

Probabilidade

Notas de Aula - Versão 0.94

Daniel Miranda
Rafael Grisi
UFABC
18 de abril de 2022

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

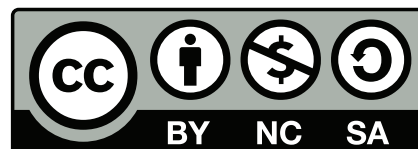
- Sherlock Holmes

Copyright © 2019

Licenciado sob a Creative Commons Attribution 4.0. Você não pode usar esse arquivo exceto em conformidade com a Licença. Você pode obter uma cópia da Licença em

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

A menos que exigido por lei aplicável ou acordado por escrito, o livro distribuído sob a Licença é distribuído “como está”, sem garantias ou condições de qualquer tipo, expressa ou implícita. Consulte a Licença para permissões específicas e limitações sob a Licença.



Sumário

Sumário	i
Observações	1
Agradecimentos	2
Notação	3
Alterações	4
1 Introdução	1
1.1 Nem sempre podemos atribuir probabilidades	1
1.2 A necessidade de álgebras	2
1.3 Álgebras, σ -álgebras e suas Amiguinhas	3
1.4 Classes de Conjuntos	8
1.5 σ -álgebra de Borel	15
2 Medidas de Probabilidade	20
2.1 Continuidade das Medidas de Probabilidade	23
2.2 Teorema de Extensão de Carathéodory	26
2.3 Classes Compactas	31
2.4 Aplicações	32
2.5 $\nexists$ Como são, como vivem e quantos são os Borelianos?	44
2.6 Medidas em Espaços Métricos	46
3 Independência	51
3.1 Independência	53
3.2 Lei 0-1 de Kolmogorov e Borel-Cantelli	56
3.3 Aplicações	59
4 Variáveis Aleatórias	67
4.1 Funções de Variáveis Aleatórias	72
4.2 Variáveis Aleatórias Geram σ -álgebras	74
4.3 Variáveis Aleatórias Induzem Medidas de Probabilidade em $\mathbb{R}$	75
4.4 Independência de Variáveis Aleatórias	75
5 Integral de Lebesgue	78

5.1	Integral de Lebesgue	78
5.2	Propriedades da Integral de Lebesgue	81
5.3	Comparando a Integral de Lebesgue e de Riemann	84
5.4	Integração e Limites	85
5.5	Espaços $\mathcal{L}^p$	88
6	Distribuições	92
6.1	Função de distribuição	92
6.2	Tipos de Distribuições	98
7	Valor Esperado	110
7.1	Momentos	110
7.2	Integração com respeito à Distribuição	115
7.3	Desigualdades de Markov e Chebyshev	118
7.4	Desigualdade Maximal de Kolmogorov	119
7.5	Outras Desigualdades	120
7.6	Funções Geradoras de Probabilidade	122
7.7	Processos de Ramificação	125
7.8	Funções Geradoras de Momento	129
7.9	Entropia	131
8	Medida Produto	137
8.1	Áreas de Seções Transversais	140
8.2	Integrais Iteradas	141
8.3	Distribuição Conjunta	142
8.4	Produtos Infinitos: Teorema de Extensão de Kolmogorov I	143
8.5	Existência de Variáveis Aleatórias Independentes	147
8.6	Convolução e Soma de Variáveis Aleatórias Independentes	150
8.7	Convolução e Aproximação por Funções Suaves	152
8.8	Aplicações	155
8.9	Teorema de Extensão de Kolmogorov II	157
9	Modos de Convergência	160
9.1	Convergência Pontual	160
9.2	Convergência Uniforme	160
9.3	Convergência em Distribuição	162
9.4	Convergência em Probabilidade	164
9.5	Convergência Quase Certa	165
9.6	Convergência em Média	168
9.7	Comparando os Modos de Convergência	170
9.8	Teorema de Representação de Skorokhod	172
9.9	Teorema de Seleção de Helly	173
9.10	Convergência Fraca de Medidas	175
10	Teoremas Limites	178
10.1	Lei Fraca	178

10.2	Lei Forte	181
10.3	Integração de Monte-Carlo	188
10.4	Polinômios de Bernstein e o Teorema de Weierstrass	189
10.5	★ Teorema de De Moivre-Laplace	193
11	Teorema do Limite Central	198
11.1	Funções Características	198
11.2	Teorema do Limite Central	208
12	Medidas com Sinal e Teorema de Radon-Nikodym	212
12.1	Medidas com Sinal	212
12.2	Decomposição de Radon-Nikodym	216
13	Acoplamento	220
13.1	Variação total	220
13.2	Acoplamento	222
13.3	Aplicações	227
14	Esperança Condicional	228
14.1	Existência e Unicidade de Esperança Condicional	230
14.2	Esperança Condicional em Relação a uma Variável Aleatória	233
14.3	Propriedades da Esperança Condicional	233
14.4	Esperança Condicional como Projeção	236
A	Integral de Riemann-Stieltjes	238
B	História da Teoria da Medida no século XX	240
B.1	Medida	240
B.2	Probabilidade	243
	Bibliografia	245

Observações

Essas notas correspondem aos apontamentos para o curso de Probabilidade (MCTB021) ministradas em 2017, 2018, 2019 e 2022. Essas notas foram escritas em tempo real e 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

asad

Agradecemos ao aluno Tiago Leite Dias, Kayo Douglas (KY), Carlo Lemos, Bruno Carvalho e Guilherme Afonso Gigeck pelas 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 $\mathcal{A}$
P	Probabilidade
μ, ν	medidas
$\mathcal{F}$	σ -álgebra
$\mathbf{P}^*, \mu^*$	medida exterior
X, Y, Z	variáveis aleatórias
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
$\hat{\mu}$	transformada de Fourier de μ
$\mathbf{E}[X \mathcal{G}]$	esperança condicional

Alterações

Versão 0.94

- 5 correções. Voltei a mexer no texto depois de um ano ☺

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.

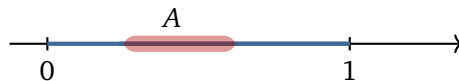
Introdução

1.1 Nem sempre podemos atribuir probabilidades

Imagine que quiséssemos sortear um ponto **uniformemente** no intervalo $[0,1]$ o que seria natural que exigíssemos?



Claramente não podemos atribuir probabilidades somente aos pontos individuais, que por serem infinitos, terão probabilidade 0. Vamos então tentar atribuir probabilidades a conjuntos. Gostaríamos de atribuir uma probabilidade à possibilidade que esse ponto pertença a um subconjunto $A \subset [0,1]$. Ou seja, queremos uma função $\mu : \mathcal{P}([0,1]) \rightarrow [0,1]$.



É razoável pedir que essa função satisfaça:

- 1** para uma sequência de conjuntos disjuntos $A_1, A_2, \dots$ queremos que $\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n)$.
(Essa propriedade é denominada sigma aditividade)
- 2** seja invariante por translação, ou seja, se denotarmos a parte decimal de um número real x por $\text{dec}(x)$, então pediremos que se $\text{dec}(A') = \text{dec}(A + b)$ então $\mu(A') = \mu(A)$.
- 3** não seja trivial, ou equivalentemente que $\mu([0,1]) = 1$.

Uma função não negativa e σ -aditiva é denominada medida, trataremos mais detalhadamente delas no Capítulo 2. Ou seja, queremos uma medida μ não nula, e invariante por translações em $[0,1]$

O problema é que isso é impossível!

1.1 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 número 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 ∞ . $\triangleleft$

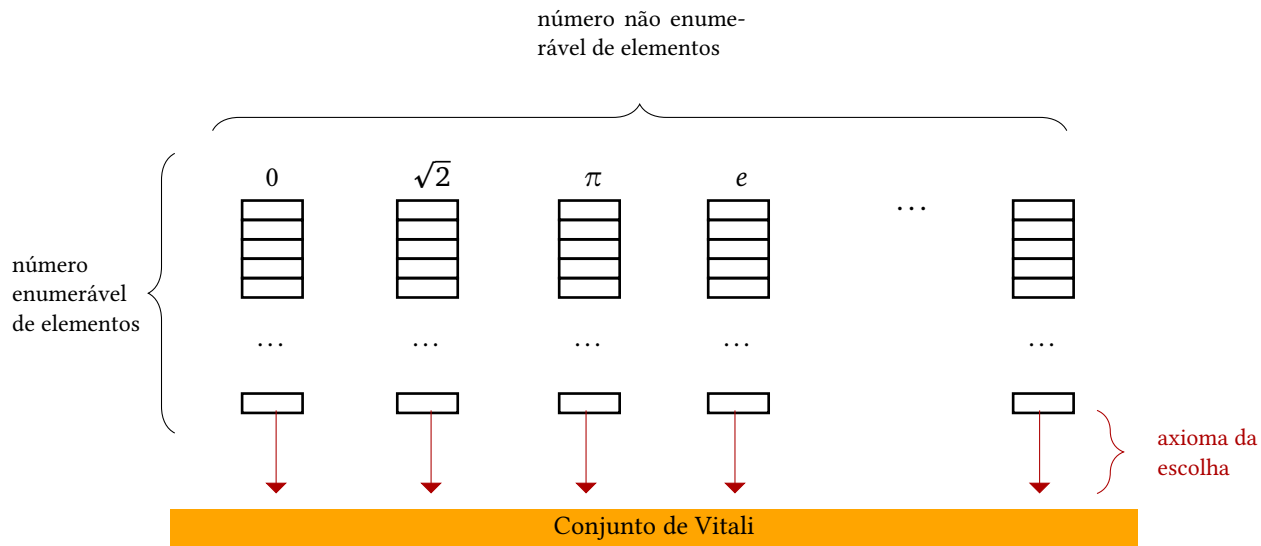
Outro modo de construir o conjunto de Vitali: Observamos que $\mathbb{R}$ é um espaço vetorial sobre $\mathbb{Q}$. E podemos ver $\mathbb{Q}$ como um subespaço unidimensional de $\mathbb{R}$. Considere V' o complemento direto de $\mathbb{Q}$ em $\mathbb{R}$.

Dessa forma podemos escrever $\mathbb{R} = \mathbb{Q} \oplus V'$. Considere $\mathcal{N}' = \text{dec}(V')$. 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;

ou aceitamos que existem subconjuntos não mensuráveis e aos quais não podemos atribuir probabilidade.

A escolha é 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

Para definirmos uma medida 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.

1.2 Definição

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

1 ao complementar

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

2 à união finita

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

Assim, é imediato a partir destas duas condições que as álgebras também estão fechadas sob intersecções finitas:

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

Demonstração. Pela Lei de De Morgan $(A \cap B)^c = A^c \cup B^c$. Logo, $A \cap B = (A^c \cup B^c)^c$. ◁

1.3 Exemplo (Álgebra Trivial)

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

◁

1.4 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!

1.5 Exemplo (Álgebra de Borel)

Considere $\Omega = \mathbb{R}$. Queremos que todos os intervalos possíveis estejam na álgebra. Por causa das propriedades da Definição 1.2, temos que a menor álgebra $\mathcal{F}$ que contém os intervalos será:

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

◁

1.6 Exemplo (Álgebra de partição)

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

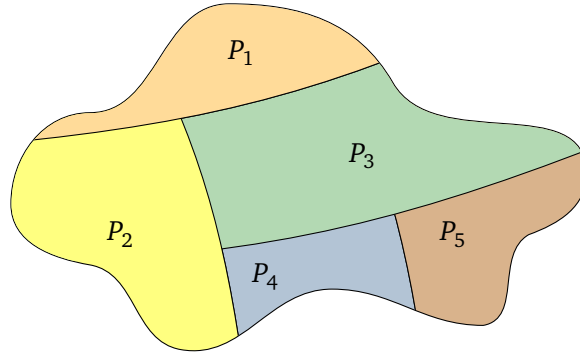
Nesse caso a família gerada por todas as combinações de átomos é uma álgebra

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

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 $Q_1, \dots, Q_m \in \mathcal{Q}$ de modo que

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

Figura 1.1: Uma partição de Ω

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.

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.

1.7 Definição (σ -álgebra)

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

1 ao complementar

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

2 a união enumerável

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

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

Se voltarmos aos nossos exemplos, quais são σ -álgebras? Os Exemplos 1.3, 1.4 e 1.6 são, mas a álgebra de Borel apresentada no Exemplo 1.5 não é.

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

1.8 Exemplo (σ -álgebra de Borel)

Considere $\Omega = \mathbb{R}$. Considere a σ -álgebra que contém todos os intervalos. Na seção 1.4 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

◁

1.9 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.

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

1.10 Definição (Limite Superior e Inferior)

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

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

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

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

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

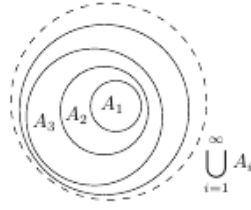


Figure 1. Sequência de intervalos crescentes.

1.11 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 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:

1.12 Definição

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

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

1.13 Exemplo

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

◁

1.14 Definição (Eventos não Crescentes e não Decrescentes)

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

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

1.15 Proposição

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 $A_n \uparrow A$).

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 $A \subset B$.

◁

1.4 Classes de Conjuntos

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

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

1.16 Definição (λ -sistema)

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

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

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

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

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.

1.17 Proposição

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

1.18 Definição (π -sistema)

Uma família não vazia de eventos $\mathcal{P} \subset \mathcal{P}(\Omega)$ é denominada π -sistema se

$$\boxed{1} \quad A, B \in \mathcal{P} \implies A \cap B \in \mathcal{P}.$$

1.19 Definição (Classe Monótona)

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

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

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

As classes monótonas 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

1.20 Teorema ($\sigma = \lambda + \pi = \mathcal{M} + f$)

São equivalentes as seguintes afirmações:

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

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

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

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 ($\implies$) são triviais. Vamos demonstrar as implicações inversas.

$$\boxed{3} \implies \boxed{1}$$

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 $A \in \mathcal{F} \implies A^c \in \mathcal{F}$ é direto

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

$$\begin{aligned} \bigcup_{k=1}^{\infty} A_k &= \bigcup_{k=1}^{\infty} \underbrace{A_k - (A_1 \cup \dots \cup A_{k-1})}_{\text{disjuntos}} \\ &= \bigcup_{k=1}^{\infty} A_k \cap A_1^c \cap \dots \cap A_{k-1}^c \end{aligned}$$

Agora $A_k^c \in \mathcal{F}$ pelas propriedades dos λ -sistemas e logo $A_k \cap A_1^c \cap \dots \cap A_{k-1}^c \in \mathcal{F}$ 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.

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

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

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. $\triangleleft$

Classes Geradas

1.21 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 1.7. Por exemplo,

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

$\triangleleft$

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

1.22 Definição

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

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

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

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

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

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

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

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

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

A σ -álgebra $\sigma\langle\mathcal{C}\rangle$ será assim a menor σ -álgebra que contém $\mathcal{C}$. 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 $\mathcal{C}$ 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.

1.23 Exemplo

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

◁

1.24 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.

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.

1.25 Teorema (Princípio dos Bons Conjuntos)

Dado $\mathcal{C}$ e $\mathcal{G}$ duas famílias 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 .

1.26 Teorema (Geradores Restringidos)

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

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

Demonstração.

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

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

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

Observe que essa inclusão é equivalente à afirmação de que, para cada $A \in \sigma\langle\mathcal{C}\rangle$, $A \cap \Omega_0 \in \sigma\langle\mathcal{C} \cap \Omega_0\rangle$. 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 $\sigma\langle\mathcal{C}\rangle \subset \mathcal{G}$.

□ $\mathcal{C} \subset \mathcal{G}$

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

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

$$\begin{aligned} \Omega_0 \subset \Omega &\implies \Omega \cap \Omega_0 = \Omega_0 \in \sigma(\mathcal{C} \cap \Omega_0) \\ &\text{pois uma } \sigma\text{-álgebra em } \Omega_0 \text{ deve conter } \Omega_0 \\ &\implies \Omega \in \mathcal{G}. \end{aligned}$$

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

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

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

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

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

◁

1.27 Teorema (Teorema da Classe Monótona de Halmos)

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

Demonstração.

Seja $\mathcal{M} = \mathcal{M}(\mathcal{F}_0)$. É suficiente provar que $\sigma(\mathcal{F}_0) \subset \mathcal{M}$. 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, $F_0 \subset G$ e, portanto, $\mathcal{M} \subset G$. Portanto, $\mathcal{M}$ é fechada sob complementação.

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

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

Então 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 tal 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, de $\mathcal{M} \subset G_1$ segue que $A \in \mathcal{M}$ e $B \subset F_0$ juntos implicam que $A \cup B \in \mathcal{M}$. 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 $A \cup B \in \mathcal{M}$. $\triangleleft$

1.28 Teorema (π - λ de Dynkin)

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

Demonstração.

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

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

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

Provaremos agora uma série de afirmativas preliminares.

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

O fato que $\mathcal{P} \subset \mathcal{G}_A$ implica que $\Omega \in \mathcal{G}_A$, e como $(B \setminus C) \cap A = (B \cap A) \setminus (C \cap A)$ e $(\bigcup_n B_n) \cap A = \bigcup_n (B_n \cap A)$ implica que $\mathcal{G}_A$ é fechado para diferenças e uniões de sequências crescentes de subconjuntos.

Desta forma $\mathcal{G}_A$ é um λ -sistema. Como $\mathcal{P}$ é fechado para intersecções finitas e $\mathcal{P} \subset \lambda\langle\mathcal{P}\rangle$ também temos $\mathcal{P} \subset \mathcal{G}_A$. 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.

Análogo.

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.

Para terminar a demonstração Considere $A, B \in \lambda\langle\mathcal{P}\rangle$. Como $A \in \lambda\langle\mathcal{P}\rangle$ e) fornece 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. $\triangleleft$

O seguinte teorema é uma consequência direta dos Teoremas 1.25 e 1.28.

1.29 Teorema (Princípio dos Bons Conjuntos 2)

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

$$\boxed{\text{a}} \quad \mathcal{P} \subset \mathcal{G};$$

$$\boxed{\text{b}} \quad \mathcal{P} \text{ é uma } \pi\text{-sistema};$$

$$\boxed{\text{c}} \quad \mathcal{G} \text{ é uma } \lambda\text{-sistema}$$

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

1.5 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.

1.30 Definição (Espaço Métrico)

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

$$d : X \times X \rightarrow R$$

é dito **espaço métrico** se

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

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

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

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

1.31 Definição (Aberto)

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

1.32 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 $\mathcal{B}((0,1]^d)$.

1.33 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

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

$\mathcal{G}_o :=$ subconjuntos abertos de Ω_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 $\subset$ deixe $B \in \mathcal{B}(\Omega) \cap \Omega_o$ e assim $B = \tilde{B} \cap \Omega_o$ onde $\tilde{B} \in \mathcal{B}(\Omega)$. Como $\Omega_o \subset \Omega$ temos que $B \in \mathcal{B}(\Omega)$ e $B \subset \Omega_o$. $\triangleleft$

1.34 Teorema (Geradores da σ -álgebra de Borel)

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

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

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

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

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

Demonstração. Mostraremos apenas uma igualdade:

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

□ 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.

1.35 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

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

não é uma σ -álgebra sobre Ω .

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

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

é uma σ -álgebra sobre Ω .

Ex. 1.6 — Prove que

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

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

2. Se $A_n \searrow$, 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

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

$$\square \mathcal{G} \text{ é uma } \sigma\text{-álgebra}$$

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 no 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 se $\mathcal{L}$ 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 se 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

$\mathcal{C} :=$ a coleção de conjuntos de $\mathcal{A}$ e seus complementos

$\mathcal{I} :=$ a coleção de intersecções finitas de $\mathcal{C}$

$\mathcal{U} :=$ a coleção de uniões finitas de $\mathcal{I}$.

então $f(\mathcal{A}) = \mathcal{U}$.

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.

Medidas de Probabilidade

Nesse capítulo vamos definir as medidas de probabilidade. Começaremos por um conceito relacionado, porém mais simples: o conceito de Probabilidade Finitamente Aditiva. Esse conceito é insuficiente para desenvolver uma teoria rica de probabilidade, mas em diversas construções que faremos ao longo do texto esse será o ponto de partida.

2.1 Definição (Probabilidade Finitamente Aditiva)

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

- 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 $\mathbf{P} : \mathcal{F}_0 \rightarrow [0,1]$ 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 $A_1, A_2, \dots \in \mathcal{F}_0$ tais que $\bigcup_{k=1}^{\infty} A_k \in \mathcal{F}_0$.

2.3 Definição

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

- 1 $\mu(\emptyset) = 0$

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

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

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

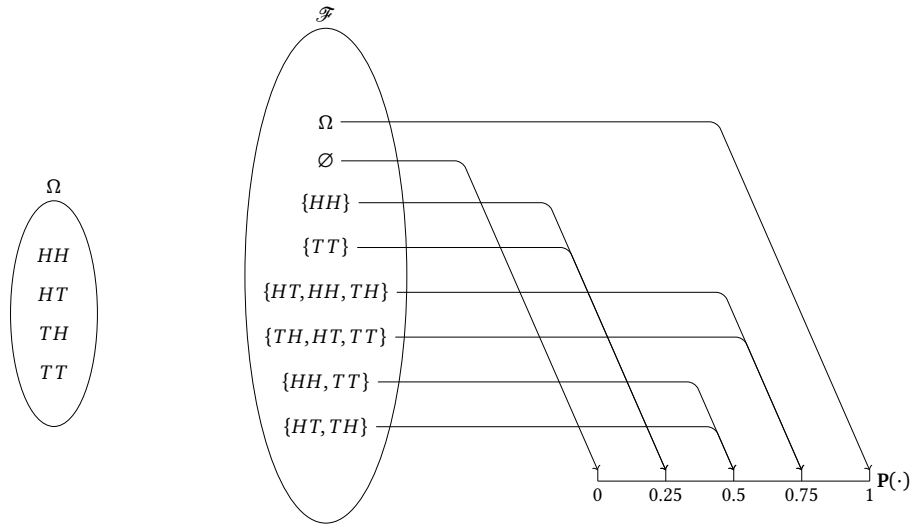


Figura 2.1: Relação entre o espaço amostral, σ -álgebra e a medida de probabilidade.

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

Note que toda medida de probabilidade é σ -finita.

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 $p : \Omega \rightarrow [0,1]$ satisfazendo $\sum_{\omega \in \Omega} p(\omega) = 1$. Definimos então

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

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

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

◁

Provaremos agora algumas propriedades elementares da medida de probabilidades.

2.7 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, $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B \setminus A) \leq \mathbf{P}(A) + \mathbf{P}(B)$. A desigualdade segue da monotonicidade.
- 4 Exercício
- 5 Exercício

◁

2.1 Continuidade das Medidas de Probabilidade

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

2.8 Teorema (Continuidade da Probabilidade)

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

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

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

Demonstração. Provaremos apenas a primeira propriedade.

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

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

◁

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

2.9 Teorema

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

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

Demonstração. Exercício.

Dica do 1. Observe as inclusões

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

◁

2.10 Teorema (Caracterizações da σ -aditividade)

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

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

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

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

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

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

($2. \implies 1.$) Basta provar a σ -aditividade. Suponha que $A_1, A_2, \dots$ 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. $\triangleleft$

2.11 Teorema (Unicidade das Medidas de Probabilidade)

Dado $\mathcal{P}$ uma família de subconjuntos de Ω . Se $\mathbf{P}$ e $\mathbf{Q}$ são duas medidas de probabilidade em $(\Omega, \sigma\langle\mathcal{P}\rangle)$ tais que

- 1** $\mathbf{P}$ e $\mathbf{Q}$ concordam em $\mathcal{P}$;
- 2** $\mathcal{P}$ é um π -sistema,

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

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

Definiremos os bons conjuntos como:

$$\mathcal{G} := \{A \subset \Omega: \mathbf{Q}[A] = \mathbf{P}[A]\}. \quad (9)$$

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

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

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

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

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

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. $\triangleleft$

2.12 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 $A \subset B$.

2.13 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.

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

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

Então

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

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

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.

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.

2.15 Definição (Medida Exterior Induzida)

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

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

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

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

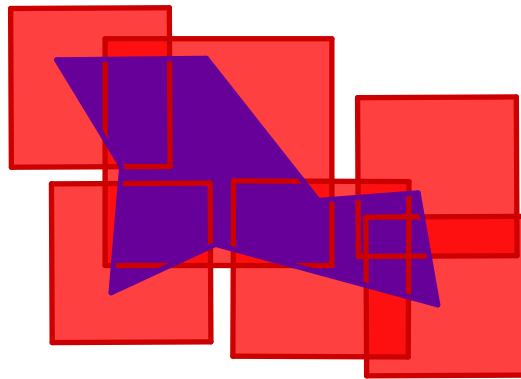
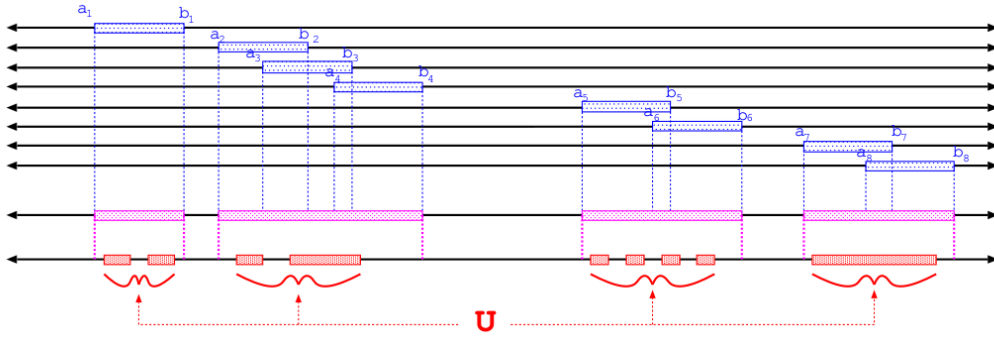


Figura 2.2: Medida externa

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

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

2.16 Definição (Medida Exterior)

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

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

2.17 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

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

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

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

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

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:

2.18 Definição (Conjuntos Mensuráveis para a Medida Exterior)

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

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

é a família de **conjuntos mensuráveis para a medida exterior Q** .

Aplicando a subaditividade de Q , a seguinte definição é equivalente:

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

2.19 Heurística ([1])

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

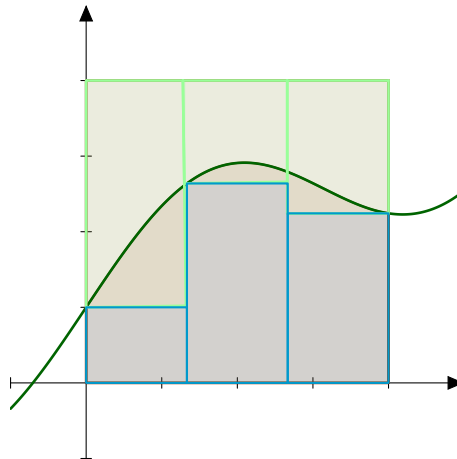
Podemos definir uma "medida interna" Q_* em Ω por

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

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

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 Caratheodory não exige que 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 $\mathcal{M}$:

2.20 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:

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

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

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

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

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

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

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

Para conveniência de notação, seja:

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

Precisamos também do seguinte fato

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

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

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

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

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

Assim

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

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

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

Tomando $N \rightarrow \infty$, obtemos

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

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

2.21 Teorema (Aditividade Enumerável da Medida Exterior)

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

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

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

A desigualdade na outra direção está implícita na subaditividade. $\triangleleft$

Resumimos o que obtivemos até agora

2.22 Corolário (Teorema de Carathéodory)

Se $\mathbf{Q}$ for uma medida exterior (2.16), então a restrição de $\mathbf{Q}$ para os conjuntos mensuráveis $\mathcal{M}$ (2.18) produz uma medida em $\mathcal{M}$.

Definindo uma Medida por Extensão

2.23 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 $\sigma(\mathcal{F}_0) =: \mathcal{F}$.

Demonstração. A unicidade segue diretamente do Teorema 2.11 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 $B_1, B_2, \dots \in \mathcal{F}_0$ cobrindo E tal que $\sum_n \mathbf{P}(B_n) \leq \mathbf{P}^*(E) + \varepsilon$. Usando as propriedades das medidas exteriores induzidas temos

$$\mathbf{P}^*(A \cap E) + \mathbf{P}^*(A^c \cap E) \leq \mathbf{P}^*\left(A \cap \bigcup_n B_n\right) + \mathbf{P}^*\left(A^c \cap \bigcup_n B_n\right) \quad (2.12)$$

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

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

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

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

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 $B_1, B_2, \dots \in \mathcal{F}_0$. Pela subaditividade enumerável e monotonicidade de $\mathbf{P}$,

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

Isto implica $\mathbf{P}(A) \leq \mathbf{P}^*(A)$; A outra desigualdade $\mathbf{P}(A) \geq \mathbf{P}^*(A)$ é automática. $\triangleleft$

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.

2.24 Definição (Classe Compacta)

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

2.25 Exemplo

A família de conjuntos compactos num espaço topológico Hausdorff é uma classe compacta.

Provamos esse fato por contradição. Suponha que nenhuma intersecção finita seja vazia.

Defina o conjunto $B_n = \bigcap_{m=1}^n A_m$. Então cada B_n não é vazio e compacto. Escolha uma sequência $x_1, x_2, \dots$, tal que $x_k \in B_k$.

Observe que esta sequência tem uma subsequência convergente porque está dentro do conjunto compacto A_1 . Podemos mostrar que o limite dessa sequência está contido em cada A_n , porque a cauda da sequência está contida em A_n e A_n é fechado. Isso é uma contradição porque provamos que $\bigcap_{n=1}^{\infty} A_n$ contém um ponto.

◁

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. ◻

2.26 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 $\mu(A_n) \rightarrow 0$. Para isso considere $\varepsilon > 0$.

Para cada n escolha $C_n \in \mathcal{K}$ com $C_n \subset A_n$ e $\mu(A_n) \leq \mu(C_n) + \frac{\varepsilon}{2^n}$. Temos então que

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

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

Como os A_n são encaixantes a equação 2.17 se reduz a

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

Consequentemente para $n > m$ temos

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

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

◁

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.

2.27 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 $(a,b] \subset (0,1]$.

2.28 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 $\mathcal{B}_0((0,1])$.

2.29 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 $0 \rightarrow -1$ obtendo a sequência de funções de Rademacher z_k .

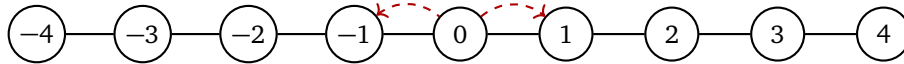
2.30 Definição

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

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

2.31 Observação

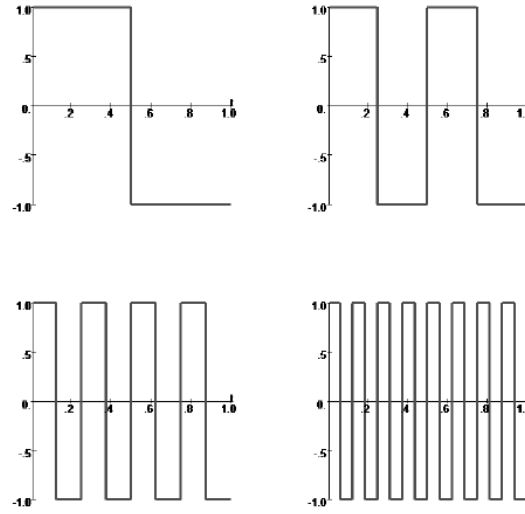
A sequência $s_n(\omega)$ definida acima pode ser vista como um passeio aleatório simétrico nos pontos de coordenadas inteiras da reta, isto é, em $\mathbb{Z}$, que começa em 0 e a cada passo move para a direita ou esquerda, +1 ou -1, com probabilidade igual. Este passeio pode ser ilustrado da seguinte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



Provaremos que

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

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.

Figura 2.4: Funções de Rademacher z_k .

2.32 Teorema (Lei Fraca dos Grandes Números - Teorema de Bernoulli)

Para todo $\varepsilon > 0$,

$$\lim_{n \rightarrow \infty} \mathbf{P}\left[\left\{\omega \in (0,1] : |s_n(\omega)/n| \geq \varepsilon\right\}\right] = 0. \quad (15)$$

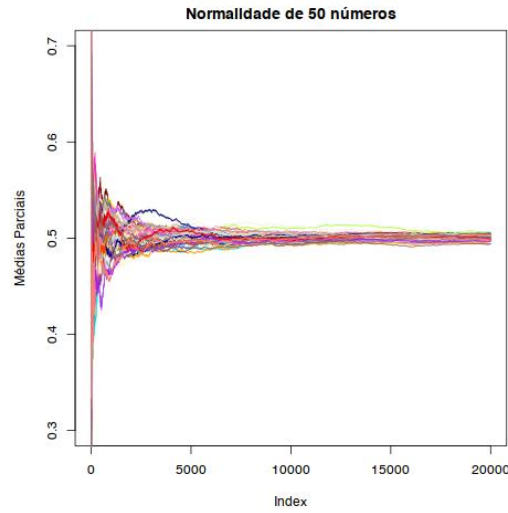
Demonstração. Qualquer evento baseado nos valores $z_1(\omega), \dots, z_n(\omega)$ pode ser descrito como uma união disjunta de intervalos diádicos da forma $(\frac{k-1}{2^n}, \frac{k}{2^n}]$. Logo $\{\omega \in (0,1] : |s_n(\omega)/n| \geq \varepsilon\} \in \mathcal{B}_0((0,1])$.

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

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

**Figure 2.**

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

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

◁

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

2.33 Definição

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

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

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

Provaremos que os números anormais formam um conjunto negligenciável.

2.34 Definição (Conjunto Negligenciável)

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

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

¹na base 2

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

2.35 Teorema (Teorema dos Números Normais de Borel)

O conjunto dos números anormais, A , é negligenciável.

Demonstração.

1 Começaremos notando as seguintes inclusões:

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

$$\begin{aligned} \{\omega : |\frac{s_{k^2}(\omega)}{k^2}| < \varepsilon_k \text{ para } k \text{ suficientemente grande}\} \\ \subset \{\omega : \lim_k \frac{s_{k^2}(\omega)}{k^2} = 0\} \\ \subset \underbrace{\{\omega : \lim_n \frac{s_n(\omega)}{n} = 0\}}_{=N} \end{aligned} \quad (16)$$

2 Agora por (1.) temos que

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

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

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

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

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.

2.36 Teorema

A aplicação $\mathbf{P} : \mathcal{B}_0((0,1]) \rightarrow [0,1]$ definida pela Definição 2.28 é uma pré-medida de probabilidade.

