x < 0; \\ 1 & , \text{ se } x \geq 0 \end{cases}$$

Observe que para $x < 0$

$$\lim_n F_{X_n}(x) = 0 = F_X(x),$$

e para $x > 0$

$$\lim_n F_{X_n}(x) = 1 = F_X(x),$$

mas para $x = 0$ temos

$$\lim_n F_{X_n}(0) = 0 \neq 1 = F_X(0).$$

Mas como $x = 0$ é ponto de descontinuidade de F_X, então ainda podemos afirmar que $X_n \xrightarrow{d} 0$.$\triangleleft$

Ao trabalhar com variáveis aleatórias que tomam valores inteiro, o seguinte teorema é frequentemente útil.

9.6 Teorema *Considere a sequência $\{X_n\}_{n=1}^{\infty}$ e a variável aleatória X. Suponha que X e X_n sejam não negativas e tomem valores inteiros.*

Então $X_n \xrightarrow{d} X$ se e somente se

$$\lim_{n\to\infty} \mathbf{P}_{X_n}(k) = \mathbf{P}_X(k), \qquad \text{para } k = 0,1,2,\cdots. \tag{9.2}$$

Demonstração. *Como X é de valor inteiro, seu função de distribuição, $F_X(x)$, é contínua em todo $x \in \mathbb{R} - \{0,1,2,\ldots\}$. Se $X_n \xrightarrow{d} X$, então*

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x), \qquad \text{para } x \in \mathbb{R} - \{0,1,2,\ldots\}. \tag{9.3}$$

Assim, para $k = 0,1,2,\cdots$, *temos*

$$\lim_{n\to\infty} \mathbf{P}_{X_n}(k) = \lim_{n\to\infty}\left[F_{X_n}\left(k+\frac{1}{2}\right) - F_{X_n}\left(k-\frac{1}{2}\right)\right] \qquad (X_n \text{ toma valores inteiros}) \tag{9.4}$$

$$= \lim_{n\to\infty} F_{X_n}\left(k+\frac{1}{2}\right) - \lim_{n\to\infty} F_{X_n}\left(k-\frac{1}{2}\right) \tag{9.5}$$

$$= F_X\left(k+\frac{1}{2}\right) - F_X\left(k-\frac{1}{2}\right) \qquad (\text{pois } X_n \xrightarrow{d} X) \tag{9.6}$$

$$= \mathbf{P}_X(k) \qquad (\text{uma vez que } X \text{ toma valores inteiros}). \tag{9.7}$$

Para provar a recíproca, suponha que

$$\lim_{n\to\infty} \mathbf{P}_{X_n}(k) = \mathbf{P}_X(k), \qquad \text{para } k = 0,1,2,\cdots. \tag{9.8}$$

Então, para todo $x \in \mathbb{R}$, *temos*

$$\lim_{n\to\infty} F_{X_n}(x) = \lim_{n\to\infty} \mathbf{P}(X_n \le x) \tag{9.9}$$

$$= \lim_{n\to\infty} \sum_{k=0}^{\lfloor x \rfloor} \mathbf{P}_{X_n}(k), \tag{9.10}$$

onde $\lfloor x \rfloor$ *denota o inteiro maior menor ou igual a* x. *Uma vez que, para qualquer* x *fixo, o conjunto* $\{0,1,\cdots,\lfloor x \rfloor\}$ *é um conjunto finito, podemos alterar a ordem do limite e a soma, então obtemos*

$$\lim_{n\to\infty} F_{X_n}(x) = \sum_{k=0}^{\lfloor x \rfloor} \lim_{n\to\infty} \mathbf{P}_{X_n}(k) \tag{9.11}$$

$$= \sum_{k=0}^{\lfloor x \rfloor} \mathbf{P}_X(k) \qquad (\text{por hipótese}) \tag{9.12}$$

$$= \mathbf{P}(X \le x) = F_X(x). \tag{9.13}$$

9.7 Teorema *Seja* $\{X_i\}_{i=1}^{\infty}$ *uma sequência de variáveis aleatórias tal que*

$$X_n \sim \text{Bin}\left(n, \frac{\lambda}{n}\right), \qquad \text{para } n \in \mathbb{N}, n > \lambda, \tag{9.14}$$

onde $\lambda > 0$ *é uma constante. Mostre que* X_n *converge em distribuição para* $\text{Po}(\lambda)$.

Basta mostrar que

$$\lim_{n\to\infty} \mathbf{P}_{X_n}(k) = P_X(k), \qquad \text{para todo } k = 0,1,2,\cdots. \tag{9.15}$$

Temos

$$\begin{aligned}
\lim_{n\to\infty} \mathbf{P}_{X_n}(k) &= \lim_{n\to\infty} \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{n-k} \\
&= \lambda^k \lim_{n\to\infty} \frac{n!}{k!(n-k)!} \left(\frac{1}{n^k}\right) \left(1-\frac{\lambda}{n}\right)^{n-k} \\
&= \frac{\lambda^k}{k!} . \lim_{n\to\infty} \left(\left[\frac{n(n-1)(n-2)...(n-k+1)}{n^k} \right] \left[\left(1-\frac{\lambda}{n}\right)^n \right] \left[\left(1-\frac{\lambda}{n}\right)^{-k} \right] \right).
\end{aligned}$$

Para k fixo, temos

$$\begin{aligned}
&\lim_{n\to\infty} \frac{n(n-1)(n-2)...(n-k+1)}{n^k} = 1, \\
&\lim_{n\to\infty} \left(1-\frac{\lambda}{n}\right)^{-k} = 1, \\
&\lim_{n\to\infty} \left(1-\frac{\lambda}{n}\right)^{n} = e^{-\lambda}.
\end{aligned}$$

Logo

$$\lim_{n\to\infty} \mathbf{P}_{X_n}(k) = \frac{e^{-\lambda}\lambda^k}{k!}.$$

Convergência em Probabilidade

Nesse modo de convergência estamos interessados em medir e controlar a probabilidade da sequência estar perto (ou longe) do limite desejado.

Tal convergência é conhecida no contexto de Teoria da Medida,como *convergência em medida*.

9.8 **Definição (Convergência em Probabilidade)** *Uma sequência de variáveis aleatórias* $\{X_n\}_{n\geq 1}$ ***converge em probabilidade*** *para a variável aleatória* X *se*

$$\forall \epsilon > 0 \quad \mathbf{P}(|X_n - X| > \epsilon) \to 0 \text{ quando } n \to \infty.$$

Denotaremos tal fato por $X_n \xrightarrow{\mathbf{P}} X$.

A ideia básica por trás desse tipo de convergência é que a probabilidade de um resultado "incomum"torna-se menor à medida que a sequência cresce.

9.9 **Exemplo** *Dado* $X_n \sim \text{Exp}(n)$*, então* $X_n \xrightarrow{\mathbf{P}} 0$.

Pois

$$\lim_{n\to\infty} \mathbf{P}\big(|X_n - 0| \geq \epsilon\big) = \lim_{n\to\infty} \mathbf{P}\big(X_n \geq \epsilon\big) \qquad (\textit{ pois } X_n \geq 0 \text{ }) \tag{9.16}$$

$$= \lim_{n\to\infty} e^{-n\epsilon} \qquad (\textit{ pois } X_n \sim \text{Exp}(n) \text{ }) \tag{9.17}$$

$$= 0, \qquad \textit{para todo } \epsilon > 0. \tag{9.18}$$

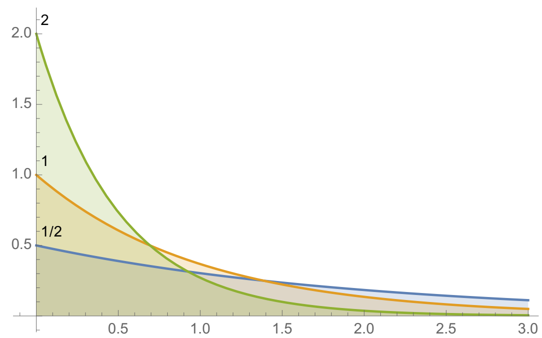


Figura 9.2: Distribuição exponenciais de parâmetros 1/2,1,2

9.10 Exemplo *Seja* X *uma variável aleatória e* $X_n = X + Y_n$*, com*

$$\mathbf{E}[Y_n] = \frac{1}{n}, \qquad \mathbf{Var}[Y_n] = \frac{\sigma^2}{n}, \tag{9.19}$$

onde $\sigma > 0$ *é uma constante. Então* $X_n \xrightarrow{\mathbf{P}} X$.

Pela desigualdade triangular para $Y_n - \mathbf{E}[Y_n] + \mathbf{E}[Y_n]$*, temos*

$$|Y_n| \leq |Y_n - \mathbf{E}[Y_n]| + \frac{1}{n}. \tag{9.20}$$

Então para todo $\epsilon > 0$*, temos*

$$\mathbf{P}\big(|X_n - X| \geq \epsilon\big) = \mathbf{P}\big(|Y_n| \geq \epsilon\big) \tag{9.21}$$

$$\leq \mathbf{P}\left(|Y_n - \mathbf{E}[Y_n]| + \frac{1}{n} \geq \epsilon\right) \tag{9.22}$$

$$= \mathbf{P}\left(|Y_n - \mathbf{E}[Y_n]| \geq \epsilon - \frac{1}{n}\right) \tag{9.23}$$

$$\leq \frac{\mathbf{Var}[Y_n]}{\left(\epsilon - \frac{1}{n}\right)^2} \qquad \textit{(pela desigualdade de Chevyshev)} \tag{9.24}$$

$$= \frac{\sigma^2}{n\left(\epsilon - \frac{1}{n}\right)^2} \longrightarrow 0 \qquad \textit{quando } n \to \infty. \tag{9.25}$$

Logo $X_n \xrightarrow{\mathbf{P}} X$.

Convergência Quase Certa

Primeiro apresentarmos a chamada convergência quase certa que, como o nome diz, é uma versão enfraquecida de uma convergência pontual, onde permitimos que a convergência falhe em um conjunto de medida nula.

Começaremos com uma leve adaptação do exemplo **??**, onde tínhamos uma sequência convergindo pontualmente.

9.11 Exemplo *Tome* $\Omega = [0,1]$, *munido da* σ*-álgebra de Borel, e* $\mathbf{P}$ *a medida de Lebesgue em* $[0,1]$. *Pra* $n \in \mathbb{N}^*$, *defina então*

$$X_n(\omega) = \begin{cases} 1 & \text{se} \quad 0 \leq \omega < 1/n \\ 0 & \text{caso contrário} \end{cases}$$

Observe que se $\omega \in (0,1]$, *então* $X_n(\omega)$ *para todo* $n > 1/\omega$, *de modo que* $X_n(\omega) \longrightarrow 0$, *quando* $n \longrightarrow \infty$.

No entanto, vale também que $X_n(0) = 1$ *para todo* $n > 0$, *de modo que*

$$\{\omega \in [0,1] : X_n(\omega) \not\to 0\} = \{0\},$$

e

$$\mathbf{P}(\{\lim_{n\to\infty} X_n = 0\} = \mathbf{P}((0,1]) = 1.$$

Dizemos assim que $X_n \longrightarrow 0$ ***quase certamente***, *quando* $n \longrightarrow \infty$. ◁

9.12 Definição (Convergência Quase Certa) *Sejam* X *uma variável aleatória e* $\{X_n\}_{n\geq 1}$ *uma sequência de variáveis aleatórias definidas no mesmo espaço de probabilidade. Dizemos que* X_n ***converge quase certamente*** *para* X, *isto é,* $X_n \xrightarrow{q.c.} X$ *se*

$$\mathbf{P}\left(\lim_{n\to\infty} X_n = X\right) = 1$$

ou, equivalentemente, se

$$\mathbf{P}\left(\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\right) = 1$$

A convergência quase certa é uma convergência pontual em um 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

9.13 Exemplo *Para* $n \in \mathbb{N}$ *definimos* $X_n(x) = \begin{cases} 1 & \text{se} \quad x \in \mathbb{Q} \\ 1/n & \text{caso contrário} \end{cases}$

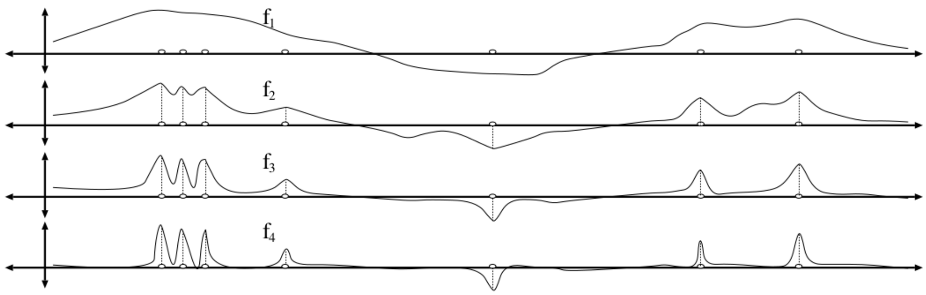


Figura 9.3: A sequência $\{X_n\}_{n=1}^{\infty}$ converge quase certamente a 0.

Essa sequência não converge pontualmente para a função constante 0*, mas converge quase certamente para* 0*.*

9.14 Heurística (math.exchange) *Considere um homem que joga três moedas todas as manhãs. Todas as tardes, ele doa um real para uma instituição de caridade para cada cara que aparecer nos lançamentos. Porém, a primeira vez que o resultado for três coroas ele irá parar de doar permanentemente.*

Seja $\{X_n\}_{n=1}^{\infty}$ *o montante diário que a instituição de caridade recebeu dele. Podemos ter quase certeza de que um dia esse valor será zero e permanecerá zero para sempre depois disso, ou seja, temos que* $X_n \xrightarrow{q.c.} 0$

Observamos que no entanto, quando consideramos qualquer número finito de dias, existe uma probabilidade diferente de zero, de a condição de término não ter ocorrido ainda.

9.15 Teorema (Critério de Convergência Quase Certa) *Sejam* $\{X_n\}_{n\geq 1}$ *e* X *variáveis aleatórias. Nestas condições, temos que* X_n *converge quase certamente para* X *se, e somente se, para todo* $\epsilon > 0$

$$\mathbf{P}\left(|X_n - X| \geq \epsilon \;\; i.v.\right) = 0$$

.

Ou, de forma equivalente, se para todo $k > 0$

$$\mathbf{P}\left(|X_n - X| \geq \frac{1}{k} \;\; i.v.\right) = 0.$$

Demonstração. *Começamos observando que as duas condições do teorema anterior são equivalentes. Para isso note que os eventos* $\{|X_n - X| \geq \epsilon \quad i.v.\}$ *são monótonos em* ϵ *e logo para demostrar que*

$$\mathbf{P}\left(|X_n - X| \geq \epsilon \quad i.v.\right) = 0,$$

para todo $\epsilon > 0$, é suficiente focar em constantes ϵ da forma $1/k$ para $k > 0$.

Posto isso, lembre que a sequência $X_n(\omega)$ converge para $X(\omega)$ se ***para todo*** *$k \geq 1$,* ***existe*** *$n \geq 1$ tal que* ***para todo*** *$m \geq n$*

$$|X_m(\omega) - X(\omega)| < \frac{1}{k}.$$

Deste modo, vale que

$$[X_n \to X] = \bigcap_{k=1}^{\infty} \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} \left[|X_m - X| < \frac{1}{k} \right]$$

Logo,

$$[X_n \nrightarrow X] = \bigcup_{k=1}^{\infty} \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} \left[|X_m - X| \geq \frac{1}{k} \right] = \bigcup_{k=1}^{\infty} \left[|X_n - X| \geq \frac{1}{k}, \quad i.v. \right]$$

Se $X_n \xrightarrow{q.c.} X$, então $\mathbf{P}[X_n \nrightarrow X] = 0$ o que implica que

$$\mathbf{P}\left(|X_n - X| \geq \frac{1}{k}, \quad i.v. \right) = 0, \quad \text{para todo } k \geq 1.$$

Por outro lado, se $\mathbf{P}\left(|X_n - X| \geq \frac{1}{k}, \quad i.v. \right) = 0, \quad$ para todo $k \geq 1$ segue que

$$\mathbf{P}[X_n \nrightarrow X] \leq \sum_{k=1}^{\infty} \mathbf{P}\left(|X_n - X| \geq \frac{1}{k}, \quad i.v. \right) = 0.$$

9.16 Teorema *Considere a sequência $\{X_n\}_{n=1}^{\infty}$. Se para todo $\epsilon > 0$, tivermos*

$$\sum_{n=1}^{\infty} \mathbf{P}\big(|X_n - X| > \epsilon \big) < \infty, \tag{9.26}$$

então $X_n \xrightarrow{q.c.} X$.

Demonstração. *A demonstração sSegue diretamente do Teorema 9.15 e do Primeiro Lema de Borel Cantelli.* □

9.17 Exemplo *Considere uma sequência $\{X_n, n = 1,2,3,\cdots\}$ tal que*

$$X_n = \begin{cases} -\frac{1}{n} & \text{com probabilidade } \frac{1}{2} \\ \frac{1}{n} & \text{com probabilidade } \frac{1}{2} \end{cases}$$

Então $X_n \xrightarrow{q.c.} 0$.

Pelo teorema anterior basta mostrar que

$$\sum_{n=1}^{\infty} \mathbf{P}\big(|X_n| > \epsilon\big) < \infty. \tag{9.27}$$

Observe que $|X_n| = \frac{1}{n}$. *Logo,* $|X_n| > \epsilon$ *se e somente se* $n < \frac{1}{\epsilon}$. *Logo temos*

$$\sum_{n=1}^{\infty} \mathbf{P}\big(|X_n| > \epsilon\big) \leq \sum_{n=1}^{\lfloor \frac{1}{\epsilon} \rfloor} \mathbf{P}\big(|X_n| > \epsilon\big) \tag{9.28}$$

$$= \lfloor \frac{1}{\epsilon} \rfloor < \infty. \tag{9.29}$$

9.18 Exemplo (Corcovas Deslizantes) *Agora apresentaremos o exemplo da "corcova deslizante". Para construir essa função o que faremos é dividir* $[0{,}1]$ *em dois intervalos* $[0{,}1/2] = I_1$ *e* $[1/2, 1] = I_2$. *Em seguida, definimos* $X_1 = \mathbb{1}_{I_1}$ *e* $X_2 = \mathbb{1}_{I_2}$. *Então, dividiremos* $[0{,}1]$ *em três intervalos, e consideraremos as funções características desses três intervalos. Sejam* X_3, X_4, *e* X_5 *as funções características desses intervalos. Repetiremos este processo para* $n = 1, 2, \ldots$.

É fácil ver que $\{X_n\}_{n=1}^{\infty}$ *converge para a função* 0 *em probabilidade. A sequência de funções não converge em quase todos os pontos (na verdade não converge em nenhum ponto) porque cada ponto pertencerá a infinito dos intervalos menores, e assim ficará fora de infinitamente muitos.*

Portanto, para cada ponto podemos encontrar subsequências de $\{X_n\}_{n=1}^{\infty}$ *que vão para* 0 *e subsequências que vão para* 1. ◁

Convergência em $\mathcal{L}^p$

O próximo modo de convergência se preocupa com o comportamento médio da sequência, sem se concentrar em probabilidades específicas.

9.19 Definição *Uma sequência de variáveis aleatórias* $\{X_n\}_{n=1}^{\infty}$ ***converge em*** $\mathcal{L}^p$ *para a variável aleatória* X *se*

$$\|X_n - X\|_p \to 0 \text{ quando } n \to \infty.$$

Em outras palavras

$$\mathbf{E}|X_n - X|^p \to 0 \text{ quando } n \to \infty.$$

9.20 Exemplo *Seja* $X_n \sim \mathrm{Uni}\left(0, \frac{1}{n}\right)$. *Mostre que* $X_n \xrightarrow{\mathcal{L}^p} 0$, *para todo* $p \geq 1$.

A função de distribuição de X_n *é dada por*

$$f_{X_n}(x) = \begin{cases} n & 0 \leq x \leq \frac{1}{n} \\ 0 & \text{caso contrário} \end{cases}$$

Então

$$\mathbf{E}\left(|X_n - 0|^p\right) = \int_0^{\frac{1}{n}} x^p n \quad dx \tag{9.30}$$

$$= \frac{1}{(p+1)n^p} \longrightarrow 0, \qquad \textit{para todo } p \geq 1. \tag{9.31}$$

9.21 Proposição *Se $q \geq p$ e $X_n \xrightarrow{\mathcal{L}^q} X$ então $X_n \xrightarrow{\mathcal{L}^p} X$*

Demonstração. *Consideramos $q = p + s$ pela desigualdade de Jensen temos que*

$$(\mathbf{E}|X_n - X|^p)^{\frac{1}{p}} \leq (\mathbf{E}|X_n - X|^q)^{\frac{1}{q}} \longrightarrow 0$$

9.22 Exemplo *Considere uma sequência $\{X_n\}$ tal que*

$$X_n = \begin{cases} n^2 & \textit{com probabilidade } \frac{1}{n} \\ \\ 0 & \textit{com probabilidade } 1 - \frac{1}{n} \end{cases} \tag{9.32}$$

Mostre que

- $X_n \xrightarrow{p} 0.$
- *X_n não converge na p-media para qualquer $p \geq 1$.* ◁

- *Para mostrar que $X_n \xrightarrow{p} 0$, podemos escrever, para qualquer $\epsilon > 0$*

$$\lim_{n\to\infty} \mathbf{P}\big(|X_n| \geq \epsilon\big) = \lim_{n\to\infty} \mathbf{P}(X_n = n^2) \tag{9.33}$$

$$= \lim_{n\to\infty} \frac{1}{n} \tag{9.34}$$

$$= 0. \tag{9.35}$$

Assim $X_n \xrightarrow{p} 0$.

- *Para qualquer $r \geq 1$, podemos escrever*

$$\lim_{n\to\infty} E\left(|X_n|^p\right) = \lim_{n\to\infty} \left(n^{2p} \cdot \frac{1}{n} + 0 \cdot \left(1 - \frac{1}{n}\right)\right) \tag{9.36}$$

$$= \lim_{n\to\infty} n^{2p-1} \tag{9.37}$$

$$= \infty \qquad (\textit{ pois } p \geq 1). \tag{9.38}$$

Portanto,X_n não converge na p-média para qualquer $p \geq 1$. Em particular, é interessante notar que, embora $X_n \xrightarrow{p} 0$, o valor esperado de X_n não converge para 0.

9.3 Comparando os Modos de Convergência

Nas seções anteriores, introduzimos várias noções de convergência de uma sequência de variáveis aleatórias (também denominados de modos de convergência). Existem várias relações entre os vários modos de convergência. No que se segue apresentamos os principais. Essas relações estão resumidas no seguinte diagrama:

$$\begin{tikzcd}
\xrightarrow{\mathcal{L}^p} \arrow[r, Rightarrow, "p>q\geq 1"] & \xrightarrow{\mathcal{L}^q} \arrow[d, Rightarrow] & \\
\xrightarrow{\text{q.c.}} \arrow[r, Rightarrow] & \xrightarrow{\mathbf{P}} \arrow[r, Rightarrow] \arrow[u, Rightarrow, bend right, "\text{momentos finitos}"'] & \xrightarrow{d}
\end{tikzcd} \tag{9.39}$$

Figura 9.4: Relação entre os modos de convergência em espaços de probabilidade.

Convergência quase certa implica convergência em probabilidade

9.23 Teorema (Convergência q.c. e em Probabilidade) *Dados X_n e X variáveis aleatórias.*

$\boxed{1}$ $X_n \to X$ *q.c. implica* $X_n \to X$ *em probabilidade.*

$\boxed{2}$ *Se* $X_n \to X$ *em probabilidade, então existe uma subsequência* $X_{n_k} \to X$ *quase certamente.*

Demonstração.

$\boxed{1}$ *Suponha* $X_n \to X$ *q.c. e tome* $\epsilon > 0$. *Sabemos então pelo Teorema 9.15 que*

$$\mathbf{P}(\limsup_n \{|X_n - X| > \epsilon\}) = \mathbf{P}(|X_n - X| > \epsilon \ \ i.v.) = 0,$$

e pelo Teorema de Continuidade,

$$\limsup_n \mathbf{P}(|X_n - X| > \epsilon) \leq \mathbf{P}(\limsup_n \{|X_n - X| > \epsilon\}) = 0.$$

Concluímos portanto que $X_n \to X$ *em probabilidade.*

$\boxed{2}$ *Suponha* $X_n \to X$ *em probabilidade. Queremos encontrar uma subsequência que converge quase certamente.*

Para isso lembre que, dado $\epsilon > 0$, *temos que*

$$\mathbf{P}(|X_{n_k} - X| > \epsilon) \to 0,$$

e portanto podemos escolher uma subsequência X_{n_k} *tal que*

$$\mathbf{P}(|X_{n_k} - X| > \epsilon) < 2^{-k} \textit{ para } k = 1,2,\ldots$$

Como $\sum_k 2^{-k}$ *converge, o Lema Borel-Cantelli nos fornece que*

$$\mathbf{P}(|X_{n_k} - X| > \frac{1}{k} \;\; i.v.) = 0.$$

Portanto, $X_{n_k} \longrightarrow X$ *quase certamente.* □

Convergência em $\mathcal{L}^p$ implica convergência em Probabilidade

9.24 Proposição $X_n \xrightarrow{\mathcal{L}^p} X$ *implica* $X_n \xrightarrow{p} X$.

Demonstração. *Usando a desigualdade de Chebyshev temos*

$$\mathbf{P}(|X_n - X| > \epsilon) \leq \mathbf{P}(|X_n - X|^p \geq \epsilon^p) \leq \frac{\mathbf{E}|X_n - X|^p}{\epsilon^p} \longrightarrow 0.$$

Vimos que se a sequência X_n converge em $\mathcal{L}^p$ para X então a sequência também converge em probabilidade, porém a recíproca nem sempre é verdadeira.

9.25 Teorema *Se* $X_n \xrightarrow{\mathbf{P}} X$ *e existe* Y *tal que* $\mathbf{E}[Y^p] < \infty$ *e* $|X_n| \leq Y$ *para todo* $n \geq 1$*, então* $X_n \xrightarrow{\mathcal{L}^p} X$

Demonstração. *Seja uma subsequência* n_k*. Escolhemos uma subsubsequência* n_{k_j} *de modo que* $X_{n_{k_j}} - X \xrightarrow{q.c.} 0$.

Como $|X_{n_{k_j}} - X|^p \leq (|X_{n_{k_j}}| + |X|)^p \overset{\text{por hipótese}}{\leq} (2Y)^p < \infty$

Assim, pelo Teorema da Convergência Dominada temos que

$$\mathbf{E}[|X_{n_{k_j}} - X|^p] \longrightarrow 0.$$

Provamos a convergência para alguma subsubsequência de uma subsequência arbitrária n_k.

E logo podemos concluir que $\mathbf{E}[|X_{n_{k_j}} - X|^p] \longrightarrow 0$.

9.26 Teorema *Seja* (X_n) *uma sequência de variáveis aleatórias não negativas tal que*

$X_n \xrightarrow{q.c} X$ *e* $\mathbf{E}[X_n] \longrightarrow \mathbf{E}[X]$*, então* $X_n \xrightarrow{\mathcal{L}^1} X$

Demonstração. *Para n suficientemente grande temos que* $\mathbf{E}[X_n] < \infty$ *e que*

$$\mathbf{E}[|X_n - X|] = \mathbf{E}[X - X_n]\mathbb{1}_{\{X_n \leq X\}} + \mathbf{E}[X_n - X]\mathbb{1}_{\{X_n > X\}} = 2\mathbf{E}[X - X_n]\mathbb{1}_{\{X_n \leq X\}} + \mathbf{E}[X_n - X].$$

Mas $0 \leq |X_n - X|\mathbb{1}_{X_n \leq X} \leq X$.

Pelo Teorema da Convergência Dominada temos que

$$\lim \mathbf{E}[X - X_n]\mathbb{1}_{\{X_n \leq X\}} = 0 \quad e \quad \mathbf{E}[X_n] \to \mathbf{E}[X],$$

Convergência em Probabilidade implica Convergência em Distribuição

9.27 **Teorema** *Se uma sequência de variáveis aleatórias* X_n *converge em probabilidade para uma variável aleatória* X*, então* X_n *também converge em distribuição para* X.

Demonstração. *Tome* $\epsilon > 0$ *e note que, como* $X_n \xrightarrow{p} X$*, então*

$$\lim_{n \to \infty} \mathbf{P}(|X_n - X| > \epsilon) = 0,$$

para todo $\epsilon > 0$.

Em outras palavras, fixado $\epsilon > 0$*, para todo* $\delta > 0$ *existe* $n_0 \geq 1$ *tal que*

$$\mathbf{P}(|X_n - X| > \epsilon) < \delta,$$

para todo $\delta > 0$.

Observe agora que se $X_n \leq x$ *e* $|X_n - X| \leq \epsilon$*, então* $X \leq x + \epsilon$. *Isso implica que se* $n \geq n_0$*, então*

$$\mathbf{P}(X_n \leq x) = \mathbf{P}(X_n \leq x; |X_n - X| > \epsilon) + \mathbf{P}(X_n \leq x; |X_n - X| \leq \epsilon) \leq \mathbf{P}(X \leq x + \epsilon) + \delta.$$

Concluímos assim que

$$\limsup_n \mathbf{P}(X_n \leq x) \leq \mathbf{P}(X \leq x + \epsilon),$$

para todo $\epsilon > 0$.

De modo análogo podemos notar que se $X \leq x - \epsilon$ *e* $|X_n - X| \leq \epsilon$*, então* $X_n \leq x$*, de modo que para* $n \geq n_0$

$$\mathbf{P}(X \leq x - \epsilon) = \mathbf{P}(X \leq x - \epsilon; |X_n - X| > \epsilon) + \mathbf{P}(X \leq x - \epsilon; |X_n - X| \leq \epsilon) \leq \mathbf{P}(X_n \leq x) + \delta.$$

Segue então que para todo $\epsilon > 0$,

$$\mathbf{P}(X \leq x - \epsilon) \leq \liminf_n \mathbf{P}(X_n \leq x) \leq \limsup_n \mathbf{P}(X_n \leq x) \leq \mathbf{P}(X \leq x + \epsilon).$$

Tomando agora x *como ponto de continuidade da função* $F_X(x) = \mathbf{P}(X \leq x)$*, e fazendo* $\epsilon \to 0$*, encontramos que*

$$\lim_n \mathbf{P}(X_n \leq x) = \mathbf{P}(X \leq x),$$

concluindo o resultado. $\square$

9.4 Teorema de Representação de Skorokhod

Na primeira parte deste capítulo estudamos os diversos modos de convergência usamos na probabilidade. Dentre eles estava a convergência em distribuição, que se concentrava no comportamento das funções de distribuição das variáveis da sequência, sem se importar com o espaço amostral onde as variáveis estavam definidas. E desse modo estas variáveis não precisavam nem estar definidas em um mesmo espaço.

Vimos que a convergência em distribuição era, de fato, a mais fraca das convergências estudadas. Isto é, se uma sequência converge em qualquer outro modo, ela converge também em distribuição.

A seguir vamos dar um primeiro passo no estudo de uma espécie de recíproca para este resultado. Vamos estudar a relação entre convergência em distribuição e convergência quase-certa.

9.28 **Teorema (Teorema de Representação de Skorokhod)** *Suponha que* $X_n \xrightarrow{d} X$*. Então, existem variáveis aleatórias* X'_n *distribuídas de forma idêntica à* X_n *e uma variável aleatória* X' *distribuída de forma idêntica à* X*, e tal que*

$$X'_n \xrightarrow{q.c.} X'$$

Demonstração. *Seja F_n a função de distribuição de X_n e F a função de distribuição de X, e suponha que $X_n \to X$ em distribuição. Ou seja, suponha que $F_n(x) \to F(x)$ quando $n \to \infty$ para todo $x \in \mathbb{R}$ ponto de continuidade de F.*

As candidatas a X'_n e X' devem ser naturais nesta altura. Dada uma variável aleatória U, definida em algum espaço de probabilidade $(\Omega, \mathcal{F}, \mathbf{P})$, uniformemente distribuída em $(0,1)$, defina

$$X' = F^*(U) \qquad e \qquad X'_n = F^*_n(U),$$

para $n \geq 1$.

Pela Proposição 6.15, sabemos que $X' \stackrel{d}{=} X$ e $X'_n \stackrel{d}{=} X_n$, para cada $n \geq 1$.

Até agora, encontramos variáveis aleatórias com as distribuições corretas. Antes de prosseguirmos, como a demonstração é bastante técnica, vamos apresentar a ideia geral por trás da demonstração que é mostrar que se $F_n \to F$ então $F^*_n \to F^*$, ao menos em um conjunto $A \in \mathcal{B}((0,1))$ com $\mathbf{P}(U \in A) = 1$, que será o complementar do conjunto de descontinuidades, que pelo Corolário 6.17 sabemos que é enumerável e logo de medida nula.

Para tanto, tome $\omega \in (0,1)$ e $\varepsilon > 0$ e escolha x um ponto de continuidade de F, de tal modo que

$$F(x) < \omega \leq F(x + \varepsilon).$$

*(**Obs:** Deixaremos a demonstração da existência de tal ponto como exercício, mas a prova é imediata da definição de $F^*(\omega)$ e do corolário 6.17).*

Por hipótese, $F_n(x) \to F(x)$ *quando* $n \to \infty$*, logo existe* $n_0 \geq 1$*, tal que*

$$F_n(x) < \omega \leq F(x+\varepsilon) \quad \text{para todo } n \geq n_0.$$

E logo, aplicando as propriedades das inversas generalizadas, temos que

$$F^*(\omega) - \varepsilon \leq x < F_n^*(\omega) \quad \text{para todo } n \geq n_0.$$

Devido às escolhas arbitrárias de ω *e* ε*, segue que*

$$\liminf_n F_n^*(\omega) \geq F^*(\omega) \quad \text{para todo } \omega \in (0,1). \tag{9.40}$$

Agora, seja $\omega' \in (\omega, 1)$*,* $\varepsilon > 0$ *e escolha* x *um ponto de continuidade de* F*, tal que*

$$F(x-\varepsilon) < \omega' \leq F(x).$$

De maneira análoga teremos que

$$F(x-\varepsilon) < \omega' < F_n(x) \text{ para todo } n \geq n_0,$$

e logo

$$F_n^*(\omega') < x < F^*(\omega') + \varepsilon \text{ para todo } n \geq n_0.$$

Além disso, como $\omega < \omega'$ *e* F^* *é não-decrescente, temos que*

$$F_n^*(\omega) < x < F^*(\omega') + \varepsilon \text{ para todo } n \geq n_0,$$

e assim

$$\limsup_n F_n^*(\omega) \leq F^*(\omega') \tag{9.41}$$

para todo $\omega < \omega' < 1$.

Assim, se F^* *é contínua em* ω*, então fazemos* $\omega' \to \omega$ *para concluir que*

$$\limsup_n F_n^*(\omega) \leq F^*(\omega). \tag{9.42}$$

Das equações 9.40 com 9.42, concluímos que

$$F^*(\omega) \leq \liminf_n F_n^*(\omega) \leq \limsup_n F_n^*(\omega) \leq F^*(\omega),$$

para todo $\omega \in D_{F^*}^c$*, onde* D_{F^*} *é conjunto dos pontos de descontinuidade de* F^**. Assim podemos concluir que*

$$\mathbf{P}(\lim_{n\to\infty} X_n' = X') = \mathbf{P}(\lim_{n\to\infty} F_n^*(U) = F^*(U)) = \mathbf{P}(U \in D_{F^*}^c).$$

E o resultado segue do corolário 6.17, observando que como D_{F^*} *é enumerável e* U *é uniforme, então*

$$\mathbf{P}(U \in D_{F^*}) = 0.$$

9.5 Princípio de Seleção de Helly

O Teorema de Seleção de Helly, nos diz que é "quase verdade"que, para cada sequência de variáveis aleatórias, existe uma subsequência que converge em distribuição. Esta é uma afirmação que é análoga à compacidade.

9.29 Teorema (Princípio de Seleção de Helly) *Seja F_n uma sequência de funções de distribuição, então existe uma subsequência F_{n_k} e uma função limitada, direita-contínua, não decrescente F tal que*

$$F_{n_k}(x) \to F(x) \text{ em todos os pontos de continuidade } x \text{ de } F.$$

Demonstração. *Começamos considerando uma enumeração dos racionais dada por $\mathbb{Q} = \{q_1, q_2, q_3, \ldots\}$. Começando pela sequência $(F_n(q_1))_{n\geq 1}$. Sabemos que $0 \leq F_n(q_1) \leq 1$, para todo $n \geq 1$, e pelo Teorema de Bolzano-Weierstrass, temos que existe uma subsequência $(n_k^1)_{k\geq 1} \subset \mathbb{N}$ tal que $F_{n_k^1}(q_1)$ converge.*

Podemos agora repetir esse argumento para a sequência $(F_{n_k^1}(q_2))_{k\geq 1}$, novamente extraindo uma subsequência $(n_k^2)_{k\geq 1} \subset (n_k^1)_{k\geq 1}$ de modo que $F_{n_k^2}(q_1)$ e $F_{n_k^2}(q_2)$ convergem. Repetindo esse procedimento recursivamente, obtemos subsequências $(n_k^m)_{k\geq 1}$, $m \geq 1$, tais que $F_{n_k^m}(q)$ converge para todo $q \in \{q_1, \ldots, q_m\}$.

Seguindo o Método da Diagonal de Cantor, defina a subsequência $(n_k^k)_{k\geq 1}$, formada pelos índices diagonais *das subsequências que construímos. Observe que, por construção, dado $m \in \mathbb{N}$ temos que $(n_k^k)_{k\geq m} \subset (n_k^m)_{k\geq 1}$, e portanto $F_{n_k^k}(q_m)$ converge.*

Deste modo encontramos uma subsequência $F_{n_k^k}$, $k \geq 1$ tal que

$$F_{n_k^k}(q) \text{ converge para todo } q \in \mathbb{Q}.$$

Denote esse limite por F_∞. Isto é,

$$F_\infty(q) := \lim_{k\to\infty} F_{n_k^k}(q), \forall q \in \mathbb{Q}.$$

Pela monotonicidade das funções F_{n_k}, podemos afirmar que F_∞ é não-decrescente.

Finalmente, precisamos estender F_∞ para todos os números reais, e para isso definimos

$$F(x) := \inf\{F_\infty(q) : q \in \mathbb{Q}, q > x\}.$$

Primeiro observe que a função F está definida em todos os reais e é não-decrescente.

Em geral, não podemos afirmar que $F_\infty(q) = F(q)$ para $q \in \mathbb{Q}$. De fato, como $F(q)$ é calculado a partir $F_\infty(q')$ para os racionais q' satisfazendo $q' > q$ e como F_∞ é não-decrescente, podemos apenas concluir que $F_\infty(q) \leq F(q)$, para $q \in \mathbb{Q}$. No entanto, para todo racional $r > 0$ temos que

$q + r > q$, e logo pelo definição de $F(x)$ temos que $F_\infty(q + r) \geq F(q)$. Como F é não-decrescente e $\mathbb{Q}$ é denso em $\mathbb{R}$, temos assim que

$$\begin{aligned}\inf_{y>x} F(y) &= \inf\{F(q) : q > x, q \in \mathbb{Q}\}\\ &= \inf\{F_\infty(q) : q > x, q \in \mathbb{Q}\}\\ &= F(x).\end{aligned}$$

De onde segue que

$$\lim_{x_n \downarrow x} F(x_n) = \inf_{y>x} F(y) = F(x),$$

e F é contínua à direita.

Assim, resta apenas demonstrar a convergência de $F_{n_k^k}(x)$. Para isso, tome $\epsilon > 0$, e note que pela definição de F, existe um $q > x$ racional que

$$F_\infty(q) < F(x) + \epsilon.$$

Como $F_{n_k^k}(q) \to F_\infty(q)$, existe $k_0 \geq 1$ tal que

$$F_{n_k^k}(q) < F(x) + \epsilon, \tag{9.43}$$

para todo $k \geq k_0$.

Assim, como $q > x$, a monotonicidade de $F_{n_k^k}$ e a equação (9.43) nos dão que

$$F_{n_k^k}(x) \leq F(x) + \epsilon$$

para todo $k \geq k_0$, e portanto

$$\limsup_k F_{n_k^k}(x) \leq F(x), \tag{9.44}$$

para todo $x \in \mathbb{R}$.

Obtido um limite superior, precisamos agora obter um limite inferior para $F_{n_k^k}(x)$. Seja então $x \in \mathbb{R}$ um ponto de continuidade de F. Pela continuidade de F em x, temos que existe $r < x$ tal que $F(x) - F(r) < \epsilon$. Ou seja,

$$F(r) > F(x) - \epsilon.$$

Escolha agora $r' \in \mathbb{Q}$ tal que $r < r' < x$, de modo que

$$F(x) - \epsilon < F(r) = \inf\{F_\infty(q); q > r, q \in \mathbb{Q}\} \leq F_\infty(r').$$

Deste modo, como $F_{n_k^k}(r') \to F_\infty(r')$, existe $k_1 \geq 1$ tal que

$$F_{n_k^k}(r') \geq F(x) - \epsilon$$

para todo $k \geq 1$, e a monotonicidade de $F_{n_k^k}$ nos dá que

$$F(x) - \epsilon \leq F_{n_k^k}(x),$$

para $k \geq k_1$.

Obtemos assim que

$$\liminf_k F_{n_k^k}(x) \geq F(x), \tag{9.45}$$

para todo $x \in \mathbb{R}$ ponto de continuidade de F.

Ao combinarmos (9.44) e (9.45), encontramos que

$$\lim_{k\to\infty} F_{n_k^k}(x) = F(x),$$

para todo $x \in \mathbb{R}$ ponto de continuidade de F, e o resultado segue. □

O princípio de seleção de Helly nos diz que se $(F_n)_{n\geq 1}$ então $F_{n_k}(x)$ converge ao longo de alguma subsequência para alguma função $F(x)$. A convergência ocorre para os pontos de continuidade de F, que é não-decrescente e contínua pela direita. Mas infelizmente, não podemos garantir que a função F dada pelo teorema seja, de fato, uma função distribuição.

Falta garantir que

$$\lim_{n\to\infty} F(x) = 1 \quad \text{e} \quad \lim_{n\to-\infty} F(x) = 0.$$

Em geral, o princípio de Helly só nos garante que

$$\lim_{n\to\infty} F(x) < 1 \quad \text{e} \quad \lim_{n\to-\infty} F(x) > 0,$$

e isso ocorre pois, a medida de que n cresce podemos ter "massa" de probabilidade escapando para $+\infty$ ou $-\infty$, como mostra o exemplo abaixo.

9.30 Exemplo *Dada uma variável aleatória U uniformemente distribuída em $[0,1]$ e constantes $\delta, \epsilon > 0$ com $\delta + \epsilon < 1$, defina*

$$X_n := -n \cdot \mathbb{1}_{\{U<\delta\}} + n \cdot \mathbb{1}_{\{U>1-\epsilon\}},$$

para $n \geq 1$.

Note que para todo $t > 0$ temos que

$$\lim_{n\to\infty} \mathbf{P}(X_n > t) = \mathbf{P}(U > 1-\epsilon) = \epsilon, \quad e$$
$$\lim_{n\to\infty} \mathbf{P}(X_n < -t) = \mathbf{P}(U < \delta) = \delta,$$

de modo que quando $n \to \infty$ uma "massa" de probabilidade ϵ está sendo perdida em $+\infty$ e δ em $-\infty$.

Como consequência, temos

$$\lim_{n\to\infty} F_{X_n}(x) = \begin{cases} \delta & \text{, para } x < 0; \\ 1-\epsilon & \text{, para } x \geq 0, \end{cases}$$

que é uma função não-decrescente e contínua pela direita, mas não é uma função distribuição. $\triangleleft$

Uma forma de resolver esse problema é garantir que a massa de probabilidade fique quase toda contida dentro de um intervalo limitado de forma uniforme ao longo da sequência. Ou seja, que para todo $\epsilon > 0$ existam $a,b \in \mathbb{R}$ tal que $\mathbf{P}(a < X_n \leq b) > 1-\epsilon$ para todo $n \geq 0$.

Esta propriedade é denominada de *tightness* e está precisamente definida abaixo.

9.31 Exemplo *Dada uma vairável aleatória U uniformemente distribuida em $[0,1]$ e constantes $\delta,\epsilon > 0$ com $\delta + \epsilon < 1$, defina*

$$X_n := -n \cdot \mathbb{1}_{\{U<\delta\}} + n \cdot \mathbb{1}_{\{U>1-\epsilon\}},$$

para $n \geq 1$.

Note que para todo $t > 0$ temos que

$$\lim_{n\to\infty} \mathbf{P}(X_n > t) = \mathbf{P}(U > 1-\epsilon) = \epsilon,$$

e

$$\lim_{n\to\infty} \mathbf{P}(X_n < -t) = \mathbf{P}(U < \delta) = \delta,$$

de modo que quando $n \to \infty$ uma "massa" de probabilidade ϵ está sendo perdida em $+\infty$ e δ em $-\infty$.

Como consequência, temos

$$\lim_{n\to\infty} F_{X_n}(x) = \begin{cases} \delta & \text{, para } x < 0; \\ 1-\epsilon & \text{, para } x \geq 0, \end{cases}$$

que é uma função não-decrescente e contínua pela direita, mas não é uma função distribuição. $\triangleleft$

Uma forma de resolver esse problema é garantir que a massa de probabilidade fique quase toda contida dentro de um intervalo limitado de forma uniforme ao longo da sequência. Ou seja, que para todo $\epsilon > 0$ existam $a,b \in \mathbb{R}$ tal que $\mathbf{P}(a < X_n \leq b) > 1-\epsilon$ para todo $n \geq 0$.

Esta propriedade é denominada de *tightness* e está precisamente definida abaixo.

9.32 Definição (Tightness) *Diremos que uma família $\{\mu_n, n \in I\}$ de medidas de probabilidade em $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ é* ***tight*** *se para todo $\epsilon > 0$ existem $a, b \in \mathbb{R}$ tal que $\mu_n([a,b]) > 1 - \epsilon$ para todo $n \geq \infty$.*

Analogamente, diremos que uma família $\{X_n, n \in I\}$ de variáveis aleatórias é tight se a família $\{\mathbf{P}_{X_n}, n \in I\}$ é tight, onde $\mathbf{P}_{X_n}$ é a distribuição de X_n.

Podemos dizer também que $\{F_n, n \in I\}$ é tight, onde $F_n : \mathbb{R} \to \mathbb{R}$ é a função distribuição de X_n.

9.33 Teorema (Teorema de Prokhorov) *Se $\{\mathbf{P}_n\}$ é uma sequência tight de medidas de probabilidade então existe uma subsequência $\{\mathbf{P}_{n_k}\}$ e uma medida $\mathbf{P}$ tal que $\{\mathbf{P}_{n_k}\}$ converge em distribuição para $\mathbf{P}$.*

Demonstração. *Seja $F_n(x) = \mathbf{P}_n((-\infty, x])$. Logo existe uma subsequência F_{n_k} e uma função F tal que $F_{n_k}(x) \to F(x)$ em todos os pontos de continuidade de F. Além disso, $0 < F < 1$. Agora afirmamos que F é, na verdade, uma função de distribuição de probabilidade, isto é, que $\lim_{x\to-\infty} F(x) = 0$ e $\lim_{x\to\infty} F(x) = 1$. Para isso considere $\varepsilon > 0$. Então como a medida é tight, podemos encontrar pontos $a < b$ que são pontos de continuidade de F, de modo que $\mathbf{P}_n((a,b]) > 1 - \varepsilon$ para todos os n. Logo*

$$\lim_{x\to\infty} F(x) - \lim_{x\to-\infty} F(x) \geq F(b) - F(a) \tag{9.46}$$

$$= \lim_{n\to\infty} [F_n(b) - F_n(a)] = \lim_{n\to\infty} \mathbf{P}_n((a,b]) > 1 - \varepsilon. \tag{9.47}$$

Isso é verdade para todo $\varepsilon > 0$, então devemos ter

$$\lim_{x\to\infty} F(x) - \lim_{x\to-\infty} F(x) = 1$$

Portanto, F é de fato uma função de distribuição de probabilidade. Assim, podemos definir a medida de probabilidade $\mathbf{P}$ por $\mathbf{P}((a,b]) = F(b) - F(a)$ para $a < b$. E $\mathbf{P}_n \underset{d}{\to} \mathbf{P}$. □

Assim, tomando X uma variável aleatória com função distribuição F, mostramos o seguinte resultado.

9.34 Proposição *Se $\{X_n, n \geq 1\}$ é uma sequência tight de variáveis aleatórias, então existe uma subsequência $(n_k)_{k\geq1}$ e uma variável aleatória X tal que $X_{n_k} \xrightarrow{d} X$ quando $k \to \infty$.*

9.35 Exemplo *Sejam $X_1, X_2, \ldots$ variáveis aleatórias independentes e identicamente distribuídas com média comum $\mathbf{E}[X_k] = \mu \in \mathbb{R}$.*

Da desigualdade triangular, temos que $\mathbf{E}[|S_n|] = \mathbf{E}[|X_1 + \cdots + X_n|] \leq \mathbf{E}[|X_1|] + \cdots \mathbf{E}[|X_n|] = n\mathbf{E}[|X_1|]$;

Com isso, segue da desigualdade de Chebyschev que, para todo $t > 0$,

$$\mathbf{P}\left(\left|\frac{S_n}{n}\right| > t\right) = \mathbf{P}(|S_n| > nt) = \leq \frac{\mathbf{E}[|S_n|]}{nt} \leq \frac{\mathbf{E}[|X_1|]}{t} \to 0,$$

quando $t \longrightarrow \infty$, uniformemente em n.

Assim, para todo $\epsilon > 0$ exite $t_0 > 0$ tal que para todo $t > t_0$,

$$\mathbf{P}\left(\left|\frac{S_n}{n}\right| \le t\right) \ge 1 - \epsilon,$$

para todo $n \ge 1$.

Concluímos assim que $S_n/n, n \ge 1$ é tight, e portanto existe uma variável aleatória Z, e uma subsequência $(n_k)_{k\ge1}$, tais que

$$\frac{S_{n_k}}{n_k} \xrightarrow{d} Z,$$

quando $k \longrightarrow \infty$. $\triangleleft$

A Proposição 9.34 garante a convergência por subsequências. Para garantir a convergência de toda a sequência, como é normal, precisamos garantir a unicidade dos limites por subsequência, como colocamos no resultado abaixo.

9.36 Proposição *Seja $\{X_n, n \ge 1\}$ é uma sequência tight de variáveis aleatórias e suponha que existe X tal que toda subsequência $X_{n_k}, k \ge 1$ convergente em distribuição converge para X. Ou seja, $X_{n_k} \xrightarrow{d} X$ sempre que $X_{n_k}, k \ge 1$ converge em distribuição. Nestas condições, $X_n \xrightarrow{d} X$ quando $n \longrightarrow \infty$*

Demonstração. *Denote por F a função distribuição de X e F_n a distribuição de X_n, e suponha que $X_n \not\to X$ em distribuição quando $n \longrightarrow \infty$. Nestas condições existem $x \in \mathbb{R}$ ponto de continuidade de F, $\epsilon > 0$ e uma subsequência $(n_k)_{k\ge1}$ tal que*

$$|F_{n_k}(x) - F(x)| \ge \epsilon,$$

para todo $k \ge 0$.

Mas como $\{X_{n_k}, k \ge 1\}$, pela proposição 9.34 e pela hipótese do teorema, existe uma subsubsequência $(n_{k_i})_{i\ge1} \subset (n_k)_{k\ge1}$, tal que $X_{n_{k_i}} \xrightarrow{d} X$.

Isso significa que $F_{n_{k_i}}(x) \longrightarrow F(x)$ e portanto

$$|F_{n_{k_i}}(x) - F(x)| < \epsilon,$$

para i suficientemente grande, o que é uma contradição. □

Exercícios

Ex. 9.1 — Deixe $X_1, X_2, X_3, \cdots$ ser uma sequência de variáveis aleatórias, de modo que

$$X_n \sim Ge\left(\frac{\lambda}{n}\right), \qquad \text{for} n = 1,2,3,\cdots, \tag{9.48}$$

onde $\lambda > 0$ é uma constante. Defina uma nova sequência Y_n como

$$Y_n = \frac{1}{n} X_n, \qquad \text{para } n = 1,2,3,\cdots. \tag{9.49}$$

Mostre que Y_n converge em distribuição para $Exp(\lambda)$.

Ex. 9.2 — **Ex. 9.3 —** Deixe $\{X_n, n = 1,2,\cdots\}$ e $\{Y_n, n = 1,2,\cdots\}$ serem duas seqüências de variáveis aleatórias, definidas no espaço amostral Ω . Suponha que nós saibamos

$$X_n \xrightarrow{p} X, \tag{9.50}$$

$$Y_n \xrightarrow{p} Y. \tag{9.51}$$

Prove que $X_n + Y_n \xrightarrow{p} X + Y$.