Demonstração. Usaremos o Teorema 2.23 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

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

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

$$\mathbf{P}\left[\bigcap_{k=1}^{N_\varepsilon} A_k\right] \leq \varepsilon. \quad (17)$$

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

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

Isso estabeleceria 2.4 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

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

◁

2.37 Teorema (Medida de Lebesgue)

Dado $\mathbf{P} : \mathcal{B}_0((0,1]) \rightarrow [0,1]$ como na Definição 2.28. Então

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

Demonstração.

□ Demonstração do item 2.37.1

O Teorema de Extensão de Carathéodory com o Teorema 2.36 mostra que existe uma única extensão $\mathbf{P} : \mathcal{B}_0((0,1]) \rightarrow [0,1]$ para $\mathbf{P} : \mathcal{B}((0,1]) \rightarrow [0,1]$ como $\mathcal{B}((0,1]) = \sigma\langle \mathcal{B}_0((0,1]) \rangle$.

□ Demonstração do item 2.37.2

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

◁

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

2.38 Corolário

Dado $\mathbf{P} : \mathcal{B}_0((0,1]) \rightarrow [0,1]$ como na Definição 2.28. 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 2.38.1

Observamos inicialmente que N e N^c estão ambas em $\mathcal{B}((0,1])$. Agora pelo Teorema 2.35 temos que N^c é negligenciável. Ou seja dado $\varepsilon > 0$, existe $B_n \in \mathcal{B}_0((0,1])$ tal que $N^c \subset \bigcup_{n=1}^{\infty} B_n$ onde $\sum_{n=1}^{\infty} \mathbf{P}[B_n] \leq \varepsilon$. Como $\bigcup_{n=1}^{\infty} B_n \in \mathcal{B}((0,1])$ temos

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

Logo

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

para todo ε . Logo $\mathbf{P}(N^c) = 0$ 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 2.48 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 $\mathcal{B}((0,1)) \subsetneq \mathcal{P}(\Omega)$.

◁

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

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

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

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

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.

2.39 Exemplo

Seja $A_n = \{\omega \mid \omega_n = 1\}$, 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 $\{0, 1\}^{\infty}$ quanto possível, para minimizar a possibilidade de incluir conjuntos patológicos aos quais as probabilidades não podem ser atribuídas. Isso nos leva a considerarmos $\mathcal{F}$ 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 $\mathbf{P}_0(A) = \frac{|B|}{2^n}$.

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

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

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

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

Esses conjuntos particionam o espaço,

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

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

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

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

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

- Dado $A \in \mathcal{B}_0^{(\mathbf{i}, \mathbf{i} + 1]}$ definimos $\mathbf{P}_i(A)$ 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 $\mathbf{P}_{bsi}$ é está bem definido).
- Depois, mostra-se que $\mathbf{P}_i$ é uma medida de probabilidade em $((\mathbf{i}, \mathbf{i} + 1], \mathcal{B}_0^{(\mathbf{i}, \mathbf{i} + 1]})$. A parte mais difícil deste passo é demonstrar a σ aditividade. No $(0, 1]$ nos usamos o fato equivalente que $\mathbf{P}_i$ é contínua por cima no $\emptyset$. O argumento também funciona em $((\mathbf{i}, \mathbf{i} + 1], \mathcal{B}_0^{(\mathbf{i}, \mathbf{i} + 1]}, \mathbf{P}_i)$.
- Finalmente utilizamos o Teorema de extensão de Caratheodory temos $((\mathbf{i}, \mathbf{i} + 1], \mathcal{B}((\mathbf{i}, \mathbf{i} + 1]), \mathbf{P}_i)$ (a unicidade segue do fato que os retângulos formam um π -sistema).
- A partir da medida de probabilidades $((\mathbf{i}, \mathbf{i} + 1], \mathcal{B}((\mathbf{i}, \mathbf{i} + 1]), \mathbf{P}_i)$ podemos definir a medida de Lebesgue $\mu_{\mathcal{L}}^d$ nos conjuntos $A \in \mathcal{B}((\mathbf{i}, \mathbf{i} + 1])$ como:

$$\mu_{\mathcal{L}}^d(A) := \sum_{\mathbf{i} \in \mathbb{Z}^d} \mathbf{P}_i((\mathbf{i}, \mathbf{i} + 1] \cap A). \quad (19)$$

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

2.40 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):

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

□ (ii) $\mu_{\mathcal{L}}^d(\emptyset) = 0$: Imediato pois $\mathbf{P}_i((i, i+1] \cap \emptyset) = 0$.

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

$$\begin{aligned} \mu_{\mathcal{L}}^d\left(\bigcup_{k=1}^{\infty} A_k\right) &= \sum_{i \in \mathbb{Z}^d} \mathbf{P}_i\left((i, i+1] \cap \bigcup_{k=1}^{\infty} A_k\right) \\ &= \sum_{i \in \mathbb{Z}^d} \mathbf{P}_i\left(\bigcup_{k=1}^{\infty} (i, i+1] \cap A_k\right) \\ &= \sum_{i \in \mathbb{Z}^d} \sum_{k=1}^{\infty} \mathbf{P}_i\left((i, i+1] \cap A_k\right) \\ &= \sum_{k=1}^{\infty} \sum_{i \in \mathbb{Z}^d} \mathbf{P}_i\left((i, i+1] \cap A_k\right) \\ &= \sum_{k=1}^{\infty} \mu_{\mathcal{L}}^d(A_k) \end{aligned} \tag{20}$$

◁

2.41 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:

$$\mu_{\mathcal{L}}^d((a_1, b_1] \times \cdots \times (a_d, b_d]) = \prod_{k=1}^d (b_k - a_k) \tag{21}$$

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 $\emptyset$. Então $\mathcal{B}(\mathbb{R}^d) = \sigma(\mathcal{P})$. Também observamos que $\mu_{\mathcal{L}}^d$ é σ -finita em $\mathcal{P}$ pois $\mu_{\mathcal{L}}^d((i, i+1]) = 1$, $\mathbb{R}^d = \bigcup_{i \in \mathbb{Z}^d} (i, i+1]$ e cada $(i, i+1] \in \mathcal{P}$. Logo o resultado decorre do Teorema 2.11. ◁

2.42 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

$$\mu_{\mathcal{L}}^d(A + x) = \mu_{\mathcal{L}}^d(A) \tag{22}$$

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,

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

Para concluir a demonstração do teorema 2.42 usaremos a mesma ideia de 2.4127 sobre a unicidade de $\mu_{\mathcal{L}}^d$.

Para isso dado x defina $\mu_x(A) := \mu_{\mathcal{L}}^d(A+x)$. Então μ_x é uma medida em $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. As medidas μ_x e $\mu_{\mathcal{L}}^d$ concordam no π -sistema dos retângulos finitos em $\mathbb{R}^d$. Logo pelo Teorema 2.41, $\mu_{\mathcal{L}}^d(A) = \mu_x(A) := \mu_{\mathcal{L}}^d(A+x)$ para todo $A \in \mathcal{B}(\mathbb{R}^d)$. $\triangleleft$

2.43 Teorema

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

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

Demonstração. Exercício $\triangleleft$

2.44 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 $A_1, A_2, \dots \in \mathcal{F}$ 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 $\mathcal{J}_c \subset \mathcal{I}$ tal que $\mu(A_k \cap B_i) > \varepsilon$ para todo $i \in \mathcal{J}_c$ e que

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

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

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

□ $\mathcal{I} = \bigcup_k \{i \in \mathcal{I} : \mu(A_k \cap B_i) > 0\}$: 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. $\triangleleft$

Se $k < d$ então $\mu_{\mathcal{L}}^d(A) = 0$ 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 $\mu_{\mathcal{L}}^d$ é invariante por translação $\mu_{\mathcal{L}}^d(A) = \mu_{\mathcal{L}}^d(A + xt)$ para todo $t \in \mathbb{R}$. Pelo Teorema 2.44, $\mu_{\mathcal{L}}^d(A) = 0$. $\triangleleft$

2.45 Definição (Conjuntos mensuráveis de Borel versus Lebesgue)

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

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

Começaremos com uma proposição simples

2.46 Proposição

- $\square$ Todo subconjunto aberto de $\mathbb{R}$ é uma união enumerável de intervalos abertos e logo é um boreliano
- $\square$ 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 $C_0 \supset C_1 \supset C_2 \supset \dots$ 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.

2.47 Proposição

O conjunto de Cantor C tem as seguintes propriedades:

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



Figura 2.5: O conjunto de Cantor

$$\boxed{3} \quad \mu(C) = 0.$$

2.48 Proposição (Cardinalidade dos conj. 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, $\mathcal{P}(C) \subset \mathcal{L}$. Mas como $|C| = |\mathbb{R}|$, segue que $\mathcal{P}(C) = \mathcal{P}(\mathbb{R})$ e, portanto, $|\mathcal{P}(\mathbb{R})| \leq |\mathcal{L}|$. $\triangleleft$

2.49 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 .

2.50 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.

2.51 Teorema (Propriedades de Σ_n e Π_n)

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

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

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

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 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 dito 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 de Σ_ω . De fato, nós temos uma sequência de conjuntos

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

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.

2.52 Teorema (Cardinalidade da Álgebra de Borel)

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

2.53 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.

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

Ex. 2.2 — [Integral de Lebesgue-Stieltjes] Dado $\Omega = \mathbb{R}$ e $\mathcal{B}_0(\mathbb{R}) := \{(-\infty, a] : -\infty < a < \infty\}$ 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 $\lim_{x \rightarrow -\infty} F(x) = 0$ e $\lim_{x \rightarrow \infty} F(x) = 1$.

Ex. 2.3 — (sub-aditividade enumerável) Mostre que

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

Ex. 2.4 — Prove que a continuidade da medida de probabilidade. Dados $A_1, A_2, \dots$ eventos, então

1. Se $A_n \nearrow A$, então $\mathbf{P}(A_n) \nearrow \mathbf{P}(A)$ quando $n \rightarrow \infty$.
2. Se $A_n \searrow A$, então $\mathbf{P}(A_n) \searrow \mathbf{P}(A)$ quando $n \rightarrow \infty$.
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. $A_n \rightarrow A$ implica que $\mathbf{P}(A_n) \rightarrow \mathbf{P}(A)$ as $n \rightarrow \infty$.

Ex. 2.5 — Considere um modelo probabilístico cujo espaço amostral é a reta real. Mostre que

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

Ex. 2.6 — (Uma condição insuficiente) Dados $\Omega = \{1, 2, 3, 4\}$, e $A = \{\{1, 2\}, \{1, 3\}\}$.

1. Mostre que a σ -álgebra $\mathcal{F}$ gerada por A é o conjunto das partes 2^{Ω} .
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. 2.7 — **Desigualdade de Bonferroni**

Definindo

$$S_1 := \sum_{i=1}^n \mathbb{P}(A_i),$$

e

$$S_2 := \sum_{1 \leq i < j \leq n} \mathbb{P}(A_i \cap A_j),$$

bem como

$$S_k := \sum_{1 \leq i_1 < \dots < i_k \leq n} \mathbb{P}(A_{i_1} \cap \dots \cap A_{i_k})$$

para todos os números inteiros de k em $\{3, \dots, n\}$.

Então, para k ímpar

$$\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{j=1}^k (-1)^{j-1} S_j$$

e para $k > 2$ par

$$\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) \geq \sum_{j=1}^k (-1)^{j-1} S_j.$$

(Se tiver com muita preguiça prove apenas para $k = 2, 3$.)

Ex. 2.8 — 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 a 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 $n \rightarrow \infty$

Ex. 2.9 — 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. 2.10 — 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. 2.11 — O conjunto de números normais e anormais está em $\mathcal{B}((0,1])$.

Ex. 2.12 — 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. 2.13 — Para todo número $\omega \in (0,1]$, denotamos por $d_k(\omega)$ o k -ésimo dígito da representação de ω . Seja $z_k(\omega) := 2d_k(\omega) - 1$ e

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

Mostre que

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

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. 2.14 — 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. 2.15 — 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

$$\mu_1(A) \leq \mu_2(A) \quad (2.19)$$

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. 2.16 — Princípio 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 $A_1, A_2, \dots$ tais que $B \subset \bigcup_{n=1}^{\infty} A_n$ e $\mu(\bigcup_{n=1}^{\infty} A_n - B) \leq \epsilon$.
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. 2.17 — 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

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

Mostre que

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

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

2. Se I é enumerável, então

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

para toda enumeração $i_1, i_2, \dots$ 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. 2.18 — Mostre que

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

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 $x \in \mathbb{R}$.

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 nos é dada a informação que num experimento o evento B ocorre, com $\mathbf{P}(B) > 0$. Nesse caso, o espaço amostral a ser considerado é reduzido de Ω para B e a probabilidade de que o evento A ocorrerá dado que B ocorreu é

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

Para explicar esta fórmula, observe que dado a informação que o evento B ocorre apenas a parte do evento A que se encontra em B pode eventualmente ocorrer e uma vez que o espaço amostral foi reduzido à B , temos que dividir por $\mathbf{P}(B)$ para termos que $\mathbf{P}(B|B) = 1$.

3.1 Definição (Probabilidade Condicional)

Dados A e B eventos com $\mathbf{P}(B) > 0$. A **probabilidade condicional** de que A ocorre dado que B ocorre, indicado por $\mathbf{P}(A|B)$, é definida como

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

3.2 Teorema

Dado $(\Omega, \mathcal{F}, \mathbf{P})$ um espaço de probabilidade e dado B com $\mathbf{P}(B) > 0$, então $(B, \mathcal{F} \cap B, \mathbf{P}(\cdot|B))$ 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 0,75 horas é o complemento do evento $T_{0,75}^c$, então a probabilidade desejada é

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

3.4 Teorema (Lei da Probabilidade Total)

Dado $(\Omega, \mathcal{F}, \mathbf{P})$ um espaço de probabilidade e deixe $\{B_1, B_2, \dots\}$ ser 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.

$$\begin{aligned} \mathbf{P}(A) &= \mathbf{P}\left(A \cap \left(\bigcup_{i=1}^{\infty} B_i\right)\right) \\ &= \mathbf{P}\left(\bigcup_{i=1}^{\infty} (A \cap B_i)\right) \\ &= \sum_{i=1}^{\infty} \mathbf{P}(A \cap B_i) \\ &= \sum_{i=1}^{\infty} \mathbf{P}(A | B_i) \mathbf{P}(B_i) \end{aligned}$$

◁

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

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

Demonstração. Da definição de probabilidade condicional temos:

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

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

Manipulando temos:

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

Dividindo por $P(A)$ temos:

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

◁

3.6 Teorema

Sejam $A_1, A_2, \dots, A_n$ eventos que formam uma partição do espaço amostral, e assumamos que $P(A_i) > 0$ para todo i . Então, para qualquer evento B tal que $P(B) > 0$, temos que

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

3.1 Independência

Um dos conceitos mais significativos na teoria da probabilidade é o conceito de independência. Intuitivamente, dois eventos A e B são independentes se a ocorrência do evento A não modifica a probabilidade de ocorrência do evento B . Isso nos leva a seguinte definição provisória.

3.7 Definição (Eventos Independentes - Definição Provisória)

Os eventos A e B são ditos **independentes** se $P(A | B) = P(A)$.

E de modo equivalente temos a seguinte definição:

3.8 Definição (Eventos Independentes)

Os eventos A e B são ditos **independentes** se

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

As duas definições são equivalentes no caso em que $P(B) > 0$. A segunda definição é preferível pois as probabilidades condicionais podem não estar definidas se $P(A)$ ou $P(B)$ é 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, $P(A) = 0$ ou $P(A) = 1$.

Demonstração.

$$P(A) = P(A \cap A) = P(A)^2$$

Resolvendo a equação do segundo grau em $P(A)$ temos que $P(A) = 0$ ou $P(A) = 1$

◁

3.10 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:

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

3.11 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

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

◁

3.12 Proposição

Se $\{E_i \mid i \in I\}$ é uma família de eventos independentes, então a família de eventos $\{E_i^c \mid i \in I\}$ também é independente.

Demonstração. Exercício. Dica utilize o Princípio de Inclusão-Exclusão. ◁

3.13 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. Seja $\Omega = \mathbb{N}$, $\mathbf{P}(n) = \frac{\zeta(s)^{-1}}{n^s}$ e deixe $\mathbf{P}$ ser 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 $k_1, \dots, k_r$,

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

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:

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

◁

3.14 Definição (Famílias Independentes)

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

3.15 Teorema

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

- 1 Subclasses** Se $\mathcal{A}_k \subset \mathcal{B}_k \subset \mathcal{F}$ para todo $k \in \mathcal{K}$ e $\{\mathcal{B}_k\}_{k \in \mathcal{K}}$ são famílias independentes então $\{\mathcal{A}_k\}_{k \in \mathcal{K}}$ são famílias independentes.
- 2 Acréscimo** $\{\mathcal{A}_k\}_{k \in \mathcal{K}}$ são famílias independentes se e somente se $\{\mathcal{A}_k \cup \{\Omega\}\}_{k \in \mathcal{K}}$ são famílias independentes.
- 3 Produto Simplificado** se $\mathcal{A}_1, \dots, \mathcal{A}_n$ são famílias de conjuntos em $\mathcal{F}$ e $\Omega \in \mathcal{A}_k$ para cada k , então $\mathcal{A}_1, \dots, \mathcal{A}_n$ são famílias independentes se e somente se

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

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

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

3.16 Definição (σ -álgebras Independentes)

Dado $(\Omega, \mathcal{F}, P)$ um espaço de probabilidade e deixe $\mathcal{A}$ e $\mathcal{B}$ serem duas sub- σ -álgebras de $\mathcal{F}$. Então $\mathcal{A}$ e $\mathcal{B}$ são ditas independentes se, dados $A \in \mathcal{A}$ e $B \in \mathcal{B}$,

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

3.17 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 $P(A) = \frac{1}{2}$ e $P(B) = \frac{1}{2}$. Para calcular a probabilidade da ocorrência de ambos eventos, listamos os casos e observamos que $P(A \cap B) = \frac{1}{4}$. Assim, A e B são independentes.

◁

3.18 Exemplo (Amostragem)

Suponha que uma urna contenha 5 bolas brancas e 5 bolas pretas. Agora escolhamos 2 bolas sem substituição.

Considere os eventos:

A = primeira bola é branca

B = 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, $P(B|A) = \frac{4}{4+5}$, e a probabilidade de B em si (quando não sabe o que acontece com a primeira bola) é $P(B) = \frac{1}{2}$ (porque $P(B) = P(B^c)$, 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.

◁

3.19 Teorema (Famílias Independentes)

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

3.20 Proposição

Considere uma família de eventos independentes $\{A_k\}_{k \in \mathcal{K}}$. Se $I \cap J = \emptyset$, então $\sigma(A_k : k \in I)$ e $\sigma(A_k : k \in J)$ são independentes.

3.2 Lei 0-1 de Kolmogorov e Borel-Cantelli

Dado uma sequência de eventos arbitrários $A_1, A_2, \dots$. Queremos definir o que significa um evento A depender somente dos eventos $A_n, A_{n+1}, \dots$, para n suficientemente grande, ou dito de outra forma um evento A não depender dos eventos iniciais $A_1, \dots, A_n$ para todo n .

3.21 Definição

A σ -álgebras caudal é

$$\mathcal{T} := \bigcap_{n=1}^{\infty} \sigma(A_n, A_{n+1}, A_{n+2}, \dots).$$

Os elementos de $\mathcal{T}$ são denominados **eventos caudais**.

Se pensarmos no índice n como o tempo, então $\sigma(A_n, A_{n+1}, A_{n+2}, \dots)$ 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.

3.22 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".

◁

3.23 Teorema (Lei 0-1 de Kolmogorov)

Dados $A_1, A_2, \dots$ eventos independentes. Então todo evento caudal A tem probabilidade 0 ou 1.

Demonstração. Considere $A_1, A_2, A_3, \dots, A_{n-1}, A_n, A_{n+1}, \dots$ eventos. Então, pelo Corolário 3.20, $\sigma\langle A_1, \dots, A_{n-1} \rangle$ e $\sigma\langle A_n, \dots, A_{n+1}, \dots \rangle$ são independentes. Note que $A \in \sigma\langle A_n, \dots, A_{n+1}, \dots \rangle$. Logo, os eventos $A_1, A_2, \dots, A_{n-1}, A$ são independentes para todo n .

Consequentemente os eventos $A_1, A_2, \dots, A$ são independentes. Então, pela Proposição 3.20, $\sigma\langle A_1, A_2, \dots \rangle$ e $\sigma\langle A \rangle$ são independentes. Mas $A \in \sigma\langle A_1, A_2, \dots \rangle$ 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.

◁

3.24 Definição (Infinitas vezes e quase sempre)

Dados $A_1, A_2, \dots$ subconjuntos de Ω . Então dizemos que os eventos ocorrem **infinitas vezes** se

$$\{A_n \text{ i.v.}\} := \bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} A_n =: \limsup_{n \rightarrow \infty} A_n$$

Dizemos que os eventos ocorrem **quase sempre** se

$$\{A_n \text{ q.s.}\} := \bigcup_{m=1}^{\infty} \bigcap_{n=m}^{\infty} A_n =: \liminf_{n \rightarrow \infty} A_n$$

3.25 Teorema (Primeiro Lema de Borel-Cantelli)

Dados $E_1, E_2, \dots$ subconjuntos em $\mathcal{F}$. Então

$$\sum_{n=1}^{\infty} \mathbf{P}(E_n) < \infty \implies \mathbf{P}(E_n \text{ i.v.}) = 0.$$

Demonstração. Suponha que:

$$\sum_{n=1}^{\infty} \mathbf{P}(E_n) < \infty.$$

Como a série converge devemos ter que

$$\sum_{n=N}^{\infty} \mathbf{P}(E_n) \rightarrow 0, \quad \text{quando } N \rightarrow \infty.$$

Logo

$$\inf_{N \geq 1} \sum_{n=N}^{\infty} \mathbf{P}(E_n) = 0.$$

Assim

$$\begin{aligned}
 \mathbf{P}\left(\limsup_{n \rightarrow \infty} E_n\right) &= \mathbf{P}(E_n \text{ i.v.}) = \mathbf{P}\left(\bigcap_{N=1}^{\infty} \bigcup_{n=N}^{\infty} E_n\right) \\
 &\leq \inf_{N \geq 1} \mathbf{P}\left(\bigcup_{n=N}^{\infty} E_n\right) \\
 &\leq \inf_{N \geq 1} \sum_{n=N}^{\infty} \mathbf{P}(E_n) \\
 &= 0
 \end{aligned}$$

◁

3.26 Teorema (Lema de Fatou para Conjuntos)

Dados $E_1, E_2, \dots$ subconjuntos de $\mathcal{F}$. Então

$$\begin{aligned}
 \mathbf{P}(E_n \text{ q.s.}) &\leq \liminf_n \mathbf{P}(E_n) \\
 &\leq \limsup_n \mathbf{P}(E_n) \leq \mathbf{P}(E_n \text{ i.v.}).
 \end{aligned}$$

Demonstração. Exercício

◁

O segundo Lema de Borel fornece uma recíproca parcial do primeiro lema de Borel-Cantelli.

3.27 Teorema (Segundo Lema de Borel-Cantelli)

Dados $E_1, E_2, \dots$ conjuntos independentes em $\mathcal{F}$. Então

$$\sum_{n=1}^{\infty} \mathbf{P}(E_n) = \infty \implies \mathbf{P}(E_n \text{ i.v.}) = 1.$$

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

$$\begin{aligned}
 &= \mathbf{P}(\{E_n \text{ i.v.}\}^c) = \mathbf{P}\left(\liminf_{n \rightarrow \infty} E_n^c\right) = \mathbf{P}\left(\left(\bigcap_{N=1}^{\infty} \bigcup_{n=N}^{\infty} E_n\right)^c\right) = \\
 &= \mathbf{P}\left(\bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} E_n^c\right) = \lim_{N \rightarrow \infty} \mathbf{P}\left(\bigcap_{n=N}^{\infty} E_n^c\right)
 \end{aligned}$$

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:

$$\begin{aligned}
 \mathbf{P}\left(\bigcap_{n=N}^{\infty} E_n^c\right) &= \prod_{n=N}^{\infty} \mathbf{P}(E_n^c) \\
 &= \prod_{n=N}^{\infty} (1 - \mathbf{P}(E_n)) \\
 &\leq \prod_{n=N}^{\infty} \exp(-\mathbf{P}(E_n)) \\
 &= \exp\left(-\sum_{n=N}^{\infty} \mathbf{P}(E_n)\right) \\
 &= 0.
 \end{aligned}$$

◁

Resumindo, dados $A_1, A_2, \dots$ eventos independentes. Então

$$\mathbf{P}(A_n \text{ i.v.}) = \begin{cases} 0 & \text{se } \sum_n \mathbf{P}(A_n) < \infty \\ 1 & \text{se } \sum_n \mathbf{P}(A_n) = \infty. \end{cases}$$

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 caras seguidas.

Seja $\ell(n)$ = ser o número de caras consecutivas após o n -ésimo lançamento. (considere que $\ell(n) = 0$ se sair coroa no n -ésimo lançamento.)

Ou seja, dado um inteiro fixo k é fácil ver que

$$\mathbf{P}(\ell(n) \geq k \text{ i.v.}) = 1.$$

E se trocássemos o inteiro fixo por uma função $k(n)$ quão rápido essa função pode crescer de modo que a probabilidade permaneça 1. Ou seja para quais funções $k(n)$ temos que

$$\mathbf{P}(\ell(n) \geq k(n) \text{ i.v.}) = 1.$$

Essa função pode crescer no máximo como $\log_2 n$ como provamos no próximo teorema:

3.28 Teorema

1 Para todo $\epsilon > 0$, $\mathbf{P}(\ell(n) \geq (1 + \epsilon) \log_2 n \text{ i.v.}) = 0$.

2 $\mathbf{P}(\ell(n) \geq \log_2 n \text{ i.v.}) = 1$.

Demonstração.

1 $\mathbf{P}(\ell(n) \geq r) = \frac{1}{2^r}$. Agora para aplicar Borel-Cantelli basta substituir $(1 + \epsilon) \log_2 n$.

$$\mathbf{P}(\ell(n) \geq (1 + \epsilon) \log_2 n) = \frac{1}{n^{1+\epsilon}}.$$

Esta série converge e logo podemos utilizar o Primeiro Lema de Borel Cantelli. Isso termina a demonstração.

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:

$$= \sum_{k=1}^{\infty} \sum_{n_k \leq n < n_{k+1}} \frac{1}{2^{r(n_k)} r(n_k)},$$

A última soma é maior ou igual à

$$\sum_{n=1}^{\infty} \frac{1}{2^{r(n)} r(n)}$$

e como $r(n_k) \leq r(n)$, 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 $\triangleleft$

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 3.28. Vamos considerar que estamos sorteando caracteres uniformemente no alfabeto $\{a, b, c, \dots, x, y, z\}$. 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"

3.29 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?

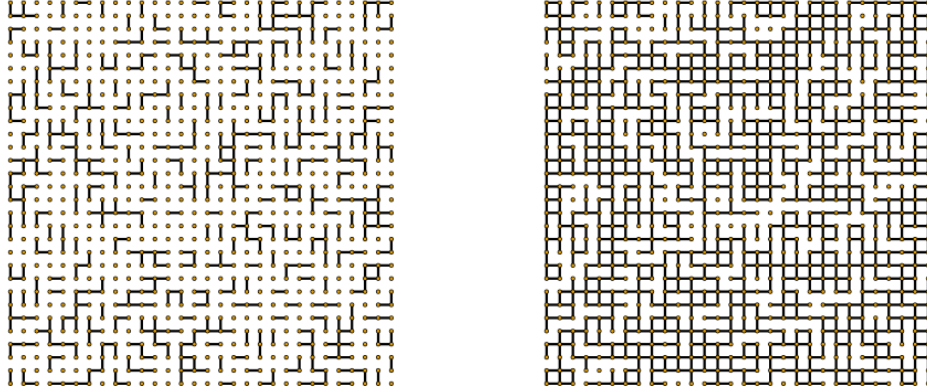
Em matemática, podemos pensar um material poroso como um grafo, digamos por exemplo que os vértices são $\mathbb{Z}^2$. Conectamos vértices vizinhos desta rede de forma independente com alguma probabilidade p . O resultado é um grafo aleatório. De modo alternativo, podemos fazer uma construção análoga com vértices em vez de arestas, e considere o subgrafo como aquele determinado pelos vértices escolhidos.

Alguma terminologia:

- ☐ As arestas são os elos
- ☐ Os vértices são sítios
- ☐ As aresta/vértices escolhidos estão 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**.



Como dissemos, a motivação física para este modelo é um líquido que percola através de uma rede sólida, onde as ligações/sítios abertos representam os vínculos/loais que o líquido pode passar. A principal questão sugerida por esta interpretação é se o líquido será 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 desse modelo. Faremos a construção para um grafo conexo $G = (V, E)$ que possui um número enumerável de arestas. Para isso definiremos o espaço amostral, a σ -álgebra e a probabilidade. Definimos

- $\Omega = \{0,1\}^{|E|}$ como o conjunto de todos os arranjos de arestas abertas, onde 0 representa uma aresta fechada e 1 representa uma aresta aberta;
- $\mathcal{F}$ como a σ -álgebras em Ω gerado pelos cilindros

$$C(F, \sigma) = \{\omega \in \Omega \mid \omega_f = \sigma_f \text{ para } f \in F\}, \quad F \subset E \text{ e } \sigma = \{0,1\}^{|F|};$$

- $\mathbf{P}$ como a medida de probabilidade em Σ 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 Caratheodory podemos definir uma probabilidade $\mathbf{P}$.

Probabilidades Críticas

Vamos analisar a percolação de elo.

Para um sítio x , deixe C_x ser o cluster aberto contendo x e deixe $\theta_x(p)$ a probabilidade de que C_x seja infinito.

3.30 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 torná-la 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 , então $\theta_x(p)$ deve estar aumentando.

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 13. <

Agora, conecte este p_c à nossa pergunta original sobre a existência de um cluster aberto infinito.

3.31 Teorema

Seja E ser o evento que existe um cluster aberto infinito. Então $\mathbf{P}(E) = 1$ se $p > p_c$ e $\mathbf{P}(E) = 0$ se $p < p_c$.

Demonstração. Se E for um evento caudal, pela lei Kolmogorov 0-1, ele só pode tomar os valores 0 e 1. Assim, se $p < p_c$, então $\mathbf{P}(E) \leq \sum_x \theta_x(p) = 0$ então $\mathbf{P}(E) = 0$. Se $p > p_c$ então $\mathbf{P}(E) \geq \theta_x(p) > 0$ para pelo menos um sítio, então $\mathbf{P}(E) = 1$.

Falta mostrar que E é um evento caudal. Para isso, enumere os elos em G como $\{e_0, e_1, \dots\}$, e para $k \geq 0$ defina

$$A_k = \{\omega \in \Omega : \omega_{e_k} = 1\}.$$

Segue que $A_0, A_1, \dots$ são independentes, e além disso $\mathcal{F} = \sigma\langle A_0, A_1, \dots \rangle$.

Dado $\omega \in \Omega$ seja $G_0(\omega)$ o subgrafo de G com vértices em V e elos em

$$\{e \in E : \omega_e = 1\} \setminus \{e_0\}.$$

Deste modo o evento

$$E_0 = \{\text{existe um cluster infinito em } G_0\},$$

é tal que $E_0 \subseteq E$ e $E_0 \in \sigma\langle A_1, A_2, \dots \rangle$.

Além disso, se $\omega \in E$, então existe um cluster infinito $C_x(\omega)$.

Se C_x não contem e_0 , então C também é infinito em $G_0(\omega)$, e segue que $\omega \in E_0$.

Se C_x contem e_0 , a retirada de e_0 dividirá C_x em, no máximo, dois clusters conexos disjuntos, e ao menos um deles ainda será infinito, e como não contem e_0 , será infinito em $G_0(\omega)$. Deste modo, $\omega \in E_0$.

Concluimos assim que $E = E_0 \in \sigma\langle A_1, A_2, \dots \rangle$.

De modo análogo, mostramos que $\sigma\langle A_n, A_{n+1}, \dots \rangle$ para todo $n \geq 0$, e portanto E é caudal.

◁

Exercícios

Ex. 3.1 — Deixe Ω ser os inteiros $\{1, 2, \dots, 9\}$ 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 recíproca 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 $Y > 0$, tal que $\mu(A_n) \geq Y$ para todo n , então existe pelo menos um ponto que pertence a infinitos conjuntos A_n .

Ex. 3.7 —

1. Para todo $k = 1, \dots, n$ deixe $\mathcal{P}_k$ ser uma partição de Ω em conjuntos enumeráveis de $\mathcal{F}$. Mostre que as σ -álgebras $\sigma\langle \mathcal{P}_1 \rangle, \dots, \sigma\langle \mathcal{P}_n \rangle$ são independentes se e somente se (3.1) vale para cada escolha de A_k em $\mathcal{P}_k, k = 1, \dots, n$.
2. Mostre que os conjuntos $A_1, \dots, A_n$ são independentes se e somente se

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

para toda escolha de B_k como A_k ou A_k^c para $k = 1, \dots, n$.

Ex. 3.8 — Deixe $\mathcal{A}_1, \dots, \mathcal{A}_n$ serem π -sistemas de conjuntos em $\mathcal{F}$ tais que

$$\mathbf{P}\left(\bigcap_{k=1}^n A_k\right) = \prod_{k=1}^n \mathbf{P}(A_k) \quad (3.1)$$

para toda escolha de $A_k \in \mathcal{A}_k$ para $k = 1, \dots, n$.

1. Mostre usando um exemplo simples que $\mathcal{A}_k$'s não precisam ser independentes.
 2. Mostre que $\mathcal{A}_k$'s serão independentes se para todo k , Ω é união enumerável de conjuntos em $\mathcal{A}_k$.
- Dica: Inclusão- Exclusão

Ex. 3.9 — Dados $A_1, A_2, \dots$ Mostre que $\mathbf{P}(A_n \text{ i.v.}_n) = 1$ se e somente se $\sum_{n=1}^{\infty} \mathbf{P}(A_n|A)$ diverge para todo A de probabilidade não nula

Dica: Mostre que $\mathbf{P}(A_n \text{ i.v.}_n) < 1 \iff \sum_{n=1}^{\infty} \mathbf{P}(A_n|A) < \infty$ 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

$$A_n(\epsilon_1, \dots, \epsilon_n) = \{\omega = (\omega_1, \omega_2, \dots) \in \Omega : \omega_1 = \epsilon_1, \dots, \omega_n = \epsilon_n\}.$$

Fixe $p \in [0, 1]$ e defina

$$\mathbf{P}_p(A_n(\epsilon_1, \dots, \epsilon_n)) = p^{\sum \epsilon_i} (1-p)^{n-\sum \epsilon_i}$$

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

$$X_n(\omega) = \omega_n, \quad \omega = (\omega_1, \omega_2, \dots)$$

para $n = 1, \dots$ 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 -álgebra em Ω gerado pelos cilindros

$$C(F, \sigma) = \{\omega \in \Omega \mid \omega_f = \sigma_f \text{ para } f \in F\},$$

$$F \subset E, |F| < \infty \text{ e } \sigma = \{0, 1\}^{|F|};$$

- $\mathbf{P}$ 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 $\sigma < \mathcal{F}_0$.

Variáveis Aleatórias

4.1 Definição

Dados $(\Omega_1, \mathcal{F}_1)$ e $(\Omega_2, \mathcal{F}_2)$ espaços mensuráveis. Uma função $f : \Omega_1 \rightarrow \Omega_2$ é dita $\mathcal{F}_1 - \mathcal{F}_2$ -**mensurável** ou simplesmente **mensurável** se a pré-imagem de todo conjunto mensurável é mensurável. Ou seja, se

$$A \in \mathcal{F}_2 \implies f^{-1}(A) \in \mathcal{F}_1.$$

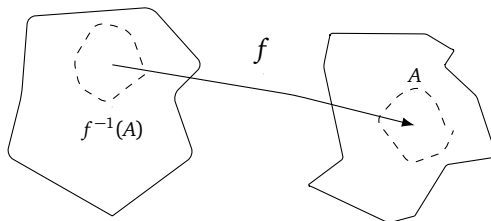


Figura 4.1: A função f é mensurável se $f^{-1}(A)$ é mensurável para todo A mensurável

4.2 Proposição (Mensurabilidade de mapas contínuos)

Seja (Ω_1, d_1) e (Ω_2, d_2) espaços métricos e seja $f : \Omega_1 \rightarrow \Omega_2$ um mapa contínuo. Então f é $\mathcal{B}(\Omega_1) - \mathcal{B}(\Omega_2)$ -mensurável.

4.3 Observação

- Dado uma função $f : \Omega_1 \rightarrow \Omega_2$ que é $\mathcal{F}_1 - \mathcal{F}_2$ -**mensurável** então se $\mathcal{F}_1 \subset \mathcal{F}'_1$ claramente f é $\mathcal{F}'_1 - \mathcal{F}_2$ -**mensurável**.

- O mesmo não ocorre se aumentarmos a σ -álgebra da imagem! Pode-se provar, por exemplo, que existem funções contínuas que não são Lebesgue mensuráveis, i.e, não são $\overline{\mathcal{B}}(\mathbb{R}) - \mathcal{B}(\mathbb{R})$ -mensuráveis.

4.4 Definição

Duas funções mensuráveis f e g que diferem num conjunto de medida nula são denominadas **equivalentes**. Ou seja ,

$$\mu(\{\omega \in \Omega_1 : f(\omega) \neq g(\omega)\}) = 0.$$

Duas dessa funções f e g são ditas **iguais quase certamente**. Tal fato será denotado por

$$f \stackrel{\text{q.c.}}{=} g,$$

4.5 Proposição

Dadas funções mensuráveis $g : \Omega_1 \rightarrow \Omega_2$ e $f : \Omega_2 \rightarrow \Omega_3$. Então a composição $f \circ g$ é mensurável.

Demonstração. Exercício

◁

Motivados pela Observação 4.3 definimos uma variável aleatória como uma função mensurável quando colocamos em $\mathbb{R}$ a σ -álgebra de Borel $\mathcal{R}$.

4.6 Definição

Uma **variável aleatória** é uma função mensurável que toma valores em $\mathbb{R}$

$$X : (\Omega, \mathcal{F}) \rightarrow (\mathbb{R}, \mathcal{R}),$$

No caso em que o espaço de chegada é $\mathbb{R}^n$, a função mensurável X é denominada **vetor aleatório**.

4.7 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

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

definida como:

$$X(H) = 1,$$

$$X(T) = 0.$$

◁

4.8 Exemplo (Função Indicadora de um Evento)

A função indicadora de um evento A , geralmente denotada por $X = 1_A$ ou por $X = I\{A\}$ é a função definida como:

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

Essa função é mensurável pois se o conjunto de Borel não contiver ou o 0 ou o 1, então a pré-imagem é vazia. Se contiver o 0 mas não o 1, então a pré-imagem é A^c . Se contiver o 1 mas não o 0, então a pré-imagem é A . Se contiver o 0 e o 1, então a pré-imagem é Ω . E assim X é mensurável.

<

4.9 Exemplo (Variáveis Aleatórias Simples)

Modificamos o exemplo anterior de modo que a imagem de X seja um conjunto finito $\{a_k\}$.

Nesse caso definimos $A_k = f^{-1}(a_k)$ e assim $X = \sum_{k=1}^n a_k 1_{A_k}$, onde a_k são alguns pesos e A_k são eventos disjuntos que particionam Ω :

$$\bigcup_{k=1}^n A_k = \Omega.$$

Se os conjuntos A_k forem mensuráveis então é fácil verificar que X é realmente uma variável aleatória.

<

Vamos guardar esse exemplo numa definição mais limpa:

4.10 Definição (Função simples)

Seja $(\Omega, \mathcal{F})$ um espaço mensurável. Uma aplicação $f : \Omega \rightarrow \mathbb{R}$ é dita função simples se houver um $n \in \mathbb{N}$ e conjuntos mensuráveis mutuamente disjuntos $A_1, \dots, A_n \in \mathcal{F}$, assim como um conjunto finito $\{a_k\} \subset \mathbb{R}$, tal que

$$\sum_{k=1}^n a_k \mathbb{1}_{A_k}.$$

Claramente uma função simples pode ser expressa de várias maneiras diferentes em termos de diferentes escolhas de conjuntos mensuráveis mutuamente disjuntos.

4.11 Exemplo

Suponha que $(\Omega, \mathcal{F}, \mathbf{P})$ é a medida de Lebesgue em $[0,1]$, então definimos as variáveis aleatórias X, Y , e Z como $X(\omega) = \omega$, $Y(\omega) = 2\omega$, e $Z(\omega) = 3\omega + 4$.

Então por exemplo $\mathbf{P}(Y < 1/3) = \mathbf{P}\{\omega; Y(\omega) < 1/3\} = \mathbf{P}\{\omega; 2\omega < 1/3\} = \mathbf{P}([0, 1/6]) = 1/6$.

<

4.12 Definição (Funções de Densidade)

Dado uma função f (que por enquanto assumiremos Riemann integrável, mas que de modo geral será Lebesgue integrável) e tal que

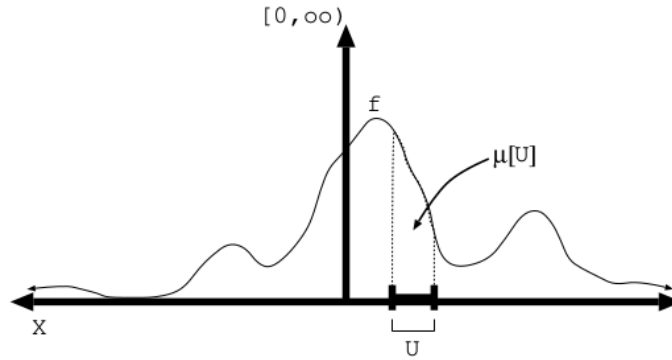
$$\int_{-\infty}^{\infty} f(y) dy = 1$$

Dizemos que uma variável aleatória X tem função de densidade f se

$$P(X \leq x) = F(x) = \int_{-\infty}^x f(y) dy$$

Nesse caso a função F é dita a distribuição de X

Figura 4.2: Densidade - [26]



Distribuição Uniforme Uma variável aleatória contínua X é uniforme no intervalo $[a, b]$, para $a < b$, se sua densidade 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 $y < z$

$$\mathbf{P}(y \leq X \leq z) = \int_y^z \frac{1}{b-a} dx = \frac{z-y}{b-a}.$$

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

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{se } x \geq 0 \\ 0 & \text{caso contrário.} \end{cases}$$

Distribuição Normal A variável aleatória X tem 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 $x \in \mathbb{R}$.

Dada uma aplicação como podemos verificar que ela é mensurável? Bem, pela definição você deve verificar para cada conjunto de Borel, mas isso é muito trabalhoso. Provaremos que na verdade, basta verificar para os geradores da σ -álgebra.

4.13 Teorema

Considere a função $f : (\Omega_1, \mathcal{F}_1, \mu_1) \rightarrow (\Omega_2, \mathcal{F}_2, \mu_2)$, onde $\mathcal{F}_2 = \sigma\langle \mathcal{A} \rangle$. Suponha que $f^{-1}(A)$ é mensurável para todo $A \in \mathcal{A}$. Ou seja,

$$A \in \mathcal{A} \Rightarrow f^{-1}(A) \in \mathcal{F}_1.$$

Então f é mensurável:

$$A \in \sigma\langle \mathcal{A} \rangle \Rightarrow f^{-1}(A) \in \mathcal{F}_1.$$

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 $A_1, A_2, \dots \in \mathcal{F}$, 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 $\mathcal{A} \subseteq \mathcal{F} = \sigma\langle \mathcal{A} \rangle$. $\triangleleft$

4.14 Teorema

Considere a função $X : (\Omega, \mathcal{F}, \mathbf{P}) \rightarrow \mathbb{R}$. Então X é a variável aleatória, se e somente se, o conjunto $\{\omega : X(\omega) \leq a\}$ é um evento para todo $a \in \mathbb{R}$.

Demonstração. Usaremos o Teorema 4.13 com

$$\begin{aligned} (\Omega_1, \mathcal{F}_1, \mu_1) &= (\Omega, \mathcal{F}, \mathbf{P}) \\ (\Omega_2, \mathcal{F}_2, \mu_2) &= (\mathbb{R}, \mathcal{R}, \mu) \end{aligned}$$

onde $\mathcal{R}$ denota a σ -álgebra $\mathcal{R} = \sigma\langle (-\infty, a], a \in \mathbb{R} \rangle$. Então $X^{-1}((-\infty, a]) = \{\omega : X(\omega) \leq a\}$. $\triangleleft$

Esta é uma maneira conveniente de verificar que X é uma variável aleatória.

4.1 Funções de Variáveis Aleatórias

4.15 Proposição

Se $X_1, \dots, X_n$ são variáveis aleatórias reais, então $(X_1, \dots, X_n)$ é um vetor aleatório.

Demonstração. Utilizando o Teorema 4.14 basta verificar a mensurabilidade do mapa

$$\omega \mapsto (X_1(\omega), \dots, X_n(\omega))$$

em uma classe de geradores da σ -álgebra de Borel em $\mathbb{R}^n$ como por exemplo os geradores

$$\{(-\infty, a_1] \times (-\infty, a_2] \times \dots \times (-\infty, a_n] : a_1, a_2, \dots, a_n \in \mathbb{R}\}.$$

Para esse fim consideraremos as pré-imagens:

$$\begin{aligned} & \{(X_1, \dots, X_n) \in (-\infty, a_1] \times (-\infty, a_2] \times \dots \times (-\infty, a_n]\} \\ &= \{X_1 \in (-\infty, a_1], \dots, X_n \in (-\infty, a_n]\} \\ &= \bigcap_{k=1}^n \{X_k \in (-\infty, a_k]\} = \bigcap_{k=1}^n \{X_k \leq a_k\}. \end{aligned}$$

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. $\triangleleft$

4.16 Teorema (Funções de Variáveis Aleatórias)

Dadas variáveis aleatórias reais $X_1, \dots, X_n$ e $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função mensurável. Então, $f(X_1, \dots, X_n)$ é a variável aleatória real.

Demonstração. A função $f(X_1, \dots, X_n)$ é composição de duas funções mensuráveis:

□ a função

$$\omega \mapsto (X_1(\omega), \dots, X_n(\omega))$$

é mensurável por 4.15.

□ a função

$$(x_1, \dots, x_n) \mapsto f(x_1, \dots, x_n)$$

é mensurável por hipótese.

Como a composição de funções mensuráveis é mensurável temos o resultado desejado. $\triangleleft$

4.17 Exemplo

Como corolário do Teorema anterior temos que se X é a variável aleatória, então:

- X^2
- $\sin(X)$
- e^X

são variáveis aleatórias.

Se $X_1, \dots, X_n$ são variáveis aleatórias, então:

$$\square X_1 + \dots + X_n$$

$$\square X_1 \cdot X_2 \cdot \dots \cdot X_n$$

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.

◁

4.2 Variáveis Aleatórias Geram σ -álgebras

4.20 Definição

A σ -álgebra em Ω gerada por X , e denotada por $\sigma\langle X \rangle$ é

$$\begin{aligned}\sigma\langle X \rangle &:= \{X^{-1}(A) : A \text{ Borel em } \mathbb{R}\} \\ &= \{X \in A\} : A \text{ Borel em } \mathbb{R}\end{aligned}$$

Observamos que $\sigma\langle X \rangle$ é a menor σ -álgebra em Ω que torna X mensurável.

4.21 Exemplo

Dados $\Omega = \{0,1\}^{\mathbb{N}}$ e $\omega = (\omega_1, \omega_2, \dots)$ um elemento de Ω . Considere a variável aleatória $X_n : \Omega \rightarrow \mathbb{R}$ definida por $X_n(\omega) = \omega_n$ e seja $\mathcal{F}_n = \sigma\langle X_1, \dots, X_n \rangle$.

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

$$X_1^{-1}(\{0\}) = A_1 = \{\omega : \omega_1 = 0\}$$

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

$$X_2^{-1}(\{0\}) = A_2 = \{\omega : \omega_2 = 0\}$$

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

$$\{\Omega, \emptyset, A_1, A_1^c, A_2, A_2^c, A_1 \cap A_2, A_1 \cap A_2^c, A_1^c \cap A_2, A_1^c \cap A_2^c, A_1 \cup A_2, A_1 \cup A_2^c, A_1^c \cup A_2, A_1^c \cup A_2^c\}$$

□

4.22 Proposição

Dado X uma variável aleatória em Ω e $\sigma\langle X \rangle$ a σ -álgebra gerada por X .

Dado 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 $f : \mathbb{R} \rightarrow \mathbb{R}$ 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.

4.3 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 da medida, o objeto central é o estudo dos espaços de medida $(\Omega, \mathcal{F}, \mu)$. Na teoria da probabilidade, em muitas situações tentamos esquecer rapidamente os espaços da medida e discutir as distribuições.

Falaremos mais sobre distribuições no Capítulo 6.

4.4 Independência de Variáveis Aleatórias

4.23 Definição

As variáveis aleatórias $X_1, \dots, X_n$ são ditas **variáveis aleatórias independentes** se

$$\mathbf{P}(X_1 \in B_1, \dots, X_n \in B_n) = \mathbf{P}(X_1 \in B_1) \cdots \mathbf{P}(X_n \in B_n)$$

para todos conjuntos de Borel $B_1, \dots, B_n$ in $\mathbb{R}$.

4.24 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.

◁

4.25 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:

4.26 Proposição

As variáveis aleatórias $X_1, \dots, X_n$ são independentes se e somente se

$$\mathbf{P}(X_1 \leq x_1, \dots, X_n \leq x_n) = \mathbf{P}(X_1 \leq x_1) \cdots \mathbf{P}(X_n \leq x_n)$$

para todo $x_1, \dots, x_n \in \mathbb{R}$.

4.27 Proposição

Dadas funções $\{f_k\}_{k=1}^n$ mensuráveis. Se $X_1, \dots, X_n$ são independentes, então $f_1(X_1), \dots, f_n(X_n)$ são independentes.

Demonstração. Primeiro observamos que

$$\mathbf{P}(f_1(X_1) \in B_1, \dots, f_n(X_n) \in B_n) = \mathbf{P}(f_1(X_1) \in B_1) \cdots \mathbf{P}(f_n(X_n) \in B_n). \quad (4.1)$$

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.

4.28 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 $(\sigma(X_n) : n \in \mathbb{N})$ 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 $\text{Exp}(\lambda)$ 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 $g : (0, 1) \rightarrow \mathbb{R}$ tal que o aleatório variável $X = g(U)$ tem distribuição $\text{Exp}(\lambda)$.

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 $\mathbf{P}(X \geq t | X \geq s) = \mathbf{P}(X \geq t - s)$ 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.

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:

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

$$\int f \, d\mu = \sum_{k=1}^n a_k \mu(A_k).$$

2 Para funções mensuráveis não negativas $f \geq 0$. Defina

$$\int f \, d\mu = \sup_{0 \leq \phi \leq f} \int \phi \, d\mu$$

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 $A_i \cap B_j \neq \emptyset$, e $f(x) = a_i = b_j$ para todo $x \in A_i \cap B_j$. Assim

$$\sum_{i=1}^n a_i \mu(A_i) = \sum_{i=1}^n \sum_{j=1}^m a_i \mu(A_i \cap B_j) \quad (5.1)$$

$$= \sum_{j=1}^m \sum_{i=1}^n b_j \mu(A_i \cap B_j) \quad (5.2)$$

$$= \sum_{j=1}^m b_j \mu(B_j) \quad (5.3)$$

◁

5.2 Proposição (Aproximação por Funções Simples)

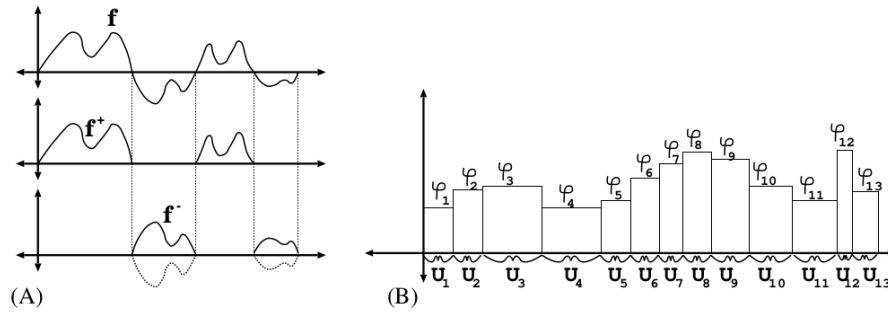
Toda função mensurável não-negativa $f : \Omega \rightarrow \mathbb{R}^+$ é o limite pontual de uma sequência monótona crescente de funções simples não negativas e se uma função f é o o limite pontual de uma sequência monótona crescente de funções simples então f é mensurável.

Demonstração.

$\Rightarrow$ 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

$$I_{n,k} = \left[\frac{k-1}{2^n}, \frac{k}{2^n} \right)$$

Figura 5.1: A- Parte Positiva e Negativa B- Função Simples - [26]



para $k = 1, 2, \dots, 2^{2n}$ e $I_{n, 2^{2n}+1} = [2^n, \infty)$.

Observe que, para n fixo, os conjuntos $I_{n,k}$ são disjuntos e cobrem a reta real não negativa.

Agora, defina os conjuntos mensuráveis

$$A_{n,k} = f^{-1}(I_{n,k}) \text{ para } k = 1, 2, \dots, 2^{2n} + 1.$$

Então, 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 $n \rightarrow \infty$. 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.

◀ Já provamos que o limite de funções mensuráveis é mensurável. ◀

Ou seja, a classe das funções mensuráveis é o fecho por limites pontuais das funções simples.

Através de uma pequena adaptação dos argumentos anteriores podemos provar que:

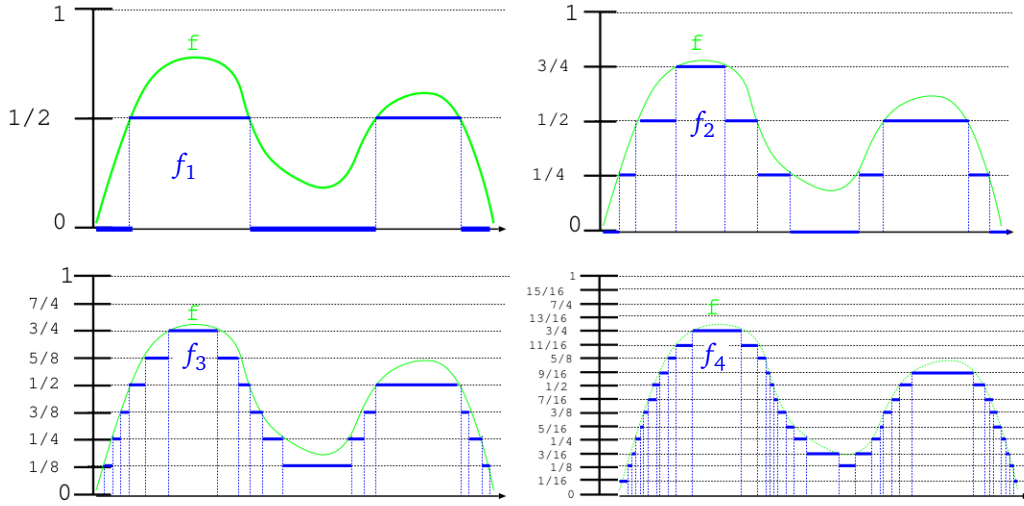
5.3 Proposição (Aproximação por Funções Simples)

Dada função mensurável não-negativa $f : E \subset \Omega \rightarrow \mathbb{R}^+$. 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

$$\int f \, d\mu = \int f^+ \, d\mu - \int f^- \, d\mu.$$

- 2 Dizemos que f é **quase integrável** se pelo menos uma das funções f^+ e f^- é integrável. E nesse caso também definimos

$$\int f \, d\mu = \int f^+ \, d\mu - \int f^- \, d\mu.$$

- 2 A integral num conjunto mensurável $A \subseteq \Omega$ é definida como

$$\int_A f \, d\mu = \int_{\Omega} f \cdot \mathbb{1}_A \, d\mu.$$

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 $\int f \, d\mu \leq \int g \, d\mu$.

2 Se $f = g$ em quase todo ponto, então as integrais são iguais.

3 (Linearidade) $\int (af + bg) \, d\mu = a \int f \, d\mu + b \int g \, d\mu$, onde $a, b \in \mathbb{R}$.

4 $\left| \int f \, d\mu \right| \leq \int |f| \, d\mu$.

Demonstração. Como $f \leq |f|$, temos $\int f \leq \int |f^+| \, d\mu$. De modo análogo, $-f \leq |f|$, logo $-\int f \, d\mu \leq \int |f^-| \, d\mu$.

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

$$\int_{\Omega} \varphi \, d\mu + \int_{\Omega} \psi \, d\mu = \sum_i a_i \mu(A_i) + \sum_j b_j \mu(B_j) \quad (5.4)$$

$$= \sum_i \sum_j a_i \mu(A_i \cap B_j) + \sum_j \sum_i b_j \mu(B_j \cap A_i) \quad (5.5)$$

$$= \sum_i \sum_j (a_i + b_j) \mu(A_i \cap B_j) \quad (5.6)$$

$$= \int_{\Omega} (\varphi + \psi) \, d\mu. \quad (5.7)$$

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

$$af_n + bg_n \uparrow af + bg$$

e usamos que

$$\int af + bg \, d\mu = \sup_{\phi} \left\{ \int \phi \, d\mu \right\} \geq \sup_{af_n + bg_n} \left\{ \int af_n + bg_n \, d\mu \right\} \quad (5.8)$$

$$= a \int f \, d\mu + b \int g \, d\mu \quad (5.9)$$

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

$$\boxed{1} \quad \int_{\Omega} f \, d\mu < \infty \text{ então } f < \infty \, \mu\text{-q.c.}$$

$$\boxed{2} \quad \int_{\Omega} f \, d\mu = 0 \iff f = 0 \, \mu\text{-q.c.}$$

$$\boxed{3} \quad \text{Se } f = g \, \mu\text{-q.c. então } \int_{\Omega} f \, d\mu = \int_{\Omega} g \, d\mu.$$

Observação. Uma consequência do Teorema 5.6 $\boxed{2}$ é que para qualquer função $f : \Omega \rightarrow \mathbb{R}$, 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 $f : \Omega \rightarrow \mathbb{R}$ que são mensuráveis $\mathcal{F}$ com : $\int_{\Omega} f^+ \, d\mu < \infty$ e $\int_{\Omega} f^- \, d\mu < \infty$;

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

$$\int f \, d\mu = \sup_{0 \leq \phi \leq f} \int \phi \, d\mu$$

converge? O próximo resultado nos tira essa esperança.

5.8 Teorema

Seja f uma função limitada em um conjunto de medida finita A . Então f é 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 $\phi_n \rightarrow \phi$. Claramente teremos que $\phi \leq f$ e $\int_A \phi \, d\mu = \int_A f \, d\mu$

Então

$$0 \leq \int_A f \, d\mu - \phi \leq \int_A \phi_n - \phi \, d\mu \rightarrow 0$$

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 Riemman-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 $f : [a, b] \rightarrow \mathbb{R}$. Se f é Riemann integrável, então f é Lebesgue integrável, e ambas as integrais concordam

Demonstração.

No que se segue denotaremos por $\int \dots \, dx$ as integrais de Riemann e por $\int \dots \, d\lambda(x)$ 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

$$\int \varphi_n \, dx \rightarrow \int f \, dx \leftarrow \int \psi_n \, dx.$$

Fazendo a troca de φ_n e ψ_n por $\max\{\varphi_1, \dots, \varphi_n\}$ e $\min\{\psi_1, \dots, \psi_n\}$, podemos assumir sem perda de generalidade que as sequências φ_n e ψ_n são crescente e decrescente respectivamente.

Para funções escadas $\gamma : [a, b] \rightarrow \mathbb{R}$, a integral de Lebesgue e a de Riemann coincidem. E assim,

$$\int |\psi_n - \varphi_n| \, d\lambda(x) = \int \psi_n - \varphi_n \, dx \rightarrow \int f \, dx - \int f \, dx = 0. \quad (5.10)$$

Pela monotonicidade, $\varphi_n \rightarrow \varphi$ e $\psi_n \rightarrow \psi$ pontualmente com $\varphi_1 \leq \varphi \leq f \leq \psi \leq \psi_1$.

Pelo Teorema da Convergência Dominada

$$\int |\psi_n - \varphi_n| \, d\lambda(x) \rightarrow \int |\psi - \varphi| \, d\lambda(x).$$

Pela Equação (5.10), temos $\int |\psi - \varphi| \, d\lambda(x) = 0$ e logo $\psi = \varphi$ quase certamente.

Como $\varphi \leq f \leq \psi$, temos $f = \varphi = \psi$ quase certamente, logo f é Lebesgue mensurável e

$$\int f \, d\lambda(x) = \int \psi \, d\lambda(x) = \lim_n \int \psi_n \, d\lambda = \lim_n \int \psi_n \, dx = \int f \, dx.$$

◁

5.11 Exemplo

Considere a função indicadora dos racionais $f = \mathbb{1}_{\mathbb{Q}}$. Como $\mathbb{Q}$ é enumerável, $\mu(\mathbb{Q}) = 0$. Logo $\int f \, d\mu = 0$. 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 $\mu(A \setminus (K_n \cup K'_n)) < \varepsilon/2^n$.

Seja

$$K = \bigcap_{n=1}^{\infty} (K_n \cup K'_n).$$

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 $f_n \nearrow f$, então $\int f_n \, d\mu \rightarrow \int f \, d\mu$.

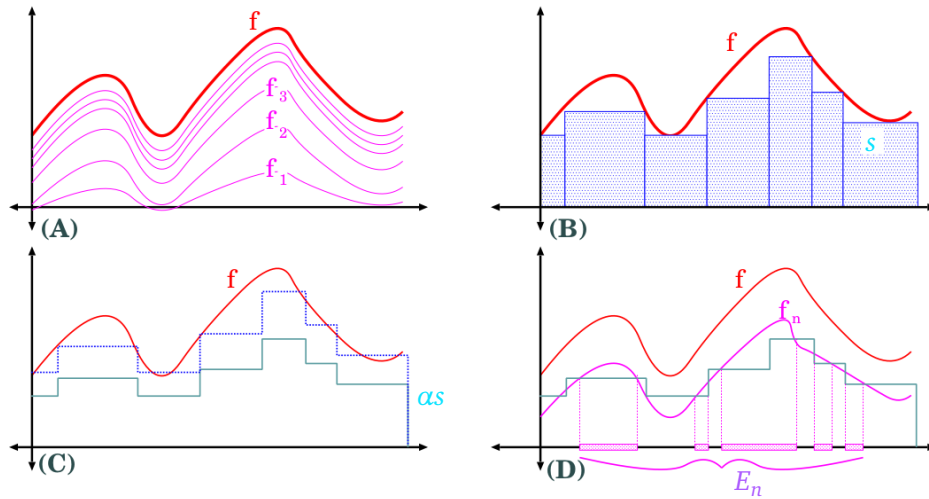
Demonstração. Como $\int f_n d\mu \leq \int f_{n+1} d\mu$ temos que existe $\alpha \in [0, \infty]$ tal que

$$\lim_{n \rightarrow \infty} \int f_n d\mu = \alpha$$

Como $f_n \leq f$ para todo n temos que $\int f_n d\mu \leq \int f d\mu$ para todo n . E logo $\alpha \leq \int f d\mu$. Seja s uma função simples mensurável tal que $0 \leq s \leq f$ e deixe c ser 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, $E_1 \subset E_2 \subset \dots$, 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,

$$\int_{\Omega} f_n d\mu \geq \int_{E_n} f_n d\mu \geq c \int_{E_n} s d\mu$$

para todo n . Fazemos $n \rightarrow \infty$ e assim obtemos que $\alpha \geq c \int_{\Omega} s d\mu$. Como esse fato é verdadeiro para todo $c < 1$ temos que

$$\alpha \geq \int_{\Omega} s d\mu$$

para toda função simples $s \leq f$. Logo, temos

$$\alpha \geq \int_{\Omega} f d\mu$$

Isso prova a desigualdade reversa. ◁

5.14 Proposição

Seja $f : \Omega \rightarrow [0, \infty]$ uma aplicação mensurável. Então:

1 $f = 0$ quase certamente se e somente se $\int_{\Omega} f \, d\mu = 0$.

2 Se $\int_{\Omega} f \, d\mu < \infty$, então $f < \infty$ quase certamente.

Demonstração. Provaremos para funções não negativas.

1 $\Rightarrow$ 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

$$0 \leq \int f \, d\mu \leq \int \infty \mathbb{1}_N \, d\mu = \lim \int n \mathbb{1}_N \, d\mu = 0$$

1 $\Leftarrow$ Seja $N_n : \{x : f(x) > \frac{1}{n}\}$. Então $N_n \uparrow N$ e

$$0 = \int f \, d\mu \geq \int \frac{1}{n} \mathbb{1}_{N_n} \, d\mu = \frac{\mu(N_n)}{n}$$

Logo $\mu(N) = 0$.

2 Seja $I : \{x : f(x) = \infty\}$. Então para todo n

$$\mu(I) = \int \mathbb{1}_I \, d\mu \leq \int \mathbb{1}_{\{f \geq n\}} \, d\mu \quad (5.11)$$

$$\leq \frac{1}{n} \int f \, d\mu \leq \frac{1}{n} \int f \mathbb{1}_{\{f \geq n\}} \, d\mu \rightarrow 0 \quad (5.12)$$

$\triangleleft$

5.15 Lema (Lema de Fatou)

Se $f_n \geq 0$, então $\int \liminf f_n \, d\mu \leq \liminf \int f_n \, d\mu$.

Demonstração. Dado $g_k(x) = \inf_{i \geq k} f_i(x)$. Então $g_k \leq f_k$, e assim

$$\int g_k \, d\mu \leq \int f_k \, d\mu$$

Mais ainda temos que $0 \leq g_1 \leq g_2 \leq \dots$, e cada g_k é mensurável com $g_k(x) \rightarrow \liminf f_n(x)$ quando $k \rightarrow \infty$. O Teorema da Convergência Monótona implica que o lado esquerdo tende ao lado direito quando $k \rightarrow \infty$. $\triangleleft$

5.16 Definição (Convergência quase todo ponto)

Dizemos que $f_n \rightarrow f$ em quase todo ponto ou quase certamente (q.c.) se $\mu(\{\omega : f_n(\omega) \not\rightarrow f(\omega)\}) = 0$.

5.17 Teorema (Teorema da Convergência Dominada)

Se $f_n \rightarrow f$ q.c. e $|f_n| \leq g$ para todo n , e g é integrável, então

$$\int f_n \, d\mu \rightarrow \int f \, d\mu.$$

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

$$\begin{aligned} \int_{\Omega} 2g \, d\mu &\leq \liminf_{n \rightarrow \infty} \int_{\Omega} (2g - |f_n - f|) \, d\mu \\ &= \int_{\Omega} 2g \, d\mu + \liminf_{n \rightarrow \infty} \left(- \int_{\Omega} |f_n - f| \, d\mu \right) \\ &= \int_{\Omega} 2g \, d\mu - \limsup_{n \rightarrow \infty} \int_{\Omega} |f_n - f| \, d\mu \end{aligned}$$

Como $\int_{\Omega} 2g \, d\mu$ é finita podemos subtraí-la

$$\limsup_{n \rightarrow \infty} \int_{\Omega} |f_n - f| \, d\mu \leq 0$$

E assim

$$\lim_{n \rightarrow \infty} \int_{\Omega} |f_n - f| \, d\mu = 0$$

Se $f \in L^1$ então $\left| \int_{\Omega} f \right| \leq \int_{\Omega} |f|$ e temos

$$\lim_{n \rightarrow \infty} \int_{\Omega} f_n \, d\mu = \int_{\Omega} f \, d\mu$$

◁

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

$$\int |f|^p \, d(\mu) < \infty$$

sob a relação de equivalência de igualdade quase certa.