Ex. 9.4 — Considere a sequência $\{X_n, n = 1,2,3,\cdots\}$ tal que

$$X_n = \begin{cases} n & \text{com probabilidade } \frac{1}{n^2} \\ 0 & \text{com probabilidade } 1 - \frac{1}{n^2} \end{cases} \tag{9.52}$$

Mostre que

1. $X_n \xrightarrow{p} 0$.
2. $X_n \xrightarrow{L^r} 0$, para $r < 2$.
3. X_n não converge para 0 na r-média para qualquer $r \geq 2$.
4. $X_n \xrightarrow{q.c;} 0$.

Ex. 9.5 — Suponha que $X_n \xrightarrow{r} X$ para $r \geq 1$. Mostre que $\mathbf{E}|X_n^r| \to \mathbf{E}|X^r|$.

Ex. 9.6 — Suponha que $X_n \xrightarrow{D} X$ e $Y_n \xrightarrow{P} c$, onde c é uma constante. Prove que $cX_n \xrightarrow{D} cX$

Ex. 9.7 — Prove o Teorema da Aplicação Contínua para $\xrightarrow{p}$:

$$X_n \xrightarrow{p} X \quad \Rightarrow \quad g(X_n) \xrightarrow{p} g(X)$$

Ex. 9.8 — Deixe X_n ser uma sequência de variáveis aleatórias independentes que converge em probabilidade para X. Mostre que X é constante quase certamente.

Ex. 9.9 — Prove que se $X_n \xrightarrow{2} X$ para $r \geq 1$. Mostre que $\mathbf{Var}[X_n] \to \mathbf{Var}[X]$.

Ex. 9.10 — Convergência em probabilidade e convergência em L^p Mostre que a convergência em L^p implica convergência em Probabilidade para todos $p > 0$. Mostre que o recíproca não é válida. (Prove por exemplo para $p = 1$).

Ex. 9.11 — Considere as variáveis aleatórias $X_1, X_2, \ldots$ com funções de distribuição $F_1, F_2, \ldots$. Suponha que $X_n \to X$ em distribuição, e a função de distribuição F de X seja contínua. Prove que $F_n \to F$ na norma sup, i.e.

$$\|F_n - F\|_\infty := \sup_{x\in\mathbb{R}} |F_n(x) - F(x)| \to 0.$$

Ex. 9.12 — Considere as variáveis aleatórias a valores inteiros $X_1, X_2, \ldots$, e uma variável aleatória X. Mostre que $X_n \to X$ na distribuição se e somente se

$$\mathbf{P}(X_n = k) \to \mathbf{P}(X = k) \qquad \text{para todo } k \in \mathbb{Z}.$$

Ex. 9.13 — Considere variáveis aleatórias independentes $X_1, X_2, \ldots$ uniformemente distribuídas em [0,1], e deixe $Z_n = \max_{k\leq n} X_k$.

1. Prove que $Z_n \to 1$ em probabilidade.
2. Prove que $Z_n \to 1$ quase certamente.

Capítulo 10

Teoremas Limites

Neste capítulo trataremos de dois teoremas de importância central tanto na teoria da probabilidade quanto na estatística: a lei fraca e forte dos grandes números e apresentaremos algumas de suas aplicações.

Ambos os teoremas possuem diversas versões ou com hipóteses diferentes, outras para processos definidos de maneira diferente da que trataremos aqui.

Nós focaremos nas versões que exploram o comportamento assintótico de uma soma $S_n = X_1 + \cdots X_n$ de uma sequência de variáveis independentes e identicamente distribuídas.

Começaremos pela **Lei dos Grandes Números** que, de certa forma, vem confirmar a noção frequentista de probabilidade, tão usada e explorada na estatística.

Mais especificamente, dada uma sequência $X_1, X_2, \ldots$ de variáveis aleatórias independentes e integráveis, com média comum $\mu = \mathbf{E}[X_k] \in \mathbb{R}$, mostraremos que

$$\frac{S_n}{n} \to \mu,$$

onde o modo de convergência varia de acordo com versão do teorema e as hipóteses.

É comum também exigir que as variáveis sejam identicamente distribuídas, mas para as versões mais simples e diretas do teorema esta não é uma hipótese necessária.

Hipóteses sobre os momentos das variáveis estão entre as mais frequentes e a hipótese de independência, que estará presente em todas as versões que demonstraremos aqui, é uma das primeiras a serem relaxadas. Em geral ela é substituída por hipóteses sobre a variância da soma ou sobre a estrutura de dependência (correlações) das variáveis a serem somadas.

10.1 Lei Fraca dos Grandes Números

A primeira versão que trabalharemos é, talvez, a mais simples delas, e é uma consequência direta da Desigualdade de Chebyshev. Em geral, o termo *fraca* indica, neste contexto, a convergência em probabilidade da sequência S_n/n, mas no primeiro teorema que apresentaremos, a convergência em probabilidade será consequência da convergência em $\mathcal{L}^2$, que seguirá de forma direta.

10.1 Teorema (Lei Fraca dos Grandes Números) *Sejam* $X_1, X_2, \ldots$ *variáveis aleatórias independentes com* $\mathbf{E}[X_k] = \mu$ *e* $\mathbf{Var}(X_k) \leq c < \infty$, *para todo* $k \geq 1$. *Nestas condições, se* $S_n = X_1 + X_2 + \cdots + X_n$, *então*

$$\frac{S_n}{n} \xrightarrow{\mathcal{L}^2} \mu \qquad e \qquad \frac{S_n}{n} \xrightarrow{p} \mu.$$

Demonstração. *Precisamos mostrar apenas que*

$$\mathbf{E}\left[\frac{S_n}{n} - \mu\right]^2 \longrightarrow 0.$$

Lembre que a convergência em probabilidade é segue da convergência em $\mathcal{L}^2$ *através da desigualdade de Chebyshev.*

Observamos primeiro que pela linearidade da esperança $\mathbf{E}\left[\frac{S_n}{n}\right] = \mu$.

Então

$$\begin{aligned}
\mathbf{E}\left[\frac{S_n}{n} - \mu\right]^2 &= \mathbf{Var}(\frac{S_n}{n}) = \frac{1}{n^2}\mathbf{Var}(S_n) \\
&= \frac{1}{n^2}(\mathbf{Var}(X_1) + \cdots + \mathbf{Var}(X_n)) \\
&\leq \frac{cn}{n^2} \\
&= \frac{c}{n} \longrightarrow 0,
\end{aligned}$$

e o resultado segue.

A demonstração anterior nos fornece também uma taxa de convergência de, no máximo, $\frac{1}{n}$. Isso ainda não é suficiente para concluirmos a convergência quase certa, uma vez que precisaríamos de uma probabilidade somável, mas também não descarta essa possibilidade.

10.2 Observação *A demonstração apresentada é válida para variáveis aleatórias* X_k *dependentes, mas não correlacionadas. Isso por que o único lugar onde usamos a independência foi para afirmar que as somas das variâncias é a variância da soma, e esse resultado ainda é válido para variáveis não correlacionadas duas a duas.* $\triangleleft$

A demonstração anterior generaliza de maneira óbvia para vetores aleatórios.

10.3 Teorema (Lei fraca de grandes números) *Para uma sequência de vetores aleatórios i.i.d.* $X^{(n)}$, $n \in \mathbb{N}$ *com quarto momento finito* $\mathbf{E}(X^{(1)^4}) < \infty$, *então*

$$\frac{1}{n}\sum_{i=1}^{n} X^{(i)} \xrightarrow{P} \mathbf{E}[X^{(1)}].$$

Um caso particular do teorema anterior, onde se considera um sequência de ensaios de Bernoulli independentes, já era conhecido quando o teorema geral foi demonstrado pela primeira vez, e trata de maneira direta a ideia frequentista de probabilidade.

10.4 Teorema (Lei dos Grandes Números de Bernoulli) *Considere* S_n *uma variável aleatória binomial com parâmetros* (n,p), *isto é,* S_n *é o número de sucessos em* n *lançamentos independentes, onde a probabilidade onde a probabilidade de sucesso em cada lançamento é* p.

Então S_n *é soma de variáveis aleatórias independentes* X_k, *que são variáveis aleatórias de Bernoulli com parâmetro* p.*Então*

$$\frac{S_n}{n} \longrightarrow p$$

em probabilidade, quando $n \longrightarrow \infty$.

10.5 Exemplo *Lançando uma moeda n vezes: Então*

$$\mathbf{P}(0.49n \leq \text{ número de caras } \leq 0.51n) \longrightarrow 1 \quad \text{quando } n \longrightarrow \infty.$$

Agora, tentaremos melhorar a Lei Fraca dos Grandes Números. Primeiramente vamos remover a condição de variância finita. Mas, então, pediremos que as variáveis aleatórias X_k sejam identicamente distribuídas.

Nos exercícios, removeremos essa hipótese.

Para lidarmos com o fato que não temos variação finita usaremos o método de truncamento que remonta a Markov.

10.6 Teorema (Método de Truncamento de Markov) *Considere* X *uma variável aleatória, com média finita. Considere* $M > 1$ *o nível de truncamento.*

Defina

$$\lfloor X \rceil^M := \left\{ \begin{array}{ll} X, & se\ |X| \leq M \\ 0, & se\ |X| > M \end{array} \right\} = X \cdot \mathbf{1}_{\{|X| \leq M\}}.$$

Então

- $|\lfloor X \rceil^M| \leq M$ *pontualmente, e* $|\lfloor X \rceil^M| \leq |X|$
- $\lfloor X \rceil^M \longrightarrow X$ *pontualmente quando* $M \longrightarrow \infty$,
- *A variável* $\lfloor X \rceil^M$ *possui todos os momentos finitos.*
- *Pelo Teorema da Convergência Dominada* $\mathbf{E}[\lfloor X \rceil^M] \longrightarrow \mathbf{E}[X]$ *quando* $M \longrightarrow \infty$. *Mais ainda,*

$$\mathbf{E}|\lfloor X \rceil^M - X| \longrightarrow 0 \text{ quando } M \longrightarrow \infty. \tag{10.1}$$

10.7 Corolário *Dado uma variável aleatória* X *com média finita*

$$\mathbf{E}\left[|X| \cdot \mathbf{1}_{\{|X|>M\}}\right] \longrightarrow 0 \text{ quando } M \longrightarrow \infty.$$

Demonstração. *Observe que*

$$X - X(M) = X \cdot \mathbf{1}_{\{|X|>M\}}.$$

10.8 Teorema (Lei Fraca dos Grandes Números) *Sejam* $X_1, X_2, \ldots$ *variáveis aleatórias independentes identicamente distribuídas com* $\mathbf{E}[X_k] = \mu$ *para todo* k.

Considere $S_n = X_1 + X_2 + \cdots + X_n$.

Então

$$\frac{S_n}{n} \longrightarrow \mu$$

em probabilidade (quando $n \longrightarrow \infty$*).*

Demonstração. *Começaremos truncando. Dado* $M > 1$, *considere*

$$\lfloor X \rceil_k^M := X_k \cdot \mathbf{1}_{\{|X_k| \leq M\}}, \qquad S_n^{(M)} = X_1^{(M)} + \cdots + X_n^{(M)}.$$

Então

$$\mathbf{Var}[X_k^{(M)}] = \mathbf{E}[(X_k^{(M)}]^2 - (\mathbf{E}[X_k^{(M)}])^2 \leq M^2.$$

Então, podemos aplicar a versão antiga da Lei dos Grandes Números com variância finita:

$$\frac{S_n^{(M)}}{n} \longrightarrow \mathbf{E}[X_1^{(M)}]$$

em $\mathcal{L}^2$ *quando* $n \longrightarrow \infty$ *esse fato implica também convergência em probabilidade.*

$$\mathbf{E}\left|\frac{S_n^{(M)}}{n} - \mathbf{E}[X_1^{(M)}]\right| \longrightarrow 0 \text{ quando } n \longrightarrow \infty.$$

Agora aproximamos $S_n^{(M)}$ por S_n e $X_1^{(M)}$ por X_1. Pela desigualdade triangular

$$\mathbf{E}\left|\frac{S_n}{n} - \mathbf{E}[X_1]\right| \leq \mathbf{E}\left|\frac{S_n^{(M)}}{n} - \mathbf{E}[X_1^{(M)}]\right| + \mathbf{E}\left|\frac{S_n - S_n^{(M)}}{n}\right| + \mathbf{E}\left[|X_1 - X_1^{(M)}|\right].$$

Agora estimaremos o primeiro termo

$$\mathbf{E}\left|\frac{S_n - S_n^{(M)}}{n}\right| \leq \frac{1}{n}\sum_{k=1}^{n}\mathbf{E}|X_k - X_k^{(n)}| = \mathbf{E}|X_1 - X_1^{(M)}|.$$

Logo

$$\mathbf{E}\left|\frac{S_n}{n} - |\mathbf{E}[X_1]\right| \leq 2\mathbf{E}|X_1 - X_1^{(M)}| + \mathbf{E}\left|\frac{S_n^{(M)}}{n} - \mathbf{E}[X_1^{(M)}]\right|$$

Tomando $n \longrightarrow \infty$ concluímos que

$$\limsup \mathbf{E}\left|\frac{S_n}{n} - \mathbf{E}[X_1]\right| \leq 2\mathbf{E}|X_1 - X_1^{(M)}|.$$

Seja $M \longrightarrow \infty$. Então $\mathbf{E}|X_1 - X_1^{(M)}| \longrightarrow 0$. □

10.2 Lei Forte

Considere $X_1, X_2, \ldots$ variáveis aleatória independentes e identicamente distribuídas com esperança finita μ. Seja $S_n := X_1 + \cdots + X_n$. A Lei Fraca dos Grandes Números afirma que

$$\frac{S_n}{n} \longrightarrow \mu \text{ em probabilidade.}$$

Por outro lado, a Lei Forte dos Grandes Números que provaremos a seguir afirma que

$$\frac{S_n}{n} \longrightarrow \mu \text{ q.c.}$$

Começaremos apresentando uma demonstração simples da Lei Forte dos Grandes Números sob o pressuposto de que o quarto momento é finito.

10.9 Teorema (Lei Forte dos Grandes Números - I) *Sejam $X_1, X_2, \ldots$ variáveis aleatórias i.i.d. com esperança μ e assuma $\mathbf{E}\left[X_k^4\right] < \infty$. Então, $S_n = X_1 + X_2 + \cdots + X_n$ satisfaz*

$$\frac{S_n}{n} \longrightarrow \mu \text{ quase certamente.}$$

Demonstração. *Primeiro assumimos que a esperança $\mu = 0$.*

De fato, fazendo $X_k' = X_k - \mu$, temos

$$\frac{\sum_{k=1}^n X_k'}{n} = \frac{\sum_{k=1}^n (X_k - \mu)}{n} = \frac{S_n}{n} - \mu,$$

e assim o resultado geral segue do caso $\mu = 0$.

Lembrando, nosso objetivo agora é mostrar que para qualquer $\varepsilon > 0$,

$$\mathbf{P}\left(\left|\frac{S_n}{n}\right| > \varepsilon \ \ i.v.\right) = 0. \tag{10.2}$$

Isso será feito a partir da desigualdade de Chebyshev, se valendo do fato das variáveis X_k *possuírem quarto momento finito.*

Para começar, calculemos primeiro o quarto momento de S_n.

$$\mathbf{E}\left[S_n^4\right] = \mathbf{E}\left[(\sum_{k=1}^{n} X_k)^4\right] = \sum_{1\le i,j,k,\ell\le n} \mathbf{E}\left[(X_iX_jX_kX_\ell)\right]$$

Observe que se i, j *são distintos, então, pela independência,* $\mathbf{E}\left[X_i^3X_j\right] = \mathbf{E}\left[X_i^3\right]\mathbf{E}[X_j] = 0$ *(lembre que* $\mu = 0$*) . Da mesma forma, temos* $\mathbf{E}\left[X_i^2X_j\right] = 0$ *e* $\mathbf{E}\left[X_iX_jX_kX_\ell\right] = 0$*, se* i, j, k, ℓ *forem diferentes.*

Desta forma os únicos termos que não são zero são da forma:

$$\mathbf{E}\left[X_k^4\right] \quad e \quad \mathbf{E}\left[X_j^2X_k^2\right].$$

Agora uma contagem simples nos mostra que existem n *termos da forma* $\mathbf{E}[X_k^4]$*, e outros* $\binom{n}{2}\binom{4}{2} = 3n(n-1)$ *termos da forma* $\mathbf{E}\left[X_j^2X_k^2\right]$.

Como as variáveis envolvidas são identicamente distribuídas, então

$$\mathbf{E}[X_k^4] = \mathbf{E}[X_1^4] \qquad e \qquad \mathbf{E}\left[X_j^2X_k^2\right] = \mathbf{E}\left[X_1^2X_2^2\right].$$

Assim, se $C = \max\{\mathbf{E}\left[X_1^2X_2^2\right], \mathbf{E}\left[X_1^4\right]\}$*, então*

$$\mathbf{E}\left[S_n^4\right] = n\mathbf{E}[X_1^4] + (3n^2 - 3n)\mathbf{E}\left[X_1^2X_2^2\right] \le C \cdot (3n^2 - 2n) \le 3Cn^2.$$

Pela desigualdade de Chebyshev, temos então que

$$\begin{aligned}\mathbf{P}\left(\left|\frac{S_n}{n}\right| > \varepsilon\right) &= \mathbf{P}(S_n^4 > \varepsilon^4 n^4)\\ &\le \frac{\mathbf{E}S_n^4}{\varepsilon^4 n^4}\\ &\le \frac{3C}{\varepsilon^4 n^2}\end{aligned}$$

E como a série $\displaystyle\sum_{n=1}^{\infty} \frac{3C}{\varepsilon^4 n^2}$ *converge, segue do Lema de Borel-Cantelli que*

$$\mathbf{P}\left(\left|\frac{S_n}{n}\right| > \varepsilon \ \ i.v.\right) = 0.$$

A demonstração anterior generaliza facilmente para vetores aleatórios

10.10 Teorema (Lei Forte dos Grandes Números) *Para uma sequência de vetores aleatórios i.i.d. $X^{(n)}$, $n \in \mathbb{N}$ com quarto momento finito $\mathbf{E}({X^{(1)}}^4) < \infty$ então*

$$\frac{1}{n}\sum_{i=1}^{n} X^{(i)} \xrightarrow{q.c.} \mathbf{E}(X^{(1)}).$$

10.11 Proposição (Esperança Finita é Necessária) *Sejam $X_1, X_2, \ldots$ variáveis aleatórias independentes e identicamente distribuídas com esperança infinita: $\mathbf{E}|X_k| = \infty$. Então $S_n = X_1 + X_2 + \cdots + X_n$ satisfaz*

$$\mathbf{P}\left(\frac{S_n}{n} \quad \textit{convergir}\right) = 0.$$

Demonstração. *Como $\mathbf{E}|X_1| = \infty$ então $\sum_{n=0}^{\infty} \mathbf{P}(|X_1| > n)$ diverge.*

Como as variáveis aleatórias X_k são identicamente distribuídas $\mathbf{P}(|X_1| > n) = \mathbf{P}(|X_n| > n)$ para $n = 1,2,\ldots$ Logo $\sum_{n=0}^{\infty} \mathbf{P}(|X_n| > n) = \infty$. E pelo Lema de Borel-Cantelli,

$$\mathbf{P}(|X_n| > n \quad i.v.) = 1.$$

Agora, definimos dois eventos:

$$A := \{|X_n| > n \quad i.v.\} \qquad (\textit{ Logo, } \mathbf{P}(A) = 1)$$

e

$$B := \left\{\frac{S_n}{n} \textit{ convergir}\right\}.$$

Provaremos que $A \cap B = \emptyset$. Se essa afirmação for verdadeira, então

$$\mathbf{P}(B) \leq \mathbf{P}(\Omega \setminus A) = 1 - \mathbf{P}(A) = 1 - 1 = 0.$$

Para provar que $A \cap B = \emptyset$, vamos supor por contradição que $A \cap B \neq \emptyset$.

Consideremos a diferença

$$\frac{S_n}{n} - \frac{S_{n+1}}{n+1} = \frac{S_n}{n(n+1)} - \frac{X_{n+1}}{n+1}.$$

Como $A \cap B \neq \emptyset$ então, para algum evento,

$$\frac{S_n}{n(n+1)} \longrightarrow 0.$$

Por outro lado,

$$\left|\frac{X_{n+1}}{n+1}\right| > 1 \quad i.v.,$$

então quando n é suficientemente grande,

$$\left|\frac{S_n}{n} - \frac{S_{n+1}}{n+1}\right| > \frac{2}{3} \quad i.v.,$$

Portanto, $\frac{S_n}{n}$ não é de Cauchy. Ou seja, $\frac{S_n}{n}$ não converge, o que é uma contradição. Portanto, a interseção de A e B está realmente vazia. □

10.12 Teorema (Lei Forte dos Grandes Números de Kolmogorov) *Sejam $X_1, X_2, \ldots X_n$ variáveis aleatórias independentes e identicamente distribuídas com esperança finita, i.e. $\mathbf{E}X_i = \mu$. Então, $S_n = X_1 + \cdots + X_n$, satisfaz*

$$\frac{S_n}{n} \longrightarrow \mu \quad \text{quase certamente.}$$

Demonstração. *Nossa demonstração seguirá alguns passos. Começaremos tomando variáveis positiva, de onde o caso geral seguirá naturalmente. Em seguida seguiremos para um truncamento das variáveis envolvidas, que garantirá a variância finita. Seguimos para um resultado técnico, sobre uma série envolvendo as variâncias. Este resultado nos permitirá verificar a convergência por subsequências escolhidas. Terminamos comparando as subsequências com o resto da sequência.*

- *Passo 0. Primeiro assumimos que $X_i \geq 0$. De fato, se escrevermos X_i*

  $$X_i = X_i^+ - X_i^-,$$

  então X_i^+ e X_i^- são integráveis e

  $$\frac{S_n}{n} = \frac{\sum_{i=1}^n X_i^+}{n} - \frac{\sum_{i=1}^n X_i^-}{n},$$

  e o resultado geral segue naturalmente do resultado para variáveis positivas.

- *Passo 1 - Truncamento.*

  Defina $\overline{X}_k = X_k \mathbb{1}_{\{X_k \leq k\}}$, e observe que $\mathbf{P}(\overline{X}_k \neq X_k) = \mathbf{P}(X_k > k)$. Assim

  $$\sum_{k=1}^{\infty} \mathbf{P}(X_k > k) = \sum_{k=1}^{\infty} \mathbf{P}(X_1 > k) \leq \int_0^{\infty} \mathbf{P}(X_1 > x)\, dx = \mathbf{E}X_1 < +\infty,$$

  e pela independência das variáveis e o Lema de Borel-Cantelli, concluímos que

  $$\mathbf{P}(\overline{X}_k \neq X_k \ \ i.v.) = 0.$$

  Assim, se demonstrarmos que

  $$\frac{\overline{X}_1 + \overline{X}_2 + \cdots + \overline{X}_n}{n} \xrightarrow{q.c.} \mu,$$

então

$$\frac{X_1 + X_2 + \cdots + X_n}{n} \xrightarrow{q.c.} \mu$$

quase certamente.

No restante da demonstração podemos assumir, portanto, que $X_k \leq k$.

- *Passo 2 - Um pequeno lema técnico.*

  Vamos mostrar que

$$\sum_{k=1}^{\infty} \frac{\mathbf{Var}(X_k)}{k^2} < +\infty.$$

Para isso note que

$$\begin{aligned}
\mathbf{Var}(X_k) &\leq \mathbf{E}(X_k^2) \\
&= \int_0^{\infty} \mathbf{P}(X_k^2 \geq x)\, dx \\
&= \int_0^{\infty} 2t\mathbf{P}(X_k \geq t)\, dt \\
&= \int_0^{\infty} 2t\mathbf{P}(X_k \geq t)\mathbb{1}_{\{t \leq k\}}\, dt \\
&= \int_0^{\infty} 2t\mathbf{P}(X_1 \geq t)\mathbb{1}_{\{t \leq k\}}\, dt,
\end{aligned}$$

onde na segunda igualdade fazemos $x = t^2$, *e ao final usamos que as variáveis são identicamente distribuídas.*

Deste modo, pelo Teorema de Fubini, segue que

$$\begin{aligned}
\sum_{k=1}^{\infty} \frac{\mathbf{Var}(X_k)}{k^2} &= \sum_{k=1}^{\infty} \int_0^{\infty} \frac{\mathbb{1}_{\{x \leq k\}}}{k^2} 2x\mathbf{P}(X_1 \geq x)\, dx \\
&= \int_0^{\infty} \sum_{k=1}^{\infty} \frac{\mathbb{1}_{\{x \leq k\}}}{k^2} 2x\mathbf{P}(X_1 \geq x)\, dx
\end{aligned}$$

Agora, sabendo que $\displaystyle\sum_{k \geq x} \frac{1}{k^2} \leq \int_x^{\infty} \frac{dt}{t^2} = \frac{1}{x}$, *temos que*

$$\sum_{k=1}^{\infty} \frac{\mathbf{Var}(X_k)}{k^2} \leq \int_0^{\infty} 2\mathbf{P}(X_1 \geq x)\, dx = 2\mathbf{E}[X_1] < +\infty.$$

- *Passo 3. - Controle ao longo de uma subsequência*

  Dado $\alpha > 1$, *defina* $k(n) = \lfloor \alpha^n \rfloor$, *para* $n \geq 1$.

É um exercício simples verificar que $k(n) = \lfloor \alpha^n \rfloor \geq \alpha^n/2$.