□ Para qualquer elemento $f \in \mathcal{L}^p(\Omega, \mathcal{F}, \mu)$ definimos a p -norma como

$$\|f\|_p = \left(\int |f|^p d(\mu) \right)^{1/p}$$

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 $f + g \in \mathcal{L}^p(\Omega, \mathcal{F}, \mu)$:

$$(f + g)^p \leq (f \vee g + f \vee g)^p = 2^p (f \vee g)^p \leq 2^p (f^p + g^p)$$

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

$$\|f + g\|_p^p = \int (f + g)^p d(\mu) = \int f(f + g)^{p-1} d(\mu) + \int g(f + g)^{p-1} d(\mu)$$

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

$$\int f(f + g)^{p-1} d(\mu) \leq \left(\int f^p d(\mu) \right)^{\frac{1}{p}} \left(\int (f + g)^{(p-1)q} d(\mu) \right)^{\frac{1}{q}} = \|f\|_p \|f + g\|_p^{p/q}$$

Aplicando este argumento ao termo $\int g(f + g)^{p-1} d(\mu)$ temos

$$\|f + g\|_p^p \leq (\|f\|_p + \|g\|_p) \cdot \|f + g\|_p^{p/q}$$

dividindo por $\|f + g\|_p^{p/q}$ e usando que $p - \frac{p}{q} = 1$ podemos concluir que $\|f + g\|_p \leq \|f\|_p + \|g\|_p$. ◁

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

$$\begin{aligned} \left\| \sum_{j=1}^{\infty} (f_{n_{j+1}} - f_{n_j}) \right\|_p &= \lim_{N \rightarrow \infty} \left\| \sum_{j=1}^N (f_{n_{j+1}} - f_{n_j}) \right\|_p \\ &\leq \lim_{N \rightarrow \infty} \sum_{j=1}^N \|f_{n_{j+1}} - f_{n_j}\|_p \\ &\leq \lim_{N \rightarrow \infty} \sum_{j=1}^N \frac{1}{2^j} < \infty \end{aligned}$$

e, portanto, podemos concluir que a soma $\sum_{j=1}^{\infty} (f_{n_{j+1}} - f_{n_j})$ 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 $\sum_{j=N}^{\infty} (f_{n_{j+1}} - f_{n_j}) < \epsilon$, então para qualquer $k \geq j \geq N$ temos

$$|f_{n_k} - f_{n_j}| = \left| \sum_{m=j}^k (f_{n_{m+1}} - f_{n_m}) \right| \leq \sum_{m=j}^k |f_{n_{m+1}} - f_{n_m}| < \epsilon$$

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 $n \geq N$,

$$\int |f - f_n|^p d\mu \leq \liminf_{j \rightarrow \infty} \int |f_{n_j} - f_n|^p d\mu \leq \sup_{m \geq n} \int |f_m - f_n|^p d\mu < \epsilon^p$$

Portanto, pela Desigualdade Minkowski, temos que $f = f_n + (f - f_n)$ está em $\mathcal{L}^p(\Omega, \mathcal{F}, \mu)$ e $f_n \xrightarrow{p} f$. $\triangleleft$

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 $A_1, A_2, \dots \in \mathcal{F}$. 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 $f : X \rightarrow \mathbb{R}^+$ 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

Distribuições

6.1 Função de distribuição

Em diversos modelos probabilísticos o domínio no qual a variável aleatória está definida não é tão importante quanto os seus valores ou seja nesses modelos as especificidades da natureza do espaço amostral Ω podem não são tão importantes quanto os valores das probabilidades para $\omega \in \Omega$.

Nesse capítulo introduziremos o conceito de distribuição que nos permitirá abstrair do espaço de probabilidade.

6.1 Definição (Distribuição)

Se X for uma variável aleatória, então X induz uma medida de probabilidade $\mathbf{P}$ em $\mathbb{R}$ definida tomando a pré-imagem para cada conjunto de Borel A :

$$\mathbf{P}(A) = \mathbf{P}(X^{-1}(A)) = \mathbf{P}(X \in A). \quad (6.1)$$

Esta medida de probabilidade $\mathbf{P}$ é denominada **distribuição** de X .

Observe que na equação 8.1, não há nenhuma referência para o espaço amostral Ω . Nessa equação estamos definindo uma nova probabilidade na reta real.

Claramente 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.

6.2 Exemplo

- 1 Para o lançamento de moedas podemos definir a variável aleatória $X(H) = 1, X(T) = 0$. Por outro lado poderíamos definir a variável aleatória $Y(H) = 0, Y(T) = 1$.

Estas variáveis são muito diferentes, essencialmente elas são opostas, mas X e Y possuem a mesma distribuição.

- 2** Considere dois lançamentos de moedas. Faça $X = 1$ se aparecer H no primeiro lançamento e $X = 0$ caso contrário. Faça $Y = 1$ se H aparecer no segundo lançamento e $Y = 0$ caso contrário.

Agora, essas variáveis aleatórias não são mais opostas e sim independentes. Mas X e Y têm a mesma distribuição.

◁

De maneira análoga podemos definir a distribuição de um vetor aleatório

6.3 Definição (Distribuição de Vetores Aleatórios)

Se X for um vetor aleatório, então X induz uma medida de probabilidade $\mathbf{P}$ em $\mathbb{R}^n$ definida tomando a pré-imagem para cada conjunto de Borel A :

$$\mathbf{P}(A) = \mathbf{P}(X^{-1}(A)) = \mathbf{P}(X \in A). \quad (6.2)$$

Esta medida de probabilidade $\mathbf{P}$ é denominada **distribuição** de X .

Se $X = (X_1, \dots, X_n)$ é um vetor aleatório então as distribuições de cada uma das variáveis aleatórias dadas pelas componente de X , X_i para $i = 1, \dots, n$ são chamadas de **distribuições marginais**.

6.4 Definição (Equivalências de Variáveis Aleatórias)

- a** Dizemos que as variáveis aleatórias X e Y são iguais **quase certamente**, fato que denotaremos por

$$X \stackrel{\text{q.c.}}{=} Y,$$

se $\mathbf{P}(X \neq Y) = 0$.

- b** Dizemos que as variáveis aleatórias X e Y são iguais em **distribuição**, fato que denotaremos por

$$X \stackrel{\text{d.}}{=} Y,$$

se eles tiverem a mesma distribuição. Equivalentemente,

$$\mathbf{P}(X \in A) = \mathbf{P}(Y \in A) \text{ para todo } A \text{ de Borel.}$$

Se $X \stackrel{\text{q.c.}}{=} Y$, então $X \stackrel{\text{d.}}{=} Y$, mas não vale a recíproca.

A distribuição de X é determinada por seus valores da forma $\mathbf{P}(X \in A)$, $A = (-\infty, a]$, $a \in \mathbb{R}$, pois tais intervalos formam uma coleção de geradores para $\mathcal{R}$, e uma medida é determinada unicamente pelos seus valores em uma coleção de geradores.

6.5 Definição (Função de Distribuição de uma Variável Aleatória)

A **função de distribuição** $F : \mathbb{R} \rightarrow [0,1]$ de uma variável aleatória X é definida como

$$F(x) := \mathbf{P}(X \leq x), \quad x \in \mathbb{R}.$$

Claramente 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 $\mathbb{R} \rightarrow \mathbb{R}$.

6.6 Exemplo

Seja X = número de H em dois lançamentos independentes de uma moeda justa. Então temos

$$X = \begin{cases} 0, & \text{com probabilidade} = \frac{1}{4} \\ 1, & \text{com probabilidade} = \frac{1}{2} \\ 2, & \text{com probabilidade} = \frac{1}{4} \end{cases}$$

Então, podemos criar a função de distribuição. Esta função apresentará descontinuidades do tipo salto nos pontos 0, 1 e 2.

$$F(X) = \mathbf{P}(X \leq x) = \begin{cases} 0, & x < 0 \\ \frac{1}{4}, & x \in [0, 1] \\ \frac{1}{4} + \frac{1}{2}, & x \in [1, 2] \\ \frac{1}{4} + \frac{1}{2} + \frac{1}{4}, & x \in [2, +\infty) \end{cases}$$

◁

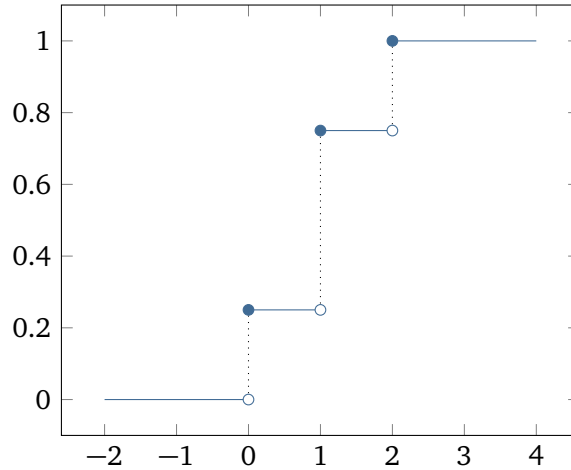


Figura 6.1: Distribuição do Exemplo 6.6

A função de distribuição de qualquer variável aleatória simples será constante por partes.

6.7 Teorema (Propriedades da Função de Distribuição)

A função de distribuição $F(x)$ de uma variável aleatória X tem as seguintes propriedades:

- 1 F é não decrescente e $0 \leq F(x) \leq 1$.
- 2 $F(x) \rightarrow 0$ quando $x \rightarrow -\infty$ e $F(x) \rightarrow 1$ quando $x \rightarrow +\infty$.

- 3 F é contínua a direita, ou seja

$$\lim_{y \rightarrow x+} F(y) = F(x).$$

- 4 F tem limites à esquerda em todos os pontos. Além disso,

$$F(x-) := \lim_{y \rightarrow x-} F(y) = \mathbf{P}(X < x).$$

Em particular,

$$\mathbf{P}(X = x) = F(x) - F(x_-).$$

- 5 F tem no máximo número enumerável de descontinuidades.

Demonstração.

- 1 Trivial

- 2 Os eventos $\{X \leq x_n\} \searrow \emptyset$ quando $x_n \rightarrow -\infty$. Pela continuidade, $\mathbf{P}(X \leq x_n) \rightarrow \mathbf{P}(\emptyset) = 0$.

Da mesma forma, $\{X \leq x_n\} \nearrow \Omega$ quando $x_n \rightarrow +\infty$. Então, $\mathbf{P}(X \leq x_n) \rightarrow \mathbf{P}(\Omega) = 1$. Isso completa a prova de que $F(x) \rightarrow 0$ como $x \rightarrow -\infty$ e $F(x) \rightarrow 1$ quando $x \rightarrow \infty$.

- 3 Para demonstrar que F é contínua a 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 $y_n \searrow x$). Nós temos

$$\mathbf{P}(X \leq y_n) \rightarrow \mathbf{P}(X \leq x), \quad n \rightarrow \infty.$$

A conclusão resulta da convergência dos conjuntos

$$\{X \leq y_n\} \searrow \{X \leq x\}$$

e da continuidade.

- 4 Agora vamos demonstrar que F tem limites à esquerda em todos os pontos, e

$$F(x-) := \lim_{y \rightarrow x-} F(y) = \mathbf{P}(X < x).$$

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 $y_n \nearrow x$, e queremos mostrar

$$\mathbf{P}(X \leq y_n) \rightarrow \mathbf{P}(X < x), \quad n \rightarrow \infty.$$

Isso decorre da convergência dos conjuntos

$$\{X \leq y_n\} \nearrow \{X < x\}$$

- 5 F tem no máximo número enumerável de descontinuidades.

Como $0 \leq F(x) \leq 1$,

- pode haver no máximo 2 saltos (ou seja, descontinuidades) de altura $\geq \frac{1}{2}$. Aqui, "saltos" significa pontos x em que $F(x) - F(x-) \geq \frac{1}{2}$.
- pode haver no máximo 3 saltos de tamanho $\in [\frac{1}{3}, \frac{1}{2})$.
- pode haver no máximo 4 saltos de tamanho $\in [\frac{1}{4}, \frac{1}{3})$.
- ...

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.

◁

6.8 Proposição

$$\mathbf{P}(X \in (a, b]) = F(b) - F(a).$$

6.9 Definição (Função de distribuição)

Uma função de monótona contínua à direita $F : \mathbb{R} \rightarrow [0, 1]$ com $\lim_{x \rightarrow -\infty} F(x) = 0$ e $\lim_{x \rightarrow \infty} F(x) = 1$ é chamada de candidata a função de distribuição de probabilidade.

Podemos começar com uma função que satisfaça as propriedades e obter uma variável aleatória:

6.10 Teorema (Existência de variável aleatória com distribuição prescrita)

Se uma função F é uma candidata a função de distribuição de probabilidade, então F é a função de distribuição de alguma variável aleatória X .

Demonstração.

A prova é construtiva. Vamos exibir a variável aleatória X para a qual F é uma função de distribuição.

Considere o espaço de probabilidade $(\Omega, \mathcal{F}, \mathbf{P})$, onde $\Omega = [0, 1]$, $\mathcal{F}$ é a coleção de conjuntos de Borel, e $\mathbf{P}$ é a medida de Lebesgue.

Bem, como faríamos essa construção se F fosse contínua e estritamente monótona? Nós pensamos no intervalo $\Omega = [0, 1]$ como o contradomínio de F . Então, definimos $X(\omega) = F^{-1}(\omega)$, pois F seria bijetiva.

Nesse caso essa é a definição correta pois a probabilidade de $(-\infty, a]$, é exatamente $F(a)$.

No caso geral fazemos

$$X(\omega) = \inf\{y : F(y) \geq \omega\}$$

A função X que definimos acima é a inversa generalizada de F .

É fácil ver então que

$$F(x) := \mathbf{P}(X \leq x), \quad x \in \mathbb{R}. \quad (6.3)$$

Para vermos isso usaremos que

$$\{\omega : X(\omega) \leq x\} = \{\omega : \omega \leq F(x)\}$$

Esse fato está demonstrado na Proposição 6.11. o resultado desejado segue imediatamente pois

$$\mathbf{P}(\omega : \omega \leq F(x)) = F(x).$$

Na fórmula acima usamos que $\mathbf{P}$ é a medida de Lebesgue em $[0,1]$.

◁

6.11 Proposição (Propriedades da Inversa Generalizada)

Se F é não decrescente e contínua a 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 $F^{-1}(F_1^{-1}(x) - \varepsilon) < x$. 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 $x_n \rightarrow x_0$. Então $\liminf_n F(x_n) \geq y$, mas como F é monótona e não-decrescente e contínua a a 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.

◁

6.12 Definição (Variáveis Aleatórias Identicamente Distribuídas)

Duas variáveis aleatórias X e Y são ditas identicamente distribuídas se

$$\mathbf{P}(X \leq x) = \mathbf{P}(Y \leq x)$$

para todo $x \in \mathbb{R}$.

6.2 Tipos de Distribuições

6.13 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

$$F(b) - F(a) = \int_a^b f(x) dx \quad \text{para todo } a < b$$

- 4 singular se $F \not\equiv 0$, F' existe e igual a 0 q.c.

Temos o seguinte teorema de decomposição

6.14 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 $\alpha + \beta + \gamma = 1$.

$$\square F_{ac} = \int_{-\infty}^x f(y) dy \quad \text{com } f(x) = F'_{ac} \text{ q.c.}$$

$\square F_d$ é uma função saltos com no máximo um número enumerável de saltos.

$\square F_s$ é singular.

Começaremos provando que toda distribuição pode ser decomposta como soma de uma discreta e uma contínua.

6.15 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

$$p(x_j) = F(x_j+) - F(x_j-) = F(x_j) - F(x_j-)$$

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 $\lim_{x \rightarrow \infty} F_d^* = \beta$ 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

$$\begin{aligned} F_c^*(x) - F_c^*(x-) &= F(x) - F_d(x) - (F(x-) - F_d(x-)) \\ &= F(x) - F(x-) - (F_d(x) - F_d(x-)) \\ &= \begin{cases} p_j - p_j = 0 & \text{se } x = x_j \\ 0 & \text{caso contrário} \end{cases} \end{aligned}$$

A função $F_c^*(x)$ satisfaz todas as propriedades de distribuição exceto possivelmente a última. Mas podemos renormalizá-la a uma distribuição. Se $\lim_{x \rightarrow \infty} F_c^*(x) = \alpha$ definimos $F_c = F_c^*/\alpha$.

Como

$$F(x) = F_c^*(x) + F_d^*(x) \tag{6.4}$$

$$= \alpha F_c + \beta F_d \tag{6.5}$$

Por último, o fato que $\alpha + \beta = 1$ decorre do fato que $F(x)$ é uma distribuição.

◁

6.16 Teorema

Toda função de distribuição contínua pode ser decomposta em uma combinação convexa de uma absolutamente contínua e uma contínua singular. Assim, se F é uma função de distribuição, 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

6.17 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
$$\int_a^b F'(x)dx \leq F(b) - F(a)$$

A demonstração desse lema será feita no Capítulo ??

Demonstração. Se definirmos $F_{ac}^*(x) = \int_{-\infty}^x F'(y)dy$ 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 $\mathbf{P}$).

6.18 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 $S = \{x_1, x_2, \dots\}$, 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	$p(x_i) = \frac{1}{n}$	$x_1, \dots, x_n$	
Bernoulli	$\text{Be}(p)$	$p(0) = q, p(1) = p$	$\{0, 1\}$
Binomial	$\text{Bin}(n, p)$	$p(k) = p^k q^{n-k}$	$k = 0, 1, \dots, n$
Geométrica	$\text{Ge}(p)$	$p(k) = pq^k$	$k \in \mathbb{N}$
Primeiro Sucesso	$\text{Po}(p)$	$p(k) = pq^{k-1}$	$k \in \mathbb{N}^*$
Poisson	$\text{Po}(m)$	$p(k) = e^{-m} \frac{m^k}{k!}$	$k \in \mathbb{N}$

A distribuição uniforme discreta assume os valores $x_1, \dots, x_n$ e $p(x_i) = \frac{1}{n}$. A distribuição $\text{Be}(p)$ descreve o resultado de um experimento de lançamento de moeda (não necessariamente honesta e a distribuição $\text{Bin}(n, p)$ é o número de sucessos em n lançamentos. A distribuição $\text{Ge}(p)$ descreve o número de fracassos antes do primeiro sucesso e a distribuição $\text{Po}(p)$ o número de lançamentos necessários para ter sucesso uma vez.

A distribuição de Poisson surge como o limite da distribuição binomial quando $n \rightarrow \infty$ e $p \rightarrow 0$ e $np \rightarrow \lambda$.

6.19 Teorema (Teorema Limite de Poisson)

Se $n \rightarrow \infty, p \rightarrow 0$, de tal modo que $np \rightarrow \lambda$ então

$$\frac{n!}{(n-k)!k!} p^k (1-p)^{n-k} \rightarrow e^{-\lambda} \frac{\lambda^k}{k!}.$$

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

$$F(x) = \int_{-\infty}^x f(y) dy \quad \text{para cada } x.$$

Claramente, é necessário que $f \geq 0$ e que

$$\int_{-\infty}^{\infty} f(x) dx = 1.$$

□ 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

$$f \geq 0 \quad \text{e} \quad \int_{-\infty}^{\infty} f(x) dx = 1,$$

e definir a função de distribuição pela fórmula

$$F(x) := \int_{-\infty}^x f(y) dy.$$

6.20 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

$$\mathbf{P}(y \leq X \leq z) = \int_y^z \frac{1}{b-a} dx = \frac{z-y}{b-a}.$$

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

$$F_X(t) = \int_0^t \lambda e^{-\lambda x} dx = 1 - e^{-\lambda x}$$

portanto $\mathbf{P}(X > a) = e^{-\lambda a}$.

Distribuição Normal A variável aleatória X tem 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 $\Phi : F \rightarrow F$, 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

$$\mu(\Phi^n(A)) = \left(\frac{2}{3} \right)^n \mu(A) \quad (6.6)$$

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

$$C = \bigcap_{n \in \mathbb{N}} C_n$$

6.21 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

$$\begin{aligned} C_0 &= [0,1] \\ C_1 &= [0,1/3] \cup [2/3,1] \\ C_2 &= [0,1/9] \cup [2/9,1/3] \cup [2/3,7/9] \cup [8/9,1] \\ C_3 &= \dots \end{aligned}$$

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 $f_n : \mathbb{R} \rightarrow \mathbb{R}$ 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 $J \in C_n$

$$\begin{aligned} \int_J f_n d\mu &= \left(\frac{3}{2}\right)^n \mu(J) \\ &= \left(\frac{3}{2}\right)^{n+k} \mu(\Phi^k(J)) \\ &= \int_J f_{n+k} d\mu \end{aligned}$$

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

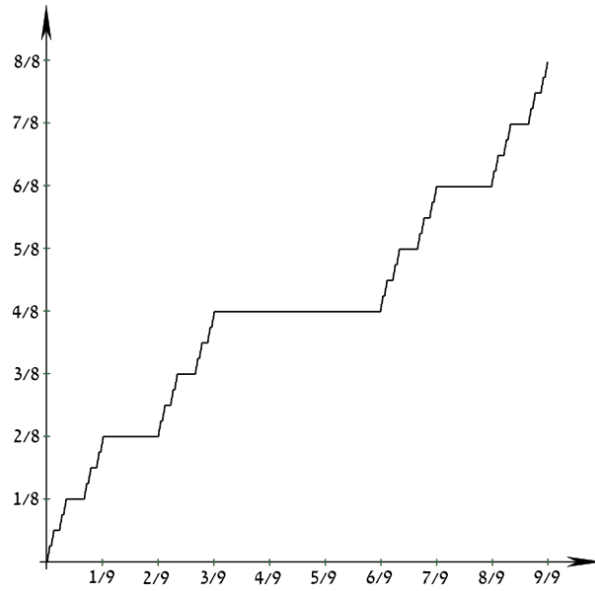
$$\begin{aligned}
|F_{n+k}(x) - F_n(x)| &= \left| \sum_{J \in \mathcal{C}_n, J \subset (-\infty, x]} F_{n+k}(J) + F_{n+k}(J \cap (-\infty, x]) \right. \\
&\quad \left. - \sum_{J \in \mathcal{C}_n, J \subset (-\infty, x]} F_n(J) - F_n(J \cap (-\infty, x]) \right| \\
&= |F_{n+k}(J \cap (-\infty, x]) - F_n(J \cap (-\infty, x])| \\
&\leq 2F_n(J) \\
&= 2\left(\frac{3}{2}\right)^n \mu(J) \\
&= 2\frac{1}{2^n}
\end{aligned}$$

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 $F : \mathbb{R} \rightarrow [0, 1]$ 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. $\square$

Outros exemplos selvagens

6.22 Exemplo

Variável aleatória uniforme nos irracionais

$$f(x) = \begin{cases} 1 & \text{para } x \in [0,1] \setminus \mathbb{Q} \\ 0 & \text{para } x \in [0,1] \cap \mathbb{Q} \end{cases}$$

◁

6.23 Exemplo

Dado $\{r_k\}$ uma enumeração dos racionais e defina

$$p(r_k) = \begin{cases} \frac{6}{\pi^2 k^2} & \text{para } x \in \mathbb{Q} \\ 0 & \text{caso contrário} \end{cases}$$

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 λ e μ respectivamente. Prove que $X + Y$ é uma variável aleatória de Poisson com parâmetro $\lambda + \mu$.
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 $(\mathbb{R}, \mathcal{B}, \mathbf{P})$ é um espaço de probabilidade.

Ex. 6.3 — Se uma função F satisfizer

□ F é não decrescente e $0 \leq F(x) \leq 1$.

□ $F(x) \rightarrow 0$ quando $x \rightarrow -\infty$ e $F(x) \rightarrow 1$ quando $x \rightarrow +\infty$.

□ F é contínua a direita, ou seja

$$\lim_{y \rightarrow x_+} F(y) = F(x).$$

então F é a função de distribuição de alguma variável aleatória X .

Ex. 6.4 — Suponha que $Y_1, Y_2, \dots$ 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 $n \rightarrow \infty$ onde $U_n := \sum_{k=1}^n 2^{-k} Y_k$.

Ex. 6.5 — Se

$$X = X^+ - X^-, \quad Y = Y^+ - Y^-.$$

Prove que

$$X \stackrel{d}{=} Y \text{ implica } X^+ \stackrel{d}{=} Y^+ \text{ e } X^- \stackrel{d}{=} Y^-.$$

*** 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 mutuamente 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 :

$$\square F_1^{-1}(x) := \sup\{x : F_X(x) < t\}, \quad t \in [0,1],$$

$$\square F_2^{-1}(x) := \inf\{x : F_X(x) \geq t\}, \quad t \in [0,1],$$

$$\square F_3^{-1}(x) := \inf\{x : F_X(x) > t\}, \quad t \in [0,1],$$

$$\square F_4^{-1}(x) := \sup\{x : F_X(x) \leq t\}, \quad t \in [0,1],$$

Valor Esperado

7.1 Momentos

No que se segue $(\Omega, \mathcal{F}, \mathbf{P})$ denotará um espaço de probabilidade.

7.1 Definição (Valor Esperado ou Esperança)

Se X é uma variável aleatória em $(\Omega, \mathcal{F}, \mathbf{P})$ integrável então definimos seu **valor esperado** ou **esperança** como

$$\mathbf{E}[X] := \int_{\Omega} X d\mathbf{P}$$

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.

De modo intuitivo, o valor esperado o valor médio "ponderado" de acordo com as probabilidades.

As seguintes propriedades da esperança seguem das propriedades da integral de Lebesgue:

7.2 Teorema

Suponha que X, Y sejam variáveis aleatórias integráveis. Então

- (Linearidade) $\mathbf{E}[aX + bY] = a\mathbf{E}[X] + b\mathbf{E}[Y]$ para quaisquer números reais a, b .
- $\mathbf{E}[b] = b$ para qualquer número real b .
- Se $X \geq 0$ então $\mathbf{E}[X] \geq 0$.
- (Monotonicidade) Se $X \geq Y$ então $\mathbf{E}[X] \geq \mathbf{E}[Y]$.
- (Desigualdade triangular) $|\mathbf{E}[X]| \leq \mathbf{E}[|X|]$
- Se $X_n \uparrow X$ então $\mathbf{E}[X] = \lim_{n \rightarrow \infty} \mathbf{E}[X_n]$

7.3 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 $X_1 + \dots + X_8$ é o número de vértices vermelhos. Por linearidade da esperança,

$$\mathbf{E}[X_1 + \dots + X_8] = \mathbf{E}[X_1] + \dots + \mathbf{E}[X_8] = 8 \cdot 0.9 = 7.2.$$

Para que o número médio de vértices seja 7.2, deve haver algum cubo com mais de 7 vértices vermelhos, 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.

7.4 Teorema (Caracterização da Esperança)

A esperança é o único operador $\mathbf{E} : \mathcal{L}^1(\Omega) \rightarrow \mathbb{R}$ que satisfaz:

- Linearidade. Para $a, b \in \mathbb{R}$, $\mathbf{E}[aX + bY] = a\mathbf{E}[X] + b\mathbf{E}[Y]$.
- Continuidade. Se $X_n \rightarrow X$ então $\mathbf{E}[X_n] \rightarrow \mathbf{E}[X]$
- 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

$$\mathbf{E}_1[X] = \lim \mathbf{E}_1[X_n] = \lim \mathbf{E}[X_n] = \mathbf{E}[X]$$

Para funções quaisquer. Consideramos a decomposição $X = X^+ + X^-$ e usamos a linearidade

◁

Para variáveis aleatórias simples a esperança se reduz a uma soma simples. Se

$$X = \sum_{k=1}^n x_k \mathbb{1}_{A_k}$$

Ou seja

$$X = x_k \text{ com probabilidade } p_k = \mathbf{P}(A_k), \quad k = 1, \dots, n.$$

Então, a definição de integral Lebesgue para funções simples nos fornece

$$\mathbf{E}[X] = \sum_{k=1}^n x_k p_k,$$

ou de outra forma

$$\sum_{k=1}^n x_k \mathbf{P}(x = x_k).$$

7.5 Teorema

Para variáveis aleatórias discretas,

$$\mathbf{E}[X] = \sum_{k=1}^{\infty} x_k \mathbf{P}(x = x_k), \quad (7.1)$$

se a série é absolutamente convergente

Demonstração. Exercício

◁

A esperança também pode ser definida para vetores aleatórios. Dado um vetor aleatório $X = (X_1, \dots, X_n)$ a **esperança do vetor aleatório** é definida como

$$\mathbf{E}[X] = (\mathbf{E}[X_1], \dots, \mathbf{E}[X_n])$$

7.6 Teorema (Distribuição Determina a Esperança)

Dado X integrável então se $X \stackrel{d}{=} Y$ então Y é integrável e $\mathbf{E}[X] = \mathbf{E}[Y]$.

Demonstração.] **Etapla 1:** provaremos para variáveis aleatórias simples X e Y . Esse fato é imediato de (7.1).

Etapla 2: provaremos para variáveis aleatórias limitadas. Suponha $X, Y \geq 0$, para $N \in \mathbb{N}$, defina as variáveis aleatórias, já usadas na demonstração de 5.2

$$X_N = \sum_{k=1}^{2^{2N}+1} \frac{k-1}{2^N} \mathbb{1}_{A_{N,k}} \text{ e } Y_N = \sum_{k=1}^{2^{2N}+1} \frac{k-1}{2^N} \mathbb{1}_{B_{N,k}}$$

onde

$$A_{N,k} = X^{-1}(I_{N,k}) \text{ para } k = 1, 2, \dots, 2^{2N} + 1.$$

$$B_{N,k} = Y^{-1}(I_{N,k}) \text{ para } k = 1, 2, \dots, 2^{2N} + 1.$$

Essas duas sequências de variáveis aleatórias tomam apenas um número finito de valores, são independentes e $X_N \uparrow X$ e $Y_N \uparrow Y$.

A parte mais interessante é que as esperanças concordam: $E[X_n] = E[Y_n]$ pois a igualdade em distribuição implica que $\mu(A_{N,k}) = \mu(B_{N,k})$. Desses fatos a igualdade das esperanças segue de imediato.

Pelo Teorema da Convergência Dominada,

$$E[X_n] \rightarrow E[X], \quad E[Y_n] \rightarrow E[Y],$$

porque $X_n \geq 0, X_n \nearrow X$.

Etapa 3: provaremos para variáveis aleatórias não-negativas $X \geq 0, Y \geq 0$. Para tanto

$$E[X] = \sup\{E[X'] : X' \leq X \text{ e } X' \text{ é limitada}\},$$

Pela etapa 2, estes supremos são iguais.

Etapa 4: Provaremos para X e Y quaisquer. Começaremos considerando a decomposição

$$X = X^+ - X^-, \quad Y = Y^+ - Y^-.$$

Mas como (veja exercício 7.3)

$$X \stackrel{d.}{=} Y \text{ implica } X^+ \stackrel{d.}{=} Y^+ \text{ e } X^- \stackrel{d.}{=} Y^-.$$

Então, usamos a etapa 3 nas partes positiva e negativa, respectivamente. ◁

7.7 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

$$m_k := E[X^k] \quad M_k := E[|X|^k] \quad \text{para } k = 1, \dots, n$$

são denominados o k -ésimo momento e k -ésimo momento absoluto de X respectivamente.

- Se X é quadrado integrável, então

$$\mathbf{Var}[X] := E[X^2] - E[X]^2$$

é a variância de X . O número $\sigma := \sqrt{\mathbf{Var}[X]}$ é denominado variância de X .

- Se X e Y são quadrado integráveis definimos a covariância de X e Y como

$$\mathbf{Cov}[X, Y] := E[(X - E[X])(Y - E[Y])]$$

As variáveis X e Y são ditas não correlacionadas se $\mathbf{Cov}[X, Y] = 0$

7.8 Teorema (Variáveis Aleatórias Independentes são Não Correlacionadas)

Sejam X, Y são variáveis aleatórias independentes integráveis. Então (XY) é integrável e

$$E[XY] = E[X]E[Y]$$

Em particular, variáveis aleatórias independentes não estão correlacionadas.

Demonstração. Assuma primeiro que as variáveis X e Y assumem um número finito de valores. Então XY também toma apenas um número finito de valores e XY é integrável.

$$\mathbf{E}[XY] = \sum_{z \in \mathbb{R}^*} z \mathbf{P}[XY = z] \quad (7.2)$$

$$= \sum_{z \in \mathbb{R}^*} \sum_{x \in \mathbb{R}^*} x \frac{z}{x} \mathbf{P}[X = x, Y = z/x] \quad (7.3)$$

$$= \sum_{y \in \mathbb{R}^*} \sum_{x \in \mathbb{R}^*} xy \mathbf{P}[X = x, Y = y] \quad (7.4)$$

$$= \sum_{y \in \mathbb{R}^*} \sum_{x \in \mathbb{R}^*} xy \mathbf{P}[X = x] \mathbf{P}[Y = y] \quad (7.5)$$

independência

$$= \sum_{x \in \mathbb{R}^*} x \mathbf{P}[X = x] \sum_{y \in \mathbb{R}^*} y \mathbf{P}[Y = y] \quad (7.6)$$

$$= \sum_{x \in \mathbb{R}^*} x \mathbf{P}[X = x] \sum_{y \in \mathbb{R}^*} y \mathbf{P}[Y = y] \quad (7.7)$$

$$(7.8)$$

Agora suponha $X, Y \geq 0$, para $N \in \mathbb{N}$, defina as variáveis aleatórias

$$X_N = \sum_{k=1}^{2^{2N}+1} \frac{k-1}{2^N} \mathbb{1}_{A_{N,k}} \text{ e } Y_N = \sum_{k=1}^{2^{2N}+1} \frac{k-1}{2^N} \mathbb{1}_{B_{N,k}}$$

Essas funções já foram usadas na Proposição 5.2.

Essas duas sequências de variáveis aleatórias tomam apenas um número finito de valores, são independentes e $X_N \uparrow X$ e $Y_N \uparrow Y$. Pelo Teorema da Convergência Dominada:

$$\mathbf{E}[XY] = \lim_{N \rightarrow \infty} \mathbf{E}[X_N Y_N] = \lim_{N \rightarrow \infty} \mathbf{E}[X_N] \mathbf{E}[Y_N] < \infty$$

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 independentes.

◁

7.9 Proposição (Propriedades da Variância)

Dado X uma variável aleatória quadrado integrável. Então:

1 $\mathbf{Var}[X] = \mathbf{E}[(X - \mathbf{E}[X])^2] \geq 0.$

2 $\mathbf{Var}[X] = 0 \iff X = \mathbf{E}[X]$ quase certamente.

3 A aplicação $f : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto \mathbf{E}[(X - x)^2]$ tem um mínimo em $\mathbf{E}[X]$ com $f(\mathbf{E}[X]) = \mathbf{Var}[X].$

Demonstração.

1 É consequência direta da linearidade da esperança.

2 observe que $\mathbf{E}[(X - \mathbf{E}[X])^2] = 0$ se e somente se $(X - \mathbf{E}[X])^2 = 0$ quase certamente.

3 Imediato.

<

7.10 Proposição

A aplicação $\mathbf{Cov} : \mathcal{L}^2(\mathbf{P}) \times \mathcal{L}^2(\mathbf{P}) \rightarrow \mathbb{R}$ 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.

<

7.11 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 houver $a, b, c \in \mathbb{R}$ com $|a| + |b| + |c| > 0$ e tal que $aX + bY + c = 0$ q.c.

Demonstração. A desigualdade de Cauchy-Schwarz é verdadeira para qualquer forma bilinear positiva semidefinida e, portanto, em particular para o mapa de covariância.

Faremos a demonstração supondo que $\mathbf{Var}[Y] > 0$. O caso no qual $\mathbf{Var}[Y] = 0$ será deixada como exercício.

Como $\mathbf{Var}[Y] > 0$, deixe $\theta := \frac{-\mathbf{Cov}[X, Y]}{\mathbf{Var}[Y]}$. Então

$$0 \leq \mathbf{Var}[X + \theta Y]\mathbf{Var}[Y] \tag{7.9}$$

$$= (\mathbf{Var}[X] + 2\theta\mathbf{Cov}[X, Y] + \theta^2\mathbf{Var}[Y])\mathbf{Var}[Y] \tag{7.10}$$

$$= \mathbf{Var}[X]\mathbf{Var}[Y] - \mathbf{Cov}[X, Y]^2 \tag{7.11}$$

com igualdade se e somente se $X + \theta Y$ for constante q.c.

Para a última parte tome $a = 1$, $b = \theta$ e $c = -\mathbf{E}[X] - b\mathbf{E}[Y]$.

<

7.2 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.

7.12 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 2.2.

7.13 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 $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbf{P}_F)$.

7.14 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.

$$\mathbf{E}_F[g] := \sum_{j=1}^m b_j \mathbf{P}_X(B_j) = \sum_{j=1}^m b_j \sum_i p_i \mathbb{1}_{t_i \in B_j} \quad (7.12)$$

$$= \sum_i p_i \sum_{j=1}^m b_j \mathbb{1}_{B_j}(t_i) \quad (7.13)$$

$$= \sum_i p_i g(t_i) \quad (7.14)$$

Para provar para as funções mensuráveis $g \geq 0$ basta utilizar o Teorema da Convergência Monótona. Se $g \geq 0$ e $g_n \uparrow g$

$$\mathbf{E}_F[g] = \lim_{n \rightarrow \infty} \mathbf{E}_F[g_n] = \lim_{n \rightarrow \infty} \sum_i p_i g_n(t_i) = \sum_i p_i g(t_i)$$

◁

7.15 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

$$\mathbf{E}_F[g] := \sum_{j=1}^m b_j \mathbf{P}(B_j) = \sum_{j=1}^m b_j \int_{B_j} f(x) dx \quad (7.15)$$

$$= \sum_{j=1}^m b_j \int_{-\infty}^{\infty} f(x) \mathbb{1}_{B_j} dx \quad (7.16)$$

$$= \int_{-\infty}^{\infty} f(x) \left[\sum_{j=1}^m b_j \mathbb{1}_{B_j} \right] dx \quad (7.17)$$

$$= \int_{-\infty}^{\infty} f(x) g(x) dx \quad (7.18)$$

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. $\triangleleft$

Utilizando as Proposições 7.14 e 7.15 podemos integrar com respeito a uma distribuição que seja combinação de uma discreta e uma absolutamente contínua.

7.16 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 $F_X(t) = \sum_i \mathbf{P}(A_i) \mathbb{1}_{a_i \leq t}$ 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. $\triangleleft$

7.17 Exemplo

Se $X \sim \mathcal{N}(0; 1)$ então $\mathbf{E}(X) = 0$.

$$\mathbf{E}[X] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-\frac{x^2}{2}} dx = -\frac{1}{\sqrt{2\pi}} \int_{*}^{*} e^u du = -\frac{1}{\sqrt{2\pi}} e^{-x^2/2} \Big|_{-\infty}^{\infty} = 0.$$

<

7.18 Exemplo

Se $X \sim \text{Exp}(\lambda)$ então $E[X^n] = \frac{n!}{\lambda^n}$.

<

7.3 Desigualdades de Markov e Chebyshev**7.19 Teorema (Desigualdade de Markov)**

Suponha que $X \geq 0$ seja uma variável aleatória. Então,

$$P(X > x) \leq \frac{E[X]}{x} \text{ para todo } x > 0. \quad (7.19)$$

Demonstração. Começamos definindo $Y := x \mathbb{1}_{\{X \geq x\}}$. Assim $Y \leq X$ pontualmente. Assim sendo,

$$E[Y] \leq E[X]. \quad (7.20)$$

e assim

$$E[Y] = x \cdot P(X \geq x). \quad (7.21)$$

Claramente as equações (7.20) e (7.21) implicam a desigualdade de Markov. <

A estimativa $\frac{1}{x}$ possui um grave defeito não é integrável.

7.20 Teorema (Desigualdade de Chebyshev)

Suponha que $X \geq 0$ seja uma variável aleatória e $p > 0$. Então

$$P(X \geq x) \leq \frac{E[x^p]}{x^p} \text{ para todo } x > 0. \quad (7.22)$$

Para que essa desigualdade não seja trivial precisamos que $X \in \mathcal{L}^p(\mathbf{P})$. Assim na desigualdade de Chebyshev fazemos exigência fortes sobre X mas somos recompensado com um lado direito integrável.

Demonstração. A prova é uma redução fácil para a desigualdade de Markov: $P(X \geq x) = P(X^p \geq x^p) \leq \frac{E[X^p]}{x^p}$, pela desigualdade de Markov. <

Podemos generalizar um pouco mais

7.21 Teorema (Desigualdade de Markov Estendida)

Se φ é uma função monótona crescente definida nos reais não positivos, e X é uma variável aleatória e $a \geq 0$, e $\varphi(a) > 0$, então

$$P(|X| \geq a) \leq \frac{E(\varphi(|X|))}{\varphi(a)}$$

Um dos corolários da desigualdade de Chebyshev é:

7.22 Corolário (Variância)

Suponha que X seja uma variável aleatória com média μ e variância σ^2 . Então

$$\mathbf{P}(|X - \mu| \geq t\sigma) \leq \frac{1}{t^2} \text{ para qualquer } t > 0. \quad (7.23)$$

Na literatura o último corolário é também conhecido como desigualdade de Chebyshev.

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,

$$\mathbf{P}(|X - \mu| \geq t\sigma) \leq \frac{\sigma^2}{(t\sigma)^2} = \frac{1}{t^2}.$$

◁

Em particular, para $t = \sqrt{2}$, cada variável aleatória X com variância σ^2 está dentro de $\sqrt{2} \cdot \sigma$ de sua média, com probabilidade $\geq \frac{1}{2}$.

7.4 Desigualdade Maximal de Kolmogorov

7.23 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 $S_n = \sum_{l=1}^n X_l$. Então para $\lambda > 0$,

$$\lambda^2 \mathbf{P}\left(\max_{1 \leq k \leq n} |S_k| \geq \lambda\right) \leq \mathbf{Var}(S_n)$$

Demonstração.

Observe que o evento $\{\max_{1 \leq k \leq n} |S_k| \geq \lambda\}$ é a união disjunta dos eventos

$A_k = \{|S_k| \geq \lambda, |S_j| < \lambda \text{ para } j < k\}$, $k = 1, \dots, n$. Além disso, temos

$$\begin{aligned} \mathbf{P}(A_k) &\leq \lambda^{-2} \mathbf{E}[S_k^2 \mathbb{1}_{A_k}] \\ &\leq \lambda^{-2} \mathbf{E}[(S_k^2 + (S_n - S_k)^2) \mathbb{1}_{A_k}] \\ &= \lambda^{-2} \mathbf{E}[(S_k^2 + 2S_k(S_n - S_k) + (S_n - S_k)^2) \mathbb{1}_{A_k}], \end{aligned}$$

onde a última linha segue, pois as variáveis $S_k \mathbb{1}_{A_k}$ e $S_n - S_k$ são independentes e $\mathbf{E}[S_n - S_k] = 0$. Assim, temos

$$\mathbf{P}(A_k) \leq \lambda^{-2} \mathbf{E}[(S_k + (S_n - S_k))^2 \mathbb{1}_{A_k}] = \lambda^{-2} \mathbf{E}[S_n^2 \mathbb{1}_{A_k}]$$

e assim somando que encontramos

$$\mathbf{P}\left(\max_{1 \leq k \leq n} |S_k| \geq \lambda\right) \leq \lambda^{-2} \mathbf{E}[S_n^2 \mathbb{1}_{\max_{1 \leq k \leq n} |S_k| \geq \lambda}] \leq \lambda^{-2} \mathbf{E}[S_n^2].$$

◁

7.5 Outras Desigualdades

7.24 Proposição (Desigualdade de Jensen)

Dado $(\Omega, \mathcal{F}, \mathbf{P})$ um espaço de probabilidade. Se X é uma variável aleatória que é integrável, e $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ 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

$$(t - t_0)\Phi \leq \varphi(t) - \varphi(t_0) \quad (7.24)$$

Agora, deixe $t = X(x)$ e $t_0 = \mathbf{E}[X]$ e a equação 7.24 se torna

$$(X(x) - \mathbf{E}[X])\Phi \leq \varphi(X(x)) - \varphi(\mathbf{E}[X]) \quad (7.25)$$

Integrando ambos os lados de 7.25 e usando que é medida de probabilidade temos

$$(\mathbf{E}[X] - \mathbf{E}[X])\Phi \leq \mathbf{E}[\varphi(X) - \varphi(\mathbf{E}[X])]$$

que ao reorganizar, torna-se

$$\varphi(\mathbf{E}[X]) \leq \mathbf{E}[\varphi(X)]$$

◁

7.25 Proposição (Desigualdade de Hölder)

Dado um espaço de probabilidade $(\Omega, \mathcal{F}, \mathbb{P})$, e X, Y variáveis aleatórias

$$\mathbf{E}[|XY|] \leq (\mathbf{E}[|X|^p])^{1/p} (\mathbf{E}[|Y|^q])^{1/q}.$$

Demonstração. Considere a medida de probabilidade $\nu := \frac{g^q}{\int_{\Omega} g^q d\mu} \cdot \mu$ e a função $h := \frac{f}{g^{q-1}}$.

Então por Jensen

$$\int_{\Omega} f g d\mu = \int_{\Omega} h g^q d\mu = \int_{\Omega} g^q d\mu \cdot \int_{\Omega} h d\nu \leq \int_{\Omega} g^q d\mu \left(\int_{\Omega} h^p d\nu \right)^{1/p} = \left(\int_{\Omega} g^q d\mu \right)^{1/q} \left(\int_{\Omega} f^p d\mu \right)^{1/p}.$$

◁

7.26 Proposição (Desigualdade de Minkowski)

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p \quad (1 \leq p \leq \infty)$$

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,

$$\left| \frac{1}{2}f + \frac{1}{2}g \right|^p \leq \left| \frac{1}{2}f + \frac{1}{2}g \right|^p \leq \frac{1}{2}|f|^p + \frac{1}{2}|g|^p.$$

Isso significa que

$$|f + g|^p \leq \frac{1}{2}|2f|^p + \frac{1}{2}|2g|^p = 2^{p-1}|f|^p + 2^{p-1}|g|^p.$$

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

$$\|f + g\|_p^p = \int_{\Omega} |f + g|^p d\mu \quad (7.26)$$

$$= \int_{\Omega} |f + g| \cdot |f + g|^{p-1} d\mu \quad (7.27)$$

$$\leq \int_{\Omega} (|f| + |g|)|f + g|^{p-1} d\mu \quad (7.28)$$

$$= \int_{\Omega} |f||f + g|^{p-1} d\mu + \int_{\Omega} |g||f + g|^{p-1} d\mu \quad (7.29)$$

$$\leq \left(\left(\int_{\Omega} |f|^p d\mu \right)^{\frac{1}{p}} + \left(\int_{\Omega} |g|^p d\mu \right)^{\frac{1}{p}} \right) \left(\int_{\Omega} |f + g|^{(p-1)\left(\frac{p}{p-1}\right)} d\mu \right)^{1-\frac{1}{p}} \quad (7.30)$$

$$(7.31)$$

(Pela Desigualdade de Hölder)

$$(7.32)$$

$$= (\|f\|_p + \|g\|_p) \frac{\|f + g\|_p^p}{\|f + g\|_p} \quad (7.33)$$

Obtemos a desigualdade de Minkowski multiplicando ambos os lados por

$$\frac{\|f + g\|_p}{\|f + g\|_p^p}.$$

◁

7.6 Funções Geradoras de Probabilidade

Considere uma variável aleatória X , tomando valores nos naturais. Seja $p_r = \mathbf{P}(X = r)$.

7.27 Definição (Função Geradora de Probabilidade)

A **função geradora de probabilidade** de X é definida como

$$p(z) = \mathbf{E}[z^X] = \sum_{r=0}^{\infty} \mathbf{P}(X = r) z^r = p_0 + p_1 z + p_2 z^2 + \cdots = \sum_{r=0}^{\infty} p_r z^r.$$

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.

7.28 Exemplo

Considere um dado justo. Então $p_r = 1/6$ para $r = 1, \dots, 6$. assim

$$p(z) = \mathbf{E}[z^X] = \frac{1}{6}(z + z^2 + \cdots + z^6) = \frac{1}{6}z \left(\frac{1 - z^6}{1 - z} \right).$$

◁

7.29 Exemplo

Neste exemplo, calculamos a função geradora de probabilidade para a distribuição de Poisson de parâmetro α . Com base na definição, temos:

$$p(z) = \sum_{j=0}^{\infty} \frac{e^{-\alpha} \alpha^j}{j!} z^j \tag{7.34}$$

$$= \sum_{j=0}^{\infty} \frac{e^{-\alpha} (\alpha z)^j}{j!} \tag{7.35}$$

$$= \frac{e^{-\alpha}}{e^{-\alpha z}} \sum_{j=0}^{\infty} \frac{e^{-\alpha z} (\alpha z)^j}{j!} \tag{7.36}$$

$$= e^{\alpha(z-1)} \tag{7.37}$$

◁

7.30 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 $(p_0, p_1, \dots)$ de $p(z)$.

◁

7.31 Teorema (Lema de Abel)

$$\mathbf{E}[X] = \lim_{z \rightarrow 1} p'(z).$$

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

$$\lim_{z \rightarrow 1} p'(z) \leq \mathbf{E}[X].$$

Por outro lado, para qualquer ε , escolhendo N suficientemente grande, temos

$$\sum_1^N r p_r \geq \mathbf{E}[X] - \varepsilon.$$

assim

$$\mathbf{E}[X] - \varepsilon \leq \sum_1^N r p_r = \lim_{z \rightarrow 1} \sum_1^N r p_r z^{r-1} \leq \lim_{z \rightarrow 1} \sum_1^{\infty} r p_r z^{r-1} = \lim_{z \rightarrow 1} p'(z).$$

Portanto, $\mathbf{E}[X] \leq \lim_{z \rightarrow 1} p'(z)$. ◁

7.32 Teorema

$$\mathbf{E}[X(X-1)] = \lim_{z \rightarrow 1} p''(z).$$

Demonstração. Exercício ◁

7.33 Exemplo

Considere a distribuição de Poisson. Então

$$p_r = \mathbf{P}(X = r) = \frac{1}{r!} \lambda^r e^{-\lambda}.$$

Então

$$p(z) = \mathbf{E}[z^X] = \sum_0^{\infty} z^r \frac{1}{r!} \lambda^r e^{-\lambda} = e^{\lambda z} e^{-\lambda} = e^{\lambda(z-1)}.$$

Temos

$$\mathbf{E}[X] = \left. \frac{d}{dz} e^{\lambda(z-1)} \right|_{z=1} = \lambda,$$

e

$$\mathbf{E}[X(X-1)] = \left. \frac{d^2}{dz^2} e^{\lambda(z-1)} \right|_{z=1} = \lambda^2$$

assim

$$\text{Var}(X) = \mathbf{E}[X^2] - \mathbf{E}[X]^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

<

7.34 Teorema

Suponha que $X_1, X_2, \dots, X_n$ sejam variáveis aleatórias independentes com funções geradoras de probabilidade $p_1, p_2, \dots, p_n$. Então, a função geradora de probabilidade de $X_1 + X_2 + \dots + X_n$ é $p_1(z)p_2(z) \cdots p_n(z)$.

Demonstração.

$$\mathbf{E}[z^{X_1 + \dots + X_n}] = \mathbf{E}[z^{X_1} \cdots z^{X_n}] = \mathbf{E}[z^{X_1}] \cdots \mathbf{E}[z^{X_n}] = p_1(z) \cdots p_n(z).$$

<

7.35 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:

$$\mathbf{P}(X + Y = r) = \sum_{i=0}^r \mathbf{P}(X = i, Y = r - i) = \sum_{i=0}^r \mathbf{P}(X = i) \mathbf{P}(Y = r - i),$$

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.

7.36 Exemplo

Sejam $X_1, X_2, \dots, X_n$ 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

$$S = X_1 + \cdots + X_N?$$

$$\begin{aligned}
 \mathbf{E}[z^S] &= \mathbf{E}[z^{X_1 + \cdots + X_N}] \\
 &= \mathbf{E}_N[\underbrace{\mathbf{E}_{X_i}[z^{X_1 + \cdots + X_N} \mid N]}_{\text{assumindo fixo } N}] \\
 &= \sum_{n=0}^{\infty} \mathbf{P}(N = n) \mathbf{E}[z^{X_1 + X_2 + \cdots + X_n}] \\
 &= \sum_{n=0}^{\infty} \mathbf{P}(N = n) \mathbf{E}[z^{X_1}] \mathbf{E}[z^{X_2}] \cdots \mathbf{E}[z^{X_n}] \\
 &= \sum_{n=0}^{\infty} \mathbf{P}(N = n) (\mathbf{E}[z^{X_1}])^n \\
 &= \sum_{n=0}^{\infty} \mathbf{P}(N = n) p(z)^n \\
 &= h(p(z))
 \end{aligned}$$

como $h(x) = \sum_{n=0}^{\infty} \mathbf{P}(N = n)x^n$.

assim

$$\begin{aligned}
 \mathbf{E}[S] &= \left. \frac{d}{dz} h(p(z)) \right|_{z=1} \\
 &= h'(p(1))p'(1) \\
 &= \mathbf{E}[N] \mathbf{E}[X_1]
 \end{aligned}$$

Para calcular a variância, use o fato de que

$$\mathbf{E}[S(S-1)] = \left. \frac{d^2}{dz^2} h(p(z)) \right|_{z=1}.$$

Então podemos encontrar que

$$\text{Var}(S) = \mathbf{E}[N] \text{Var}(X_1) + \mathbf{E}[X_1^2] \text{Var}(N).$$

◁

7.7 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 $X_0, X_1, \dots$, 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

$$X_{n+1} = Y_1^n + Y_2^n + \dots + Y_{X_n}^n,$$

onde Y_i^n são variáveis aleatórias independentes e identicamente distribuídas, como X_1 .

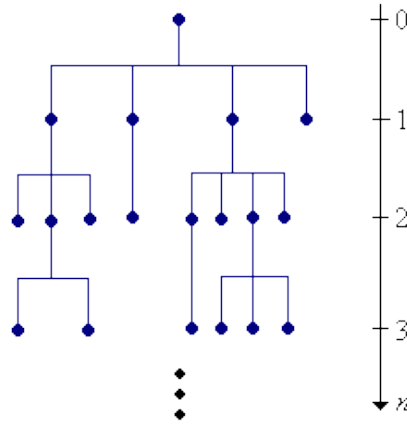


Figura 7.1: Processo de ramificação [14]

Como veremos é útil considerar a 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

7.37 Teorema

$$F_{n+1}(z) = F_n(F(z)) = F(F(F(\dots F(z) \dots))) = F(F_n(z)).$$

Demonstração.

$$\begin{aligned}
 F_{n+1}(z) &= \mathbf{E}[z^{X_{n+1}}] \\
 &= \mathbf{E}[\mathbf{E}[z^{X_{n+1}} | X_n]] \\
 &= \sum_{k=0}^{\infty} \mathbf{P}(X_n = k) \mathbf{E}[z^{X_{n+1}} | X_n = k] \\
 &= \sum_{k=0}^{\infty} \mathbf{P}(X_n = k) \mathbf{E}[z^{Y_1^n + \dots + Y_k^n} | X_n = k] \\
 &= \sum_{k=0}^{\infty} \mathbf{P}(X_n = k) \mathbf{E}[z^{Y_1}] \mathbf{E}[z^{Y_2}] \dots \mathbf{E}[z^{Y_n}] \\
 &= \sum_{k=0}^{\infty} \mathbf{P}(X_n = k) (\mathbf{E}[z^{X_1}])^k \\
 &= \sum_{k=0}^{\infty} \mathbf{P}(X_n = k) F(z)^k \\
 &= F_n(F(z))
 \end{aligned}$$

◁

7.38 Teorema

Suponha que

$$\mathbf{E}[X_1] = \sum k p_k = \mu$$

e

$$\text{Var}(X_1) = \mathbf{E}[(X - \mu)^2] = \sum (k - \mu)^2 p_k < \infty.$$

Então

$$\mathbf{E}[X_n] = \mu^n, \quad \text{Var}X_n = \sigma^2 \mu^{n-1} (1 + \mu + \mu^2 + \dots + \mu^{n-1}).$$

Demonstração.

$$\begin{aligned}
 \mathbf{E}[X_n] &= \mathbf{E}[\mathbf{E}[X_n | X_{n-1}]] \\
 &= \mathbf{E}[\mu X_{n-1}] \\
 &= \mu \mathbf{E}[X_{n-1}]
 \end{aligned}$$

por indução, $\mathbf{E}[X_n] = \mu^n$ (pois $X_0 = 1$).

Para calcular a variância, observe que

$$\text{Var}(X_n) = \mathbf{E}[X_n^2] - (\mathbf{E}[X_n])^2$$

e, portanto

$$\mathbf{E}[X_n^2] = \text{Var}(X_n) + (\mathbf{E}[X_n])^2$$

Então, calculamos

$$\begin{aligned}
 \mathbf{E}[X_n^2] &= \mathbf{E}[\mathbf{E}[X_n^2 \mid X_{n-1}]] \\
 &= \mathbf{E}[\text{Var}(X_n) + (\mathbf{E}[X_n])^2 \mid X_{n-1}] \\
 &= \mathbf{E}[X_{n-1} \text{Var}(X_1) + (\mu X_{n-1})^2] \\
 &= \mathbf{E}[X_{n-1} \sigma^2 + (\mu X_{n-1})^2] \\
 &= \sigma^2 \mu^{n-1} + \mu^2 \mathbf{E}[X_{n-1}^2].
 \end{aligned}$$

assim

$$\begin{aligned}
 \text{Var} X_n &= \mathbf{E}[X_n^2] - (\mathbf{E}[X_n])^2 \\
 &= \mu^2 \mathbf{E}[X_{n-1}^2] + \sigma^2 \mu^{n-1} - \mu^2 (\mathbf{E}[X_{n-1}])^2 \\
 &= \mu^2 (\mathbf{E}[X_{n-1}^2] - \mathbf{E}[X_{n-1}]^2) + \sigma^2 \mu^{n-1} \\
 &= \mu^2 \text{Var}(X_{n-1}) + \sigma^2 \mu^{n-1} \\
 &= \mu^4 \text{Var}(X_{n-2}) + \sigma^2 (\mu^{n-1} + \mu^n) \\
 &= \dots \\
 &= \mu^{2(n-1)} \text{Var}(X_1) + \sigma^2 (\mu^{n-1} + \mu^n + \dots + \mu^{2n-3}) \\
 &= \sigma^2 \mu^{n-1} (1 + \mu + \dots + \mu^{n-1}).
 \end{aligned}$$

◁

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 $A_1 \subseteq A_2 \subseteq A_3 \subseteq \dots$, sabemos que

$$q = \mathbf{P}(A) = \lim_{n \rightarrow \infty} \mathbf{P}(A_n) = \lim_{n \rightarrow \infty} \mathbf{P}(X_n = 0).$$

Mas

$$\mathbf{P}(X_n = 0) = F_n(0),$$

Como $F_n(0) = \sum \mathbf{P}(X_n = k)z^k$. Logo

$$F(q) = F\left(\lim_{n \rightarrow \infty} F_n(0)\right) = \lim_{n \rightarrow \infty} F(F_n(0)) = \lim_{n \rightarrow \infty} F_{n+1}(0) = q.$$

assim

$$F(q) = q.$$

Alternativamente, usando a lei da probabilidade total

$$q = \sum_k \mathbf{P}(X_1 = k) \mathbf{P}(\text{extinção} \mid X_1 = k) = \sum_k p_k q^k = F(q),$$

onde a segunda igualdade vem do fato de que, para que toda a população se extinga, cada população individual deve se extinguir.

Isso significa que para encontrar a probabilidade de extinção, precisamos encontrar um ponto fixo de F .

7.39 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 q é a menor raiz, deixe α ser 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

$$q = \lim_{n \rightarrow \infty} F_n(0) \leq \alpha.$$

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. $\triangleleft$

7.8 Funções Geradoras de Momento

7.40 Definição

A **função geradora de momento** $\phi_X(t)$ é definida como

$$\phi_X(t) = \mathbf{E}[e^{tX}]$$

desde que $\mathbb{E}(e^{tX})$ exista no intervalo $(-h, h)$ para $h > 0$.

Para variáveis discretas ou absolutamente contínuas temos que

$$\phi_X(t) = \mathbf{E}[e^{tX}] = \begin{cases} \sum_i e^{tx_i} p(x_i) & \text{se } X \text{ é discreto} \\ \int_x e^{tx} f(x) dx & \text{se } X \text{ for contínua} \end{cases}$$

7.41 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:

$$\sum_{x_i=c} e^{x_i t} = e^{ct}.$$

$\triangleleft$

7.42 Teorema

Seja X uma variável aleatória que possua função geradora de momento. Então

$$M_X^{(n)}(t) = \frac{d^n}{dt^n} \mathbf{E}(e^{tX}) = \mathbf{E}\left(\frac{d^n}{dt^n} e^{tX}\right) = \mathbf{E}(X^n e^{tX}).$$

7.43 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.

$$\begin{aligned}\phi_{X+Y}(t) &= \mathbf{E}[e^{t(X+Y)}] \\ &= \mathbf{E}[e^{tX}e^{tY}] \\ &= \mathbf{E}[e^{tX}] \mathbf{E}[e^{tY}] \\ &= \phi_X(t)\phi_Y(t).\end{aligned}$$

◁

7.44 Teorema

A função geradora de momento determina de forma única a distribuição de probabilidade.

7.45 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.

$$\begin{aligned}\phi_Z(t) &= \mathbf{E}[e^{tX}] \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{tx} e^{-(x-\mu)^2/(2\sigma^2)} dx \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \exp\left(\frac{-(x^2 - 2\mu x + \mu^2 - 2\sigma^2 tx)}{2\sigma^2}\right) dx\end{aligned}$$

Completando o quadrado:

$$\begin{aligned}x^2 - 2\mu x + \mu^2 - 2\sigma^2 tx &= x^2 - 2(\mu + \sigma^2 t)x + \mu^2 \\ &= (x - (\mu + \sigma^2 t))^2 - (\mu + \sigma^2 t)^2 + \mu^2 \\ &= (x - (\mu + \sigma^2 t))^2 - \sigma^4 t^2 - 2\mu\sigma^2 t.\end{aligned}$$

Então voltando ao cálculo da função geradora de momento

$$\begin{aligned}\phi_Z(t) &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \exp\left(\frac{-((x - (\mu + \sigma^2 t))^2 - \sigma^4 t^2 - 2\mu\sigma^2 t)}{2\sigma^2}\right) dx \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{\sigma^4 t^2 + 2\mu\sigma^2 t}{2\sigma^2}\right) \int_{-\infty}^{\infty} \exp\left(\frac{-(x - (\mu + \sigma^2 t))^2}{2\sigma^2}\right) dx \\ &= \exp\left(\frac{\sigma^4 t^2 + 2\mu\sigma^2 t}{2\sigma^2}\right) \\ &= \exp(\mu t + \sigma^2 t^2/2)\end{aligned}$$

◁

7.46 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.

$$\begin{aligned}\phi_{Z_1+Z_2}(t) &= \phi_{Z_1}(t)\phi_{Z_2}(t) \\ &= \exp(\mu_1 t + \sigma_1^2 t^2/2) \exp(\mu_2 t + \sigma_2^2 t^2/2) \\ &= \exp((\mu_1 + \mu_2)t + (\sigma_1^2 + \sigma_2^2)t^2/2)\end{aligned}$$

<

7.9 Entropia**Informação**

Gostaríamos de desenvolver uma medida da informação que obtemos de observar a ocorrência de um evento de probabilidade p . Nossa primeira redução será ignorar quaisquer características específicas do evento e apenas observar se ou não aconteceu. Assim, pensamos no evento como a observação de um símbolo cuja probabilidade de ocorrência é p .

A abordagem que vamos adotar é axiomática: Queremos que a medida de informação $I(p)$ tenha as seguintes propriedades

7.47 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** $I(p^2) = I(p \cdot p) = I(p) + I(p) = 2 \cdot I(p)$
- 2** Assim, além disso, $I(p^n) = n \cdot I(p)$
Epor indução

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

$$I(p^{n/m}) = \frac{n}{m} \cdot I(p)$$

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,

1 $I(p) \geq 0$

2 $I(p_1 \cdot p_2) = I(p_1) + I(p_2)$

3 $I(p)$ é monótona e contínua em p

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 $\log_2(p)$.

7.48 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 $\{a_1, a_2, \dots, a_n\}$ e que alguma fonte está nos fornecendo um fluxo desses símbolos. Suponha ainda que a fonte emite os símbolos com probabilidades $\{p_1, p_2, \dots, p_n\}$, respectivamente. Por enquanto, também assumiremos que os símbolos são emitidos de forma independente

Qual é a quantidade média de informação que obtemos de cada símbolo que vemos no fluxo?

O que realmente queremos aqui é uma média ponderada. Se observarmos o símbolo a_i , nós obteremos $\log(1/p_i)$ informação dessa observação particular. A longo prazo, digamos N de observações, veremos aproximadamente $N \cdot p_i$ ocorrências do símbolo a_i . Assim, após N observações obteremos a informação total I de

$$I = \sum_{i=1}^n (N \cdot p_i) \cdot \log(1/p_i).$$

Mas, então, a média de informações que obtemos por símbolo observado será

$$\begin{aligned} I/N &= (1/N) \sum_{i=1}^n (N \cdot p_i) \cdot \log(1/p_i) \\ &= \sum_{i=1}^n p_i \cdot \log(1/p_i) \end{aligned} \quad (7.38)$$

Acima adotamos a convenção que defina $0 \cdot \log(1/0)$ é 0.

Isso nos leva a seguinte definição.

7.49 Definição

Seja uma distribuição de probabilidade $P = \{p_1, p_2, \dots, p_n\}$. Definimos **entropia** da distribuição P por:

$$H(P) = \sum_{i=1}^n p_i \cdot \log(1/p_i).$$

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.

$$H(P) = E[I(p)].$$

Em outras palavras, a entropia de uma distribuição de probabilidade é o valor esperado da informação da distribuição.

7.50 Teorema (Desigualdade de Gibbs)

Considere as distribuições de probabilidade, $P = \{p_1, p_2, \dots, p_n\}$ e $Q = \{q_1, q_2, \dots, q_n\}$, onde $p_i, q_i \geq 0$ e $\sum_i p_i = \sum_i q_i = 1$. Então

$$-\sum p_i \log q_i \geq -\sum p_i \log p_i$$

com igualdade somente quando $p_i = q_i$ para todos i .

Demonstração. A demonstração baseia-se na desigualdade $\log x \leq x - 1$.

$$\begin{aligned}
\sum_{i=1}^n p_i \ln \left(\frac{q_i}{p_i} \right) &\leq \sum_{i=1}^n p_i \left(\frac{q_i}{p_i} - 1 \right) = \sum_{i=1}^n (q_i - p_i) \\
&= \sum_{i=1}^n q_i - \sum_{i=1}^n p_i = 1 - 1 = 0,
\end{aligned} \tag{7.39}$$

◁

Podemos usar a desigualdade de Gibbs para encontrar a distribuição de probabilidade que maximiza a função entropia. Suponha que $P = \{p_1, p_2, \dots, p_n\}$ é uma distribuição de probabilidade. Nós temos

$$H(P) - \log(n) = \sum_{i=1}^n p_i \log(1/p_i) - \log(n) \tag{7.40}$$

$$= \sum_{i=1}^n p_i \log(1/p_i) - \log(n) \sum_{i=1}^n p_i \tag{7.41}$$

$$= \sum_{i=1}^n p_i \log(1/p_i) - \sum_{i=1}^n p_i \log(n) \tag{7.42}$$

$$= \sum_{i=1}^n p_i (\log(1/p_i) - \log(n)) \tag{7.43}$$

$$= \sum_{i=1}^n p_i (\log(1/p_i) + \log(1/n)) \tag{7.44}$$

$$= \sum_{i=1}^n p_i \log \left(\frac{1/n}{p_i} \right) \tag{7.45}$$

$$\leq 0, \tag{7.46}$$

com igualdade somente quando $p_i = \frac{1}{n}$ para todos i .

Exercícios

Ex. 7.1 — Mostre que para variáveis aleatórias discretas,

$$\mathbf{E}X = \sum_{k=1}^{\infty} x_k \mathbf{P}(X = x_k), \tag{7.47}$$

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 Lebesgue tal que $\mathbf{P}(X > Y) > 1/2$, mas $\mathbf{E}(X) < \mathbf{E}(Y)$.

Ex. 7.3 — Mostre que

$$X \stackrel{d}{=} Y \text{ implica } X^+ \stackrel{d}{=} Y^+ \text{ e } X^- \stackrel{d}{=} Y^-.$$

Ex. 7.4 — Suponha que $X_1, X_2, \dots$ é uma sequência de variáveis independentes em (Ω, F, P) . Mostre que as duas famílias $X_1, X_3, X_5, \dots$ e $X_2, X_4, X_6, \dots$ são independentes.

Ex. 7.5 — Deixe $X_1, X_2, \dots$ 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 ν , com N independente de todas as X_i . Deixe $S = X_1 + \dots + X_N = \sum_{i=1}^{\infty} X_i \mathbf{1}_{N \geq i}$. Calcule $\text{Var}(S)$.

Ex. 7.6 — Prove a Desigualdade de Cauchy-Schwarz: dadas duas variáveis aleatórias X e Y , temos

$$|\mathbb{E}XY| \leq \sqrt{\mathbb{E}[X^2]\mathbb{E}[Y^2]}, \quad (7.48)$$

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 $\text{Corr}(X, Y)$.
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 $\mathbb{E}[X]$ e $\text{Var}[X]$. (esse problema já apareceu numa forma ligeiramente diferente na Lista 2)

Ex. 7.9 — Se $\mathbb{E}|X_1| = \infty$ então $\sum_{n=0}^{\infty} \mathbb{P}(|X_1| > n)$ diverge.

Ex. 7.10 — Dada uma variável aleatória X , então, para todo $\epsilon > 0$, existe uma variável aleatória limitada X_ϵ , tal que $\mathbb{P}(X \neq X_\epsilon) < \epsilon$

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

$$\text{Var}(X) = P''(1) + P'(1) - P'(1)^2.$$

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

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\left|\frac{S_n}{n} - \frac{1}{2}\right| < \epsilon\right) = 1$$

para todo $\epsilon > 0$.