Queremos agora verificar que

$$\frac{S_{k(n)}}{k(n)} \xrightarrow{q.c.} \mu.$$

Para isso tome $\varepsilon > 0$, *e note que pela desigualdade de Chebyshev e pela independência das variáveis,*

$$\begin{aligned}
\sum_{n=1}^{\infty} \mathbf{P}\left(\left|\frac{S_{k(n)}}{k(n)} - \mu\right| > \varepsilon\right) &= \sum_{n=1}^{\infty} \mathbf{P}\left(|S_{k(n)} - \mathbf{E}[S_{k(n)}]| > \varepsilon \cdot k(n)\right) \\
&\leq \sum_{n=1}^{\infty} \frac{\mathbf{Var}\left(S_{k(n)}\right)}{(\varepsilon k(n))^2} \\
&= \frac{1}{\varepsilon^2} \sum_{n=1}^{\infty} \frac{1}{k(n)^2} \sum_{k=1}^{k(n)} \mathbf{Var}(X_k) \\
&= \frac{1}{\varepsilon^2} \left[\sum_{k=1}^{\infty} \mathbf{Var}(X_k) \sum_{n:k(n)\geq k} \frac{1}{k(n)^2}\right].
\end{aligned}$$

Mas como $k(n) \geq \alpha^n/2$, *então*

$$\begin{aligned}
\sum_{n:\alpha^n \geq k} \frac{1}{k(n)} &= \sum_{n:\alpha^n \geq k} \frac{1}{\lfloor \alpha^n \rfloor^2} \\
&\leq \sum_{n:\alpha^n \geq k} \frac{4}{\alpha^{2n}} \\
&\leq \sum_{n \geq \log_\alpha k} \frac{4}{\alpha^{2n}} \\
&\leq \frac{4\alpha^2}{(\alpha^2 - 1)\, k^2}
\end{aligned}$$

Segue que

$$\frac{1}{\varepsilon^2} \sum_{k=1}^{\infty} \left[\mathbf{Var}(X_k) \sum_{n:k(n)\geq k} \frac{1}{k(n)^2}\right] \leq \frac{4\alpha^2}{\varepsilon^2\, (\alpha^2 - 1)} \sum_{k=1}^{\infty} \frac{\mathbf{Var}(X_k)}{k^2},$$

e assim

$$\sum_{n=1}^{\infty} \mathbf{P}\left(\left|\frac{S_{k(n)}}{k(n)} - \mu\right| > \varepsilon\right) \leq \frac{4\alpha^2}{\varepsilon^2\, (\alpha^2 - 1)} \sum_{k=1}^{\infty} \frac{\mathbf{Var}(X_k)}{k^2} < \infty$$

Logo, pelo Lema de Borel-Cantelli

$$\mathbf{P}\left(\left|\frac{S_{k(n)}}{k(n)} - \mu\right| > \varepsilon \;\; i.v.\right) = 0$$

e portanto

$$\frac{S_{k(n)}}{k(n)} \xrightarrow{q.c.} \mu.$$

□ *Passo 4. - Preenchendo lacunas na sequência.*

Começaremos observando que se $k(n) < k < k(n+1)$ então

$$S_{k(n)} \leq S_k \leq S_{k(n+1)},$$

e portanto, como as variáveis X_i são positivas,

$$\frac{k(n)}{k(n+1)} \frac{S_{k(n)}}{k(n)} = \frac{S_{k(n)}}{k(n+1)} \leq \frac{S_k}{k} \leq \frac{S_{k(n+1)}}{k(n)} = \frac{k(n+1)}{k(n)} \frac{S_{k(n+1)}}{k(n)}.$$

Como $k(n) = \lfloor \alpha^n \rfloor$, temos que $\alpha^n - 1 < k(n) \leq \alpha^n$, e portanto

$$\lim_{n\to\infty} \frac{k(n+1)}{k(n)} = \alpha.$$

Consequentemente,

$$\frac{\mu}{\alpha} \leq \liminf_k \frac{S_k}{k} \leq \limsup_k \frac{S_k}{k} \leq \alpha\mu$$

com a probabilidade um.

Uma vez que $\alpha > 1$ é arbitrário, podemos fazer $\alpha \longrightarrow 1$.Assim, o limite existe e

$$\lim_{k\to\infty} \frac{S_k}{k} = \mu$$

quase certamente. □

10.3 Integração de Monte-Carlo

Considere uma função integrável $f : [0,1] \to \mathbb{R}$.

O problema que colocamos nessa seção é descrever um método numérico para integrar f. Ou seja, calcular $\int_0^1 f(x)\,dx$. Ressaltamos que a escolha do intervalo [0,1] aqui é arbitrária.

A ideia clássica seria aproximar a integral por somas de Riemann, ou seja fazer uma partição equidistribuída de pontos em [0,1]. Ou seja escolhemos uma sequência de pontos equidistantes $x_1, \ldots, x_n$. Em cada ponto x_k, determinamos o valor da função nesse ponto. Então esperamos que integral possa ser bem-aproximada pela média:

$$\int_0^1 f(x)\,dx \approx \frac{1}{n} \sum_{k=1}^{n} f(x_k).$$

Esta é uma integral Riemann, porque o tamanho da malha $\frac{1}{n}$ vai para zero. Em situações muito simples, funciona. Mas quase sempre falha nos problemas difíceis da computação científica. Por que? Existe um problema com essa abordagem.

A priori não conhecemos nada sobre esta função, exceto que é integrável. Em particular, se f é complicada, podemos sempre escolher os pontos errados. Ou seja a estrutura de f pode levar a informações erradas sobre a integral dos pontos $f(x_k)$. Por exemplo, existem muitas funções "oscilantes"(de senos e cossenos).

O resultado que vamos provar permite a integração de funções arbitrárias f. Não iremos tomar pontos equidistribuídos mas em vez disso, usaremos pontos aleatórios.

Consideramos variáveis aleatórias $x_1, \ldots, x_n$ independentes e distribuídas uniformemente em [0,1].

Então, esperamos que

$$I_n = \frac{1}{n} \sum_{k=1}^{n} f(x_k)$$

seja uma boa aproximação de $\int_0^1 f(x)\,dx$.

Bem, isso é apenas uma reformulação da Lei Fraca dos Grandes Números. Em outras palavras,

10.13 Teorema (Integração de Monte-Carlo) *Se cada x_k é uma variável aleatória independente uniforme em $[0,1]$ então $I_n \to \int_0^1 f(x)\,dx$ em probabilidade.*

Demonstração. *Começamos observando que f é uma variável aleatória no espaço de probabilidade $\Omega = [0,1]$ com a σ-álgebra de Borel e a medida de probabilidade uniforme.*

Então $f \stackrel{d.}{=} F(x_k)$

Uma vez que vemos isso é verdade, então $\mathbf{E}f = \int_0^1 f(x)\,dx$. Então, a Lei Fraca dos Grandes Números aplicada às variáveis aleatórias $f(x_1), \ldots, f(x_n)$ completa a prova. □

Este é um dos primeiros algoritmos randomizados em computação científica. A principal força do método é que não exige nenhum conhecimento sobre f. Exigimos apenas que f seja integrável.

10.4 Polinômios de Bernstein e o Teorema de Weierstrass

Esta seção é adaptada de [23]

Polinômios de Bernstein

Os polinômios de Bernstein são definidos como

$$b_{n,j}(x) = \binom{n}{j} x^j (1-x)^{n-j}$$

para $j = 0,1,\ldots,n$.

- Os polinômios de Bernstein $b_{n,j}$ são uma família de $n+1$ polinômios de grau n.
- Cada polinômio de Bernstein com $j \neq 0, n$ tem um valor máximo único que ocorre em j/n.
- Os polinômios de Bernstein de grau n formam a base para os polinômios de grau n.
- Os polinômios de Bernstein possuem a simetria $b_{n,j}(x) = b_{n,n-j}(1-x)$, positividade $b_{i,n}(x) \geq 0$ em $[0,1]$ e normalização $\sum_{j=0}^{n} b_{n,j}(x) = 1$.

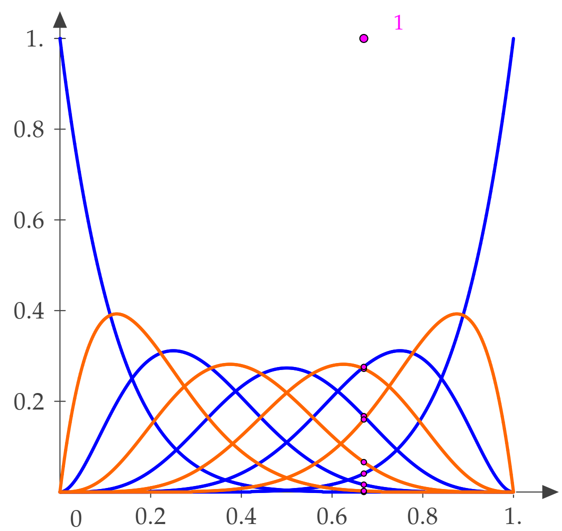


Figura 10.1: Polinômios de Bernstein $b_{i,8}$ para $i = 1,\ldots,8$. Em pontilhado o polinômio $b_{3,8}$. Observe que as somas dos valores em um ponto é 1.

Teorema de Aproximação de Weierstrass

Agora apresentaremos uma demonstração elegante Teorema de Aproximação de Weierstrass. Este teorema afirma que se f é uma função contínua, então a sequência de polinômios de Bernstein

$$B_n[f](\mathbf{x}) = \sum_{0 \leq i_1,\ldots,i_d \leq n} f\left(\frac{i_1}{n},\ldots,\frac{i_d}{n}\right) \prod_{k=1}^{d} \binom{n}{i_k} x_k^{i_k} (1-x_k)^{n-i_k}$$

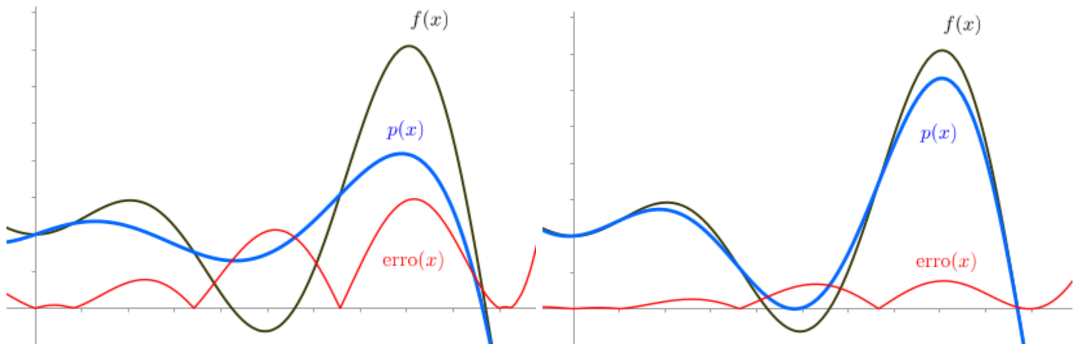


Figura 10.2: Aproximação com $n = 10{,}50$

aproxima f.

A estrutura e a origem dos polinômios tornam-se muito mais claras quando é apresentada a interpretação probabilística subjacente. O coração dessa conexão é a fórmula

$$B_n[f](x) = \mathbf{E}\left[f\left(\frac{\mathbf{S}_n}{n}\right)\right]$$

onde $\mathbf{S}_n$ é um vetor binomial.

A ideia geral da demonstração é a seguinte: seja X_{ij} uma variável aleatória de Bernoulli, para $1 \leq i \leq d$ e $j \geq 1$ tal que

$$\mathbf{P}\left[\mathrm{X}_{ij} = 1\right] = x_i \text{ e } \mathbf{P}\left[\mathrm{X}_{ij} = 0\right] = 1 - x_i.$$

Seja

$$S_{in} = \sum_{j=1}^{n} \mathrm{X}_{ij}$$

e

$$\mathbf{S}_n = (S_1, \ldots, S_n).$$

Pela Lei Fraca ou pela Lei Forte de Grandes números podemos afirmar que $\mathbf{S_n}/n \approx \mathbf{x}$ em algum sentido. Por continuidade, $f(\mathbf{S}_n/n) \approx f(\mathbf{x})$. Então, temos $\mathbf{E}[f(\mathbf{S}_n/n)] \approx f(\mathbf{x})$. Mas esse fato é equivalente a afirmar que $B_n[f](\mathbf{x}) \approx f(\mathbf{x})$, e podemos ver que $B_n[f](\mathbf{x})$ é um polinômio em $\mathbf{x}$.

10.14 Teorema (Teorema de Aproximação de Weierstrass) *Seja d ser um número inteiro positivo e $f : [0,1]^d \to \mathbb{C}$ uma função contínua. Defina*

$$B_n[f](\mathbf{x}) = \sum_{0 \leq i_1, \ldots, i_d \leq n} f\left(\frac{i_1}{n}, \ldots, \frac{i_d}{n}\right) \prod_{k=1}^{d} \binom{n}{i_k} x_k^{i_k} (1 - x_k)^{n - i_k}$$

para $n = 1,2,\ldots$ e $\mathbf{x} = (x_1, \ldots, x_n)$ em $[0,1]^d$, então $B_n[f] \to f$ uniformemente em $[0,1]^d$.

Demonstração. *Considere um número inteiro positivo n e $\mathbf{x} = (x_1, \dots, x_d)$ em $[0,1]^d$. Seja $X_j = (X_{1j}, \dots X_{dj})$ ser uma sequência finita de vetores aleatórios definidos em um espaço de probabilidade, de modo que $\{X_{ij}\}$ sejam independentes e cada uma e assuma valores 0 ou 1 de acordo com a distribuição de Bernoulli: $\mathbf{P}\left[X_{ij} = 1\right] = x_i$ e $\mathbf{P}\left[X_{ij} = 0\right] = 1 - x_i$.*

Escreva $S_{in} = \sum_{j=1}^{n} X_{ij}$ e $\mathbf{S}_n = (S_{1n}, \dots, S_{dn})$. Então,

$$\mathbf{P}\left[S_{kn}/n = i_k/n\right] = \binom{n}{i_k} x_k^{i_k}(1 - x_k)^{n-i_k}$$

Assim se

$$B_n[f](\mathbf{x}) = \sum_{0 \le i_1, \dots, i_d \le n} f\left(\frac{i_1}{n}, \dots, \frac{i_d}{n}\right) \prod_{k=1}^{n} \binom{n}{i_k} x_k^{i_k}(1 - x_k)^{n-i_k}$$

então

$$B_n[f](x) = \mathbf{E}\left[f\left(\frac{\mathbf{S}_n}{n}\right)\right]$$

Provaremos esse fato agora utilizando a independência e a definição da esperança. Começamos observando que

$$\sigma_{j,n}^2 = \mathbf{Var}[S_{kn}/n] = \frac{x_k(1 - x_k)}{n} \le \frac{1}{4n}.$$

Também temos que $\mathbf{E}\left[\mathbf{S}_n\right] = n\mathbf{x}$, de onde temos

$$B_n[f](\mathbf{x}) - f(\mathbf{x}) = \mathbf{E}\left[f\left(\frac{\mathbf{S}_n}{n}\right) - f(\mathbf{x})\right] = \mathbf{E}\left[f\left(\frac{\mathbf{S}_n}{n}\right) - f(\mathbf{E}\left[\frac{\mathbf{S}_n}{n}\right])\right].$$

Definimos $\|\mathbf{x}\| = \max_k\{|x_k|\}$ para $\mathbf{x} \in \mathbb{R}^d$. O evento que o máximo das coordenadas é maior do que um valor está contido na união dos eventos em que cada coordenada é maior do que o valor.

Então, aplicando a desigualdade de Chebyshev, e a independência, obtemos para qualquer $\delta > 0$ o limite superior

$$\mathbf{P}\left[\left\|\frac{\mathbf{S}_n}{n} - \mathbf{x}\right\| \ge \delta\right] \le \sum_{1 \le k \le d} \mathbf{P}\left[\left|\frac{S_{kn}}{n} - x_k\right| \ge \delta\right] \le \sum_{1 \le k \le n} \frac{\sigma_{kn}^2}{\delta^2} \le \frac{d}{4n\delta^2}.$$

$$
\begin{aligned}
\mathbf{P}\left[\left\|\frac{\mathbf{S}_n}{n}-\mathbf{x}\right\| \geq \delta\right] &= 1-\mathbf{P}\left[\left\|\frac{\mathbf{S}_n}{n}-\mathbf{x}\right\| < \delta\right] \\
&= 1-\left(\prod_{1\leq k\leq d}\mathbf{P}\left[\left|\frac{S_{kn}}{n}-x_k\right| < \delta\right]\right) \\
&= 1-\left(\prod_{1\leq k\leq d}\left(1-\mathbf{P}\left[\left|\frac{S_{kn}}{n}-x_k\right| \geq \delta\right]\right)\right) \\
&= \sum_{1\leq k\leq d}\mathbf{P}\left[\left|\frac{S_{kn}}{n}-x_k\right| \geq \delta\right] - \sum_{1\leq k,l\leq d}\left(\mathbf{P}\left[\left|\frac{S_{kn}}{n}-x_k\right| \geq \delta\right]\right) \\
&\cdot \left(\mathbf{P}\left[\left|\frac{S_{ln}}{n}-x_l\right| \geq \delta\right]\right) + \dots \\
&\leq \sum_{1\leq k\leq d}\mathbf{P}\left[\left|\frac{S_{kn}}{n}-x_k\right| \geq \delta\right]
\end{aligned}
$$

Agora seja $\varepsilon/2 > 0$. Pela continuidade uniforme de f em $[0,1]^d$, existe um $\delta > 0$ tal que $|f(\mathbf{x}) - f(\mathbf{y})| < \varepsilon/2$ sempre que $\|\mathbf{x}-\mathbf{y}\| < \delta$. Então

$$
\int_{\|\mathbf{S}_n/n-\mathbf{x}\|<\delta}\left|f\left(\frac{\mathbf{S}_n}{n}\right)-f(\mathbf{x})\right| d\mathbf{P} \leq \frac{\varepsilon}{2}.
$$

Então, no evento complementar, usando a desigualdade triangular e estimativa da probabilidade do evento complementar:

$$
\begin{aligned}
\int_{\|\mathbf{S}_n/n-\mathbf{x}\|\geq\delta}\left|f\left(\frac{\mathbf{S}_n}{n}\right)-f(\mathbf{x})\right| d\mathbf{P} &\leq 2\max_{\mathbf{x}\in[0,1]^d}|f(\mathbf{x})|\cdot\mathbf{P}\left[\left\|\frac{\mathbf{S_n}}{n}-\mathbf{x}\right\|\geq\delta\right] \\
&\leq \frac{2d}{4n\delta^2}\max_{\mathbf{x}\in[0,1]^d}|f(\mathbf{x})|.
\end{aligned}
$$

Observamos que, para

$$
n > \frac{d}{\varepsilon\delta^2}\max_{\mathbf{x}\in[0,1]^d}|f(\mathbf{x})|
$$

O lado direito é inferior a $\varepsilon/2$.

Finalmente, considere

$$
\left|\mathbf{E}\left[f\left(\frac{\mathbf{S}_n}{n}\right)-f\left(\mathbf{E}\left[\frac{\mathbf{S}_n}{n}\right]\right)\right]\right|.
$$

Usando a desigualdade triangular para integrais para trazer o módulo para dentro da integral, e dividindo o domínio de integração em domínios complementares $\|\mathbf{S}_n/n-\mathbf{x}\| \geq \delta$ e $\|\mathbf{S}_n/n-\mathbf{x}\| < \delta$. Agora aplicaremos as estimativas estabelecidas acima. Então, para

$$
n > \frac{d}{\varepsilon\delta^2}\max_{\mathbf{x}\in[0,1]^d}|f(\mathbf{x})|
$$

independente de x, temos

$$\left|\mathbf{E}\left[f\left(\frac{\mathbf{S}_n}{n}\right) - f\left(\mathbf{E}\left[\frac{\mathbf{S}_n}{n}\right]\right)\right]\right| < \varepsilon$$

e o teorema é estabelecido. □

Exercícios

Ex. 10.1 — Prove o Método de Tuncamento de Markov:
Considere X uma variável aleatória, com média finita. Considere $M > 1$ o nível de truncamento.
Defina

$$\lfloor X \rceil^M := \left\{ \begin{array}{ll} X, & \text{se } |X| \le M \\ 0, & \text{se } |X| > M \end{array} \right\} = X \cdot \mathbf{1}_{\{|X| \le M\}}.$$

Então

- □$|\lfloor X \rceil^M| \le M$ pontualmente, e $|\lfloor X \rceil^M| \le |X|$
- □$\lfloor X \rceil^M \to X$ pontualmente quando $M \to \infty$,
- □A variável $\lfloor X \rceil^M$ possui todos os momentos finitos.
- □Pelo Teorema da Convergência Dominada $\mathbf{E}[\lfloor X \rceil^M] \to \mathbf{E}[X]$ quando $M \to \infty$.

Ex. 10.2 — Sejam $X_1, X_2, \ldots$ variáveis aleatórias independentes com distribuição comum No(0,1). Prove que

$$\frac{X_1^2 + \cdots + X_n^2}{(X_1 - 1)^2 + \cdots + (X_n - 1)^2}$$

converge quase certamente e determine para qual valor.

Ex. 10.3 — Sejam $X_1, X_2, \ldots$ variáveis aleatórias independentes e identicamente distribuídas, com $X_1 \sim$ Uni[0,1]. Prove que a média geométrica

$$\left(\prod_{k=1}^{n} X_k\right)^{1/n}$$

converge quase certamente e determine para qual valor.
Dica: Aplique ln.

Ex. 10.4 — Demonstração da Lei Fraca usando Funções Geradoras
O objetivo aqui é provar a lei (fraca) de grandes números usando funções geradoras de momento.
Em particular, deixe $X_1, X_2, \ldots, X_n$ ser i.i.d. variáveis aleatórias com valor esperado

$$\mathbf{E}[X_i] = \mu < \infty$$

e funções geradoras de momento $M_X(s)$ que é finito em algum intervalo $[-c, c]$ onde $c > 0$ é uma constante. Seja

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}. \tag{10.3}$$

Prove

$$\lim_{n\to\infty} M_{\overline{X}}(s) = e^{s\mu}, \qquad \text{para todo } s \in [-c, c]. \tag{10.4}$$

Uma vez que esta é a função geradora de momento da variável aleatória constante μ, concluímos que a distribuição de $\overline{X}$ converge para μ.
Dica: use que para uma variável aleatória X com funções geradoras de momento bem definida, $M_X(s)$, temos

$$\lim_{n\to\infty} \left[M_X\left(\frac{s}{n}\right)\right]^n = e^{s\mathbf{E}[X]}.$$

Ex. 10.5 — Sejam $X_1, X_2, \ldots$ variáveis aleatórias independentes com $E[X_i] = 0$ e $\sigma_i = \mathbf{Var}(X_i) < \infty$ para todo i, e seja $S_n = \sum X_i$. Prove que se $\sum \sigma_k^2/k^2 < \infty$, então $S_n/n \xrightarrow{q.c} 0$.
[Esta é uma Lei Forte de Grandes Números para somas independentes, mas não necessariamente identicamente distribuídas.]

Ex. 10.6 — Seja $X_1, X_2, \ldots$ variáveis aleatórias i.i.d. com $\mathbf{P}[X_k = 0] = \mathbf{P}[X_k = 2] = 1/2$.

1. Prove que $Z = \sum_{k=1}^{\infty} X_k/3^k$ converge quase certamente.
2. Prove que Z tem distribuição de Cantor.
3. Use os itens anteriores para calcular a média e variância da distribuição Cantor.

Ex. 10.7 — Integração de Monte Carlo

Sejam f uma função contínua com imagem no intervalo [0,1] e X_1,c_1,X_2,$c_2 \cdots$ uma sequência de variáveis aleatórias independentes uniformes no [0,1]. Tomamos

$$Y_i = \begin{cases} 1, & \text{se } f(X_i) > c_i \\ 0, & \text{se } f(X_i) \le c_i. \end{cases}$$

Prove que:

$$\frac{1}{n}\sum_{i=1}^{n} Y_i \to \int_0^1 f(x)dx, \quad \text{quase certamente}$$

CAPÍTULO 11

Teorema do Limite Central

Nos capítulos anteriores vimos que se X_1, X_2 é uma sequência de variáveis aleatórias independentes e identicamente distribuídas com média comum $\mu = \mathbf{E}[X_1]$ finita, então

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n}(X_k - \mu) = 0,$$

com probabilidade 1. Em particular, se $\mu = 0$ então

$$\frac{S_n}{n} \xrightarrow{q.c.} 0,$$

quando $n \to \infty$, onde $S_n = X_1 + \cdots + X_n$. Ou seja $|S_n| << n$ quando $n \to \infty$.

A pergunta então passa a ser qual a verdadeira taxa de convergência de $|S_n|$? Em outras palavras, queremos saber se existe alguma sequência $(a_n)_{n\geq 1}$, tal que

$$\frac{S_n}{a_n} \to Z \neq 0,$$

quando $n \to \infty$, de algum modo (não necessariamente quase-certamente), para alguma variável Z.

Para buscar uma candidata a sequência a_n, vamos começar supondo que $\mathrm{Var}[X_k] = \sigma^2 < \infty$. Pela independência das variáveis, segue que

$$\mathrm{Var}\left[\frac{S_n}{a_n}\right] = \frac{\mathrm{Var}[S_n]}{a_n^2} = \frac{n}{a_n^2}\sigma^2,$$

e se fizermos $a_n = \sqrt{n}$, teremos

$$\mathrm{Var}\left[\frac{S_n}{a_n}\right] = \sigma^2,$$

para todo $n \geq 1$.

Com isso, vemos que se o limite de $S_n/\sqrt{n}$ existir de algum modo, este limite não deve ser constante, o que nos indica que podemos ver algo interessante nesta escala.

O mais simples neste caso seria investigar a convergência em distribuição, e é neste sentido que segue o Teorema do Limite Central.

Tal teorema, de papel central na teoria da probabilidade e na estatística, nos revela não apenas que a sequência $S_n/\sqrt{n}$ possui um limite em distribuição, mas que tal limite é universal, não dependendo da distribuição das variáveis individuais, pedindo apenas que sejam quadrado integráveis.

É essa universalidade do limite que nos permite, por exemplo, calcular o erro por trás de diversas estimativas feitas na estatística, além de explicar o comportamento "Gaussiano" de diversas observações definidas pela ação sucessiva de diversos efeitos independentes similares.

Assim como a Lei do Grandes Números, o Teorema do Limite Central possui diversas versões diferentes, alterando especialmente as hipóteses sobre a independência das variáveis e suas variâncias.

11.1 Funções Características

Nesta seção, estudaremos a convergência fraca de variáveis aleatórios mais cuidadosamente e em especial a relação entre funções características e convergência fraca.

Nosso objetivo é desenvolver apenas o suficiente da teoria de funções características para provar o Teorema do Limite Central.

A motivação para a teoria que estamos prestes a desenvolver deriva da intuição que a maior parte do comportamento de uma distribuição de probabilidade em $\mathbb{R}$ é capturada por seus momentos. Grosseiramente partiremos do seguinte paradigma: se pudéssemos colocar a informação sobre todos os momentos da distribuição em um único pacote então o pacote resultante poderia "caracterizar" a distribuição de probabilidade. Uma abordagem ingênua inicial poderia ser definir uma função $f(t) = \sum_{n=0}^{\infty} M_n t^n$ onde M_n indica o n-ésimo momento. Infelizmente, essa abordagem falha miseravelmente, pois é muito raro que os momentos diminuam rapidamente o suficiente para que a série de potência formal definida por $f(t)$ convirja.

Uma abordagem melhor seria redimensionar os momentos para dar alguma chance para que a série convirja. Por exemplo, poderíamos definir

$$f(t) = \int e^{tX} d\mathbf{P} = \sum_{n=0}^{\infty} \frac{M_n}{n!} t^n$$

Essa ideia tem muito mais mérito que a anterior e pode ser usada de forma eficaz, como já vimos na Seção 7.10 mas tem a forte desvantagem de que só funciona para distribuições que possuam momentos de todas as ordens.

A ideia que exploraremos neste capítulo é que ao passarmos para o domínio dos números complexos, obtemos um caracterização da distribuição que é sempre definida e (pelo menos, conceitualmente) captura todos os momentos.

Ou seja, como usual, nos complexos as coisas são mais simples.

Especificamente, definimos

$$f(t) = \int e^{itX} d\mathbf{P}$$

ou seja a *Transformada de Fourier* da distribuição de probabilidade e dessa forma, como veremos, conseguimos um objeto que determina exclusivamente a distribuição.

11.1 Definição (Função Característica) *Dado uma variável aleatória* X *definimos a **função característica** de* X*, denotada por* φ_X*, como:*

$$\begin{cases} \varphi_X \colon \mathbb{R} \to \mathbb{C} \\ \varphi_X(t) = \mathbf{E}\left[e^{itX}\right] = \int_{\mathbb{R}} e^{itx} \; \mathrm{d}F_X(x) = \int_{\mathbb{R}} e^{itx} f_X(x) \; \mathrm{d}\mu \end{cases}$$

Como veremos a seguir, a convergência em distribuição poderá então ser descrita como convergência **pontual** de funções características e através dessa conexão, obteremos uma demonstração do Teorema do Limite Central.

$$\begin{tikzcd}
\text{problema} \arrow[r, "\text{Transformada de Fourier}"] \arrow[d, dotted, "\text{resolução difícil}"] & \text{problema transformado} \arrow[d, "\text{resolução facil}"] \\
\text{solução} & \text{solução transformada} \arrow[l, "\text{Transformada Inversa}"']
\end{tikzcd}$$

Neste capítulo, usaremos integrais de funções mensuráveis complexas. A teoria é completamente análoga a teoria de integração real que apresentamos no Capítulo 5 Vamos nos limitar a apresentar as definições básicas e estabelecer alguns fatos principais:

Estamos considerando o conjunto dos números complexos com os Borelianos provenientes da norma complexa. Como espaço métrico, esquecendo qualquer outra estrutura, temos que $\mathbb{C}$ e $\mathbb{R}^2$ são isométricos e assim podemos identificar a σ-álgebra de Borel de $\mathbb{C}$ com a σ-álgebra de Borel de $\mathbb{R}^2$.

Como consequência da identificação temos a seguinte caracterização da mensurabilidade de funções complexas.

11.2 Proposição *Uma função $f : (\Omega, \mathcal{F}, \mu) \to \mathbb{C}$ é mensurável se e somente se $f = g + ih$ onde $g, h : (\Omega, \mathcal{F}, \mu) \to \mathbb{R}$ são mensuráveis.*

Partindo da medida em Ω podemos construir uma integral para as funções $f(\Omega, \mathcal{F}, \mu) \to \mathbb{C}$ através de pequenas adaptações do maquinário que apresentamos no capítulo 5.

Outra abordagem, mais direta é definir a integral complexa como soma de duas integrais reais. As duas abordagens são equivalentes.

11.3 Definição *Seja $f(\Omega, \mathcal{F}, \mu) \to \mathbb{C}$ uma função mensurável, então $f = g + ih$ onde $g,h : (\Omega, \mathcal{F}, \mu) \to \mathbb{R}$ são mensuráveis. Definimos a integral de f como:*

$$\int f \, \mathrm{d}\mu = \int g \, \mathrm{d}\mu + i \int h \, \mathrm{d}\mu$$

11.4 Proposição

- *Sejam f, g funções integráveis e $a, b \in \mathbb{C}$ então $af + bg$ é integrável e*

$$\int (af + bg) \, \mathrm{d}\mu = a \int f \, \mathrm{d}\mu + b \int g \, \mathrm{d}\mu$$

- *Se f for integrável, então $\overline{\int f \, \mathrm{d}\mu} = \int \overline{f} \, \mathrm{d}\mu$*

- *Se $\mu(A) < \infty$, então $\left| \int f \, \mathrm{d}\mu \right| \leq \int |f| \, \mathrm{d}\mu$.*

Demonstração. *Demonstraremos apenas (3.). Pela desigualdade triangular para a norma complexa, sabemos que dados $z, w \in \mathbb{C}$ e $t \in [0,1]$, $|(1-t)z + tw| \leq (1-t)|z| + t|w|$ e, portanto, a norma complexa é convexa. Então, pela desigualdade de Jensen, temos $\left| \int f \, \mathrm{d}\mu \right| \leq \int |f| \, \mathrm{d}\mu$.* □

11.5 Definição *Dada uma variável aleatória X, definimos a **função característica** de X como a função $\phi_X : \mathbb{R} \to \mathbb{C}$ dada por*

$$\phi_X(t) = \mathbf{E}[e^{itX}] = \mathbf{E}[\cos(tX)] + i\mathbf{E}[\operatorname{sen}(tX)].$$

Por se tratar de uma esperança, a função característica depende unicamente da distribuição da variável X. Ou seja, se $X \stackrel{d}{=} Y$, então $\phi_X(t) = \phi_Y(t)$. Assim, qualquer informação que tirarmos dela será sobre a distribuição de X, e não sobre a variável especificamente.

11.6 Exemplo *Vamos calcular a função característica de uma variável aleatória $X \sim \text{Bernoulli}(p)$. Para isso basta notar que*

$$\phi_X(t) = \mathbf{E}[e^{itX}] = e^{it}\mathbf{P}(X = 1) + 1 \cdot \mathbf{P}(X = 0),$$

e portanto

$$\phi_X(t) = 1 + p(e^{it} - 1).$$

Outro resultado importante, geralmente estudado em teoria da medida, trata da derivação dentro do sinal de integral. Enunciaremos o resultado abaixo sem demonstrar, e a demonstração ficará como exercício para o leitor.

11.7 Proposição *Seja $f : \mathbb{R}^2 \to \mathbb{R}$ tal que para todo $s \in \mathbb{R}$, $\frac{\partial}{\partial t} f(t,s)$ exista $a < t < b$. Dada uma variável aleatória X, suponha que exista uma variável Y integrável e tal que*

$$\left|\frac{\partial}{\partial t} f(t,X)\right| \leq Y,$$

para todo $a < t < b$. Nestas condições

$$\frac{\partial}{\partial t}\mathbf{E}[f(t,X)] = \mathbf{E}\left[\frac{\partial}{\partial t} f(t,X)\right]$$

Vamos agora estudar algumas propriedades das funções característica. Pra começar, é importante observar que como cos e sen são funções limitadas, então $\phi_X(t)$ está bem definida para todo t e toda variável aleatória X.

Mais do que isso, pela desigualdade de Jensen,

$$|\phi_X(t)| = \sqrt{(\mathbf{E}[\cos(tX)])^2 + (\mathbf{E}[\text{sen}(tX)])^2} \leq \sqrt{\mathbf{E}[\cos^2(tX) + \text{sen}^2(tX)]} = 1$$

No início desse capítulo motivamos a definição da função característica considerando como podemos codificar informações sobre os momentos de uma medida de probabilidade. Para mostrar que tivemos sucesso, precisamos mostrar como extrair os momentos da função característica. Assim se $\mathbf{E}|X|^k < \infty$, então $\mathbf{E}[|X^j|] < \infty$ para todo $0 < j < k$, e aplicando a proposição 11.7 k vezes para as funções $X^j \cos(tX)$ e $X^j \text{sen}(tX)$, $j = 0, \ldots, k-1$, descobrimos o seguinte resultado.

11.8 Proposição *Seja X uma variável aleatória com função caraterística $\phi(t) = \mathbf{E}[e^{itX}]$. Nestas condições, se $\mathbf{E}[|X|^k] < \infty$, então $\phi(t)$ é k vezes diferenciável e*

$$\phi^{(k)}(t) = \frac{d^k}{dt^k}\phi(t) = i^k\mathbf{E}[X^k e^{itX}]$$

E em particular

$$\phi^{(k)}(0) = i^k \mathbf{E}[X^k].$$

Como

$$\phi_{X+Y}(t) = \mathbf{E}[e^{it(X+Y)}] = \mathbf{E}[e^{itX}e^{itY}] = \mathbf{E}[e^{itX}]\mathbf{E}[e^{itY}] = \phi_X(t)\phi_Y(t).$$

temos a seguinte proposição

11.9 Proposição *Sejam* X *e* Y *variáveis aleatórias independentes. Então* $\varphi_{X+Y}(u) = \varphi_X(u)\varphi_Y(u)$.

11.10 Exemplo *Para calcular a função característica de uma variável Binomial de parâmetros n e $p \in (0,1)$, tome $X_1,\ldots,X_n$ variáveis aleatórias independentes com distribuição $Bernoulli(p)$, e faça $X = X_1 + \cdots + X_n$. Segue que $X \sim \mathrm{Bin}(n,p)$, e*

$$\phi_X(t) = \phi_{X_1}(t)\cdots\phi_{X_n}(t) = \left[\phi_{X_1}(t)\right]^n = [1 + p(e^{it} - 1)]^n.$$

11.11 Exemplo *Seja* X *uma variável aleatória $N(0,1)$. Então $\varphi_X(u) = e^{\frac{-u^2}{2}}$. A maneira menos técnica de ver isso exige um pequeno truque. Primeiro observe que como $\operatorname{sen} ux$ é uma função ímpar temos que*

$$\begin{aligned}
\varphi_X(u) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{iux} e^{\frac{-x^2}{2}} \, \mathrm{d}x \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\frac{-x^2}{2}} \cos ux \, \mathrm{d}x + \frac{i}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\frac{-x^2}{2}} \operatorname{sen} ux \, \mathrm{d}x \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\frac{-x^2}{2}} \cos ux \, \mathrm{d}x
\end{aligned}$$

Por outro lado, pelo Lema **??** *e pelo fato que $x \cos ux$ é uma função par que temos*

$$\begin{aligned}
\frac{d\varphi_X(u)}{\mathrm{d}u} &= \frac{i}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{iux} e^{\frac{-x^2}{2}} \, \mathrm{d}x \\
&= \frac{i}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{\frac{-x^2}{2}} \cos ux \, \mathrm{d}x - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{\frac{-x^2}{2}} \operatorname{sen} ux \, \mathrm{d}x \\
&= \frac{-1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{\frac{-x^2}{2}} \operatorname{sen} ux \, \mathrm{d}x
\end{aligned}$$

Esta última integral pode ser integrada por partes (faça $df = xe^{\frac{-x^2}{2}} \, \mathrm{d}x$ e $g = \operatorname{sen} ux$, portanto $f = -e^{\frac{-x^2}{2}}$ e $dg = u \cos ux$) para obter

$$\frac{d\varphi_X(u)}{\mathrm{d}u} = \frac{-u}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\frac{-x^2}{2}} \cos ux \, \mathrm{d}x$$

e, portanto, mostramos que a função característica satisfaz a equação diferencial de primeira ordem simples $\frac{d\varphi_X(u)}{\mathrm{d}u} = -u\varphi_X(u)$ que tem a solução geral $\varphi_X(u) = Ce^{\frac{-u^2}{2}}$ para alguma constante C.

E como $\phi_X(0) = \mathbf{E}[e^{i0X}] = 1$*, segue que*

$$\phi_X(t) = e^{-\frac{t^2}{2}}.$$

Antes de seguirmos vamos definir a transformada de Fourier para medidas.

11.12 Definição *Seja* μ *uma medida de probabilidade em* R*. A* ***transformada de Fourier****, denotada por* $\widehat{\mu}$*, é a função complexa em* R *definida por*

$$\widehat{\mu}(u) = \int e^{iux}\ \mathrm{d}\mu(x) = \int \cos(ux)\ \mathrm{d}\mu(x) + i\int \operatorname{sen}(ux)\ \mathrm{d}\mu(x)$$

A primeira coisa que faremos será estabelecer as propriedades básicas da transformada de Fourier de uma medida de probabilidade, incluindo o fato de que a definição anterior realmente faz sentido.

11.13 Teorema *Seja* μ *uma medida de probabilidade, então* $\widehat{\mu}$ *existe e é uma função limitada uniformemente contínua com* $\widehat{\mu}(0) = 1$*.*

Demonstração. *Para ver que* $\widehat{\mu}$ *existe, usaremos representação*

$$\widehat{\mu}(u) = \int \cos(ux)\ \mathrm{d}\mu(x) + i\int \operatorname{sen}(ux)\ \mathrm{d}\mu(x)$$

e que $|\cos\theta| \le 1$ *e* $|\operatorname{sen}\theta| \le 1$ *para concluir que ambas as integrais são limitadas.*

Para ver que $\widehat{\mu}(0) = 1$*, basta calcular*

$$\widehat{\mu}(0) = \int \cos 0\ \mathrm{d}\mu(x) + i\int \operatorname{sen} 0\ \mathrm{d}\mu(x) = \int\ \mathrm{d}\mu(x) = 1$$

A limitação é um cálculo simples

$$\left|\widehat{\mu}(u)\right| \le \int \left|e^{iux}\right|\ \mathrm{d}\mu(x) = \int\ \mathrm{d}\mu(x) = 1$$

Por último, para provar a continuidade uniforme, primeiro observe que para qualquer $u, v \in R$*, temos*

$$\begin{aligned}
\left|e^{iux} - e^{ivx}\right|^2 &= \left|e^{i(u-v)x} - 1\right|^2 \\
&= (\cos((u-v)x) - 1)^2 + \operatorname{sen}^2((u-v)x) \\
&= 2(1 - \cos((u-v)x)) \\
&\le ((u-v)x)^2 && \textit{usando } 1 - \cos(x) = 2\operatorname{sen}^2(x/2) \\
&\le \|u-v\|^2\|x\|^2 && \textit{por Cauchy Schwartz}
\end{aligned}$$

Por outro lado, é claro a partir da desigualdade triangular que

$$\left|e^{iux} - e^{iux}\right| \le \left|e^{iux}\right| + \left|e^{iux}\right| \le 2$$

e, portanto, temos o seguinte limitante $\left|e^{iux} - e^{iux}\right| \leq \|u - v\|\|x\| \wedge 2$.

Observe agora que, em um ponto, $x \in R$, $\lim_{n\to\infty} \frac{1}{n}\|x\| \wedge 2 = 0$ *e trivialmente* $\frac{1}{n}\|x\| \wedge 2 \leq 2$ *então o Teorema da Convergência Dominada mostra que* $\lim_{n\to\infty} \displaystyle\int \frac{1}{n}\|x\| \wedge 2 \;\mathrm{d}\mu(x) = 0$.

Finalmente, dado um $\epsilon > 0$*, escolha* $N > 0$ *tal que* $\displaystyle\int \frac{1}{N}\|x\| \wedge 2 \;\mathrm{d}\mu(x) < \epsilon$ *então para* $\|u - v\| \leq \frac{1}{N}$,

$$\begin{aligned}\left|\widehat{\mu}(u) - \widehat{\mu}(v)\right| &\leq \int \left|e^{iux} - e^{iux}\right| \;\mathrm{d}\mu(x)\\ &\leq \int \|u - v\|\|x\| \wedge 2 \;\mathrm{d}\mu(x)\\ &\leq \int \frac{1}{N}\|x\| \wedge 2 \;\mathrm{d}\mu(x) < \epsilon\end{aligned}$$

provando continuidade uniforme. □

Unicidade de Fourier

Agora demonstraremos que a função característica caracteriza completamente as medidas de probabilidade.

Suponha que μ_1 e μ_2 tenham a mesma função característica, ou seja,

$$\int_{-\infty}^{\infty} e^{itx} d\mu_1(x) = \int_{-\infty}^{\infty} e^{itx} \mu_2(dx) \text{ para todo } t \in \mathbb{R}$$

Vamos demonstrar que $\mu_1 = \mu_2$. Observamos que é suficiente mostrar que

$$\int_{-\infty}^{\infty} f(x) d\mu_1(x) = \int_{-\infty}^{\infty} f(x) d\mu_2(x) \tag{11.1}$$

para qualquer função contínua limitada f.

Faremos isso aproximando f por combinações lineares de funções da família $\left\{e^{2\pi inx/T} : n \in \mathbb{Z}\right\}$.

A inspiração aqui são as séries de Fourier: qualquer função periódica contínua $f(x)$ com período T (ou seja, $f(x+1) = f(x)$) admite uma expansão em série de Fourier

$$f(x) \sim \sum_{n=-\infty}^{\infty} c_n e^{2\pi inx/T}, \quad x \in [0,1],$$

onde $c_n = \int_0^1 f(x) e^{2\pi inx/T} dx$ é o coeficiente de Fourier.

No que segue estamos apresentando uma heurística e estaremos sendo descuidados e várias passagens não estão sendo justificadas. Como $\int_{-T}^{T} e^{itx} d\mu_1(x) = \int_{-T}^{T} e^{itx} \mu_2(dx)$ para todo $t \in \mathbb{R}$, tomando $t = 2\pi n/T$ temos que $\int_{-\infty}^{\infty} e^{2\pi i n x/T} d\mu_1(x) = \int_{-\infty}^{\infty} e^{2\pi i n x/T} \mu_2(dx)$ para todo n e logo

$$\int_{-T}^{T} f(x) d\mu_1(x) = \sum_{n=-\infty}^{\infty} \int_{-T}^{T} c_n e^{2\pi i n x/T} d\mu_1(x) = \sum_{n=-\infty}^{\infty} \int_{-T}^{T} c_n e^{2\pi i n x/T} d\mu_2(x) = \int_{-T}^{T} f(x) d\mu_2(x)$$

que é quase o que queremos demonstrar, veja a equação 11.1. Precisamos retirar a hipótese de periódica e poder tomar o T tendendo ao infinito, mas ambos procedimentos podem ser feitos com ferramentas usuais de análise.

Mas ao invés de seguirmos por esse caminho, utilizando o Teorema de Fourier, vamos utilizar um resultado mais simples que qualquer função f pode ser aproximada por combinações lineares de funções da família $\left\{e^{2\pi i n x/T} : n \in \mathbb{Z}\right\}$, cuja demonstração é consequência quase imediata do Teorema de Stone-Weierstrass

11.14 Teorema (Teorema de Stone-Weierstrass para funções periódicas) *Seja $T > 0$. Defina C_T como o espaço de funções periódicas contínuas $f : \mathbb{R} \to \mathbb{C}$ com período T. Seja $\mathcal{A}$ um subconjunto de C_T satisfazendo as seguintes propriedades:*

- $\mathcal{A}$ é uma álgebra: $f, g \in \mathcal{A}, a, b \in \mathbb{R} \Longrightarrow af + bg, f \cdot g \in \mathcal{A}$;
- $\mathcal{A}$ não se anula em nenhum ponto: para qualquer $x \in [0, T)$, existe $f \in \mathcal{A}$ tal que $f(x) \neq 0$;
- $\mathcal{A}$ separa pontos: para qualquer $x \neq y \in [0, T)$, existe $f \in \mathcal{A}$ tal que $f(x) \neq f(y)$.

Então $\mathcal{A}$ é denso em C_T com relação à convergência uniforme em $[0, T]$. Mais precisamente, para qualquer função periódica $f \in C_T$ e $\varepsilon > 0$, existe $g \in \mathcal{A}$ tal que

$$\sup_{t \in [0,T]} |f(t) - g(t)| < \varepsilon$$

Agora provamos o resultado da unicidade para a função característica.

11.15 Teorema (Teorema da Unicidade de Fourier) *Sejam μ e ν medidas de probabilidade tais que a função característica é igual, então $\mu = \nu$.*

Demonstração. *Sejam μ_1, μ_2 duas medidas de probabilidade com a mesma função característica.*

Seja $T > 0$ um número positivo arbitrário e defina C_T como o espaço das funções periódicas $f : \mathbb{R} \to \mathbb{C}$ com período T. Seja $\mathcal{A}_T \subseteq C_T$ o espaço vetorial gerado pela família $\left\{e^{2\pi i n x/T} : n \in \mathbb{Z}\right\}$ de funções. A família $\mathcal{A}_T$ satisfaz as hipóteses do Teorema de Stone-Weierstrass. Portanto, $\mathcal{A}_T$ é denso em C_T com relação à convergência uniforme em $[0, T]$. Além disso, é fácil ver que para qualquer função $h(t) \in \mathcal{A}_T$ temos que

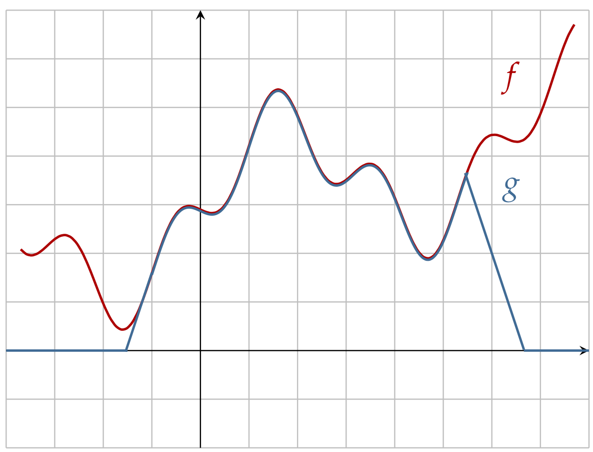
$$\int h(t) d\mu_1 = \int h(t) d\mu_1 .$$

Em seguida, queremos estender essa aproximação para toda função contínua limitada f. A ideia é substituir f por uma função periódica de período grande. Dado um $\varepsilon > 0$ arbitrário, existe $M > 0$ tal que

$$\mu_i\left([-M, M]^c\right) < \varepsilon \quad \textit{for } i = 1{,}2.$$

Seja $g : [-M-1, M+1] \to \mathbb{R}$ a função contínua dada por

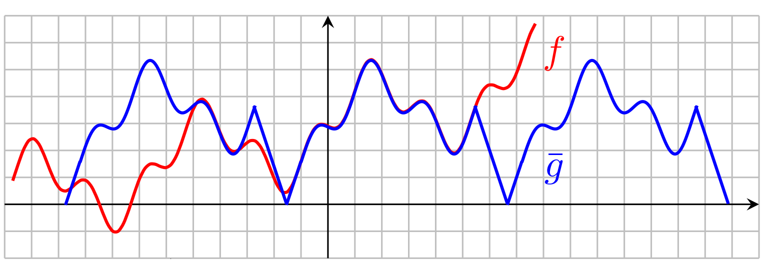
$$g(x) := \begin{cases} f(x), & x \in [-M, M] \\ 0, & x \in (-\infty, -M-1) \cup (M+1, \infty) \\ \textit{interpolação linear}, & x \in [-M-1, -M] \textit{ ou } x \in [M, M+1] \end{cases}$$



Por definição temos $g(-M-1) = g(M+1)$ e

$$|g(x)| \leqslant \|f\|_\infty := \sup_{y \in \mathbb{R}} |f(y)| \quad \textit{para todos } x \in [-M-1, M+1]$$

Seja $\bar{g} : \mathbb{R} \to \mathbb{R}$ a extensão periódica de g a $\mathbb{R}$ com período $T = 2M + 2$.



Do passo anterior sabemos pelo Teorema de Stone Weierstrass que $\bar{g}$ pode ser aproximadas por funções em $\mathcal{A}_T$. Como também temos $f = \bar{g}$ em $[-M, M]$, segue que

$$\begin{aligned}
&\left|\int f d\mu_1 - \int f d\mu_2\right| \\
&\leqslant \left|\int f d\mu_1 - \int \bar{g} d\mu_1\right| + \underbrace{\left|\int \bar{g} d\mu_1 - \int \bar{g} d\mu_2\right|}_{\boxed{1}} + \left|\int \bar{g} d\mu_2 - \int f d\mu_2\right| \\
&= \left|\int f d\mu_1 - \int \bar{g} d\mu_1\right| + \left|\int \bar{g} d\mu_2 - \int f d\mu_2\right| \\
&\leqslant 2\|f\|_\infty \cdot (\mu_1([-M, M]^c) + \mu_2([-M, M]^c)) \\
&< 4\|f\|_\infty \varepsilon
\end{aligned}$$

Para ver que $\boxed{1}$ é verdadeiro, seja $g_n \in \mathcal{A}_T$ tal que g_n converge uniformemente a $\bar{g}$ em compactos e assim,

$$\begin{aligned}
\left|\int \bar{g} d\mu_1 - \int \bar{g} d\mu_2\right| &= \left|\int \lim_{n\to\infty} \bar{g}_n d\mu_1 - \int \lim_{n\to\infty} \bar{g}_n d\mu_2\right| \\
&\lim_{n\to\infty} \left|\int \bar{g}_n d\mu_1 - \int \bar{g}_n d\mu_2\right| = 0
\end{aligned}$$

Como ε é arbitrário, temos que para toda função contínua limitada f

$$\int f d\mu_1 = \int f d\mu_2$$

e logo $\mu_1 = \mu_2$. □

Convergência de Funções Características

Nesta seção vamos estudar como podemos usar funções características para estudar as distribuições de uma sequência de variáveis aleatórias.

11.16 Proposição *Seja $\{X_n; n \geq 1\}$ uma sequência de variáveis aleatórias e denote por $\phi_n(t)$ a função característica de X_n, $n \geq 1$. Nestas condições, se existe uma função $g(t)$ contínua $t \in (-t_0,t_0)$, tal que $\phi_n(t) \longrightarrow g(t)$ para todo $t \in (-t_0,t_0)$, então $\{X_n; n \geq 1\}$ é tight.*

Demonstração. *Para essa demonstração vamos nos valer do fato de que $g(t)$ é contínua para $t \in (-t_0,t_0)$ com $g(0) = \lim_n \phi_n(0) = 1$. Isso significa que, dado $\epsilon > 0$, podemos encontrar $a \in (0,t_0)$ tal que*

$$|1 - g(t)| < \epsilon/4,$$

e portanto

$$\left|\int_{-a}^{a} (1 - g(t))\, \mathrm{d}t\right| \leq \int_{-a}^{a} |1 - g(t)|\, \mathrm{d}t < a\epsilon/2.$$

Além disso, como $\phi_n(t) \to g(t)$ para $-t_0 < t < t_0$ e $|\phi_n(t)| \leq 1$ então, pelo Teorema da Convergência Dominada

$$\lim_{n\to\infty} \int_{-a}^{a} (1-\phi_n(t))\,\mathrm{d}t = \int_{-a}^{a} (1-g(t))\,\mathrm{d}t,$$

para todo $0 < a < t_0$. Isso significa que existe $n_0 \geq 1$ tal que para $n > n_0$ temos que

$$\left|\frac{1}{a}\int_{-a}^{a}(1-\phi_n(t))\,\mathrm{d}t - \frac{1}{a}\int_{-a}^{a}(1-g(t))\,\mathrm{d}t\right| < \epsilon/2,$$

e portanto

$$\left|\frac{1}{a}\int_{-a}^{a}(1-\phi_n(t))\,\mathrm{d}t\right| \leq \left|\frac{1}{a}\int_{-a}^{a}(1-\phi_n(t))\,\mathrm{d}t - \frac{1}{a}\int_{-a}^{a}(1-g(t))\,\mathrm{d}t\right| + \left|\frac{1}{a}\int_{-a}^{a}(1-g(t))\,\mathrm{d}t\right| < \epsilon,$$

para todo $n \geq n_0$.