Ex. 7.14 — Demonstração Probabilística do Teorema de Weierstrass

Utilize a desigualdade de Chebyshev para mostrar que para toda função contínua $f : [0, 1] \rightarrow \mathbb{R}$,

$$\sum_{k=0}^n f\left(\frac{k}{n}\right) \binom{n}{k} x^k (1-x)^{n-k} \rightarrow f(x)$$

uniformemente em $x \in [0, 1]$ quando $n \rightarrow \infty$.

Dica: Sejam $X_1, X_2, \dots, X_n$ variáveis aleatórias independentes cada uma assumindo os valores 0 e 1 com probabilidade p e $1-p$ respectivamente. Seja $S_n = X_1 + X_2 + \dots + X_n$ 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:

$$\mathbb{P}(|X - \mu| \geq t\sigma) = \frac{1}{t^2}.$$

Ex. 7.16 — Deixe X ser uma variável aleatória não-negativa, tal que $\mathbb{E}X$ existe.

1. Mostre através de um exemplo que $\mathbb{E}(X^2)$ pode não existir.
2. Considere o truncamento $X_n = \min(X, n)$. Prove que para todo $p > 2$,

$$\sum_{n=1}^{\infty} n^{-p} \mathbb{E}(X_n^2) < \infty.$$

3. Prove a propriedade anterior para $p = 2$.

Ex. 7.17 — Deixe $X_1, \dots, X_n$ serem variáveis aleatórias reais independentes e deixe $S_k = X_1 + \dots + X_k$ para $k = 1, \dots, n$. Mostre que para $t > 0$ a desigualdade de Etemadi é válida:

$$\mathbb{P}\left[\max_k |S_k| \geq t\right] \leq 3 \max_k \mathbb{P}[|S_k| \geq t/3].$$

Dica: Considere os conjuntos

$$A_j := \left\{ \max_{1 \leq k < j} |S_k| < 3r, |S_j| \geq 3r \right\}, \quad j = 1, \dots, n$$

e observe que

$$\left\{ \max_{1 \leq j \leq n} |S_j| \geq 3r \right\} = \bigcup_{j=1}^n A_j.$$

Medida Produto

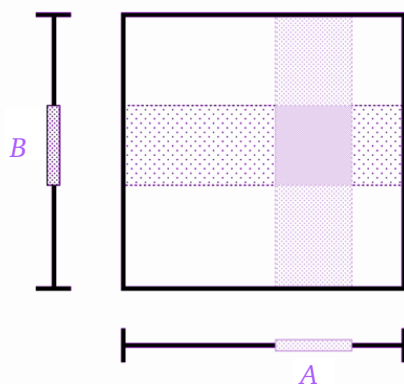
Nesse capítulo generalizamos o conceito de medidas para o produto de espaços. A integral que obtemos como subproduto dessas medidas são os análogos abstratos das integrais de várias variáveis e como tais podem ser calculadas como integrais iteradas.

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

$$\mathcal{F}_1 \otimes \mathcal{F}_2 = \sigma \{ \{A \times B : A \in \mathcal{F}_1, B \in \mathcal{F}_2\} \}$$

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

$$\pi_{\Omega_1} : \Omega_1 \times \Omega_2 \rightarrow \Omega_1 \quad \pi_{\Omega_2} : \Omega_1 \times \Omega_2 \rightarrow \Omega_2.$$

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 $\mathcal{F}_1$:

$$\pi_{\Omega_1}^*(\mathcal{F}_1) := \{\pi_{\Omega_1}^{-1}(E) : E \in \mathcal{F}_1\} = \{E \times \Omega_2 : E \in \mathcal{F}_1\}.$$

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 -álgebras $\mathcal{F}_1 \otimes \mathcal{F}_2$ 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. $\triangleleft$

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 $E_1, E_2, \dots \in \mathcal{M}$, então $(\bigcup E_n)^y = \bigcup E_n^y \in \mathcal{F}_1$. Logo $\mathcal{M}$ é fechado em relação a união enumerável.
- Se $E \in \mathcal{M}$, e $F = E^c$, então $F^y = \{x \in \Omega_1 : (x, y) \notin E\} = \Omega_1 \setminus E^y \in \mathcal{F}_1$. Logo $\mathcal{M}$ é fechada sob complementação.

◁

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.,

$$\mathcal{F}_0 = \mathcal{f} \{ \{A \times B : A \in \mathcal{F}_1, B \in \mathcal{F}_2\} \}$$

Como $\mathcal{F}_0$ é uma álgebra, e $\mathcal{F} = \sigma(\mathcal{F}_0)$, 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

$$\mu_1(A)\mu_2(B) = \sum_{i=1}^{\infty} \mu_1(A_i)\mu_2(B_i)$$

Consequentemente

$$\mu_1(A)\mu_2(B) = \sum_i \mu_1(A_i)\mu_2(B_i)$$

Pelo Teorema de Extensão de Caratheodory temos que existe uma única medida definida na σ -álgebra $\mathcal{F}_1 \otimes \mathcal{F}_2$. $\triangleleft$

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 $f : \Omega_1 \times \Omega_2 \rightarrow Z$ é mensurável, então as funções $f^y : \Omega_1 \rightarrow Z$, $f^x : \Omega_2 \rightarrow Z$ obtidas fixando uma variável também são mensuráveis.

Demonstração. Provaremos apenas para f^y . Seja $I^y : \Omega_1 \rightarrow \Omega_1 \times \Omega_2$ a aplicação de inclusão $I^y(x) = (x, y)$. Dado $E \in \mathcal{F}_1 \otimes \mathcal{F}_2$, por 8.3 $(I^y)^{-1}(E) = E^y \in \mathcal{F}_1$, logo I^y é uma função mensurável. Mas $f^y = X \circ I^y$. $\triangleleft$

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 a $\mathcal{F}_1 \otimes \mathcal{F}_2$, 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 é mensurável.

Então $\mathcal{M}$ contém a álgebra das uniões finita disjuntas de retângulos mensuráveis.

- Se E_n são conjuntos crescentes em $\mathcal{M}$ então $\mu((\bigcup_n E_n)^y) = \mu(\bigcup_n E_n^y) = \lim_{n \rightarrow \infty} \mu(E_n^y)$ é mensurável, e logo $\bigcup_n E_n \in \mathcal{M}$.

De modo análogo, se E_n são conjuntos decrescentes 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 (1.27), $\mathcal{M} = \mathcal{F}_1 \otimes \mathcal{F}_2$.

Para demonstrar o caso de medida infinita seja $\Omega_n \nearrow \Omega$ com $\mu(\Omega_n) < \infty$, e reaplique o argumento trocando μ por uma medida finita $\mu_n(F) = \mu(F \cap \Omega_n)$. Então $\mu(E^y) = \lim_{n \rightarrow \infty} \mu_n(E^y) \quad \triangleleft$

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 Caratheodory, 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 $\mu \otimes \nu: \mathcal{F}_1 \otimes \mathcal{F}_2 \rightarrow [0, \infty]$, 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. Essas 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 2.11 temos que $\lambda_1 = \lambda_2 = \mu \otimes \nu$ para todo $\mathcal{F}_1 \otimes \mathcal{F}_2$. $\triangleleft$

8.9 Teorema (Fubini)

Sejam $(\Omega_1, \mathcal{F}_1, \mu)$ e $(\Omega_2, \mathcal{F}_2, \nu)$ espaços de medida. Se $X: \Omega_1 \times \Omega_2 \rightarrow \mathbb{R}$ é $\mu \otimes \nu$ -integrável, então

$$\int_{\Omega_1 \times \Omega_2} X d(\mu \otimes \nu) = \int_{x \in \Omega_1} \left[\int_{y \in \Omega_2} X(x, y) d\nu \right] d\mu = \int_{y \in \Omega_2} \left[\int_{x \in \Omega_1} X(x, y) d\mu \right] d\nu.$$

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. $\triangleleft$

8.10 Teorema (Tonelli)

Sejam $(\Omega_1, \mathcal{F}_1, \mu)$ e $(\Omega_2, \mathcal{F}_2, \nu)$ espaços de medida, e $f : \Omega_1 \times \Omega_2 \rightarrow \mathbb{R}$ uma função $\mu \otimes \nu$ -mensurável. Então f é $\mu \otimes \nu$ -integrável se e somente se

$$\int_{x \in \Omega_1} \left[\int_{y \in \Omega_2} |f(x, y)| \, d\nu \right] d\mu < \infty$$

Demonstração. Segue de imediato do Teorema de Fubini aplicado à função não negativa $|f|$. $\triangleleft$
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 .

 $\triangleleft$ **8.3 Distribuição Conjunta**

Lembramos a definição de distribuição (conjunta) para vetores aleatórios:

8.12 Definição (Distribuição de Vetores Aleatórios)

Se $X = (X_1, \dots, X_n)$ for um vetor aleatório, então X induz uma medida de probabilidade $\mathbf{P}$ em $\mathbb{R}^n$ definida tomando a pré-imagem para cada conjunto de Borel A :

$$\mathbf{P}(A) = \mathbf{P}(X^{-1}(A)) = \mathbf{P}(X \in A). \quad (8.1)$$

Esta medida de probabilidade $\mathbf{P}$ é denominada **distribuição** de X .

8.13 Definição (Função de Distribuição Conjunta)

Se $X = (X_1, \dots, X_n)$ for um vetor aleatório, então a função de distribuição conjunta de $(X_1, \dots, X_n)$

$$F_{(X_1, \dots, X_n)}(x_1, \dots, x_n) = \mathbf{P}(X \leq x_1, \dots, X_n \leq x_n). \quad (8.2)$$

A partir da definição da independência vemos que a função de distribuição conjunta fatora como o produto de funções de distribuição.

$$F_{(X_1, \dots, X_n)}(x_1, \dots, x_n) = F_{X_1}(x_1) \times \dots \times F_{X_n}(x_n),$$

8.14 Teorema

As variáveis aleatórias $X_1, \dots, X_n$ 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 $\implies$, 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 $A = A_1 \times \cdots \times A_n$, onde A_k é Borel. Então, pela independência,

$$\mathbf{P}(A_1 \times \cdots \times A_n) = \mathbf{P}((X_1, \dots, X_n) \in A_1 \times \cdots \times A_n) \quad (8.3)$$

$$= \mathbf{P}(X_1 \in A_1, \dots, X_n \in A_n) \quad (8.4)$$

$$= \mathbf{P}(X_1 \in A_1) \times \cdots \times \mathbf{P}(X_n \in A_n) \quad (8.5)$$

$$= \mathbf{P}_1(A_1) \times \cdots \times \mathbf{P}_n(A_n). \quad (8.6)$$

Na outra direção ($\impliedby$), para verificar a independência de $X_1, \dots, X_n$, é suficiente considerarmos os semi-intervalos. Ou seja, se $\mathbf{P}(X_1 \leq x_1, \dots, X_n \leq x_n)$ fatorar, o teorema estará demonstrado.

Mas este último termo é (pela definição de distribuição conjunta)

$$\mathbf{P}((-\infty, x_1] \times \cdots \times (-\infty, x_n]) = \mathbf{P}_1(-\infty, x_1] \times \cdots \times \mathbf{P}_n(-\infty, x_n] \quad (8.7)$$

$$= \mathbf{P}(X_1 \leq x_1) \times \cdots \times \mathbf{P}(X_n \leq x_n). \quad (8.8)$$

E assim terminamos. $\triangleleft$

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 8.9, 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 no 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 $\mathcal{F}_n$.

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 $\sigma(\mathcal{A})$ satisfazendo $\mu_n = \mu|_{\mathcal{F}_n}$.

8.15 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 $\sigma(\bigcup_{n=1}^{\infty} \mathcal{F}_n)$.

Demonstração.

Defina μ na álgebra $\mathcal{A} = \bigcup_{n=1}^{\infty} \mathcal{F}_n$ por $\mu(E) = \mu_n(E)$ para $E \in \mathcal{F}_n$. A condição de consistência de Kolmogorov garante que μ está bem definida.

É simples também 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 $i = 1, \dots, n$ então existe m tal que $A_i \in \mathcal{F}_m$ para todo $i = 1, \dots, n$.

Agora, se $A \in \mathcal{A}$, então existe n tal que $A \in \mathcal{F}_n$ e

$$\begin{aligned} \mu(A) &= \mu_n(A) \\ &= \sup\{\mu_n(C) : C \in \mathcal{K} \cap \mathcal{F}_n \text{ e } C \subset A\} \\ &\leq \sup\{\mu_n(C) : C \in \mathcal{K} \text{ e } C \subset A\} \\ &\leq \mu(A). \end{aligned}$$

Pelo Teorema de Extensão Compacta (2.26) temos que μ é pré-medida. Finalmente pelo Teorema de Extensão de Caratheodory podemos estender a uma probabilidade em $\sigma(\mathcal{A})$. $\triangleleft$

Postas estas considerações iniciais, podemos nos concentrar no objeto central da 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 temos um candidato natural.

Nós equiparemos $\mathbb{R}^{\mathbb{N}}$ com a σ -álgebra produto gerada pelos retângulos finito dimensionais

$$\mathcal{C} = \{A : A = \{\omega : \omega_i \in (a_i, b_i] \text{ para } i = 1, \dots, k \quad k \in \mathbb{N}\},$$

onde $-\infty \leq a_i < b_i \leq \infty$.

Esta é, de fato, a σ -álgebra com a qual trabalhamos intuitivamente nos cursos mais básicos de probabilidade. Lembre, por exemplo, que quando queremos estudar o “lançamento de infinitas moedas”, sempre descrevemos os eventos de interesse a partir de eventos que dependem do resultado de finitos lançamentos. Mesmo eventos do tipo $A = \{\text{observamos cara infinitas vezes}\}$ é escrito como

$$A = \bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n,$$

com $A_n = \{\text{observamos cara na } n\text{-ésima jogada}\}$.

Dando sequência, vamos começar a preparar o caminho para usar o teorema 8.15. Assim, 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 $\mathcal{C} \subseteq \mathcal{A}$.

8.16 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 $\mathcal{G}_n = \{A \in \mathcal{B}(\mathbb{R}^n) : A \times \mathbb{R}^{\mathbb{N}-n} \in \sigma\langle \mathcal{C} \rangle\}$. Como $\mathcal{G}_n$ é uma σ -álgebra contendo os retângulos do tipo

$$\prod_{k=1}^n (a_k, b_k] \subset \mathbb{R}^n,$$

então $\mathcal{G}_n = \mathcal{B}(\mathbb{R}^n)$. Segue que $\mathcal{F}_n \subseteq \sigma\langle \mathcal{C} \rangle$ para todo $n \geq 1$ e

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

◁

Dado $A \in \mathcal{A}$ não vazio, defina

$$\dim(A) := \min\{n \geq 1 : A \in \mathcal{F}_n\},$$

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 $m < n$

$$(B_1 \times \mathbb{R}^{\mathbb{N}-m}) \cap (B_2 \times \mathbb{R}^{\mathbb{N}-n}) = ((B_1 \times \mathbb{R}^{n-m}) \cap B_2) \times \mathbb{R}^{\mathbb{N}-n} \in \mathcal{F}_n.$$

8.17 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 $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^{\mathbb{N}}))$ com

$$\mathbf{P}(\omega : \omega_i \in (a_i, b_i], 1 \leq i \leq n) = \mathbf{P}_n((a_1, b_1] \times \cdots \times (a_n, b_n])$$

Demonstração.

Para começar, dado $n \geq 1$ defina a medida de probabilidade Q_n em $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ por

$$Q_n(A) = \mathbf{P}_{n+1}(A \times \mathbb{R}).$$

Como Q_n coincide com $\mathbf{P}_n$ no π -sistema dos retângulos 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 daí 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 $B \in \mathcal{F}_m$,

$$\mu_n(B \times \mathbb{R}^{\mathbb{N}-m}) = \mu_n(B \times \mathbb{R}^{n-m} \times \mathbb{R}^{\mathbb{N}-n}) = \mathbf{P}_n(B \times \mathbb{R}^{n-m}) = \mathbf{P}_m(B) = \mu_m(B \times \mathbb{R}^{\mathbb{N}-m}).$$

Concluimos assim que a sequência $\{(\mathcal{F}_n, \mu_n)\}_{n \geq 1}$ é Kolmogorov consistente.

Considere agora as classes de cilindros

$$\mathcal{K}_n = \{K \times \mathbb{R}^{\mathbb{N}-n}, K \in \mathcal{B}(\mathbb{R}^n) \text{ compacto} \}$$

e

$$\mathcal{K} = \{K \times \mathbb{R}^{\mathbb{N}-n}, K \in \mathcal{B}(\mathbb{R}^n) \text{ compacto}, n \geq 1\} = \bigcup_{n \geq 1} \mathcal{K}_n.$$

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.

Para ver isso considere uma sequência de conjuntos $\{C_i\}_{i=1}^{\infty} \in \mathcal{K}$ 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 $j \geq 1$,

$$\pi_j(B_m) \neq \emptyset.$$

Temos agora dois casos: $\sup\{\dim(B_m); m \geq 1\} = d < \infty$ e $\sup\{\dim(B_m); m \geq 1\} = \infty$.

No primeiro caso, como a sequência $(\dim(B_m))_{m \geq 1}$ 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 $B_m = K_m \times \mathbb{R}^{\mathbb{N}-d}$. Vale assim que

$$\bigcap_{m=m_0}^{\infty} K_m \neq \emptyset,$$

e

$$\bigcap_{i=1}^{\infty} C_i = \bigcap_{m=m_0}^{\infty} B_m = \left(\bigcap_{m=m_0}^{\infty} K_m \right) \times \mathbb{R}^{\mathbb{N}-d} \neq \emptyset.$$

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.

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 de 8.15.

◁

8.5 Existência de Variáveis Aleatórias Independentes

8.18 Teorema

Dadas probabilidades $\mathbf{P}_1, \dots, \mathbf{P}_n$ em $\mathbb{R}$, então existem variáveis aleatórias independentes $X_1, \dots, X_n$ com distribuições $\mathbf{P}_1, \dots, \mathbf{P}_n$.

Demonstração. Seja $\mathbf{P} = \mathbf{P}_1 \otimes \dots \otimes \mathbf{P}_n$ 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 6.10.

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 $X = (X_1, \dots, X_n)$ então as funções componentes $X_1, \dots, X_n$ são independentes, pelo teorema anterior.

◁

Podemos generalizar o resultado anterior para sequências de variáveis aleatórias i.i.d.

8.19 Teorema (Sequências de Variáveis i.i.d.)

Seja F uma distribuição em $\mathbb{R}$, então existe uma sequência de variáveis aleatórias independentes $X_1, \dots, X_n, \dots$ com distribuição F .

Demonstração. Exercício. Repita o argumento do teorema anterior usando o Teorema de Extensão de Kolmogorov. $\triangleleft$

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 $\mathbf{P}(X \in A) = \int_A f(x) \, dx$, 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 f está definido em $\mathbb{R}^n$ e A é um Boreliano de $\mathbb{R}^n$.
- 3** Densidade conjunta de n variáveis aleatórias $f_1, \dots, f_n$ é a densidade do vetor aleatório $(X_1, \dots, X_n)$.

8.20 Teorema

Sejam $X_1, \dots, X_n$ variáveis aleatórias com densidades $f_1, \dots, f_n$ e densidade conjunta f . Então $X_1, \dots, X_n$ são independentes se e somente se $f(x_1, \dots, x_n) = f_1(x_1) \cdots f_n(x_n)$ para quase todo $(x_1, \dots, x_n)$ 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:

8.21 Teorema (Teorema de Diferenciação de Lebesgue)

Seja $f : \mathbb{R} \rightarrow \mathbb{R}$ uma função Lebesgue integrável. Então

$$\frac{1}{\epsilon} \int_x^{x+\epsilon} f(y) \, dy$$

vai convergir para $f(x)$, 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.

8.22 Teorema

Se $f : \mathbb{R}^n \rightarrow \mathbb{R}$ é Lebesgue integrável, então

$$\frac{1}{\text{vol}(Q)} \int_Q f(y) \, dy \rightarrow f(x)$$

para quase todos os x , onde Q é um cubo contendo x .

Demonstração. [Prova do Teorema 8.20] 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 $P(X \in A) = \int_A f(x) dx$, $P(Y \in B) = \int_B g(y) dy$ e $P((X, Y) \in A \times B) = \int_{A \times B} \varphi(x, y) dx dy$

Então, vamos começar a demonstrar a implicação. Então supomos que X e Y são independentes. Então, $\int_{A \times B} \varphi(x, y) dx dy$ deve ser igual a, por definição, o produto das probabilidades individuais $\int_A f(x) dx \times \int_B g(y) dy$. 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 $\epsilon \rightarrow 0$, 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.

$$P(X \in A, Y \in B) = \int_{A \times B} \varphi(x, y) dx dy \quad (8.9)$$

$$= \int_{A \times B} f(x)g(y) dx dy \quad (8.10)$$

$$= \int_A f(x) dx \int_B g(y) dy, \text{ por Fubini} \quad (8.11)$$

$$= P(X \in A) \cdot P(Y \in B). \quad (8.12)$$

Então, por definição, X e Y são independentes. $\triangleleft$

Vejamos alguns exemplos.

8.23 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

$$\varphi(x, y) = \begin{cases} 1, & \text{se } (x, y) \in [0, 1] \times [0, 1] \\ 0, & \text{caso contrário.} \end{cases}$$

$\triangleleft$

8.24 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 $(X_1, \dots, X_n)$ é

$$\varphi(x_1, \dots, x_n) = \frac{1}{(2\pi)^{1/2}} e^{-\frac{1}{2}|x|^2},$$

onde $|x| = x_1^2 + x_2^2 + \dots + x_n^2$.

◁

Faremos outra demonstração de que para variáveis independentes a esperança do produto é o produto das esperanças. Compare com o Teorema 7.8.

8.25 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

$$\mathbf{E}[XY] = \int xy \, d(\mathbf{P}_X \times \mathbf{P}_Y)$$

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, 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, denotando cara por 1 e coroa por -1 .

Considere agora $X + X$ e $X + (-X)$. Os quatro somandos anteriores têm 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 $X + Y$.

8.26 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 $\mathbf{P}_X \otimes \mathbf{P}_Y$, 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

$$H(z) = \int_{\{(x,y) \in \mathbb{R}^2 : x+y \leq z\}} d(\mathbf{P}_X \times \mathbf{P}_Y) \quad (8.13)$$

$$\int_{\mathbb{R}} \int_{\mathbb{R}} \mathbf{1}_{\{(x,y) \in \mathbb{R}^2 : x+y \leq z\}} d(\mathbf{P}_X \times \mathbf{P}_Y) \quad (8.14)$$

Por Fubini,

$$\begin{aligned} &= \int_{\mathbb{R}} d\mathbf{P}_Y \cdot \int_{\mathbb{R}} \mathbf{1}_{\{(x,y) \in \mathbb{R}^2 : x+y \leq z\}} d\mathbf{P}_X \\ &= \int_{\mathbb{R}} d\mathbf{P}_Y \cdot \int_{\mathbb{R}} \mathbf{1}_{\{(x,y) \in \mathbb{R}^2 : x \leq z-y\}} d\mathbf{P}_X \\ &= \int_{\mathbb{R}} d\mathbf{P}_Y \cdot \int_{\mathbb{R}} \mathbf{P}_X(X \in (-\infty, z-y]) d\mathbf{P}_Y \\ &= \int_{\mathbb{R}} F(z-y) d\mathbf{P}_Y(y) \end{aligned}$$

Agora observamos que a integral em relação a uma medida é a mesma que a integral Stieltjes em relação a uma função de distribuição.

$$\text{e assim} = \int_{\mathbb{R}} F(z-y) dG(y). \quad \triangleleft$$

8.27 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. $\triangleleft$

8.28 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

$$Z(x) = \int_{\mathbb{R}} \mathbf{1}_{[0,1]}(x-y) \mathbf{1}_{[0,1]}(y) dy$$

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 $\int_{x-1}^1 dy = 2-x$.

◁

8.7 Convolução e Aproximação por Funções Suaves

Nesta seção discutimos uma técnica para aproximar funções integráveis arbitrárias por funções suaves.

8.29 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

$$\text{supp}(f) := \overline{\{x \in X \mid f(x) \neq 0\}} = \overline{f^{-1}(\{0\}^c)}.$$

é um conjunto compacto.

Para começar, estabelecemos a existência de uma função teste em $[-1,1]$.

8.30 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.

8.31 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 $f : \mathbb{R} \rightarrow \mathbb{R}$ 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, $|\frac{1}{h}(\rho_n(x+h-y) - \rho_n(x-y))f(y)| \leq M|f(y)|$ e pela integrabilidade de $f(y)$ no intervalo $[-n, n]$ podemos usar a Convergência Dominada para concluir que

$$\begin{aligned} f'_n(x) &= \lim_{h \rightarrow 0} \frac{f_n(x+h) - f_n(x)}{h} \\ &= \lim_{h \rightarrow 0} \frac{1}{h} \int_{-n}^n (\rho_n(x+h-y) - \rho_n(x-y))f(y)dy \\ &= \int_{-n}^n \lim_{h \rightarrow 0} \frac{1}{h} (\rho_n(x+h-y) - \rho_n(x-y))f(y)dy \\ &= \int_{-n}^n \rho'_n(x-y)f(y)dy \end{aligned}$$

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 quaisquer $x, y \in K$ 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 $n \geq \max(N_1, N_2)$

$$\begin{aligned}
 |f_n(x) - f(x)| &= \left| \int_{-n}^n (\rho_n(x-y)f(y) - \rho_n(x-y)f(x)) dy \right| \\
 &= \left| \int_{x-\frac{1}{n}}^{x+\frac{1}{n}} (\rho_n(x-y)f(y) - \rho_n(x-y)f(x)) dy \right| \quad \text{pois } n > N_2 \\
 &\leq \int_{x-\frac{1}{n}}^{x+\frac{1}{n}} \rho_n(x-y) |f(y) - f(x)| dy \\
 &\leq \varepsilon \int_{x-\frac{1}{n}}^{x+\frac{1}{n}} \rho_n(x-y) dy \quad \text{pois } \frac{1}{n} < \delta \\
 &\leq \varepsilon \quad \text{pois } \rho_n \text{ é positivo e } \int_{-\infty}^{\infty} \rho_n(x) dx = 1
 \end{aligned}$$

A última coisa a provar é a desigualdade da norma caso f seja limitada.

$$\begin{aligned}
 |f_n(x)| &\leq n \int_{-n}^n \rho(n(x-y)) |f(y)| dy \quad \text{pois } \rho \text{ é positivo} \\
 &\leq n \|f\|_{\infty} \int_{-\infty}^{\infty} \rho(n(x-y)) dy = \|f\|_{\infty}
 \end{aligned}$$

◁

Testando Convergência com Funções Suaves

8.32 Lema

Seja $\{X_i\}_{i=1}^{\infty}$ uma sequência de variáveis aleatórias então $X_n \xrightarrow{d} X$ se e somente se $\lim_{n \rightarrow \infty} \mathbf{E}[f(X_n)] = \mathbf{E}[f(X)]$ para todas as funções $f \in C_c^{\infty}(\mathbb{R}, \mathbb{R})$.

Demonstração. Como qualquer $f \in C_c^{\infty}(\mathbb{R}, \mathbb{R})$ é temos que $X_n \xrightarrow{d} X$ implica $\lim_{n \rightarrow \infty} \mathbf{E}[f(X_n)] = \mathbf{E}[f(X)]$.

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 convergência uniforme de 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

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbf{P}[|X_n| > R] &= 1 - \lim_{n \rightarrow \infty} \mathbf{E}[\mathbb{1}_{|X_n| \leq R}] \\ &\leq 1 - \lim_{n \rightarrow \infty} \mathbf{E}[\psi_R(X_n)] \\ &= 1 - \mathbf{E}[\psi_R(X)] \\ &\leq 1 - \mathbf{E}\left[\mathbb{1}_{|X| \leq \frac{R}{2}}\right] \\ &= \mathbf{P}\left[|X| > \frac{R}{2}\right] \end{aligned}$$

Por outro lado, sabemos que $\lim_{R \rightarrow \infty} \mathbb{1}_{|X| \leq \frac{R}{2}} = 0$ q.c. e, portanto, pelo Teorema da Convergência Monótona, $\lim_{R \rightarrow \infty} \mathbf{P}\left[|X| > \frac{R}{2}\right] = 0$. Escolha $R > 0$ tal que

$$\mathbf{P}[|X| > R] \leq \mathbf{P}\left[|X| > \frac{R}{2}\right] \leq \frac{\varepsilon}{4\|f\|_\infty}$$

Então podemos escolher $N_1 > 0$ tal que $\mathbf{P}[|X_n| > R] \leq \frac{\varepsilon}{2\|f\|_\infty}$ 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,

$$\begin{aligned} |\mathbf{E}[f_k(X) - f(X)]| &\leq \mathbf{E}[|f_k(X) - f(X)|; |X| \leq R] + \mathbf{E}[|f_k(X) - f(X)|; |X| > R] \\ &\leq \varepsilon \mathbf{P}[|X| \leq R] + 2\|X\|_\infty \mathbf{P}[|X| > R] \\ &\leq \varepsilon + \frac{\varepsilon}{2} < 2\varepsilon \end{aligned}$$

e pelo mesmo cálculo, por $n > N_1$

$$|\mathbf{E}[f_k(X_n) - f(X_n)]| \leq \varepsilon + 2\|X_n\|_\infty \mathbf{P}[|X_n| > R] \leq 2\varepsilon$$

Para terminar a prova, escolha $k > K$ e então podemos encontrar $N_2 > 0$ tal que para todo $n > N_2$, temos $|\mathbf{E}[f_k(X_n) - f_k(X)]| < \varepsilon$. Juntando essas três estimativas, temos $n > \max(N_1, N_2)$,

$$|\mathbf{E}[f(X_n) - f(X)]| \leq 5\varepsilon$$

◁

8.8 Aplicações

Lei 0 – 1 de Hewitt-Savage

8.33 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 $A = \pi(A)$.

8.34 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 $A_n = \{\omega : (X_1, \dots, X_n) \in B_n\}$,

$$\mathbf{P}(A \Delta A_n) \rightarrow 0, n \rightarrow \infty. \quad (8.15)$$

Como as variáveis aleatórias $\{X_i\}_{i \in \mathbb{N}}$ 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

$$\mathbf{P}(A \Delta A_n) = \mathbf{P}(X \in B \Delta B_n) = \mathbf{P}(\pi_n(X) \in B \Delta B_n). \quad (8.16)$$

Consequentemente

$$\begin{aligned} \mathbf{P}(\pi_n(X) \in B \Delta B_n) &= \mathbf{P}\{\pi_n(X) \in B\} \Delta \{\pi_n(X) \in B_n\} \\ &= \mathbf{P}\{(X \in B) \Delta (\pi_n(X) \in B_n)\} = \mathbf{P}\{A \Delta \pi_n(A_n)\}. \end{aligned} \quad (8.17)$$

Por 8.16 e 8.17 temos que

$$\mathbf{P}(A \Delta A_n) = \mathbf{P}(A \Delta \pi_n(A_n)). \quad (8.18)$$

Logo por 8.15 temos que

$$\mathbf{P}(A \Delta (A_n \cap \pi_n(A_n))) \rightarrow 0, n \rightarrow \infty. \quad (8.19)$$

Assim, por 8.15, 8.18 e 8.19, obtemos

$$\mathbf{P}(A_n) \rightarrow \mathbf{P}(A) \text{ e } \mathbf{P}(\pi_n(A)) \rightarrow \mathbf{P}(A)$$

$$\mathbf{P}(A_n \cap \pi_n(A_n)) \rightarrow \mathbf{P}(A).$$

Além disso, como $\{X_i\}$ são independentes, escolhendo π_n como a permutação que leva $(X_1, \dots, X_n)$ para $(X_{n+1}, \dots, X_{2n})$

temos

$$\mathbf{P}(A_n \cap \pi_n(A_n)) = \mathbf{P}\{(X_1, \dots, X_n) \in B_n, (X_{n+1}, \dots, X_{2n}) \in B_n\}$$

$$\begin{aligned}
&= \mathbf{P}\{(X_1, \dots, X_n) \in B_n\} \cdot \mathbf{P}\{(X_{n+1}, \dots, X_{2n}) \in B_n\} \\
&= \mathbf{P}(A_n) \mathbf{P}(\pi_n(A_n))
\end{aligned}$$

Logo $\mathbf{P}(A) = \mathbf{P}^2(A)$ e, portanto, $\mathbf{P}(A) = 0$ ou 1 . $\triangleleft$

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 $\mathbb{R}^n$

8.35 Teorema

Se S_n é um passeio aleatório em $\mathbb{R}$ então uma das seguintes alternativas ocorre

- ☐ $S_n = 0$ para todo n
- ☐ $S_n \rightarrow \infty$
- ☐ $S_n \rightarrow -\infty$
- ☐ $-\infty = \liminf S_n < \limsup S_n = \infty$

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 $\limsup S_n$ 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, $\liminf S_n \in \{-\infty, \infty\}$. $\triangleleft$

8.9 ★ 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 $\pi_\beta : \Omega_A \rightarrow \Omega_\beta$, que leva x_A para x_β . Também podemos definir a composta de tais aplicações de projeção de coordenadas.

Por exemplo, dados $B \subset A$, podemos definir a projeção parcial $\pi_B : \Omega_A \rightarrow \Omega_B$ de modo natural. Mais geralmente se $C \subset B \subset A$, podemos definir $\pi_{C \leftarrow B} : \Omega_B \rightarrow \Omega_C$. Então temos as leis de composição

$$\pi_{D \leftarrow C} \circ \pi_{C \leftarrow B} = \pi_{D \leftarrow B}$$

se $D \subset C \subset B \subset A$.

Podemos definir a σ -álgebras 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

$$\mathcal{B}_A = \prod_{\beta \in A} \mathcal{B}_\beta := \left(\bigcup_{\beta \in A} \pi_\beta^*(\mathcal{B}_\beta) \right).$$

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 $f : \Omega_A \rightarrow [0, +\infty]$ é mensurável, isso significa que existe um conjunto enumerável $B \subset A$ e uma função mensurável em $\mathcal{B}_B$ $f_B : \Omega_B \rightarrow [0, +\infty]$ 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 $\prod_{\alpha \in A} \Omega_\alpha$ 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 $C \subset B \subset A$.

8.36 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 $X_1, \dots, X_n, \dots$ 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 $\langle \cdot, \cdot \rangle : V \times V \rightarrow F$ uma forma bilinear não degenerada em V , e sejam $v_1, \dots, v_n$ vetores não-nulos em V . Mostre que as variáveis aleatórias $\langle X, v_1 \rangle, \dots, \langle X, v_n \rangle$ são independentes se e somente se os vetores $v_1, \dots, v_n$ são linearmente independentes.