Precisamos agora relacionar $\int_{-a}^{a}(1-\phi_n(t))\,\mathrm{d}t$ com a distribuição de X_n. Para isso note que

$$\begin{aligned}
\int_{-a}^{a}(1-\phi_n(t))\,\mathrm{d}t &= \int_{-a}^{a}\mathbf{E}[1-\cos(tX_n) - i\,\mathrm{sen}(tX_n)]\,\mathrm{d}t && \textit{(por Fubini)}\\
&= \mathbf{E}\left[\int_{-a}^{a}(1-\cos(tX_n))\,\mathrm{d}t\right]\\
&\geq \mathbf{E}\left[\int_{-a}^{a}(1-\cos(tX_n))\mathbb{1}_{\{|X_n|\geq 2/a\}}\,\mathrm{d}t\right]\\
&= \mathbf{E}\left[\left(2a - \frac{2\,\mathrm{sen}(aX_n)}{X_n}\right)\mathbb{1}_{\{|X_n|\geq 2/a\}}\right]\\
&= 2a\,\mathbf{E}\left[\left(1 - \frac{\mathrm{sen}(a|X_n|)}{a|X_n|}\right)\mathbb{1}_{\{|X_n|\geq 2/a\}}\right] && \left(\frac{\mathrm{sen}(x)}{x} = \frac{\mathrm{sen}(|x|)}{|x|}\right)\\
&\geq 2a\,\mathbf{E}\left[\left(1 - \frac{1}{a|X_n|}\right)\mathbb{1}_{\{|X_n|\geq 2/a\}}\right] && (\mathrm{sen}(x) \leq 1)\\
&\geq a\,\mathbf{E}[\mathbb{1}_{\{|X_n|\geq 2/a\}}]\\
&= a\,\mathbf{P}(|X_n| \geq 2/a)
\end{aligned}$$

Fazendo então $b = 2/a$, mostramos assim que

$$\mathbf{P}(-b < X_n < b) = 1 - \mathbf{P}(|X_n| \geq b) \geq 1 - \frac{1}{a}\int_{-a}^{a}(1-\phi_n(t))\,\mathrm{d}t > 1-\epsilon,$$

o que garante que $\{x_n, n \geq 1\}$ é tight. □

11.17 Teorema *Dadas variáveis aleatórias $X, X_1, X_2, \ldots$ com funções características $\phi, \phi_1, \phi_2 \ldots$, respectivamente, então $X_n \xrightarrow{d} X$ se, e somente se, $\phi_n(t) \to \phi(t)$, quando $n \to \infty$, para todo $t \in \mathbb{R}$.*

Demonstração. *Primeiro note que se* $X_n \xrightarrow{d} X$, *pelo teorema de representação de Skorohod, existem variáveis* $X', X_1', X_2', \ldots$, *com* $X' \stackrel{d}{=} X$ *e* $X_n' \stackrel{d}{=} X_n$ *para todo* $n \geq q$, *tais que* $X_n' \xrightarrow{q.c.} X'$. *Isso implica que*

$$\phi_{X'}(t) = \mathbf{E}[e^{itX'}] = \mathbf{E}[e^{itX}]\phi(t), \quad e \tag{11.2}$$

$$\phi_{X_n'}(t) = \mathbf{E}[e^{itX_n'}] = \mathbf{E}[e^{itX_n}] = \phi_n(t). \tag{11.3}$$

Note também que fixado $t \in \mathbb{R}$, *as funções* $\cos(tx)$ *e* $\operatorname{sen}(tx)$ *são contínuas em* $x \in \mathbb{R}$ *e limitadas, de modo que*

$$\cos(tX_n') \xrightarrow{q.c.} \cos(tX'), \qquad \operatorname{sen}(tX_n') \xrightarrow{q.c.} \operatorname{sen}(tX'),$$

e o Teorema da Convergência Dominada garante que

$$\lim_{n\to\infty} \phi_n(t) = \lim_{n\to\infty} \mathbf{E}[\cos(tX_n')] + i\mathbf{E}[\operatorname{sen}(tX_n')] = \mathbf{E}[\cos(tX')] + i\mathbf{E}[\operatorname{sen}(tX')] = \phi(t).$$

Para a recíproca, primeiro note que $\phi(t)$ *é uma função contínua com* $\phi(0) = 1$, *e pela proposição 11.16,* $X_n; n \geq 1$ *é tight, e pela proposição 9.34, existe uma subsequência* $(n_k)_{k\geq 1}$ *e uma variável aleatória* $\tilde{X}$ *tal que* $X_{n_k} \xrightarrow{d} \tilde{X}$, *com função característica* $\phi_{\tilde{X}}$.

Pela primeira parte da demonstração, temos que $\phi_{n_k}(t) \longrightarrow \phi_{\tilde{X}}(t)$, *e portanto* $\phi_{\tilde{X}}(t) = \phi(t)$ *para todo* $t \in \mathbb{R}$. *E pelo Teorema 11.15 temos que*

$$\tilde{X} \stackrel{d}{=} X \quad e \quad X_{n_k} \xrightarrow{d} X.$$

Como a última igualdade vale para qualquer subsequência que converge em distribuição, temos que

$$X_n \xrightarrow{d} X.$$

11.2 Teorema do Limite Central

11.18 Teorema (Teorema do Limite Central) *Sejam* $X, X_1, X_2, \ldots$ *variáveis aleatórias i.i.d. com* $\mu = \mathbf{E}[X]$ *e* $\sigma = \operatorname{Var}[X_n] < \infty$, *então*

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right) \xrightarrow{d} N(0, \sigma^2)$$

Demonstração. *Primeiramente observamos que usando fazendo a mudança de variável* $X_i' = \frac{X_i - \mu}{\sigma}$, *podemos supor que* $\mu = 0$ *e* $\sigma = 1$. *Assim, só temos que mostrar que* $\frac{1}{\sqrt{n}} \sum_{k=1}^{n} X_k \xrightarrow{d} N(0,1)$.

Defina $S_n = \sum_{k=1}^n X_k$. Pelo Teorema 11.17 é suficiente mostrar que

$$\lim_{n\to\infty} \mathbf{E}\left[e^{itS_n/\sqrt{n}}\right] = e^{t^2/2}$$

Para calcular o limite, observamos que, por independência e i.i.d., temos que

$$\mathbf{E}\left[e^{itS_n/\sqrt{n}}\right] = \prod_{k=1}^n \mathbf{E}\left[e^{itX_k/\sqrt{n}}\right] = \left[\mathbf{E}\left[e^{itX/\sqrt{n}}\right]\right]^n$$

Para calcular o limite, tomamos a expansão de Taylor da função exponencial

$$e^{ix} = 1 + ix - \frac{1}{2}x^2 + R(x)$$

Pelo Resto de Lagrange e pelo fato que $\left|\frac{d}{\mathrm{d}x}e^{ix}\right| \leq 1$, temos que $|R(x)| \leq \frac{1}{6}|x|^3$. Observe que essa estimativa não é muito boa para valores grandes de $|x|$, mas pela desigualdade triangular temos para todo x que

$$\left|e^{ix} - 1 - ix + \frac{1}{2}x^2\right| \leq 2 + |x| + \frac{1}{2}x^2 \leq \frac{7}{2}x^2$$

que é uma estimativa melhor para $|x| > 1$.

Portanto, temos o limite $|R(x)| \leq \frac{7}{2}(|x|^3 \wedge x^2)$. Aplicando a expansão de Taylor obtemos

$$\mathbf{E}\left[e^{itS_n/\sqrt{n}}\right] = \left(1 - \frac{t^2}{2n} + \mathbf{E}\left[R\left(\frac{tX}{\sqrt{n}}\right)\right]\right)^n$$

Pela estimativa do resto, podemos ver que

$$n\left|\mathbf{E}\left[R\left(\frac{tX}{\sqrt{n}}\right)\right]\right| \leq \frac{7}{2}\mathbf{E}\left[\frac{t^3|X|^3}{\sqrt{n}} \wedge t^2X^2\right] \tag{11.4}$$

$$\leq \frac{7}{2}\mathbf{E}\left[t^2X^2\right] = \frac{7t^2}{2} \tag{11.5}$$

Pela desigualdade 11.4 e pelo Teorema da Convergência Dominada, podemos concluir que

$$\lim_{n\to\infty} n\left|\mathbf{E}\left[R\left(\frac{tX}{\sqrt{n}}\right)\right]\right| = 0$$

então se definimos $\epsilon_n = \frac{2n}{t^2}\left|\mathbf{E}\left[R\left(\frac{tX}{\sqrt{n}}\right)\right]\right|$ então temos $\lim_{n\to\infty}\epsilon_n = 0$ e

$$\lim_{n\to\infty} \mathbf{E}\left[e^{itS_n/\sqrt{n}}\right] e = \lim_{n\to\infty}\left(1 - \frac{t^2}{2n}(1+\epsilon_n)\right)^n = \lim_{n\to\infty} e^{n\log(1-\frac{t^2}{2n}(1+\epsilon_n))} = e^{-t^2/2}$$

Exercícios

Funções Características

Ex. 11.1 — Prove a seguinte tabela:

Distribuição	Função Característica
Binomial	$[(1-p)+pe^{it}]^n$
Poisson	$e^{\lambda(e^{it}-1)}$
Normal	$e^{i\mu t-\dfrac{(\sigma t)^2}{2}}$

Ex. 11.2 — Se X e Y são duas variáveis aleatórias independentes então $\varphi_{X+Y} = \varphi_X(t)\varphi_Y(t)$.

Ex. 11.3 — Prove que se $X = aY + b$, então $\varphi_X(t) = \varphi_{aY+b}(t) = e^{itb}\varphi_Y(at)$

Ex. 11.4 — Sejam $X_1, X_2, \ldots, X_n$ variáveis aleatórias independentes e identicamente distribuídas, de Poisson de parâmetro λ Calcule a distribuição de $X_1 + X_2 + \cdots + X_n$

Ex. 11.5 — O objetivo neste problema é provar o Teorema do Limite Central usando funções Funções Características.
Seja $\{X_i\}_{i\geq 1}$ sequência de variáveis aleatórias i.i.d. com $\mathbf{E}[X_1] < \infty$, $S_n = X_1 + \cdots + X_n$ e $\mathbf{E}[X_1] = \mu$. Então $\dfrac{S_n}{n} \xrightarrow{\mathbf{P}} \mu$ tal que para todo $\varepsilon > r0$

$$\mathbf{P}\left(\left|\frac{S_n}{n} - m\right| \geq \varepsilon\right) \xrightarrow{n\to\infty} 0$$

1. Mostre que $\varphi_{S_n/n}(t) = \left[\varphi\left(\frac{t}{n}\right)\right]^n$
2. Mostre que se $\mathbf{E}|X|^1 < \infty$, então:

$$\varphi(t) = 1 + it\mu + r(t), \qquad \text{quando } t \to 0$$

   Com $\dfrac{r(t)}{t} \xrightarrow{t\to 0} 0$
   Dica: Polinômio de Taylor
3. Conclua que

$$\varphi_{S_n/n}(t) = \left[1 + i\frac{t}{n}\mu + r(\frac{1}{n})\right]^n \to e^{it\mu}$$

4. Conclua o resultado.

Teorema do Limite Central

Ex. 11.6 — Você convidou 64 convidados para uma festa. Você precisa fazer sanduíches para os convidados. Você acredita que um convidado pode precisar de 0, 1 ou 2 sanduíches com probabilidades $\frac{1}{4}$, $\frac{1}{2}$ e $\frac{1}{4}$, respectivamente. Você assume que o número de sanduíches que cada hóspede precisa é independente dos outros hóspedes. Quantos sanduíches você deve fazer para que você tenha 95% de certeza de que não haverá escassez?

Ex. 11.7 — Use o Teorema do Limite Central para mostrar que

$$\lim_{n\to\infty} \sum_{k=0}^{n} e^{-}n\frac{n^k}{k!} = 1/2$$

Ex. 11.8 — Seja $M_X(s)$ finita para todo $s \in [-c,c]$. Prove que

$$\lim_{n\to\infty} \left[M_X(\frac{s}{n})\right]^n = e^{s\mathbf{E}[X]}.$$

Dica: Calcule

$$\lim_{n\to\infty} n \ln \left(M_X(\frac{s}{n})\right) = s\mathbf{E}[X].$$

usando L'Hopital.

Ex. 11.9 — O objetivo neste problema é provar o teorema do limite central usando funções geradoras de momento.

Em particular, sejam $X_1, X_2, ..., X_n$ variáveis aleatórias i.i.d. com valor esperado $\mathbf{E}[X_i] = \mu < \infty$,$\mathbf{Var}(X_i) = \sigma^2 < \infty$, e funções geradoras de momento $M_X(s)$ que é finito em algum intervalo $[-c, c]$, onde $c > 0$ é uma constante. Considere

$$Z_n = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} = \frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sqrt{n}\sigma}. \tag{11.6}$$

Prove

$$\lim_{n\to\infty} M_{Z_n}(s) = e^{\frac{s^2}{2}}, \qquad \text{para todo } s \in [-c, c]. \tag{11.7}$$

Uma vez que este é a funções geradoras de momento de uma variável aleatória normal padrão, concluímos que a distribuição de Z_n converge para a variável aleatória normal padrão.

Dica: use o resultado anterior.

CAPÍTULO 12

Acoplamento

É frequentemente necessário comparar as distribuições de duas variáveis aleatórias X e Y. Uma vez que X e Y podem não estar definidas no mesmo espaço amostral Ω, é, a princípio, impossível comparar X e Y. Uma técnica extremamente útil e moderna é construir cópias X' e Y', de X e Y respectivamente, no mesmo espaço amostral Ω e, em seguida, comparar X' e Y'. Esta abordagem é conhecida como acoplamento, e tem diversas aplicações importantes. Como veremos há mais do que um possível acoplamento de um par X e Y, e o segredo do sucesso na utilização da técnica de acoplamento é encontrar o acoplamento que é adequado para a aplicação particular.

Vejamos um exemplo.

Imagine duas moedas com probabilidades de sucesso $p, p \in (0, 1)$ satisfazendo $p < p'$. Claramente, em um lançamento é menos provável que a p-moeda seja bem sucedida do que a p'-moeda. No entanto, se você jogar as duas moedas de forma independente, pode acontecer que a p-moeda seja bem sucedida enquanto a p'-moeda não o seja. Podemos jogar as duas moedas de um modo que o resultado seja sempre ordenado?

A resposta é sim. Para ver esse fato considere $p'' = (p' - p)/(1 - p) \in (0, 1)$. Agora considere uma terceira moeda com probabilidade de sucesso p''. Jogue a p-moeda e a p''-moeda de forma independente. Seja X o resultado da p-moeda e X' o resultado da p''-moeda, e defina $X' = X \vee X''$. Como $p' = p + (1 - p)p''$, X' tem a mesma distribuição que a p'-moeda Mas $X' \geq X$, como queríamos.

12.1 Variação total

Dada uma medida com sinal limitada μ em um espaço mensurável $(\Omega, \mathcal{F})$ tal que $\mu(E) = 0$, a **norma de variação total** de μ é definida como

$$\|\mu\|_{\mathrm{VT}} := 2 \sup_{A \in \mathcal{F}} \mu(A)$$

A **distância de variação total** entre duas medidas de probabilidade μ e ν é a diferença máxima entre elas quando calculadas num mesmo evento. Formalmente:

$$||\mu - \nu||_{\mathrm{VT}} = 2 \sup_{A \in \mathcal{F}} |\mu(A) - \nu(A)|$$

12.1 Exemplo *Para medidas num espaço finito temos*

$$||\mu - \nu||_{\mathrm{VT}} = \|\mu - \nu\|_1 = \sum_{\omega \in \Omega} |\mu(\omega) - \nu(\omega)|$$

Esse exemplo motiva a escolha do fator 2 na definição de distância de variação total entre duas medidas de probabilidade

Por considerar todos os subconjuntos possíveis de Ω, pode não ser simples calcular essa distância. As próximas duas proposições mostram meios alternativos de encontrar essa medida.

Dado a medida $\mu - \nu$, considere $\Omega = D^+ + D^-$ a decomposição de Hahn de $\mu - \nu$ então definimos a medida

$$|\mu - \nu| = \mathbb{1}_{D^+}(\mu - \nu) - \mathbb{1}_{D^-}(\mu - \nu)$$

12.2 Proposição *Sejam μ e ν duas medidas de probabilidade em Ω. Assim,*

$$||\mu - \nu||_{\mathrm{VT}} = \int_{\Omega} d|\mu(x) - \nu(x)|$$

Demonstração. *Claramente*

$$||\mu - \nu||_{\mathrm{VT}} = 2 \sup_{A \in \mathcal{F}} |\mu(A) - \nu(A)| = 2(\mu(D^+) - \nu(D^+))$$

Como $\mu(D^+) - \nu(D^+) = \nu(D^-) - \mu(D^-)$

$$||\mu - \nu||_{\mathrm{VT}} = 2 \sup_{A \in \mathcal{F}} |\mu(A) - \nu(A)| \tag{12.1}$$
$$= (\mu(D^{)} - \nu(D^{+})) + (\nu(D^{-}) - \mu(D^{-})) \tag{12.2}$$
$$= \int_{\Omega} d|\mu(x) - \nu(x)| \tag{12.3}$$

12.3 Proposição *Seja μ e ν duas medidas de probabilidade em Ω. Então a distância de variação total satisfaz:*

$$||\mu - \nu||_{\mathrm{VT}} = sup \left\{ \int_{\Omega} f(x)\mu(x) - \int_{\Omega} f(x)\nu(x) \right. \tag{12.4}$$
$$\left. f \text{ satisfazendo } \max_{x \in \Omega} |f(x)| \leq 1 \right\} \tag{12.5}$$

Demonstração. *Exercício. Dica: Use a decomposição de Hahn e a proposição anterior.* □

12.4 Observação *Da Proposição e da desigualdade triangular, temos, para distribuições μ, ν e Υ, a seguinte desigualdade:*

$$||\mu - \nu||_{\mathrm{VT}} \leq ||\mu - \Upsilon||_{\mathrm{VT}} + ||\Upsilon - \nu||_{\mathrm{VT}}$$

12.2 Acoplamento

Um acoplamento de duas distribuições (ou seja, duas medidas de probabilidade) não é mais que uma distribuição conjunta delas. Mais precisamente:

12.5 Definição (Acoplamento)

- *Um* ***acoplamento*** *de duas medidas de probabilidade* $\mathbf{P}$ *e* $\mathbf{P}'$ *no mesmo espaço mensurável* $(\Omega, \mathcal{F})$ *é qualquer medida de probabilidade* $\hat{\mathbf{P}}$ *no espaço mensurável produto* $(\Omega \times \Omega, \mathcal{F} \otimes \mathcal{F})$ *cujas marginais são* $\mathbf{P}$ *e* $\mathbf{P}'$ *, isto é,*

$$\mathbf{P} = \mathbf{P} \circ \pi^{-1},$$

$$\mathbf{P}' = \mathbf{P} \circ \pi'^{-1},$$

onde π e π' são as projeções, definidas por

$$\pi(x, x') = x,$$

$$\pi'(x, x') = x',$$

para todo $(x, x') \in \Omega \times \Omega$.

- *Um* ***acoplamento de variáveis aleatórias*** (X,Y) *definidas no mesmo espaço de probabilidade* $\Omega,\mathcal{F}$ *é qualquer par de variáveis aleatórias* $(\hat{X},\hat{Y})$ *que tomam valores em* $(\Omega \times \Omega, \mathcal{F} \otimes \mathcal{F})$ *e cujas marginais têm a mesma distribuição que* (X,Y)*, isto é,* $X \xrightarrow{d} \hat{X}$ *e* $Y \xrightarrow{d} \hat{Y}$

Vamos denotar por $\mathbb{C}(\mu, \nu)$ o conjunto de todos os acoplamentos de μ e ν. Obviamente, a medida do produto $\mu \otimes \nu$ é um acoplamento de μ e ν, que é chamado de acoplamento independente. Este acoplamento é muito simples mas pelo menos indica a existência de acoplamentos.

12.6 Exemplo (Acoplamento de Variáveis de Bernoulli)

Sejam X *e* Y *variáveis aleatórias de Bernoulli com parâmetros* $0 \leq q < r \leq 1$*, respectivamente. Ou seja,*

$$\mathbf{P}[X = 1] = q \text{ e } \mathbf{P}[X = 0] = 1 - q$$

e

$$\mathbf{P}[Y = 1] = r \text{ e } \mathbf{P}[Y = 0] = 1 - r$$

com $\Omega = \{0, 1\}$ *e* $\mathcal{F} = 2^{\Omega}$.

- *(Acoplamento independente) Um acoplamento de* X *e* Y *é* (X', Y') *onde* $X' \stackrel{d}{=} X$ *e* $Y' \stackrel{d}{=} Y$ *e* X' *e* Y' *são independentes. Sua lei é*

$$\mathbf{P}[(X', Y') = (i,j)] = \begin{pmatrix} (1-q)(1-r) & (1-q)r \\ q(1-r) & qr \end{pmatrix}$$

- *(Acoplamento monótono) Outra possibilidade é escolher* U *uniformemente em* $[0, 1]$ *e definir* $X'' = \mathbb{1}_{U\ leqq}$ *e* $Y'' = \mathbb{1}_{U\ leqr}$*. A lei do acoplamento* (X'', Y'') *é então*

$$\mathbf{P}[(X', Y') = (i,j)] = \begin{pmatrix} (1-r) & r-q \\ 0 & q \end{pmatrix}$$

Em aplicações, o desafio é encontrar um acoplamento que torne $||\mathbf{P} - \mathbf{P}'||_{VT}$ tão pequeno quanto possível.

12.7 Exemplo *Os acoplamentos não são únicos. Dois exemplos triviais são:*

- $\hat{P} = P \times P'$ *com* P, P' *arbitrárias se e somente se* X, X' *são independentes,*
- $P = P'$ *e* $\hat{P}$ *a probabilidade na diagonal se e somente se* $X = X'$. ◁

12.8 Exemplo (Passeio aleatório enviesado em $\mathbb{Z}$) *Para $p \in [0,1]$, seja $(S_n(p))$ o passeio aleatório em $\mathbb{Z}$ iniciado no 0 e com probabilidade p de saltar para a direita e probabilidade $1-p$ de saltar para a esquerda. Ou seja, dado uma sequência infinita de Bernoulli independentes em $\{-1,1\}$ e de parâmetro p $X_1, X_2, \cdots$:*

$$S_n(p) = X_1 + \cdots + X_n$$

Assuma $0 \le q < r \le 1$. Usando o acoplamento das variáveis Bernoulli acima podemos construir um acoplamento entre $S(q)$ e $S(r)$

Seja $(X_i'', Y_i'')_{i=1}^{\infty}$ uma sequência infinita de acoplamentos monótonos de Bernoulli i.i.d. com parâmetros q e r, respectivamente.

Seja $\hat{S}_n(p) = X_1'' + \cdots + X_n''$ e $\hat{S}_n(q) = Y_1'' + \cdots + Y_n''$, então $(\hat{S}_n(p), \hat{S}_n(q))$ é um acoplamento de $(S_n(p), S_n(q))$ de modo que $\hat{S}_n(q) \le \hat{S}_n(p)$

12.9 Exemplo (Acoplamento de medidas discretas) *Dadas duas duas medidas discretas μ e ν.*

Um acoplamento de variáveis aleatórias (X,Y) satisfazendo

$$\mathbf{P}\{X = x\} = \mu(x) \text{ e } \mathbf{P}\{Y = y\} = \nu(y).$$

Dado um acoplamento (X,Y) de μ e ν, se q é a distribuição conjunta de (X,Y) em $\Omega \times \Omega$, ou seja $q(x,y) = \mathbf{P}\{X = x, Y = y\}$, então:

$$\sum_{y\in\Omega} q(x,y) = \sum_{y\in\Omega} \mathbf{P}\{X = x, Y = y\} = \sum_{y\in\Omega} \mathbf{P}\{X = x\}\mathbf{P}\{Y = y\} = \mu(x)$$

e

$$\sum_{x\in\Omega} q(x,y) = \sum_{x\in\Omega} \mathbf{P}\{X = x, Y = y\} = \sum_{x\in\Omega} \mathbf{P}\{X = x\}\mathbf{P}\{Y = y\} = \nu(y)$$

Assim, dada uma probabilidade de distribuição q no espaço produto $\Omega\times\Omega$ que satisfaz $\sum_{y\in\Omega} q(x,y) = \mu(x)$ e $\sum_{x\in\Omega} q(x,y) = \nu(y)$, existe um par de variáveis aleatórias (X,Y) que tem q como uma distribuição conjunta e, consequentemente, esse par (X,Y) é um acoplamento de μ e ν.