Modos de Convergência

Nesse capítulo apresentaremos e compararemos diferentes noções de convergência. É de fundamental importância ressaltar que estamos tratando de **convergência em espaços de probabilidade**. Alguns resultados desse capítulo são falsos se a medida não for finita.

9.1 Convergência Pontual

Seja Ω um conjunto e $X_n : \Omega \rightarrow \mathbb{R}$ uma sequência de funções que compartilham do mesmo domínio Ω . Dizemos que $X_n(x)$ **converge pontualmente** para uma função $X : \Omega \rightarrow \mathbb{R}$ se:

$$\lim_{n \rightarrow \infty} X_n(x) = X(x),$$

para todo $x \in \Omega$.

9.1 Exemplo

Para $n \in \mathbb{N}$ definimos $X_n(x) = \begin{cases} 1 & \text{se } 0 < x < 1/n \\ 0 & \text{caso contrário} \end{cases}$

Essa sequência converge pontualmente para a função constante.

◁

9.2 Convergência Uniforme

A **distância uniforme** entre duas funções X e Y é definida como

$$\|X - Y\|_u = \sup |X(x) - Y(x)|$$

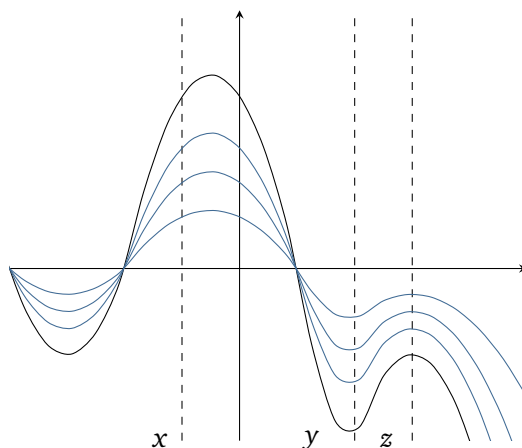
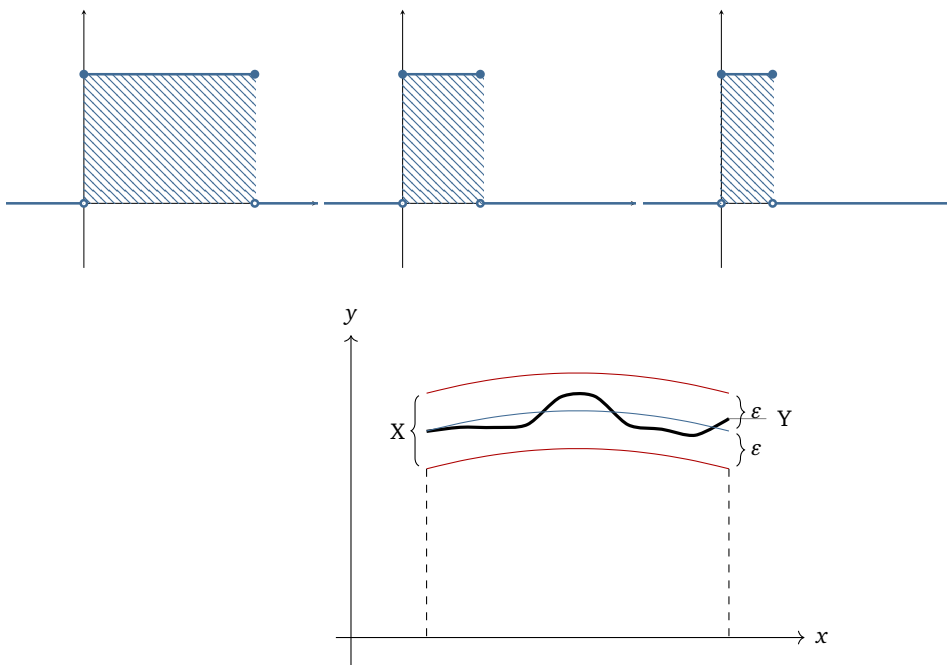


Figura 9.1: A sequência $\{X_n\}_{n=1}^{\infty}$ converge pontualmente para a função constante 0. Assim para todo ponto $x, y, z \in X$, temos $\lim X_n(x) = 0$, $\lim X_n(y) = 0$ e $\lim X_n(z) = 0$.



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 $X_n(x) = \begin{cases} 1 & \text{se } 0 < x < 1/n \\ 0 & \text{caso contrário} \end{cases}$

Essa sequência converge pontualmente para a função constante, mas não uniformemente.

<

9.3 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.

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

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x), \quad (9.1)$$

para todo x tal que $F_X(x)$ é contínua.

9.5 Exemplo

O exigência de que apenas os pontos de continuidade de F devem ser considerados na definição acima é essencial.

Considere X_n uniformemente distribuídas nos intervalos $(0, 1/n)$, então essa sequência converge em distribuição para uma variável aleatória degenerada $X = 0$. De fato, $F_n(x) = 0$ para todo n quando $x \leq 0$ e $F_n(x) = 1$ para todo $x \geq 1/n$ quando $n > 0$.

No entanto, para a variável aleatória limite $F(0) = 1$, enquanto que $F_n(0) = 0$ para todo n .

Assim, a convergência de distribuição falha no ponto $x = 0$ em que F é descontínua.

<

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

$$\lim_{n \rightarrow \infty} P_{X_n}(k) = P_X(k), \quad \text{para } k = 0, 1, 2, \dots. \quad (9.2)$$

Demonstração.

Como X é de valor inteiro, sua função de distribuição, $F_X(x)$, é contínua em todo $x \in \mathbb{R} - \{0, 1, 2, \dots\}$. Se $X_n \xrightarrow{d} X$, então

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x), \quad \text{para } x \in \mathbb{R} - \{0, 1, 2, \dots\}. \quad (9.3)$$

Assim, para $k = 0, 1, 2, \dots$, temos

$$\lim_{n \rightarrow \infty} \mathbf{P}_{X_n}(k) = \lim_{n \rightarrow \infty} \left[F_{X_n} \left(k + \frac{1}{2} \right) - F_{X_n} \left(k - \frac{1}{2} \right) \right] \quad (X_n \text{ toma valores inteiros}) \quad (9.4)$$

$$= \lim_{n \rightarrow \infty} F_{X_n} \left(k + \frac{1}{2} \right) - \lim_{n \rightarrow \infty} F_{X_n} \left(k - \frac{1}{2} \right) \quad (9.5)$$

$$= F_X \left(k + \frac{1}{2} \right) - F_X \left(k - \frac{1}{2} \right) \quad (\text{pois } X_n \xrightarrow{d} X) \quad (9.6)$$

$$= \mathbf{P}_X(k) \quad (\text{uma vez que } X \text{ toma valores inteiros}). \quad (9.7)$$

Para provar a recíproca, suponha que

$$\lim_{n \rightarrow \infty} \mathbf{P}_{X_n}(k) = \mathbf{P}_X(k), \quad \text{para } k = 0, 1, 2, \dots \quad (9.8)$$

Então, para todo $x \in \mathbb{R}$, temos

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = \lim_{n \rightarrow \infty} \mathbf{P}(X_n \leq x) \quad (9.9)$$

$$= \lim_{n \rightarrow \infty} \sum_{k=0}^{\lfloor x \rfloor} \mathbf{P}_{X_n}(k), \quad (9.10)$$

onde $\lfloor x \rfloor$ denota o inteiro maior menor ou igual a x . Uma vez que, para qualquer x fixo, o conjunto $\{0, 1, \dots, \lfloor x \rfloor\}$ é um conjunto finito, podemos alterar a ordem do limite e a soma, então obtemos

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = \sum_{k=0}^{\lfloor x \rfloor} \lim_{n \rightarrow \infty} \mathbf{P}_{X_n}(k) \quad (9.11)$$

$$= \sum_{k=0}^{\lfloor x \rfloor} \mathbf{P}_X(k) \quad (\text{por hipótese}) \quad (9.12)$$

$$= \mathbf{P}(X \leq x) = F_X(x). \quad (9.13)$$

◁

9.7 Teorema

Seja $\{X_i\}_{i=1}^{\infty}$ uma sequência de variáveis aleatórias tal que

$$X_n \sim \text{Bin} \left(n, \frac{\lambda}{n} \right), \quad \text{para } n \in \mathbb{N}, n > \lambda, \quad (9.14)$$

onde $\lambda > 0$ é uma constante. Mostre que X_n converge em distribuição para $\text{Po}(\lambda)$.

Basta mostrar que

$$\lim_{n \rightarrow \infty} \mathbf{P}_{X_n}(k) = P_X(k), \quad \text{para todo } k = 0, 1, 2, \dots \quad (9.15)$$

Temos

$$\begin{aligned}\lim_{n \rightarrow \infty} \mathbf{P}_{X_n}(k) &= \lim_{n \rightarrow \infty} \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \lambda^k \lim_{n \rightarrow \infty} \frac{n!}{k!(n-k)!} \left(\frac{1}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \frac{\lambda^k}{k!} \cdot \lim_{n \rightarrow \infty} \left(\left[\frac{n(n-1)(n-2)\dots(n-k+1)}{n^k} \right] \left[\left(1 - \frac{\lambda}{n}\right)^n \right] \left[\left(1 - \frac{\lambda}{n}\right)^{-k} \right] \right).\end{aligned}$$

Para k fixo, temos

$$\begin{aligned}\lim_{n \rightarrow \infty} \frac{n(n-1)(n-2)\dots(n-k+1)}{n^k} &= 1, \\ \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^{-k} &= 1, \\ \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^n &= e^{-\lambda}.\end{aligned}$$

Logo

$$\lim_{n \rightarrow \infty} \mathbf{P}_{X_n}(k) = \frac{e^{-\lambda} \lambda^k}{k!}.$$

9.4 Convergência em Probabilidade

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

$$\forall \epsilon > 0 \quad \mathbf{P}(|X_n - X| > \epsilon) \rightarrow 0 \text{ quando } n \rightarrow \infty.$$

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

$$\lim_{n \rightarrow \infty} \mathbf{P}(|X_n - 0| \geq \epsilon) = \lim_{n \rightarrow \infty} \mathbf{P}(X_n \geq \epsilon) \quad (\text{pois } X_n \geq 0) \quad (9.16)$$

$$= \lim_{n \rightarrow \infty} e^{-n\epsilon} \quad (\text{pois } X_n \sim \text{Exp}(n)) \quad (9.17)$$

$$= 0, \quad \text{para todo } \epsilon > 0. \quad (9.18)$$

◁

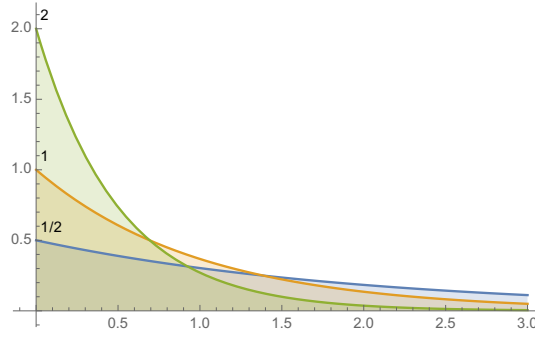


Figura 9.2: Distribuição exponenciais de parâmetros 1/2,1,2

9.10 Exemplo

Seja X uma variável aleatória e $X_n = X + Y_n$, com

$$\mathbf{E}[Y_n] = \frac{1}{n}, \quad \mathbf{Var}[Y_n] = \frac{\sigma^2}{n}, \quad (9.19)$$

onde $\sigma > 0$ é uma constante. Então $X_n \xrightarrow{P} X$.

Pela desigualdade triangular para $Y_n - \mathbf{E}[Y_n] + \mathbf{E}[Y_n]$, temos

$$|Y_n| \leq |Y_n - \mathbf{E}[Y_n]| + \frac{1}{n}. \quad (9.20)$$

Então para todo $\epsilon > 0$, temos

$$\mathbf{P}(|X_n - X| \geq \epsilon) = \mathbf{P}(|Y_n| \geq \epsilon) \quad (9.21)$$

$$\leq \mathbf{P}\left(|Y_n - \mathbf{E}[Y_n]| + \frac{1}{n} \geq \epsilon\right) \quad (9.22)$$

$$= \mathbf{P}\left(|Y_n - \mathbf{E}[Y_n]| \geq \epsilon - \frac{1}{n}\right) \quad (9.23)$$

$$\leq \frac{\mathbf{Var}[Y_n]}{\left(\epsilon - \frac{1}{n}\right)^2} \quad (\text{pela desigualdade de Chebyshev}) \quad (9.24)$$

$$= \frac{\sigma^2}{n\left(\epsilon - \frac{1}{n}\right)^2} \rightarrow 0 \quad \text{quando } n \rightarrow \infty. \quad (9.25)$$

Logo $X_n \xrightarrow{P} X$.

◁

9.5 Convergência Quase Certa

9.11 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

$$\mathbf{P}\left(\lim_{n \rightarrow \infty} X_n = X\right) = 1$$

ou, equivalentemente, se

$$\mathbf{P}\left(\omega \in \Omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\right) = 1$$

A convergência quase certa é uma convergência pontual em um conjunto de medida 1, ou seja, $X_n(\omega) \rightarrow X(\omega)$ para quase todo ω , exceto aqueles dentro de um conjunto de medida nula

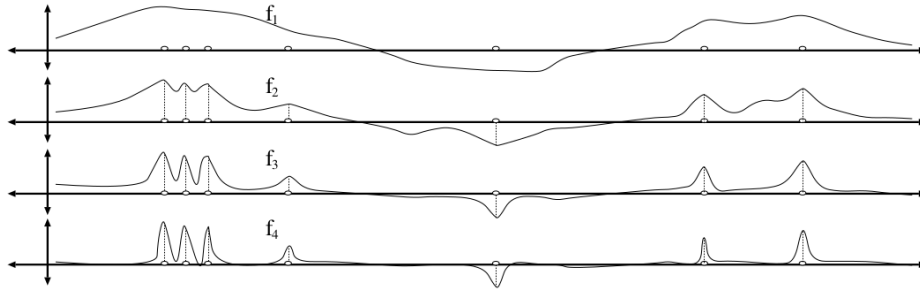


Figura 9.3: A sequência $\{X_n\}_{n=1}^{\infty}$ converge quase certamente a 0.

9.12 Exemplo

Para $n \in \mathbb{N}$ definimos $X_n(x) = \begin{cases} 1 & \text{se } x \in \mathbb{Q} \\ 1/n & \text{caso contrário} \end{cases}$

Essa sequência não converge pontualmente para a função constante 0, mas converge quase certamente para 0.

◁

9.13 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.

No entanto, quando consideramos qualquer número finito de dias, existe uma probabilidade diferente de zero, de a condição de término não ter ocorrido ainda.

9.14 Teorema (Critério de Convergência Quase Certa)

Sejam $\{X_n\}_{n \geq 1}$ e X variáveis aleatórias. Então X_n converge quase certamente para X se, e somente se, para todo $k \geq 1$

$$\mathbf{P}\left(|X_n - X| \geq \frac{1}{k} \text{ i.v.}\right) = 0$$

Demonstração.

Um sequência de funções converge se para todo $k \geq 1$, existe $n \geq 1$ tal que $m \geq n$ implica

$$|X_m(\omega) - X(\omega)| < \frac{1}{k}$$

equivalentemente

$$[X_n \rightarrow X] = \bigcap_{k=1}^{\infty} \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} \left[|X_m - X| < \frac{1}{k} \right]$$

Logo,

$$[X_n \not\rightarrow X] = \bigcup_{k=1}^{\infty} \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} \left[|X_m - X| \geq \frac{1}{k} \right] = \bigcup_{k=1}^{\infty} \left[|X_n - X| \geq \frac{1}{k}, \text{ i.v.} \right]$$

Se $X_n \xrightarrow{q.c.} X$, então $\mathbf{P}[X_n \not\rightarrow X] = 0$ o que implica que

$$\mathbf{P}\left(|X_n - X| \geq \frac{1}{k}, \text{ i.v.}\right) = 0, \quad \text{para todo } k \geq 1$$

Por outro lado, se $\mathbf{P}\left(|X_n - X| \geq \frac{1}{k}, \text{ i.v.}\right) = 0$, para todo $k \geq 1$ então

$$\mathbf{P}[X_n \not\rightarrow X] \leq \sum_{k=1}^{\infty} \mathbf{P}\left(|X_n - X| \geq \frac{1}{k}, \text{ i.v.}\right) = 0$$

◁

9.15 Teorema

Considere a sequência $\{X_n\}_{n=1}^{\infty}$. Se para todo $\epsilon > 0$, tivermos

$$\sum_{n=1}^{\infty} \mathbf{P}(|X_n - X| > \epsilon) < \infty, \quad (9.26)$$

então $X_n \xrightarrow{q.c.} X$.

Demonstração.

Pelo Teorema e pelo Primeiro Lema de Borel Cantelli

◁

9.16 Exemplo

Considere uma sequência $\{X_n, n = 1, 2, 3, \dots\}$ 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

$$\sum_{n=1}^{\infty} \mathbf{P}(|X_n| > \epsilon) < \infty. \quad (9.27)$$

Observe que $|X_n| = \frac{1}{n}$. Logo, $|X_n| > \epsilon$ se e somente se $n < \frac{1}{\epsilon}$. Logo temos

$$\sum_{n=1}^{\infty} \mathbf{P}(|X_n| > \epsilon) \leq \sum_{n=1}^{\lfloor \frac{1}{\epsilon} \rfloor} \mathbf{P}(|X_n| > \epsilon) \quad (9.28)$$

$$= \lfloor \frac{1}{\epsilon} \rfloor < \infty. \quad (9.29)$$

◁

9.17 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 $n = 1, 2, \dots$.

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.

◁

9.6 Convergência em Média

9.18 Definição

Uma sequência de variáveis aleatórias $\{X_n\}_{n=1}^{\infty}$ **converge em $\mathcal{L}^p$** para a variável aleatória X se

$$\|X_n - X\|_p \rightarrow 0 \text{ quando } n \rightarrow \infty.$$

Em outras palavras

$$\mathbf{E}|X_n - X|^p \rightarrow 0 \text{ quando } n \rightarrow \infty.$$

9.19 Exemplo

Seja $X_n \sim \text{Uni}(0, \frac{1}{n})$. 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

$$\mathbf{E}(|X_n - 0|^p) = \int_0^{\frac{1}{n}} x^p n \, dx \quad (9.30)$$

$$= \frac{1}{(p+1)n^p} \rightarrow 0, \quad \text{para todo } p \geq 1. \quad (9.31)$$

◁

9.20 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

$$(\mathbf{E}|X_n - X|^p)^{\frac{1}{p}} \leq (\mathbf{E}|X_n - X|^q)^{\frac{1}{q}} \rightarrow 0$$

◁

9.21 Exemplo

Considere uma sequência $\{X_n\}$ tal que

$$X_n = \begin{cases} n^2 & \text{com probabilidade } \frac{1}{n} \\ 0 & \text{com probabilidade } 1 - \frac{1}{n} \end{cases} \quad (9.32)$$

Mostre que

$$\square X_n \xrightarrow{p} 0.$$

$$\square X_n \text{ não converge na } p\text{-media para qualquer } p \geq 1.$$

$$\square \text{ Para mostrar que } X_n \xrightarrow{p} 0, \text{ podemos escrever, para qualquer } \epsilon > 0$$

$$\lim_{n \rightarrow \infty} \mathbf{P}(|X_n| \geq \epsilon) = \lim_{n \rightarrow \infty} \mathbf{P}(X_n = n^2) \quad (9.33)$$

$$= \lim_{n \rightarrow \infty} \frac{1}{n} \quad (9.34)$$

$$= 0. \quad (9.35)$$

Assim $X_n \xrightarrow{p} 0$.

$$\square \text{ Para qualquer } r \geq 1, \text{ podemos escrever}$$

$$\lim_{n \rightarrow \infty} \mathbf{E}(|X_n|^p) = \lim_{n \rightarrow \infty} \left(n^{2p} \cdot \frac{1}{n} + 0 \cdot \left(1 - \frac{1}{n}\right) \right) \quad (9.36)$$

$$= \lim_{n \rightarrow \infty} n^{2p-1} \quad (9.37)$$

$$= \infty \quad (\text{pois } p \geq 1). \quad (9.38)$$

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.

◁

9.7 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:

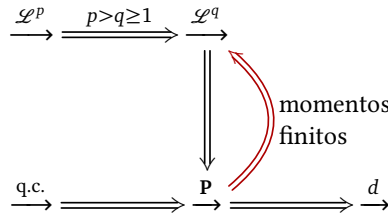


Figura 9.4: Relação entre os modos de convergência em espaços de probabilidade.

Convergência quase certa implica convergência em probabilidade

9.22 Teorema (Convergência q.c. e em Probabilidade)

Dados X_n e X variáveis aleatórias.

- 1 $X_n \rightarrow X$ q.c. implica $X_n \rightarrow X$ em probabilidade.
- 2 Se $X_n \rightarrow X$ em probabilidade, então existe uma subsequência $X_{n_k} \rightarrow X$ quase certamente.

Demonstrações

- 1 Suponha $X_n \rightarrow X$ q.c. Seja $\epsilon > 0$. Então $0 = \mathbf{P}(|X_n - X| > \epsilon \text{ i.v.}) = \mathbf{P}(\limsup\{|X_n - X| > \epsilon\})$. Pelo teorema de continuidade, $\mathbf{P}(\limsup\{|X_n - X| > \epsilon\}) \geq \limsup \mathbf{P}(|X_n - X| > \epsilon)$. Portanto, $X_n \rightarrow X$ em probabilidade.

- 2 Suponha $X_n \rightarrow X$ em probabilidade. Queremos encontrar uma subsequência que converge quase certamente. Tomaremos uma sequência que decresça rapidamente.

Podemos escolher uma subsequência X_{n_k} tal que

$$\mathbf{P}(|X_{n_k} - X| > \frac{1}{k}) < 2^{-k} \text{ para } k = 1, 2, \dots$$

Como $\sum_k 2^{-k}$ converge, o Lema Borel-Cantelli nos fornece que

$$\mathbf{P}(|X_{n_k} - X| > \frac{1}{k} \text{ i.v.}) = 0.$$

Portanto, $X_{n_k} \rightarrow X$ quase certamente.

◁

Convergência em Média implica convergência em Probabilidade

9.23 Proposição

$X_n \xrightarrow{\mathcal{L}^p} X$ implica $X_n \xrightarrow{p} X$.

Demonstração. Usando a desigualdade de Chebyshev

$$\mathbf{P}(|X_n - X| > \epsilon) \leq \mathbf{P}(|X_n - X|^p \geq \epsilon^p) \leq \frac{\mathbf{E}|X_n - X|^p}{\epsilon^p} \rightarrow 0.$$

◁

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.

9.24 Teorema

Se $X_n \xrightarrow{p} X$ 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 $|X_{n_{k_j}} - X|^p \leq (|X_{n_{k_j}}| + |X|)^p \stackrel{\text{por hipótese}}{\leq} (2Y)^p < \infty$

Assim, pelo Teorema da Convergência Dominada temos que

$$\mathbf{E}[|X_{n_{k_j}} - X|^p] \rightarrow 0.$$

Provamos a convergência para alguma subsubsequência de uma subsequência arbitrária n_k .

E logo podemos concluir que $\mathbf{E}[|X_{n_{k_j}} - X|^p] \rightarrow 0$.

◁

9.25 Teorema

Seja (X_n) uma sequência de variáveis aleatórias não negativas tal que

$$X_n \xrightarrow{q.c.} X \text{ e } \mathbf{E}[X_n] \rightarrow \mathbf{E}[X], \text{ então } X_n \xrightarrow{\mathcal{L}^1} X$$

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

$$\lim \mathbf{E}[X - X_n] \mathbb{1}_{\{X_n \leq X\}} = 0 \quad \text{e} \quad \mathbf{E}[X_n] \rightarrow \mathbf{E}[X],$$

◁

Convergência em Probabilidade implica Convergência em Distribuição

9.26 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. Exercício.

◁

9.8 Teorema de Representação de Skorokhod

9.27 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 . Suponha que $X_n \rightarrow X$ em distribuição. Seja

$$X'_n(\omega) = \inf\{x : F_n(x) \geq \omega\}, \text{ e } X'(\omega) = \inf\{x : F(x) \geq \omega\},$$

de modo que

$$\{F_n(x) \geq \omega\} = \{X'_n(\omega) \leq x\}, \text{ e } \{F(x) \geq \omega\} = \{X'(\omega) \leq x\},$$

logo

$$F'_n(x) = \mathbf{P}'(X'_n \leq x) = \mathbf{P}'\{\omega : \omega \in (0, F_n(x)]\} = F(x),$$

$$F'(x) = \mathbf{P}'(X' \leq x) = \mathbf{P}'\{\omega : \omega \in (0, F(x)]\} = F(x).$$

Até agora, encontramos variáveis aleatórias com as distribuições corretas. Para provar a convergência, seja $\omega \in (0, 1)$ e $\varepsilon > 0$ e tome x um ponto de continuidade de F , de tal modo que

$$F(x) < \omega \leq F(x + \varepsilon)$$

Por hipótese, $F_n(x) \rightarrow F(x)$ quando $n \rightarrow \infty$, logo existe n_0 , tal que

$$F_n(x) < \omega \leq F(x + \varepsilon) \text{ para todo } n \geq n_0.$$

E logo

$$X'_n(\omega) - \varepsilon \leq x < X'_n(\omega) \text{ para todo } n \geq n_0.$$

Devido às escolhas arbitrárias de ω e ε , segue que

$$\liminf X'_n(\omega) \geq X'(\omega) \text{ para todo } \omega. \quad (9.39)$$

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

$$F(x - \varepsilon) < \omega^* < F_n(x) \text{ para todo } n \geq n_0,$$

e logo

$$X'_n(\omega^*) < x < X'_n(\omega^*) + \varepsilon \text{ para todo } n \geq n_0.$$

Além disso, o fato de que $\omega < \omega^*$, implica que

$$X'_n(\omega) < x < X'_n(\omega^*) + \varepsilon \text{ para todo } n \geq n_0,$$

e assim

$$\limsup X'_n(\omega) \leq X'(\omega^*) \text{ para todo } \omega^* > \omega. \quad (9.40)$$

Das equações 9.39) com 9.40, temos que $X_n(\omega) \rightarrow X(\omega)$ quando $n \rightarrow \infty$ para todo ω ponto de continuidade de X . Para concluir observamos que X possui no máximo um número enumerável de descontinuidade e assim temos convergência quase certa. $\triangleleft$

9.9 Teorema de Seleção de Helly

O Teorema de Seleção de Helly, nos diz que é quase verdade que, para cada sequência de variáveis aleatórias, há uma subsequência que converge em distribuição. Esta é uma afirmação que é análoga à compacidade.

9.28 Teorema (Teorema de Seleção de Helly)

Seja F_n uma sequência de funções de distribuição, existe uma subsequência F_{n_k} e uma função limitada, direita-contínua, não decrescente F tal que

$$F_{n_k}(x) \rightarrow F(x) \text{ em todos os pontos de continuidade } x \text{ de } F.$$

Observamos que F não é necessariamente uma função de distribuição. Os limites em $\pm\infty$ podem não ser os corretos. Ou seja, $F(x) \not\rightarrow 0$ quando $x \rightarrow -\infty$ e $F(x) \not\rightarrow 1$ quando $x \rightarrow \infty$.

Demonstração. Começamos considerando uma enumeração dos racionais. Isto é $\mathbb{Q} = \{q_1, q_2, q_3, \dots\}$.

Seja q_1 , então como $0 \leq F_n(q_1) \leq 1$ pelo Teorema de Bolzano-Weierstrass, há uma subsequência n_k^1 para a qual $F_{n_k^1}(q_1)$ converge. Podemos repetir esse argumento, extraíndo uma subsequência n_k^2 de modo que $F_{n_k^1}(q_1)$ e $F_{n_k^2}(q_2)$ converge. Repetindo, obtemos subsequências $F_{n_k^m}$ que convergem no conjunto $\{x_1, \dots, x_m\}$.

Pelo Método da Diagonal de Cantor, podemos encontrar uma subsequência $F_{n_k^k}$ tal que

$$F_{n_k^k}(q) \text{ converge para todo } q \in \mathbb{Q}.$$

Nós chamamos esse limite F_∞ . Isto é,

$$F_{n_k^k}(q) \rightarrow F_\infty(q), \forall q \in \mathbb{Q}.$$

Para estender a todos os reais definimos

$$F(x) := \inf\{F_\infty(q) : q \in \mathbb{Q}, q > x\}.$$

A função F está definida em $\mathbb{R}$ e é não decrescente. E finalmente, F é contínua à direita. Pois

$$\lim_{x_n \searrow x} F(x_n) = \inf_{y > x} F(y),$$

e como F não é decrescente temos

$$= \inf\{F_\infty(q) : q > x\} = F(x).$$

Logo a única coisa que precisamos demonstrar é a convergência. Pela definição de F , existe um $q > x$ racional que

$$F_\infty(q) < F(x) + \epsilon. \quad (9.41)$$

Então, pela definição de $F_\infty(q)$,

$$F_{n_k^k}(q) \rightarrow F_\infty(q).$$

Como $q > x$, $F_{n_k^k}(x) \leq F_{n_k^k}(q)$. Esse fato juntamente com a equação (9.41) implicam que

$$F_{n_k^k}(x) \leq F(x) + \epsilon$$

para k suficientemente grande.

Agora precisamos de um limite inferior para $F_{n_k^k}(x)$.

Seja x um ponto de continuidade de F . Ou seja, pela continuidade esquerda de F , existe $r < x$ tal que

$$F(r) > F(x) - \epsilon.$$

Escolha $r' \in \mathbb{Q}$ tal que $r < r' < x$. E logo

$$F_{n_k}(r) \rightarrow F_\infty(r').$$

Agora, queremos comparar isso com x . $F_\infty(r') \geq F(r) \geq F(x) - \epsilon$ pois $F(r)$ é o mínimo de todos os valores. Além disso, $F_{n_k}(x) \geq F_{n_k}(r')$ porque F_{n_k} não é decrescente. Assim, provamos o limite inferior

$$F_{n_k}(x) \geq F(r) - \epsilon \text{ para } k \text{ suficientemente grande.}$$

Ao combinar nossas desigualdades, temos:

$$F(x) - \epsilon \leq F_{n_k}(x) \leq F(x) + \epsilon.$$

Logo $\limsup F_{n_k}(x) \leq F(x) + \epsilon$ e $\liminf F_{n_k}(x) \geq F(x) - \epsilon$. Como $\epsilon > 0$ é arbitrário, concluímos

$$\lim F_{n_k}(x) = F(x).$$

◁

9.10 Convergência Fraca de Medidas

9.29 Definição

Dadas as medidas de probabilidade de Borel $\mu_1, \mu_2, \mu_3, \dots$ dizemos que μ_n converge fracamente para μ , se

$$\int_{\mathbb{R}} f \, d\mu_n \rightarrow \int_{\mathbb{R}} f \, d\mu$$

para todas as funções contínuas limitadas $f : \mathbb{R} \rightarrow \mathbb{R}$.

Destacamos que na definição anterior a convergência precisa ser testada apenas para funções contínuas. Ou seja, a topologia de $\mathbb{R}$ está desempenhando um papel importante aqui, e não somente propriedades de espaços mensuráveis.

Medidas Tight

9.30 Definição

Dizemos que uma coleção $\{\mu_n\}$ de medidas de probabilidade em $\mathbb{R}$ é **tight** se para todo $\epsilon > 0$, existem $a < b$ com $\mu_n([a, b]) > 1 - \epsilon$ para todo $n \in \mathbb{N}^*$.

Isto é, todas as medidas possuem a maioria de sua massa no mesmo intervalo finito. A massa não "escapa para o infinito".

9.31 Teorema (Teorema de Prokhorov)

Se $\{\mu_n\}$ é uma sequência tight de medidas de probabilidade então existe uma subsequência $\{\mu_{n_k}\}$ e uma medida μ tal que $\{\mu_{n_k}\}$ converge fracamente para μ .

Demonstração. Seja $F_n(x) = \mu_n((-\infty, x])$. Logo existe uma subsequência F_{n_k} e uma função F tal que $F_{n_k}(x) \rightarrow F(x)$ 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 $\lim_{x \rightarrow -\infty} F(x) = 0$ e $\lim_{x \rightarrow \infty} F(x) = 1$. 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 $\mu_n((a, b]) > 1 - \varepsilon$ para todos os n . Logo

$$\lim_{x \rightarrow \infty} F(x) - \lim_{x \rightarrow -\infty} F(x) \geq F(b) - F(a) \quad (9.42)$$

$$= \lim_{n \rightarrow \infty} [F_{n_k}(b) - F_{n_k}(a)] = \lim_{n \rightarrow \infty} \mu_{n_k}((a, b]) > 1 - \varepsilon. \quad (9.43)$$

Isso é verdade para todo $\varepsilon > 0$, então devemos ter

$$\lim_{x \rightarrow \infty} F(x) - \lim_{x \rightarrow -\infty} F(x) = 1$$

Portanto, F é de fato uma função de distribuição de probabilidade. Assim, podemos definir a medida de probabilidade μ por $\mu((a, b]) = F(b) - F(a)$ para $a < b$. E $\mu_n \rightarrow \mu$.

Claramente temos que $\{\mu_{n_k}\}$ converge fracamente para μ , pois a esperança só depende da distribuição, pelo Teorema 7.6.

◁

Exercícios

Ex. 9.1 — Deixe $X_1, X_2, X_3, \dots$ ser uma sequência de variáveis aleatórias, de modo que

$$X_n \sim Ge\left(\frac{\lambda}{n}\right), \quad \text{for } n = 1, 2, 3, \dots, \quad (9.44)$$

onde $\lambda > 0$ é uma constante. Defina uma nova sequência Y_n como

$$Y_n = \frac{1}{n} X_n, \quad \text{para } n = 1, 2, 3, \dots. \quad (9.45)$$

Mostre que Y_n converge em distribuição para $Exp(\lambda)$.

Ex. 9.2 — **Ex. 9.3** — Deixe $\{X_n, n = 1, 2, \dots\}$ e $\{Y_n, n = 1, 2, \dots\}$ serem duas sequências de variáveis aleatórias, definidas no espaço amostral Ω . Suponha que nós saibamos

$$X_n \xrightarrow{p} X, \quad (9.46)$$

$$Y_n \xrightarrow{p} Y. \quad (9.47)$$

Prove que $X_n + Y_n \xrightarrow{p} X + Y$.

Ex. 9.4 — Considere a sequência $\{X_n, n = 1, 2, 3, \dots\}$ tal que

$$X_n = \begin{cases} n & \text{com probabilidade } \frac{1}{n^2} \\ 0 & \text{com probabilidade } 1 - \frac{1}{n^2} \end{cases} \quad (9.48)$$

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 $E|X_n^r| \rightarrow E|X^r|$.