Desigualdades

12.10 Proposição *Seja μ e ν duas medidas de probabilidade em Ω. Então*

$$||\mu - \nu||_{VT} = \inf\{\mathbf{P}\{X \neq Y\} \ : \ (X,Y) \text{ é um acoplamento de } \mu \text{e } \nu\}$$

Demonstração. *Para todo* $A \in \mathcal{F}$,

$$\mu(A) - \nu(A) = \mathbf{P}[X \in A] - \mathbf{P}[Y \in A] \tag{12.6}$$
$$= \mathbf{P}[X \in A, X = Y] + \mathbf{P}[X \in A, X \neq Y] \tag{12.7}$$
$$- \mathbf{P}[Y \in A, X = Y] - \mathbf{P}[Y \in A, X \neq Y] \tag{12.8}$$
$$= \mathbf{P}[X \in A, X \neq Y] - \mathbf{P}[Y \in A, X \neq Y] \tag{12.9}$$
$$\leq \mathbf{P}[X \neq Y], \tag{12.10}$$

12.11 Exemplo (Acoplamento de variáveis aleatórias de Poisson) *Dado* $X \sim \mathrm{Po}(\lambda_1)$ *e* $Y \sim \mathrm{Po}(\lambda_2)$ *com* $\lambda_1 > \lambda_2$. *Faremos o seguinte acoplamento: sejam* $\hat{Y} \sim \mathrm{Po}(\lambda_2)$, $\hat{Z} \sim \mathrm{Po}(\lambda_1 - \lambda_2)$ *independente de* Y *e* $\hat{X} = \hat{Y} + \hat{Z}$. *Como a soma de variáveis aleatórias de Poisson é de Poisson, então* $(\hat{X},\hat{Y})$ *é um acoplamento e*

$$||\mu_X - \mu_Y||_{VT} \leq \mathbf{P}[\hat{X} \neq \hat{Y}] = \mathbf{P}[\hat{Z} > 0] = 1 - e^{-(\lambda_1 - \lambda_2)} \leq \lambda_1 - \lambda_2$$

Acoplamentos Maximais

12.12 Definição (Acoplamento Maximal) *Dadas duas* μ *e* ν *duas medidas de probabilidade em* Ω *e um acoplamento* (X,Y) *de* μ *e* ν *é dito* ***acoplamento maximal*** *se*

$$||\mu - \nu||_{VT} = \mathbf{P}\{X \neq Y\}$$

Acoplamentos maximais existem. Por simplicidade, provamos isso apenas no caso finito.

12.13 Teorema (Acoplamentos Maximais) *Suponha que* Ω *seja finito e seja* $\mathcal{F} = 2^{\Omega}$. *Sejam* μ *e* ν *duas medidas de probabilidade em* Ω. *Então,*

$$||\mu - \nu||_{VT} = \mathbf{P}\{X \neq Y\} : (X,Y) \text{ acoplamento de } \mu \text{ e } \nu$$

Demonstração. *Seja* $A = \{x \in \Omega : \mu(x) > \nu(x)\}, B = \{x \in \Omega : \mu(x) \leq \nu(x)\}$ *e*

$$p := \sum_{x \in \Omega} \mu(x) \wedge \nu(x), \quad \alpha := \sum_{x \in A} [\mu(x) - \nu(x)], \quad \beta := \sum_{x \in B} [\nu(x) - \mu(x)]$$

Começaremos provando dois fatos:

- $p = \sum_{x \in \Omega} \mu(x) \wedge \nu(x) = 1 - ||\mu - \nu||_{VT}$
- $\alpha - \beta = ||\mu - \nu||_{VT} = 1 - p$ □

Começaremos demonstrando o primeiro fato:

$$
\begin{aligned}
2||\mu - \nu||_{\mathrm{VT}} &= \sum x \in \Omega \, |\mu(x) - \nu(x)| && (12.11)\\
&= \sum x \in A[\mu(x) - \nu(x)] + \sum x \in B[\nu(x) - \mu(x)] && (12.12)\\
&= \sum x \in A\mu(x) + \sum x \in B\nu(x) - \sum_{x \in \Omega} \mu(x) \wedge \nu(x) && (12.13)\\
&= 2 - \sum x \in B\mu(x) + \sum x \in A\nu(x) - \sum_{x \in \Omega} \mu(x) \wedge \nu(x) && (12.14)\\
&= 2 - 2\sum_{x \in \Omega} \mu(x) \wedge \nu(x) && (12.15)\\
& && (12.16)
\end{aligned}
$$

A demonstração do segundo fato. A primeira igualdade é imediata e a segunda é a segunda linha da demonstração anterior.

Acoplamento maximal: *O acoplamento maximal é definido como:*

- *Com probabilidade p escolha $\mathrm{X} = \mathrm{Y}$ de $\gamma_{\min}$ onde*

$$\gamma_{\min}(x) = \frac{1}{p}\mu(x) \wedge \nu(x), \quad x \in S$$

- *Caso contrário escolha X de γ_A onde*

$$\gamma_A(x) = \frac{\mu(x) - \nu(x)}{1-p}, \quad x \in A$$

E independente, escolha Y de

$$\gamma_B(x) = \frac{\nu(x) - \mu(x)}{1-p}, \quad x \in B$$

Dessa forma a distribuição de X é

$$p\gamma_{\min}(x) + (1-p)\gamma_A(x) = \mu(x)$$

De modo análogo temos a distribuição de Y

Finalmente $\mathbf{P}[\mathrm{X} \neq \mathrm{Y}] = 1 - p = ||\mu - \nu||_{\mathrm{VT}}$

12.14 Exemplo (Acoplamento maximal de variáveis de Bernoulli) *Sejam X e Y variáveis aleatórias de Bernoulli com parâmetros $0 \le q < r \le 1$, respectivamente. Ou seja,*

$$\mathbf{P}[\mathrm{X} = 1] = q \text{ e } \mathbf{P}[\mathrm{X} = 0] = 1 - q$$

e

$$\mathbf{P}[Y = 1] = r \text{ e } \mathbf{P}[Y = 0] = 1 - r$$

com $\Omega = \{0, 1\}$ *e* $\mathcal{F} = 2^{\Omega}$.

Para construir o acoplamento maximal como acima, notamos que: $A = \{x \in \Omega : \mu(x) > \nu(x)\} = \{0\}$, $B = \{x \in \Omega : \mu(x) \leq \nu(x)\} = \{1\}$ *e*

$$p = \sum_{x \in \Omega} \mu(x) \wedge \nu(x) = 1 - r + q, \quad \alpha = \beta = 1 - p = r - q$$

então

$$\gamma_{\min}(0) = \frac{1 - r}{1 - r + q}$$

$$\gamma_{\min}(1) = \frac{q}{1 - r + q}$$

$$\gamma_A(0) = 1$$

$$\gamma_B(1) = 1$$

Logo a distribuição do acoplamento maximal (X'', Y'') *é*

$$\mathbf{P}[(X', Y') = (i, j)] = \begin{pmatrix} p\gamma_{\min}(0) & (1 - p)\gamma_A(0)\gamma_B(1) \\ 0 & p\gamma_{\min}(1) \end{pmatrix} \tag{12.17}$$

$$= \begin{pmatrix} (1 - r) & r - q \\ 0 & q \end{pmatrix} \tag{12.18}$$

12.3 Aplicações

Passeio Aleatório

Capítulo 13

Esperança Condicional

Seja $(\Omega, \mathcal{F}, \mathbf{P})$ um espaço de probabilidade. Suponha que X seja uma variável aleatória $\mathcal{F}$-mensurável e seja $\mathcal{G} \subset \mathcal{F}$ ser uma σ-álgebra. Associado a sub-σ-álgebra $\mathcal{G}$ temos um novo espaço de probabilidade $(\Omega, \mathcal{G}, \mathbf{P}|_{\mathcal{G}})$ com $\mathbf{P}|_{\mathcal{G}}$ a restrição de $\mathbf{P}$ para $\mathcal{G}$.

O conjunto de funções mensuráveis e integráveis em relação a $(\Omega, \mathcal{G}, \mathbf{P}|_{\mathcal{G}})$ é menor que o conjunto de funções mensuráveis e integráveis em relação a $(\Omega, \mathcal{F}, \mathbf{P})$. Assim, de modo geral, a variável X não precisa ser $\mathcal{G}$ mensurável. E consequentemente a existência das integrais da forma $\int_G X \; d\mathbf{P}|_{\mathcal{G}}$ onde $G \in \mathcal{G}$ não pode,em geral, ser verificada.

Queremos encontrar uma variável aleatória Y que seja $\mathcal{G}$-mensurável e que em algum sentido seja a mais próxima de X. Para isso considere $G \in \mathcal{G}$, para esse conjunto podemos calcular as integrais

$$\int_G X \; d\mathbf{P} \qquad \int_G Y \; d\mathbf{P}$$

como essas integrais representam as médias locais sobre G, uma condição natural seria exigir que

$$\int_G X \; d\mathbf{P} = \int_G Y \; d\mathbf{P}$$

A função Y é uma versão desfocada de X e representa a "melhor estimativa possível"do valor de X, dada a informação limitada fornecida por $\mathcal{G}$. A função Y está definida sobre uma coleção de eventos menor, mas para esses eventos ela fornece “a mesma informação” que X.

$$\text{restrição a } \mathcal{G} : (\Omega, \mathcal{G}, \mathbf{P}) \longrightarrow (\Omega, \mathcal{G}, \mathbf{P}|_{\mathcal{G}}) \tag{13.1}$$

$$\int_G \mathrm{X}\ \mathrm{d}\mathbf{P} \longmapsto \int_G \mathrm{Y}\ \mathrm{d}\mathbf{P} \tag{13.2}$$

13.1 Definição *Seja $(\Omega, \mathcal{G}, \mathbf{P})$ um espaço de probabilidade. Suponha que X seja uma variável aleatória mensurável $\mathcal{G}$-mensurável. Seja $\mathcal{G} \subset \mathcal{G}$ ser uma σ-álgebra. A* ***esperança condicional de X dado $\mathcal{G}$****, denotada $\mathbf{E}[\mathrm{X}|\mathcal{G}]$ é uma variável aleatória Y tal que:*

$\boxed{1}$ *Y é $\mathcal{G}$-mensurável.*

$\boxed{2}$ *Para $G \in \mathcal{G}$,*

$$\int_G \mathrm{X}\, d\mathbf{P} = \int_G \mathrm{Y}\, d\mathbf{P}.$$

Vamos começar provando a unicidade.

13.2 Teorema *A esperança condicional de uma variável aleatória integrável X se existir é única (a menos de um conjunto nulo).*

Demonstração. *Suponha que Y e Y' satisfaçam as propriedades da esperança condicional. Por (2.), temos*

$$\int_A \mathrm{Y}\, d\mathbf{P} = \int_A \mathrm{Y}'\, d\mathbf{P} \text{ para todo } A \in \mathcal{G}.$$

Então $\int_A (\mathrm{Y} - \mathrm{Y}')\, d\mathbf{P} = 0$ para todo $A \in \mathcal{G}$. Considere o conjunto A_ϵ em que $\mathrm{Y} - \mathrm{Y}' > \epsilon$ para $\epsilon > 0$. Então

$$0 = \int_{A_\epsilon} (\mathrm{Y} - \mathrm{Y}')\, d\mathbf{P} \geq \mathbf{P}(A_\epsilon) \cdot \epsilon.$$

Então, $\mathbf{P}(A_\epsilon) = 0$ para todo $\epsilon > 0$. Como $\{\mathrm{Y} - \mathrm{Y}' > 0\} = \bigcap_{n=1}^{\infty} A_{\frac{1}{n}}$. Então, a propriedade de continuidade das medidas de probabilidade implica em

$$\mathbf{P}(\mathrm{Y} - \mathrm{Y}' > 0) = 0.$$

Da mesma forma, trocando as funções de Y e Y', $\mathbf{P}(\mathrm{Y}' - \mathrm{Y} > 0) = 0$. Logo $\mathrm{Y} = \mathrm{Y}'$ quase certamente

Demonstraremos a existência da esperança condicional de modo mais geral na seção seguinte. Antes disso vamos ver alguns exemplos nos quais conseguimos mostrar a existência exibindo a variável aleatória.

13.3 Exemplo *Seja G ser um evento tal que $\mathbf{P}(G) > 0$ e $\mathcal{G} = \sigma(\{G\}) = \{\emptyset, G, G^c, \Omega\}$. Então,*

$$\mathbf{E}[\mathrm{X}|\mathcal{G}](\omega) = \left\{ \begin{array}{ll} \mathbf{E}[\mathrm{X}|G], & \text{se } \omega \in G \\ \mathbf{E}[\mathrm{X}|G^c], & \text{se } \omega \in G^c. \end{array} \right\} = \mathrm{Y}(\omega)$$

A variável aleatória definida acima é uma esperança condicional.

Demonstração. *A variável* Y *é* $\mathcal{G}$*-mensurável, porque* $\{\omega : Y(\omega) \in B\}$ *está em* $\mathcal{G}$ *para todo conjunto de Borel B em* $\mathbb{R}$.

É fácil ver que

$$\int_G X\, d\mathbf{P} = \int_G Y\, d\mathbf{P}$$

e

$$\int_{G^c} X\, d\mathbf{P} = \int_{G^c} Y\, d\mathbf{P}.$$

Esse exemplo pode ser generalizado:

13.4 Exemplo *Considere* $(\Omega, \mathcal{G}, \mathbf{P})$ *um espaço de probabilidade e* X *uma variável aleatória em* Ω. *Suponha que* P *seja uma partição de* Ω *e que* $\sigma\langle P\rangle$ *seja a sigma-álgebra correspondente.*

Então $\mathbf{E}[X|\sigma\langle P\rangle]$ *é uma função simples, constante em todo conjunto da partição. E tal que*

$$\mathbf{E}[X|\mathcal{G}](\omega) = \mathbf{E}[X|A], \text{ onde } A \text{ é o conjunto tal que } \omega \in A$$

Veja o exemplo unidimensional da Figura 13.1. ◁

13.5 Heurística *Imagine que* X *seja a variável aleatória que mede o quanto de dinheiro uma pessoa tem na sua conta corrente ao final de cada dia.*

Podemos pensar que controlar a quantia de dinheiro diariamente é desnecessário e dessa forma podemos querer trocar a escala para semanal. Então Y *será a variável que representa a quantia de dinheiro em cada semana.*

Veja que agora temos fundir a informação de vários dias de modos para determinar quanto dinheiro temos em cada semana. Nesse caso a esperança condicional é o único modo de aglutinar essas informações preservando as esperanças.

13.1 Existência e Unicidade de Esperança Condicional

Existência em $\mathcal{L}^2$

Começaremos provando a existência em um sentido um pouco mais fraco. A vantagem dessa abordagem é tornar claro a interpretação geométrica da esperança condicional.

Considere o espaço de Hilbert $\mathcal{L}^2(\Omega, \mathcal{F}, \mathbf{P})$. A norma no espaço das variáveis aleatórias é

$$\|X\|_2 = (\mathbf{E}[X^2])^{1/2}.$$

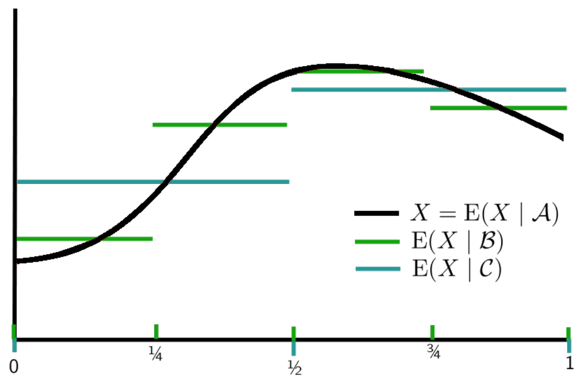
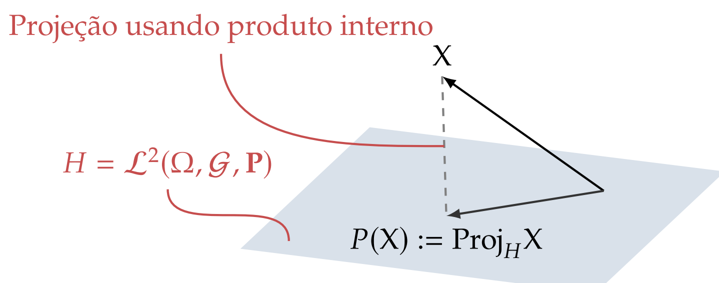


Figura 13.1: A esperança condicional em relação a uma σ-álgebra: neste exemplo, o espaço de probabilidade $(\Omega, \mathcal{G}, \mathbf{P})$ é o intervalo [0,1] com a medida Lebesgue. Definimos as seguintes σ-álgebras: $\mathcal{A} = \mathcal{G}$; $\mathcal{B}$ é a σ-álgebra gerada pelos intervalos com pontos finais 0, 1/4, 1/2, 3/4, 1; e $\mathcal{C}$ é a σ-álgebra gerada pelos intervalos com pontos finais 0, 1/2, 1. Nesse exemplo a esperança condicional é efetivamente a média sobre os conjuntos minimais da σ-álgebra.

Como é um espaço de Hilbert, a norma é realmente dada por um produto interno, que é

$$\langle X, Y \rangle = \mathbf{E}[XY].$$

13.6 Lema *$\mathcal{L}^2(\Omega, \mathcal{G}, \mathbf{P})$ é um subspaço fechado de $\mathcal{L}^2(\Omega, \mathcal{F}, \mathbf{P})$.* ◁



13.7 Teorema *Seja $(\Omega, \mathcal{G}, \mathbf{P})$ um espaço de medida, e $\mathcal{G}$ uma sub-σ-algebra de $\mathcal{G}$. Seja*

$$P : \mathcal{L}^2(\Omega, \mathcal{F}, \mathbf{P}) \to \mathcal{L}^2(\Omega, \mathcal{G}, \mathbf{P})$$

a aplicação de projeção ortogonal. Então, para qualquer $X \in \mathcal{L}^2(\Omega, \mathcal{G}, \mathbf{P})$, temos:

$$\mathbf{E}[X, \mathcal{G}] = P(X).$$

Demonstração. *A projeção em $\mathcal{L}^2(\Omega, \mathcal{G}, \mathbf{P})$ é por definição uma função $\mathcal{G}$-mensurável.*

Além disso para $G \in \mathcal{G}$,

$$\int_G X \, d\mathbf{P} = \int_\Omega X \mathbb{1}_G \, d\mathbf{P} = \int_\Omega Y \mathbb{1}_G \, d\mathbf{P} = \int_G Y \, d\mathbf{P}.$$

pois $X = Y + Z$ *com* $Z \perp \mathcal{L}^2(\Omega, \mathcal{G}, \mathbf{P})$.

Existência em $\mathcal{L}^1$

Provaremos agora que a esperança condicional existe e é única em $\mathcal{L}^1$. Esse fato é consequência do Teorema de Radon-Nikodym.

13.8 **Teorema** *A esperança condicional de uma variável aleatória integrável* X *existe e é única (a menos de um conjunto nulo).*

Demonstração. *Para provar a existência, aplicaremos o Teorema de Radon-Nikodym. Em primeiro lugar, faremos o caso para* $X \geq 0$. *Então, na notação do Teorema de Radon-Nikodym para* $(\Omega, \mathcal{G})$, $\mu = \mathbf{P}$. *E a medida* ν *é definida como*

$$\nu(A) := \int_A X \, d\mathbf{P} \text{ para } A \in \mathcal{G}.$$

Na verdade, ν *é uma medida* σ*-finita. Claramente* $\nu \ll \mu$: *se* A *for um conjunto nulo com respeito a* μ, *então a integração sobre o conjunto* A *dá zero.*

Assim, estamos nas hipóteses do Teorema de Radon-Nikodym. Concluímos que existe uma função mensurável f *tal que*

$$\nu(A) = \int_A f \, d\mathbf{P} \text{ para todo } A \in \mathcal{G}.$$

Esta função será nossa variável aleatória: $Y := f$. *Essa função satisfaz ambas as propriedades que definem a esperança condicional. Isso prova a existência, para* X *não negativa.*

Para uma função X *arbitrária, decompomos* $X = X^+ - X^-$ *com* $X^+ \geq 0$ *e* $X^- \geq 0$. *Encontramos a esperança condicional de cada parte. Consideramos*

$$\begin{aligned} Y_1 &:= \mathbf{E}[X^+|\mathcal{G}] \\ Y_2 &:= \mathbf{E}[X^-|\mathcal{G}], \end{aligned}$$

que existem pelo acima. Então, defina $Y = Y_1 - Y_2$. *Então, claro,* Y *é* $\mathcal{G}$*-mensurável. A segunda propriedade segue por linearidade:*

$$\int_A Y \, 1 \, d\mathbf{P} - \int_A Y_2 \, d\mathbf{P} - \int_A X^+ \, d\mathbf{P} - \int_A X^- \, d\mathbf{P} = \int_A X \, d\mathbf{P}.$$

Alguns exemplos:

13.9 **Exemplo**

1 *A σ-álgebra trivial $\mathcal{G} = \{\emptyset, \Omega\}$. Nesse caso uma função ser $\mathcal{G}$-mensurável é uma condição forte. A condição 1. implica* Y *é constante (quase certamente). Na verdade,* $Y = \mathbf{E}X$.

Na verdade, verificamos a condição 2., ou seja, verificamos que

$$\int_A (\mathbf{E}X)\, d\mathbf{P} = \int_A X\, d\mathbf{P}$$

para $A = \Omega$. Isso é verdade, uma vez que o lado direito é uma esperança. Assim, ambos os lados são iguais a $\mathbf{E}X$.

2 *Considere a σ-álgebra completa $\mathcal{G} = \mathcal{F}$. Então*

$$\mathbf{E}[X|\mathcal{G}] = X.$$

3 *A σ-álgebra discreta. Suponha que $\Omega = \bigcup_k \Omega_k$ é uma partição de Ω tal que $\mathbf{P}(\Omega_k) > 0$. Gere $\mathcal{G}$ por esses conjuntos: $\mathcal{G} = \sigma(\Omega_1, \Omega_2, \ldots)$. Então,*

$$\mathbf{E}[X|\mathcal{G}](\omega) = \mathbf{E}[X|\Omega_k](\omega) \text{ para } \omega \in \Omega_k.$$

A verificação deste item será deixada como exercício.

13.2 Esperança Condicional em Relação a uma Variável Aleatória

13.10 Definição *Dadas duas variáveis aleatórias definimos*

$$\mathbf{E}[X|Y] := \mathbf{E}[X|\sigma(Y)].$$

onde $\sigma(Y) = \{\{\omega : Y(\omega) \in B\}, B \text{ Borel }\}$.

13.11 Exemplo *Por exemplo, se Y é uma variável aleatória que toma um conjunto finito de valores $y_1, \ldots, y_n$. Então $\sigma(Y) = \sigma(\Omega_1, \ldots, \Omega_n)$, onde $\Omega_k = \{Y = y_k\}$. Então, $\mathbf{E}[X|Y](\omega) = \mathbf{E}[X|\Omega_k](\omega)$ para $\omega \in \Omega_k$.*

Em outras palavras, $\mathbf{E}[X|Y](\omega) = \mathbf{E}[X|Y = y_k]$, para $\omega \in \Omega_k$.

13.3 Propriedades da Esperança Condicional

13.12 Proposição (Linearidade) $\mathbf{E}[aX + bY|\mathcal{G}] = a\mathbf{E}[X|\mathcal{G}] + b\mathbf{E}[Y|\mathcal{G}]$.

Demonstração. *Iremos verificar que o lado direito da expressão é a esperança condicional de $aX+bY$ dada a σ-álgebra $\mathcal{G}$. Verificaremos as propriedades:*

$\boxed{1}$ *$a\mathbf{E}[X|\mathcal{G}]+b\mathbf{E}[Y|\mathcal{G}]$ é $\mathcal{G}$-mensurável, pois é uma combinação linear de duas funções $\mathcal{G}$-mensuráveis.*

$\boxed{2}$ *Para todo $A \in \mathcal{G}$,*

$$\begin{aligned}\int_A (a\mathbf{E}[X|\mathcal{G}] + b\mathbf{E}[Y|\mathcal{G}])\, d\mathbf{P} &= a\int_A \mathbf{E}[X|\mathcal{G}]\, d\mathbf{P} + b\int_A \mathbf{E}[Y|\mathcal{G}]\, d\mathbf{P} \text{ por linearidade}\\ &= a\int_A X\, d\mathbf{P} + b\int_A Y\, d\mathbf{P} \text{ por definição}\\ &= \int_A (aX + bY)\, d\mathbf{P} \text{ por linearidade.}\end{aligned}$$

13.13 Corolário *A esperança condicional é um operador linear em $L^1(\Omega,\mathcal{F},\mathbf{P})$.*

Uma observação fundamental a ser feita aqui é a diferença entre a esperança, que é um funcional linear, e a esperança condicional é um operador linear.

13.14 Proposição (Monotonicidade) *Se $X \leq Y$ q.c. , então $\mathbf{E}[X|\mathcal{G}] \leq \mathbf{E}[Y|\mathcal{G}]$ q.c.*

Demonstração. *Considere um evento $A \in \mathcal{G}$. Então*

$$\begin{aligned}\int_A \mathbf{E}[X|\mathcal{G}]\, d\mathbf{P} &= \int_A X\, d\mathbf{P}\\ &\leq \int Y\, d\mathbf{P}\\ &= \int_A \mathbf{E}[Y|\mathcal{G}]\, d\mathbf{P}\end{aligned}$$

Portanto, $\int_A (\mathbf{E}[Y|\mathcal{G}] - \mathbf{E}[X|\mathcal{G}])\, d\mathbf{P} \geq 0$ para todo$A \in \mathcal{G}$. Portanto, $\mathbf{E}[Y|\mathcal{G}] - \mathbf{E}[X|\mathcal{G}] \geq 0$ q.c.. □

13.15 Corolário $\left|\mathbf{E}[X|\mathcal{G}]\right| \leq \mathbf{E}[|X|\,|\mathcal{G}]$.

Demonstração. *Observe $X \leq |X|$ e $-X \leq |X|$.* □

Temos também um Teorema de Convergência Dominada para Esperanças Condicionais:

13.16 Teorema (Teorema da Convergência Dominada) *Suponha que as variáveis aleatórias X_n convergem para X q.c. e $|X_n| \leq Y$ q.c. , com Y integrável. Então,*

$$\mathbf{E}[X_n|\mathcal{G}] \longrightarrow \mathbf{E}[X|\mathcal{G}] \text{ q.c.}$$

Demonstração.

$$\begin{aligned}|\mathbf{E}[X_n|\mathcal{G}] - \mathbf{E}[X|\mathcal{G}]| &= |\mathbf{E}[X_n - X|\mathcal{G}]| \\ &\leq \mathbf{E}[|X_n - X| \,\big|\, \mathcal{G}] \\ &\leq \mathbf{E}[Z_n|\mathcal{G}], \textit{ onde } Z_n = \sup_{k\geq n} |X_k - X| \textit{ e claramente } Z_n \downarrow .\end{aligned}$$

Na verdade, $Z_n \downarrow 0$ *por hipótese. Queremos mostrar que* $\mathbf{E}[Z_n|\mathcal{G}] \downarrow 0$ *q.c.*

Pela monotonicidade das esperanças condicionais, temos que $\mathbf{E}[Z_n|\mathcal{G}]$ *não aumenta q.c. Uma vez que não é negativa, converge para um limite:*

$$\mathbf{E}[Z_n|\mathcal{G}] \downarrow Z \textit{ q.c.}$$

Se $Z \geq 0$, $\mathbf{E}[Z] = 0$ *implica* $Z = 0$. *Então, basta mostrar que* $\mathbf{E}[Z] = 0$. *Para a esperança, podemos usar o Teorema de Convergência Dominada:*

Dessa forma temos $0 \leq Z \leq Y + |X| \leq 2Y$. *Portanto*

$$\begin{aligned}\mathbf{E}Z &= \int_\Omega Z \, d\mathbf{P} \\ &\leq \int \mathbf{E}[Z_n|\mathcal{G}] \, d\mathbf{P} \\ &= \mathbf{E}Z_n, \textit{ pela definição de esperança condicional}.\end{aligned}$$

Como $Z_n \downarrow 0$, *o Teorema de Convergência Monótona fornece*

$$\mathbf{E}[Z_n] \longrightarrow 0.$$

Portanto, $\mathbf{E}[Z] = 0$. □

13.17 Proposição *Se* X *for* $\mathcal{G}$*-mensurável, então*

$$\mathbf{E}[XY|\mathcal{G}] = X \cdot \mathbf{E}[Y|\mathcal{G}]$$

se XY *e* X *forem integráveis.*

Demonstração. *Verificamos que* $X \cdot \mathbf{E}[Y|\mathcal{G}]$ *é a esperança condicional de* XY *dado* $\mathcal{G}$.

$\boxed{1}$ $X \cdot \mathbf{E}[Y|\mathcal{G}]$ *é* $\mathcal{G}$*-mensurável, porque* X *e* $\mathbf{E}[Y|\mathcal{G}]$ *são* $\mathcal{G}$*-mensurável.*

$\boxed{2}$ *Para todo* $A \in \mathcal{G}$, *precisamos verificar que*

$$\int_A XE[Y|\mathcal{G}] \, d\mathbf{P} = \int_A XY \, d\mathbf{P}. \tag{13.3}$$

a)Se $X = \mathbb{1}_B$ *para* $B \in \mathcal{G}$*, então* (13.3) *se torna*

$$\int_{A\cap B} \mathbf{E}[Y|\mathcal{G}]\, d\mathbf{P} = \int_{A\cap B} Y\, d\mathbf{P}$$

o que é verdadeiro porque $A \cap B \in \mathcal{G}$.

b)Se X *é uma variável aleatória simples, então* (13.3) *é verdadeiro por linearidade (da integral).*

c)Se $X \geq 0$*, então existem variáveis aleatórias simples* $X_n \uparrow X$ *q.c. Em seguida, usamos* (13.3) *por* X_n *e apliquemos o Teorema de Convergência Monótona para completar o argumento e alcançar* (13.3) *por* X.

d)Para X *arbitrário usaremos a decomposição* $X = X^+ - X^-$ *e a linearidade para provar* (13.3).

Isso verifica a propriedade 2.. □

Essa propriedade se mostrará útil quando lidarmos com martingales.

13.4 Esperança Condicional como Projeção

13.18 Proposição (Suavização/Elevação) *Se* $\mathcal{G}_1 \subseteq \mathcal{G}_2$*, então*

$\boxed{1}$ $\mathbf{E}[\mathbf{E}[X|\mathcal{G}_1]|\mathcal{G}_2] = \mathbf{E}[X|\mathcal{G}_1]$.

$\boxed{2}$ $\mathbf{E}[\mathbf{E}[X|\mathcal{G}_2]|\mathcal{G}_1] = \mathbf{E}[X|\mathcal{G}_1]$.

Demonstração. *Seja* $\mathcal{G}_1 \subseteq \mathcal{G}_2$.

$\boxed{1}$ $Y = \mathbf{E}[X|\mathcal{G}_1]$ *é* $\mathcal{G}_1$*-mensurável.* $\mathcal{G}_1 \subseteq \mathcal{G}_2$*, portanto, também é* $\mathcal{G}_2$*-mensurável. Consequentemente*

$$\mathbf{E}[Y|\mathcal{G}_2] = Y$$

$\boxed{2}$ *Queremos mostrar que* $\mathbf{E}[X|\mathcal{G}_1]$ *é a esperança condicional de* $\mathbf{E}[X|\mathcal{G}_2]$ *dado* $\mathcal{G}_1$.

a)Na verdade, $\mathbf{E}[X|\mathcal{G}_1]$ *é* $\mathcal{G}_1$*-mensurável.*

b)Para $A \in \mathcal{G}_1$*, queremos mostrar que*

$$\int_A \mathbf{E}[X|\mathcal{G}_1]\, d\mathbf{P} = \int_A \mathbf{E}[X|\mathcal{G}_2]. \tag{13.4}$$

O lado esquerdo é a integral de X *em* A*. O lado direito também pois* $A \in \mathcal{G}_1 \subseteq \mathcal{G}_2$*, a definição de esperança condicional mostra que ambos os lados de* (13.4) *são iguais à*

$$\int_A X\, d\mathbf{P}.$$

Assim, (1.) e (2.) são verdadeiras. □

13.19 Corolário $\mathbf{E}[\mathbf{E}[X|\mathcal{G}]|\mathcal{G}] = \mathbf{E}[X|\mathcal{G}]$.

Ou seja, o corolário anterior mostra que a esperança condicional $P : X \mapsto \mathbf{E}[X|\mathcal{G}]$ é uma projeção linear, isto é $P^2 = P$ em $L^1(\Omega,\mathcal{G},\mathbf{P})$ o subespaço de todas as variáveis aleatórias $\mathcal{G}$-mensuráveis.

13.20 Teorema (A esperança condicional é o melhor estimador) *Para qualquer variável aleatória $\mathcal{G}$-mensurável Y, temos*

$$\mathbf{E}[X - \mathbf{E}[X|\mathcal{G}]]^2 \leq \mathbf{E}[X - Y]^2. \tag{13.5}$$

Em outras palavras, para qualquer outra variável aleatória $\mathcal{G}$-mensurável, o erro quadrático será pior.

Começaremos provando alguns fatos geométricos Primeiro provaremos que o erro $X - \mathbf{E}[X|\mathcal{G}]$ é ortogonal à $\mathcal{L}^2(\Omega,\mathcal{G},\mathbf{P})$.

13.21 Proposição (Ortogonalidade) $X - \mathbf{E}[X|\mathcal{G}]$ *é ortogonal a* $\mathcal{L}^2(\Omega,\mathcal{G},\mathbf{P})$*, ou seja, para todo* $Z \in \mathcal{L}^2(\Omega,\mathcal{G},\mathbf{P})$*, temos* $\mathbf{E}Z(X - \mathbf{E}[X|\mathcal{G}]) = 0$.

Demonstração. *Na verdade,* $\mathbf{E}Z(X - \mathbf{E}[X|\mathcal{G}]) = \mathbf{E}ZX - \mathbf{E}[Z\mathbf{E}[X|\mathcal{G}])$. *Uma vez que* Z *é* $\mathcal{G}$-*mensurável, podemos colocá-lo no interior do lado direito (uma vez que atua como uma constante). Assim, a expressão é* $= \mathbf{E}ZX - \mathbf{E}[\mathbf{E}[ZX|\mathcal{G}])$. *Pela propriedade de suavização, temos que* $= \mathbf{E}ZX - \mathbf{E}ZX = 0$. □

Como consequência temos uma nova demonstração do seguinte fato:

13.22 Corolário *A esperança condicional* $\mathbf{E}[\,\cdot\,|\,\mathcal{G}]$ *é a projeção ortogonal em* $\mathcal{L}^2(\Omega,\mathcal{F},\mathbf{P})$ *para o subespaço* $\mathcal{L}^2(\Omega,\mathcal{G},\mathbf{P})$ *de todas as variáveis aleatórias* $\mathcal{G}$-*mensuráveis.*

Demonstração. *Agora provaremos o Teorema 13.20*

Seja $Y \in \mathcal{L}^2(\Omega,\mathcal{G},\mathbf{P})$. *Desejamos mostrar que* $\mathbf{E}[X - Y]^2 \geq \mathbf{E}[X - \mathbf{E}[X|\mathcal{G}])^2$. *Como*

$$\begin{aligned}
\mathbf{E}[X - Y]^2 &= \mathbf{E}[X - \mathbf{E}[X|\mathcal{G}] + Z)^2, \quad Z = \mathbf{E}[X|\mathcal{G}] - Y \in \mathcal{L}^2(\Omega,\mathcal{G},\mathbf{P}) \\
&= \mathbf{E}[X - \mathbf{E}[X|\mathcal{G}])^2 + \mathbf{E}Z^2 + 2\mathbf{E}Z(X - \mathbf{E}[X|\mathcal{G}])
\end{aligned}$$

O segundo termo não é negativo e o terceiro termo já demonstramos ser zero. portanto

$$\mathbf{E}[X - Y]^2 \geq \mathbf{E}[X - \mathbf{E}[X|\mathcal{G}])^2.$$

Isso completa a prova. □

Acabamos de demonstrar o Teorema de Pitágoras.

Apêndice A

Integral de Riemann-Stieltjes

Lembre-se do cálculo de que se f for contínua no intervalo $[a,b]$, então definimos a integral Riemann de f em $[a,b]$como:

$$\int_a^b f(x)\,dx = \lim_{n\to\infty}\sum_{k=1}^{n} f(x_k^*)\Delta x \tag{A.1}$$

onde $\Delta x = \frac{b-a}{n}$ e $x_k^*[x_{k-1}, x_k]$ para cada subintervalo $k \in \{1, 2, \ldots, n\}$. Se o limite acima existe, escrevemos $\int_a^b f(x)\,dx = A$.

Geometricamente, se f for uma função positiva e contínua em $[a,b]$, então $\int_a^b f(x)dx = A$ representa a área abaixo do f, acima do eixo x, e delimitado pelas retas $x = a$ e $x = b$.

Vamos agora olhar introduzir um conceito mais geral de integral denominada Riemann-Stieltjes.

A.1 Definição (Riemann-Stieltjes) *Deixe $I = [a,b]$ ser um intervalo fechado e deixe f, α serem funções definidas em $[a,b]$. Além disso, deixe $P = \{a = x_0, x_1, \ldots, x_n = b\} \in \mathcal{P}[a,b]$ e para cada $k \in \{1, 2, \ldots, n\}$ deixe $t_k \in [x_{k-1}, x_k]$ (o k^{th} subintervalo de $[a,b]$ em relação à partição P) e deixe $\alpha_k = \alpha(x_k) - \alpha(x_{k-1})$.*

A soma de Riemann-Stieltjes com relação à partição P e as funções f e α é definida como

$$S(P, f, \alpha) = \sum_{k=1}^{n} f(t_k)\Delta\alpha_k.$$

A função f é dita ser Riemann-Stieltjes integrável em relação a α em $[a,b]$ se existir $L \in \mathbb{R}$ tal que para cada $\epsilon > 0$ existe uma partição $P_\epsilon \in \mathcal{P}[a,b]$ tal que para todas as partições P mais finas do que P_ϵ ($P_\epsilon \subseteq P$) e para qualquer escolha de $t_k \in [x_{k-1}, x_k]$ nós temos que $| S(P,f,\alpha) - L | < \epsilon$. Se tal $L \in \mathbb{R}$ existe, dizemos que esta integral de Riemann-Stieltjes existe e escreve $\displaystyle\int_a^b f(x)\,d\alpha(x) = L$.

Às vezes, a notação "$f \in R(\alpha)$ em $[a,b]$"será usada para indicar que f é Riemann-Stieltjes integrável em relação a α em $[a,b]$.

Se o limite L existe, a notação abreviada $\displaystyle\int_a^b f\,d\alpha = L$ também será usada.

Além disso, a notação $f \in R(\alpha)$ em $[a,b]$ significa que f é Riemann-Stieltjes integrável em relação a α no intervalo $[a,b]$.

Apêndice B

História da Teoria da Medida no século XX

Esse texto é a tradução de um trecho do artigo História da Teoria da Medida no século XX de Maurice Sion [34].

B.1 Medida

A noção matemática de medida pretende representar conceitos como comprimento, área, volume, massa, carga elétrica, etc., do mundo físico. Os objetos a serem medidos são representados por conjuntos e uma medida é uma função aditiva de conjuntos, ou seja, o valor que atribui à união de dois conjuntos disjuntos é a soma dos valores que atribui a cada um dos conjuntos.

Exemplos concretos de medidas e de métodos para calcular a medida de conjuntos específicos são tão antigos quanto a história registrada. A teoria foi introduzida pela primeira vez pelos gregos antigos como parte do desenvolvimento de um sistema numérico. Uma teoria mais sistemática apareceu na forma de integração no cálculo de Newton e Leibniz, na segunda metade do século 17. Nesta teoria o gráfico de uma função f é usado para descrever a fronteira de um conjunto cuja medida é a integral de f. O teorema fundamental do cálculo estabeleceu a conexão entre a integral e a derivada, ou seja, da área com a taxa de variação, e forneceu uma nova e poderosa ferramenta para a computação e o estudo das propriedades de medidas.

No século 19, motivado em grande parte pelo trabalho de Fourier sobre a teoria do calor que exigia a integração de expressões mais complicadas do que as até então consideradas, foi levado a cabo um programa para reexaminar as noções de função, continuidade, integral e derivadas. Tal tarefa foi empreendida por alguns dos principais matemáticos da época. Isso levou a uma definição geral de integral por Riemann no meio do século, mas também gerou a percepção de que o teorema fundamental do cálculo, bem como outros teoremas extensivamente usados no intercâmbio da ordem de integração com as operações limite não eram válidos, sem inúmeras hipóteses. Esse estudo levou também ao interesse por conjuntos muito mais complicados do que os que já haviam sido considerados, conjuntos estes que não eram descritos por condições geométricas ou físicas intuitivas mas sim indiretamente por expressões analíticas. Por exemplo, o conjunto de todos os pontos em que alguns função dada é descontínua.

Como resultado, da generalidade das funções e conjuntos que agora precisavam ser incluídos, a teoria da medida ou a integração já não tinham a simplicidade e facilidade de aplicação que prevalecia na definição mais elementar do passado. O trabalho nesta área, na segunda metade do século, estava preocupado principalmente com a busca de uma noção mais adequada de medida e integral, que renderia teoremas mais simples e mais poderosos.

Foi apenas nos últimos anos do século 19 que E. Borel, finalmente, introduziu uma noção de medida na reta real, o que levaria a uma das teorias mais belas e mais amplamente utilizada na matemática. O desenvolvimento desta teoria, principalmente por H. Lebesgue nos estágios iniciais, e suas aplicações subsequentes para quase todos os ramos de análise e às principais áreas da física constituem uma das partes mais importantes da história da matemática no século 20.

Medida de Borel

Em 1898 Borel estendeu a noção de tamanho de intervalos para o conceito de uma medida sobre uma ampla classe de conjuntos na reta real, que tem propriedades particularmente desejáveis. Esta medida é agora denominada como medida de Borel. Sua principal ideia nova foi a noção de aditividade enumerável. Uma função de uma família de conjuntos é enumeravelmente aditiva se o valor que esta atribui a união de uma sequência infinita de conjuntos disjuntos é a soma dos valores que atribui a cada um dos elementos da sequência.

Começando com a família de intervalos e a função que atribui a cada intervalo de seu comprimento, ele utilizou recursão para ampliar passo a passo o domínio de definição da função, adicionando em cada cenários conjuntos cujo complemento foi previamente definido ou que são a união de uma sequência de conjuntos disjuntos anteriormente definidos.

A família resultante no final é estável sob as operações de complementação e uniões enumeráveis e a medida resultante é contavelmente aditiva. Agora, qualquer família de conjuntos em qualquer espaço com estas propriedades é chamado de uma família Borel.

Medida e Integração de Lebesgue

Em sua famosa tese publicada em 1902, Henri Lebesgue simplificou e generalizou a definição de medida de Borel e desenvolveu uma teoria de integração e diferenciação em que grande parte da análise dias atuais repousa.

Restringindo-se primeiramente aos subconjuntos de intervalo unitário e partindo de que um conjunto aberto é uma união disjunta de uma seqüência de intervalos, ele definiu a medida de um conjunto aberto como a soma dos comprimentos desses intervalos. Como um conjunto fechado é o complemento de um conjunto aberto, ele definiu a medida de um conjunto fechado como 1 menos a medida de seu complemento. Em seguida, ele definiu a medida exterior de qualquer conjunto como o ínfimo das medidas de conjuntos abertos que o contenham e a medida interior do conjunto como o supremo das medidas de conjuntos fechados nele contidos.

Se a medida exterior e interior de um conjunto coincidem, este é chamado mensurável e sua medida é este valor comum. Ele mostrou que a família de conjuntos mensuráveis contém os conjuntos de Borel e que sua medida concorda com medida de Borel sobre a conjuntos de Borel. Sua medida também é contável aditiva e a família de conjuntos mensuráveis é um sigma álgebra. Ele, então, estendeu essa medida para toda a linha real e, por analogia, introduziu medidas similares em espaços euclidianos de dimensão superior para representar a área no plano, o volume no espaço tridimensional, etc.

Voltando a integração, ele primeiro definiu a integral de uma função positiva sobre os reais como a medida da região bidimensional sob o seu gráfico, em seguida, a integral de qualquer função como a diferença das integrais da parte positiva e negativa. Ele chamou uma função de mensurável se a imagem inversa de um intervalo é um conjunto mensurável e, em seguida, provou que a integral de uma função mensurável limitada em um intervalo limitado existe. Assim, ele ampliou a integral de Riemann adequando o processo de integração a uma classe mais ampla de funções.

Depois de verificar propriedades algébricas elementares da integral, ele provou um dos resultados mais profundos da análise, o Teorema da Convergência Dominada de Lebesgue: se uma seqüência de funções integráveis converge para uma função integrável então a integral do limite é o limite das integrais

Em 1904, voltando sua atenção para a diferenciação, ele recuperou o teorema fundamental do cálculo de uma forma particularmente simples. Dizemos que uma afirmação é verdadeira quase sempre, se o conjunto de pontos em que não é verdadeira tem medida zero. De posse desse conceito Lebesgue primeiro provou que uma função monótona, e, portanto, a soma ou a diferença de duas funções monótonas, é diferenciável em quase todo ponto. Em seguida, ele mostrou que a integral indefinida de uma função é diferenciável em quase todo ponto.

Finalmente, ele caracterizou todas as funções que podem ser expressas como integrais indefinidas: essas funções são absolutamente contínuas, funções estas introduzidas anteriormente por G. Vitali. Uma função é absolutamente contínua se a sua variação total num conjunto aberto tende à zero à medida que a medida do conjunto aberto tende à zero.

Agora, uma função cuja variação total é limitada é a diferença de duas funções monótonas e, portanto, é diferenciável quase todos os pontos. No entanto, a integral indefinida de sua derivada não precisa ser igual a ela. A diferença é caracterizada pelo fato de que sua derivada é zero em quase todos os pontos. Funções com esta propriedade são chamadas singulares e não precisam ser constante, e na verdade, elas podem assumir todos os valores em um intervalo. Lebesgue mostrou que uma função de variação limitada pode ser decomposta em uma soma de uma função absolutamente contínua e de uma função singular. Este é o famoso Teorema de decomposição de Lebesgue.

Assim, nos primeiros anos do século, Lebesgue estabeleceu as bases de uma teoria da medida e integração na reta real, que alargou o âmbito do cálculo de uma forma inimaginável antes dele e que recuperou a liberdade para operar com integrais, derivados e limites com um mínimo de restrições naturais e simples. As próximas décadas foram para testemunhar o crescimento explosivo do campo, as aplicações cada vez mais amplas destas ideias para outras áreas, e o desenvolvimento novos ramos da matemática inspirado por essas noções e resultados.

B.2 Probabilidade

Até o advento da teoria da medida o campo de probabilidade foi principalmente uma coleção de problemas de azar e de métodos ad hoc para resolvê-los. O interesse de matemáticos nestes problemas remonta a Pascal e Fermat no século 17. No entanto, não havia nenhuma teoria geral, nem boas definições fundamentais em que se basearem. Esta é uma das principais realizações do século 20 e é inteiramente devido ao desenvolvimento da teoria da medida.

O uso de uma medida para descrever a probabilidade foi feita por E. Borel na virada do século. Em 1920-1923 N. Wiener construiu uma medida sobre o espaço de curvas contínuas para representar a probabilidade que descreve o movimento de partículas suspensas num fluido conhecido como movimento Browniano. Este fenômeno tinha sido objeto de estudo por físicos há algum tempo e de tratamentos matemáticos por A. Einstein e por M. Smoluchowski por volta de 1905. A ideia de usar uma medida em um espaço tão complicado para descrever um fenômeno físico abriu totalmente novas perspectivas. Esta medida, agora referida como medida de Wiener, eventualmente, levou ao estabelecimento de uma teoria geral da probabilidade e de processos estocásticos nas próximas duas décadas.

Em 1930-1933 A. Kolmogorov lançou as bases formais para a teoria da probabilidade O espaço básico é um conjunto que representa todas as possibilidades do fenômeno a ser estudado. Um evento é um subconjunto deste espaço e uma probabilidade é uma medida positiva em uma álgebra de eventos que atribui o valor 1 para todo o espaço. A variável aleatória é uma função mensurável neste espaço e a sua expectativa média é a integral dessa função. Assim, a teoria das probabilidades foi incorporada na teoria da medida e, assim, entrou na corrente principal da matemática. Por um lado, isso permitiu probabilidade de crescer e ampliar seu campo de aplicações em novas direções, desde à estatísticas até a teoria de potencial da física matemática, enquanto por outro lado, motivou o estudo de problemas e utilização de métodos de campos totalmente diferentes.

A medida de Wiener e o trabalho pioneiro relacionado de Paul Levy na década de 1920 e 30, eventualmente, levou ao nascimento de um ramo de probabilidade conhecida como processos estocásticos, estabelecida inicialmente principalmente por JL Doob. Trata-se de medidas de probabilidade em espaços de dimensão infinita, como espaços de curvas. Esta área tem experimentado extraordinário crescimento na segunda metade do século e tem encontrado aplicações a problemas em quase todos os ramos de análise, nas ciências sociais, bem como nas ciências físicas. Ela está preocupada principalmente com a evolução de um sistema no tempo. O sistema é representado por um ponto em um espaço S, o qual pode ser bastante complexa dependendo do fenômeno em estudo, e a evolução no tempo por uma curva no espaço S. A lei que rege esta evolução é representado por uma medida sobre o espaço das curvas. A construção de uma dessas medidas a partir de algumas observações primitivas ou pressupostos é geralmente um grande problema fundamental. Uma vez obtido, esta medida torna-se uma das ferramentas mais eficazes para estudar o fenômeno.

Respostas dos Exercícios

2.2

Bibliografia

[1] Mark (https://mathoverflow.net/users/7392/mark). *Demystifying the Carathéodory Approach to Measurability*. MathOverflow. URL:https://mathoverflow.net/q/34029 (version: 2010-07-31).

[2] Charalambos D Aliprantis e Kim Border. *Infinite Dimensional Analysis: A Hitchhiker's Guide*. Springer Science & Business Media, 2006.

[3] Robert Ash. *Real Analysis and Probability*. Vol. 6. Academic Press, 1972, p. 70.

[4] Robert G Bartle. *The elements of integration and Lebesgue measure*. John Wiley & Sons, 2014.

[5] Andrej Bauer. *Why do roots of polynomials tend to have absolute value close to 1?* MathOverflow `https://mathoverflow.net/q/182412`. 2014.

[6] Heinz Bauer e R... B. Burckel. *Probability theory and elements of measure theory*. Academic Press London, 1981.

[7] V Bentkus et al. *Limit theorems of probability theory*. Springer Science & Business Media, 2013.

[8] Patrick Billingsley. *Convergence of probability measures*. John Wiley & Sons, 2013.

[9] Patrick Billingsley. *Probability and measure*. John Wiley & Sons, 2008.

[10] Salomon Bochner. *Harmonic analysis and the theory of probability*. Courier Corporation, 2013.

[11] Vladimir I Bogachev. *Measure theory*. Vol. 1. Springer Science & Business Media, 2007.

[12] Leo Breiman. *Probability*. 1992.

[13] Marek Capinski e Peter E Kopp. *Measure, integral and probability*. Springer Science & Business Media, 2013.

[14] Jair Donadelli. *Processos de Ramificação*. `https://anotacoesdeaula.wordpress.com`. 2016.

[15] Rick Durrett. *Probability: theory and examples*. Cambridge university press, 2010.

[16] Terrence L Fine. *Theories of probability: An examination of foundations*. Academic Press, 2014.

[17] Gerald B Folland. *Real analysis: modern techniques and their applications*. John Wiley & Sons, 2013.

[18] Geoffrey Grimmett e Dominic Welsh. *Probability: an introduction*. Oxford University Press, 2014.

[19] Allan Gut. *Probability: a graduate course*. Springer Science & Business Media, 2012.

[20] Frank den Hollander. "Probability theory: The coupling method". Em: *Lectures Notes-Mathematical Institute, Leiden University* (2012).

[21] Alan Karr. *Probability*. Springer Science & Business Media, 1992.

[22] Achim Klenke. *Probability theory: a comprehensive course*. Springer Science & Business Media, 2013.

[23] Emmanuel Kowalski. "Bernstein polynomials and Brownian motion". Em: *American Mathematical Monthly* 113.10 (2006), pp. 865–886.

[24] Emmanuel Lesigne. *Heads or tails: An introduction to limit theorems in probability*. Vol. 28. American Mathematical Soc., 2005.

[25] KR Parthasarathy. *CONVERGENCE OF PROBABILITY MEASURES*. 1970.

[26] Marcus Pivato. "Analysis, Measure, and Probability: A visual introduction". Em: (2003).

[27] David Pollard. *A user's guide to measure theoretic probability*. Vol. 8. Cambridge University Press, 2002.

[28] Sidney I Resnick. *A probability path*. Springer Science & Business Media, 2013.

[29] Jeffrey S Rosenthal. *A first look at rigorous probability theory*. World Scientific Publishing Co Inc, 2006.

[30] Halsey Lawrence Royden e Patrick Fitzpatrick. *Real analysis*. Vol. 198. 8. Macmillan New York, 1988.

[31] Luis A Santaló. *Integral geometry and geometric probability*. Cambridge university press, 2004.

[32] Klaus D Schmidt. "The Cantor set in probability theory". Em: (1991).

[33] Albert N Shiryaev. *Probability*. Springer-Verlag, New York, 1996.

[34] Maurice Sion. "HISTORY on MEASURE THEORY IN THE TWENTIETH CENTURY". Em: (1990).

[35] Jan Von Plato. *Creating modern probability: Its mathematics, physics and philosophy in historical perspective*. Cambridge University Press, 1994.