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 $\xrightarrow{p}$:

$$X_n \xrightarrow{p} X \Rightarrow g(X_n) \xrightarrow{p} g(X)$$

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 $\text{Var}[X_n] \rightarrow \text{Var}[X]$.

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 recíproco não é válida. (Prove por exemplo para $p = 1$).

Ex. 9.11 — Considere as variáveis aleatórias $X_1, X_2, \dots$ com funções de distribuição $F_1, F_2, \dots$. Suponha que $X_n \rightarrow X$ em distribuição, e a função de distribuição F de X seja contínua. Prove que $F_n \rightarrow F$ na norma sup, i.e.

$$\|F_n - F\|_\infty := \sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \rightarrow 0.$$

Ex. 9.12 — Considere as variáveis aleatórias a valores inteiros $X_1, X_2, \dots$, e uma variável aleatória X . Mostre que $X_n \rightarrow X$ na distribuição se e somente se

$$P(X_n = k) \rightarrow P(X = k) \quad \text{para todo } k \in \mathbb{Z}.$$

Ex. 9.13 — Considere variáveis aleatórias independentes $X_1, X_2, \dots$ uniformemente distribuídas em $[0, 1]$, e deixe $Z_n = \max_{k \leq n} X_k$.

1. Prove que $Z_n \rightarrow 1$ em probabilidade.
2. Prove que $Z_n \rightarrow 1$ quase certamente.

Teoremas Limites

10.1 Lei Fraca

10.1 Teorema (Lei Fraca dos Grandes Números)

Sejam $X_1, X_2, \dots$ variáveis aleatórias independentes com $\mathbf{E}[X_k] = \mu$ para todo k , e $\mathbf{Var}(X_k) \leq c < \infty$. Considere $S_n = X_1 + X_2 + \dots + X_n$. Então

$$\frac{S_n}{n} \xrightarrow{\mathbf{P}} \mu.$$

Logo a média de n variáveis aleatórias independentes converge para a média comum.

Demonstração. Provaremos, na verdade, algo mais forte: que $\frac{S_n}{n} \xrightarrow{\mathcal{L}^2} \mu$. Isto é

$$\mathbf{E} \left[\frac{S_n}{n} - \mu \right]^2 \rightarrow 0? \quad (10.1)$$

Observamos que pela linearidade da esperança $\mathbf{E} \left[\frac{S_n}{n} \right] = \mu$.

Então

$$\mathbf{E} \left[\frac{S_n}{n} - \mu \right]^2 = \mathbf{Var} \left(\frac{S_n}{n} \right) = \frac{1}{n^2} \mathbf{Var}(S_n) \quad (10.2)$$

$$= \frac{1}{n^2} (\mathbf{Var}(X_1) + \dots + \mathbf{Var}(X_n)) \leq \frac{cn}{n^2} = \frac{c}{n} \rightarrow 0. \quad (10.3)$$

◁

A demonstra anterior nos fornece também a taxa de convergência: $\frac{1}{n}$.

10.2 Observação

- 1 Como já observamos a convergência é em $\mathcal{L}^2$, em vez de simplesmente em probabilidade.
- 2 O teorema é válido para variáveis aleatórias X_k não correlacionadas e não independentes. O único lugar onde precisamos que as variáveis sejam não correlacionado é para demonstrar que as somas das variâncias é a variância da soma.

A demonstração anterior generaliza de maneira óbvia para vetores aleatórios.

10.3 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 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

$$\frac{S_n}{n} \rightarrow p$$

em probabilidade, quando $n \rightarrow \infty$.

10.4 Exemplo

Lançando uma moeda n vezes: Então

$$P(0.49n \leq \text{número de caras} \leq 0.51n) \rightarrow 1 \quad \text{quando } n \rightarrow \infty.$$

◁

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.

10.5 Teorema (Método de Truncamento de Markov)

Considere X uma variável aleatória, com média finita. Considere $M > 1$ o nível de truncamento.

Defina

$$[X]^M := \begin{cases} X, & \text{se } |X| \leq M \\ 0, & \text{se } |X| > M \end{cases} = X \cdot \mathbf{1}_{\{|X| \leq M\}}.$$

Então

- $|[X]^M| \leq M$ pontualmente, e $|[X]^M| \leq |X|$
- $[X]^M \rightarrow X$ pontualmente quando $M \rightarrow \infty$,
- A variável $[X]^M$ possui todos os momentos finitos.

□ Pelo Teorema da Convergência Dominada $E[|X|^M] \rightarrow E[X]$ quando $M \rightarrow \infty$. Mais ainda,

$$E[|X|^M - X] \rightarrow 0 \text{ quando } M \rightarrow \infty. \quad (10.4)$$

10.6 Corolário

Dado uma variável aleatória X com média finita

$$E[|X| \cdot \mathbf{1}_{\{|X|>M\}}] \rightarrow 0 \text{ quando } M \rightarrow \infty.$$

Demonstração. Observe que

$$X - X(M) = X \cdot \mathbf{1}_{\{|X|>M\}}.$$

◁

10.7 Teorema (Lei Fraca dos Grandes Números)

Sejam $X_1, X_2, \dots$ variáveis aleatórias independentes identicamente distribuídas com $E[X_k] = \mu$ para todo k .

Considere $S_n = X_1 + X_2 + \dots + X_n$.

Então

$$\frac{S_n}{n} \rightarrow \mu$$

em probabilidade (quando $n \rightarrow \infty$).

Demonstração. Começaremos truncando. Dado $M > 1$, considere

$$[X]_k^M := X_k \cdot \mathbf{1}_{\{|X_k| \leq M\}}, \quad S_n^{(M)} = X_1^{(M)} + \dots + X_n^{(M)}.$$

Então

$$\text{Var}[X_k^{(M)}] = E[(X_k^{(M)})^2] - (E[X_k^{(M)}])^2 \leq M^2.$$

Então, podemos aplicar a versão antiga da Lei dos Grandes Números com variância finita:

$$\frac{S_n^{(M)}}{n} \rightarrow E[X_1^{(M)}]$$

em $\mathcal{L}^2$ quando $n \rightarrow \infty$ esse fato implica também convergência em probabilidade.

$$E \left| \frac{S_n^{(M)}}{n} - E[X_1^{(M)}] \right| \rightarrow 0 \text{ quando } n \rightarrow \infty.$$

Agora aproximamos $S_n^{(M)}$ por S_n e $X_1^{(M)}$ por X_1 . Pela desigualdade triangular

$$E \left| \frac{S_n}{n} - E[X_1] \right| \leq E \left| \frac{S_n^{(M)}}{n} - E[X_1^{(M)}] \right| + E \left| \frac{S_n - S_n^{(M)}}{n} \right| + E[|X_1 - X_1^{(M)}|].$$

Agora estimaremos o primeiro termo

$$\mathbf{E} \left| \frac{S_n - S_n^{(M)}}{n} \right| \leq \frac{1}{n} \sum_{k=1}^n \mathbf{E} |X_k - X_k^{(M)}| = \mathbf{E} |X_1 - X_1^{(M)}|.$$

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 $n \rightarrow \infty$. Então

$$\limsup \mathbf{E} \left| \frac{S_n}{n} - \mathbf{E}[X_1] \right| \leq 2\mathbf{E} |X_1 - X_1^{(M)}|.$$

Seja $M \rightarrow \infty$. Então $\mathbf{E} |X_1 - X_1^{(M)}| \rightarrow 0$. $\triangleleft$

10.2 Lei Forte

Considere $X_1, X_2, \dots$ variáveis aleatórias independentes e identicamente distribuídas com esperança finita μ . Seja $S_n := X_1 + \dots + X_n$. A Lei Fraca dos Grandes Números afirma que

$$\frac{S_n}{n} \rightarrow \mu \text{ em probabilidade.}$$

Por outro lado, a Lei Forte dos Grandes Números que provaremos a seguir afirma que

$$\frac{S_n}{n} \rightarrow \mu \text{ q.c.}$$

Começaremos apresentando uma demonstração simples da Lei Forte dos Grandes Números sob o pressuposto de que o quarto momento é finito.

10.8 Teorema (Lei Forte dos Grandes Números - I)

Sejam $X_1, X_2, \dots$ variáveis aleatórias i.i.d. com esperança μ e assumamos $\mathbf{E}[X_k^4] < \infty$. Então, $S_n = X_1 + X_2 + \dots + X_n$ satisfaz

$$\frac{S_n}{n} \rightarrow \mu \text{ quase certamente.}$$

Demonstração.

- Passo 1. Primeiro, podemos assumir, sem perda de generalidade, que a esperança $\mu = 0$. Na verdade, faça $X'_k = X_k - \mu$. Então,

$$\frac{\sum_{k=1}^n (X_k - \mu)}{n} = \frac{S_n}{n} - \mu.$$

Portanto, podemos supor que $\mu = 0$. Vamos provar que para qualquer $\varepsilon > 0$,

$$\mathbf{P} \left(\left| \frac{S_n}{n} \right| > \varepsilon \text{ i.v.} \right) = 0. \quad (10.5)$$

□ Etapa 2. Consideramos o quarto momento de S_n .

$$\mathbf{E}[S_n^4] = \mathbf{E}\left[\left(\sum_{k=1}^n X_k\right)^4\right] = \sum_{1 \leq i, j, k, \ell \leq n} \mathbf{E}[X_i X_j X_k X_\ell] \quad (10.6)$$

Temos n^4 termos, mas apenas $3n^2 - 2n$ são diferentes de zero.

De fato, se i, j, k, ℓ são distintos, então, pela independência, $\mathbf{E}[X_i^3 X_j] = \mathbf{E}[X_i^3] \mathbf{E}[X_j] = 0$ porque assumimos que o valor médio é 0. Da mesma forma, temos $\mathbf{E}[X_i^2 X_j] = 0$. Claramente $\mathbf{E}[X_i X_j X_k X_\ell] = 0$ se i, j, k, ℓ forem diferentes. Desta forma os únicos termos que não são zero são da forma:

$$\mathbf{E}[X_k^4] \quad \text{e} \quad \mathbf{E}[X_j^2 X_k^2].$$

Agora uma contagem simples nos mostra que existem apenas $3n^2 - 2n$ termos dessa forma.

Observe, temos n termos da forma $\mathbf{E}[X_k^4]$. Para contar os termos da forma $\mathbf{E}[X_j^2 X_k^2]$, observe que esses termos podem ser formados escolhemos 2 elementos de um alfabeto de n letras, então a combinação é $\binom{n}{2}$.

Assim, temos

$$\left(\sum_{k=1}^n X_k\right) \left(\sum_{k=1}^n X_k\right) \left(\sum_{k=1}^n X_k\right) \left(\sum_{k=1}^n X_k\right)$$

Uma vez que escolhemos j e k , o número de maneiras de escolher dois j e dois k s é $\binom{4}{2}$ pois estamos escolhendo um termo em cada fator na expressão acima.

Como

$$\binom{n}{2} \binom{4}{2} = 3n(n-1),$$

temos $3n^2 - 2n$ termos diferentes de zero.

□ Etapa 3. Cada termo diferente de zero é limitado por $C = \mathbf{E}[X_k^4]$.

Pelo teorema de Cauchy-Schwarz,

$$\mathbf{E}[X_j^2 X_k^2] \leq \mathbf{E}[X_j^4]^{\frac{1}{2}} \mathbf{E}[X_k^4]^{\frac{1}{2}} = C.$$

Portanto, $\mathbf{E}[S_n^4] \leq C \cdot (3n^2 - 2n) \leq 3Cn^2$.

□ Etapa 4. Pela desigualdade de Chebyshev, temos

$$\mathbf{P}\left(\left|\frac{S_n}{n}\right| > \varepsilon\right) = \mathbf{P}(S_n^4 > \varepsilon^4 n^4) \quad (10.7)$$

$$\leq \frac{\mathbf{E}S_n^4}{\varepsilon^4 n^4} \text{ por Chebyshev} \quad (10.8)$$

$$\leq \frac{3C}{\varepsilon^4 n^2} \text{ pela Etapa 3} \quad (10.9)$$

A série $\sum_{n=1}^{\infty} \frac{3C}{\varepsilon^4 n^2}$ converge, então pelo Lema de Borel-Cantelli,

$$\mathbf{P}\left(\left|\frac{S_n}{n}\right| > \varepsilon \text{ i.v.}\right) = 0.$$

◁

A demonstração anterior generaliza facilmente para vetores aleatórios

10.9 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

$$\frac{1}{n} \sum_{i=1}^n X^{(i)} \xrightarrow{q.c.} \mathbf{E}(X^{(1)}).$$

10.10 Proposição (Esperança Finita é Necessária)

Sejam $X_1, X_2, \dots$ variáveis aleatórias independentes e identicamente distribuídas com esperança infinita: $\mathbf{E}|X_k| = \infty$. Então $S_n = X_1 + X_2 + \dots + X_n$ satisfaz

$$\mathbf{P}\left(\frac{S_n}{n} \text{ convergir}\right) = 0.$$

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 $n = 1, 2, \dots$. Logo $\sum_{n=0}^{\infty} \mathbf{P}(|X_n| > n) = \infty$. E pelo Lema de Borel-Cantelli,

$$\mathbf{P}(|X_n| > n \text{ i.v.}) = 1.$$

Agora, definimos dois eventos:

$$A := \{|X_n| > n \text{ i.v.}\} \quad (\text{Logo, } \mathbf{P}(A) = 1)$$

e

$$B := \left\{ \frac{S_n}{n} \text{ convergir} \right\}.$$

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,

$$\frac{S_n}{n(n+1)} \rightarrow 0.$$

Por outro lado,

$$\left| \frac{X_{n+1}}{n+1} \right| > 1 \text{ i.v.,}$$

então quando n é suficientemente grande,

$$\left| \frac{S_n}{n} - \frac{S_{n+1}}{n+1} \right| > \frac{2}{3} \text{ i.v.,}$$

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. $\triangleleft$

10.11 Teorema (Lei Forte dos Grandes Números de Kolmogorov)

Sejam $X_1, X_2, \dots, X_n$ variáveis aleatórias independentes e identicamente distribuídas com esperança finita, i.e. $\mathbf{E}X_i = \mu$. Então, $S_n = X_1 + \dots + X_n$, satisfaz

$$\frac{S_n}{n} \rightarrow \mu \quad \text{quase certamente.}$$

Demonstração.

□ Passo 0. Podemos assumir que $X_i \geq 0$. Na verdade, se dividirmos X_i

$$X_i = X_i^+ + X_i^-$$

em uma parte positiva e uma parte negativa, então $\mathbf{E}|X_i^+| < \infty$ e $\mathbf{E}|X_i^-| < \infty$.

□ Passo 1 (truncamento). Definimos

$$\bar{X}_k = \begin{cases} X_k & \text{se } X_k \leq k \\ 0 & \text{se } X_k > k \end{cases}$$

Agora provaremos que $\mathbf{P}(\bar{X}_k \neq X_k \text{ i.v.}) = 0$.

Observamos que $\mathbf{P}(\bar{X}_k \neq X_k) = \mathbf{P}(X_k > k)$. Calculando

$$\sum_{k=1}^{\infty} \mathbf{P}(X_k > k) = \sum_{k=1}^{\infty} \mathbf{P}(X_1 > k) \leq \int_0^{\infty} \mathbf{P}(X_1 > x) dx = \mathbf{E}X_1 < +\infty.$$

Pelo Lema de Borel-Cantelli, a afirmação é verdadeira.

Se demonstrarmos que

$$\frac{\bar{X}_1 + \bar{X}_2 + \dots + \bar{X}_n}{n} \rightarrow \mu \text{ quase certamente,}$$

então

$$\frac{X_1 + X_2 + \cdots + X_n}{n} \rightarrow \mu$$

quase certamente.

No restante da demonstração vamos assumir que $X_k \leq k$. Ou seja, estamos truncando as variáveis aleatórias.

□ Passo 2.

10.12 Lema

$$\sum_{k=1}^{\infty} \frac{\text{Var}(X_k)}{k^2} < +\infty.$$

Demonstração.

$$\begin{aligned} \text{Var}(X_k) &= \mathbf{E}(X_k^2) - (\mathbf{E}X_k)^2 \\ &\leq \mathbf{E}(X_k^2) \\ &= \int_0^{\infty} \mathbf{P}(X_k \geq x) d(x^2) \\ &= \int_0^{\infty} 2x \mathbf{P}(X_k \geq x) dx \\ &= \int_0^{\infty} \mathbb{1}_{\{x \leq k\}} 2x \mathbf{P}(X_k \geq x) dx. \end{aligned}$$

Então, calculando o somatório

$$\begin{aligned} \sum_{k=1}^{\infty} \frac{\text{Var}(X_k)}{k^2} &\stackrel{\text{i.i.d.}}{=} \sum_{k=1}^{\infty} \int_0^{\infty} \frac{\mathbb{1}_{\{x \leq k\}}}{k^2} 2x \mathbf{P}(X_1 \geq x) dx \\ &\stackrel{\text{Fubini}}{=} \int_0^{\infty} \sum_{k=1}^{\infty} \frac{\mathbb{1}_{\{x \leq k\}}}{k^2} 2x \mathbf{P}(X_1 \geq x) dx \end{aligned}$$

usaremos agora que $\sum_{k \geq x} \frac{1}{k^2} \leq \int_x^{\infty} \frac{dt}{t^2} = \frac{1}{x}$. Consequentemente

$$\sum_{k=1}^{\infty} \frac{\text{Var}(X_k)}{k^2} \leq \int_0^{\infty} 2\mathbf{P}(X_1 \geq x) dx = 2\mathbf{E}[X_1] < +\infty.$$

◁

□ Passo 3. Controle ao longo de uma subsequência

Seja $k(n)$ ser a subsequência de k , tal que $k(n) = \lfloor \alpha^n \rfloor$, onde $\alpha > 1$ e $\lfloor \alpha^n \rfloor \geq \alpha^n/2$. Então, afirmamos que:

$$\frac{S_{k(n)}}{k(n)} \rightarrow \mu \text{ q.c.}$$

Para a prova dessa afirmação, também usamos o Lema de Borel-Cantelli. Para cada $\varepsilon > 0$, calculamos

$$\begin{aligned} \sum_{n=1}^{\infty} \mathbf{P} \left(\left| \frac{S_{k(n)}}{k(n)} - \mu \right| > \varepsilon \right) &= \sum_{n=1}^{\infty} \mathbf{P}(|S_{k(n)} - \mathbf{E}[S_{k(n)}]| > \varepsilon \cdot k(n)) \\ &\leq \sum_{n=1}^{\infty} \frac{\mathbf{Var}(S_{k(n)})}{(\varepsilon k(n))^2} \end{aligned}$$

pela desigualdade de Chebyshev. Finalmente, o último termo pode ser escrito como

$$\frac{1}{\varepsilon^2} \sum_{n=1}^{\infty} \frac{1}{k(n)^2} \sum_{k=1}^{k(n)} \mathbf{Var}(X_k).$$

Por outro lado, considere

$$\frac{1}{\varepsilon^2} \sum_{k=1}^{\infty} \mathbf{Var}(X_k) \sum_{n: k(n) \geq k} \frac{1}{k(n)^2}.$$

Aqui, mudamos a ordem de soma, e por enquanto não temos certeza de que esses dois valores são os mesmos. Para garantir que sejam iguais, precisamos verificar a convergência absoluta. Verificamos que a soma interna

$$\sum_{n: k(n) \geq k} \frac{1}{k(n)^2}.$$

é limitada por (substituindo $k(n)$ por $\lfloor \alpha^n \rfloor$)

$$\sum_{n: \alpha^n \geq k} \frac{1}{\lfloor \alpha^{2n} \rfloor} \leq \sum_{n: \alpha^n \geq k} \frac{1}{\alpha^{2n}} \quad (10.10)$$

$$\leq \sum_{n \geq \log_{\alpha} k} \frac{1}{\alpha^{2n}} \quad (10.11)$$

$$\stackrel{\text{progressão geométrica}}{\leq} \frac{2\alpha^2}{(\alpha^2 - 1)k^2} \quad (10.12)$$

$$(10.13)$$

Na progressão

$$\frac{1}{\varepsilon^2} \sum_{k=1}^{\infty} \mathbf{Var}(X_k) \sum_{n: k(n) \geq k} \frac{1}{k(n)^2} \leq \frac{2\alpha^2}{\varepsilon^2(\alpha^2 - 1)} \sum_{k=1}^{\infty} \frac{\mathbf{Var}(X_k)}{k^2}$$

Logo

$$\sum_{n=1}^{\infty} \mathbf{P} \left(\left| \frac{S_{k(n)}}{k(n)} - \mu \right| > \varepsilon \right) \leq \frac{2\alpha^2}{\varepsilon^2(\alpha^2 - 1)} \sum_{k=1}^{\infty} \frac{\mathbf{Var}(X_k)}{k^2} < \infty$$

Logo, pelo Lema de Borel-Cantelli

$$\mathbf{P} \left(\left| \frac{S_{k(n)}}{k(n)} - \mu \right| > \varepsilon \text{ i.v.} \right) = 0$$

assim

$$\frac{S_{k(n)}}{k(n)} \rightarrow \mu \text{ quase certamente.}$$

□ Passo 4. (Preenchendo lacunas na sequência).

Começamos observando que se $k(n) < k < k(n+1)$ então

$$S_{k(n)} \leq S_k \leq S_{k(n+1)}$$

Então, temos a seguinte fórmula:

$$\frac{1}{\alpha} \cdot \frac{S_{k(n)}}{k(n)} = \frac{S_{k(n)}}{k(n+1)}$$

Como $k(n) = \alpha^n$. E

$$\frac{1}{\alpha} \cdot \frac{S_{k(n)}}{k(n)} = \frac{S_{k(n)}}{k(n+1)} \leq \frac{S_k}{k} \leq \frac{S_{k(n+1)}}{k(n)} \leq \alpha \cdot \frac{S_{k(n+1)}}{k(n+1)}$$

portanto

$$\frac{1}{\alpha} \cdot \frac{S_{k(n)}}{k(n)} \rightarrow \frac{\mu}{\alpha} \text{ quase certamente.}$$

Consequentemente,

$$\frac{\mu}{\alpha} \leq \liminf \frac{S_k}{k} \leq \limsup \frac{S_k}{k} \leq \alpha\mu$$

com a probabilidade um.

Uma vez que $\alpha > 1$ é arbitrário, podemos fazer $\alpha \rightarrow 1$. Então, o limite $\lim \frac{S_k}{k}$ existe, e é igual a μ quase certamente.

◁

10.3 Integração de Monte-Carlo

Considere uma função integrável $f : [0,1] \rightarrow \mathbb{R}$.

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 $x_1, \dots, x_n$. 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 $x_1, \dots, x_n$ 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 $I_n \rightarrow \int_0^1 f(x) dx$ 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 $Ef = \int_0^1 f(x) dx$. Então, a Lei Fraca dos Grandes Números aplicada às variáveis aleatórias $f(x_1), \dots, f(x_n)$ completa a prova. $\triangleleft$

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 $j = 0, 1, \dots, n$.

- 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 $\sum_{j=0}^n b_{n,j}(x) = 1$.

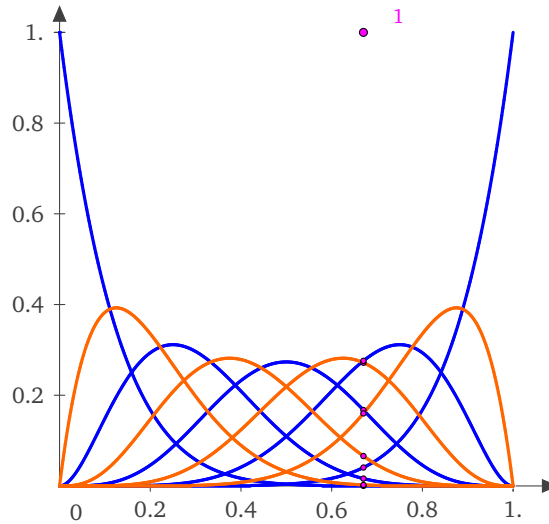


Figura 10.1: Polinômios de Bernstein $b_{i,8}$ para $i = 1, \dots, 8$. Em pontilhado o polinômio $b_{3,8}$. Observe que as somas dos valores em um ponto é 1.

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

$$B_n[f](\mathbf{x}) = \sum_{0 \leq i_1, \dots, i_d \leq n} f\left(\frac{i_1}{n}, \dots, \frac{i_d}{n}\right) \prod_{k=1}^d \binom{n}{i_k} x_k^{i_k} (1 - x_k)^{n-i_k}$$

aproxima f .

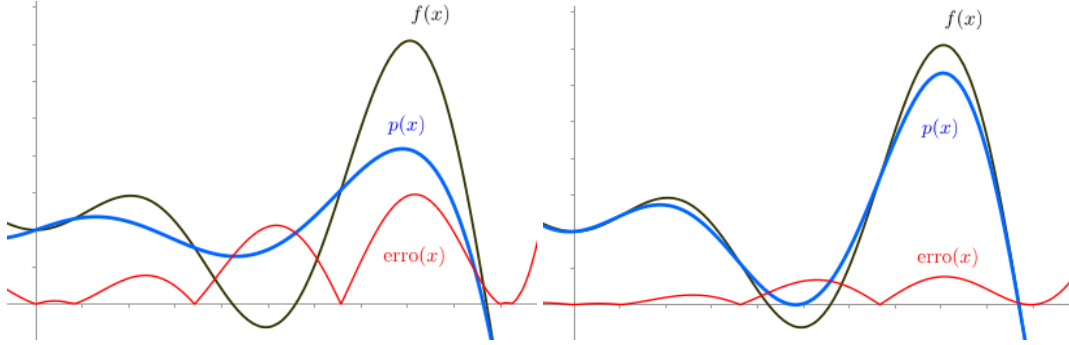


Figura 10.2: Aproximação com $n = 10,50$

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

$$\mathbf{P}[X_{ij} = 1] = x_i \text{ e } \mathbf{P}[X_{ij} = 0] = 1 - x_i.$$

Seja

$$S_{in} = \sum_{j=1}^n X_{ij}$$

e

$$\mathbf{S}_n = (S_{1n}, \dots, S_{dn}).$$

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 $f : [0,1]^d \rightarrow \mathbb{C}$ uma função contínua. Defina

$$B_n[f](\mathbf{x}) = \sum_{0 \leq i_1, \dots, i_d \leq n} f\left(\frac{i_1}{n}, \dots, \frac{i_d}{n}\right) \prod_{k=1}^d \binom{n}{i_k} x_k^{i_k} (1 - x_k)^{n-i_k}$$

para $n = 1, 2, \dots$ e $\mathbf{x} = (x_1, \dots, x_n)$ em $[0, 1]^d$, então $B_n[f] \rightarrow f$ 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 $\mathbf{X}_j = (X_{1j}, \dots, X_{dj})$ ser uma sequência finita de vetores aleatórios definidos em um espaço de probabilidade, de modo que $\{X_{ij}\}$ sejam independentes e cada uma assume valores 0 ou 1 de acordo com a distribuição de Bernoulli: $\mathbf{P}[X_{ij} = 1] = x_i$ e $\mathbf{P}[X_{ij} = 0] = 1 - x_i$.

Escreva $S_{in} = \sum_{j=1}^n X_{ij}$ e $\mathbf{S}_n = (S_{1n}, \dots, S_{dn})$. Então,

$$\mathbf{P}[S_{kn}/n = i_k/n] = \binom{n}{i_k} x_k^{i_k} (1 - x_k)^{n-i_k}$$

Assim se

$$B_n[f](\mathbf{x}) = \sum_{0 \leq i_1, \dots, i_d \leq n} f\left(\frac{i_1}{n}, \dots, \frac{i_d}{n}\right) \prod_{k=1}^d \binom{n}{i_k} x_k^{i_k} (1 - x_k)^{n-i_k}$$

então

$$B_n[f](\mathbf{x}) = \mathbf{E}\left[f\left(\frac{\mathbf{S}_n}{n}\right)\right]$$

Provaremos esse fato agora utilizando a independência e a definição da esperança. Começamos observando que

$$\sigma_{j,n}^2 = \mathbf{Var}[S_{kn}/n] = \frac{x_k(1-x_k)}{n} \leq \frac{1}{4n}.$$

Também temos que $\mathbf{E}[\mathbf{S}_n] = n\mathbf{x}$, de onde temos

$$B_n[f](\mathbf{x}) - f(\mathbf{x}) = \mathbf{E}\left[f\left(\frac{\mathbf{S}_n}{n}\right) - f(\mathbf{x})\right] = \mathbf{E}\left[f\left(\frac{\mathbf{S}_n}{n}\right) - f\left(\mathbf{E}\left[\frac{\mathbf{S}_n}{n}\right]\right)\right].$$

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

$$\mathbf{P}\left[\left\|\frac{\mathbf{S}_n}{n} - \mathbf{x}\right\| \geq \delta\right] \leq \sum_{1 \leq k \leq d} \mathbf{P}\left[\left|\frac{S_{kn}}{n} - x_k\right| \geq \delta\right] \leq \sum_{1 \leq k \leq d} \frac{\sigma_{kn}^2}{\delta^2} \leq \frac{d}{4n\delta^2}.$$

$$\begin{aligned}
\mathbf{P}\left[\left\|\frac{\mathbf{S}_n}{n} - \mathbf{x}\right\| \geq \delta\right] &= 1 - \mathbf{P}\left[\left\|\frac{\mathbf{S}_n}{n} - \mathbf{x}\right\| < \delta\right] \\
&= 1 - \left(\prod_{1 \leq k \leq d} \mathbf{P}\left[\left|\frac{S_{kn}}{n} - x_k\right| < \delta\right]\right) \\
&= 1 - \left(\prod_{1 \leq k \leq d} \left(1 - \mathbf{P}\left[\left|\frac{S_{kn}}{n} - x_k\right| \geq \delta\right]\right)\right) \\
&= \sum_{1 \leq k \leq d} \mathbf{P}\left[\left|\frac{S_{kn}}{n} - x_k\right| \geq \delta\right] - \sum_{1 \leq k, l \leq d} \left(\mathbf{P}\left[\left|\frac{S_{kn}}{n} - x_k\right| \geq \delta\right]\right. \\
&\quad \cdot \left.\left(\mathbf{P}\left[\left|\frac{S_{ln}}{n} - x_l\right| \geq \delta\right]\right) + \dots\right. \\
&\leq \sum_{1 \leq k \leq d} \mathbf{P}\left[\left|\frac{S_{kn}}{n} - x_k\right| \geq \delta\right]
\end{aligned}$$

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 $\mathbf{x}$, 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. $\triangleleft$

10.5 ★ Teorema de De Moivre-Laplace

Suponha que $\{X_n\}_{n=1}^{\infty}$ sejam variáveis aleatórias i.i.d. com $\mathbf{P}(X_1 = 1) = \mathbf{P}(X_1 = -1) = 1/2$ e seja $S_n = X_1 + \cdots + X_n$.

10.15 Teorema (Teorema de Moivre-Laplace)

Se $a < b$, então quando $n \rightarrow \infty$

$$\lim_{n \rightarrow \infty} \mathbf{P}_n[|S_n - n| < x\sqrt{n}/2] = \frac{1}{\sqrt{2\pi}} \int_{-x}^x e^{-t^2/2} dt.$$

Então a expressão que queremos avaliar e estimar é

$$\mathbf{P}_{2n}[|S_{2n} - n| < x\sqrt{2n}/2].$$

que pode ser calculada:

$$\sum_{|k-n| < x\sqrt{n}/2} 2^{-2n} \binom{2n}{k} = \sum_{|j| < x\sqrt{n}/2} 2^{-2n} \binom{2n}{n+j}.$$

Seja $P_n = 2^{-2n} \binom{2n}{n}$ ser o termo binomial central e então escreva cada probabilidade binomial em termos desta probabilidade central P_n , especificamente

$$2^{-2n} \binom{2n}{n+j} = P_n \cdot \frac{n(n-1)\cdots(n-j+1)}{(n+j)\cdots(n+1)}.$$

Nomeie o fator fracionário acima como $D_{j,n}$ e reescreva-o como

$$D_{j,n} = \frac{1}{(1+j/n)(1+j/(n-1))\cdots(1+j/(n-j+1))}$$

e depois

$$\log(D_{j,n}) = - \sum_{k=0}^{j-1} \log(1 + j/(n-k)).$$

Agora use a expansão assintótica comum de dois termos para a função logaritmo $\log(1+x) = x(1+\epsilon_1(x))$. Observe que

$$\epsilon_1(x) = \frac{\log(1+x)}{x} - 1 = \sum_{k=2}^n \frac{(-1)^{k+1} x^k}{k}$$

logo $-x/2 < \epsilon_1(x) < 0$ e $\lim_{x \rightarrow 0} \epsilon_1(x) = 0$.

$$\log(D_{j,n}) = - \sum_{k=0}^{j-1} \frac{j}{n-k} \left(1 + \epsilon_1\left(\frac{j}{n-k}\right) \right).$$

Seja

$$\epsilon_{1,j,n} \sum_{k=0}^{j-1} \frac{j}{n-k} = \sum_{k=0}^{j-1} \frac{j}{n-k} \epsilon_1 \left(\frac{j}{n-k} \right).$$

Então podemos escrever

$$\log(D_{j,n}) = -(1 + \epsilon_{1,j,n}) \sum_{k=0}^{j-1} \frac{j}{n-k}.$$

Observe que j está restrito ao intervalo $|j| < x\sqrt{n/2}$ assim

$$\frac{j}{n-k} < \frac{x\sqrt{n/2}}{n-x\sqrt{n/2}} = \frac{x}{\sqrt{2n}-x}$$

e depois

$$\epsilon_{1,j,n} = \max_{|j| < x\sqrt{n/2}} \epsilon_1 \left(\frac{j}{n-k} \right) \rightarrow 0 \text{ quando } n \rightarrow \infty.$$

Escreva

$$\frac{j}{n-k} = \frac{j}{n} \cdot \frac{1}{1-k/n}$$

e, em seguida, expanda $\frac{1}{1-k/n} = 1 + \epsilon_2(k/n)$ onde $\epsilon_2(x) = 1/(1-x) - 1 = \sum_{k=1}^{\infty} x^k$ então $\epsilon_2(x) \rightarrow 0$ quando $x \rightarrow 0$. Mais uma vez k é restrito ao intervalo $|k| \leq |j| < x\sqrt{n/2}$ então

$$\frac{k}{n} < \frac{x\sqrt{n/2}}{n} = \frac{x}{\sqrt{2n}}$$

de modo que

$$\epsilon_{2,j,n} = \max_{|k| < x\sqrt{n/2}} \epsilon_2 \left(\frac{k}{n} \right) \rightarrow 0 \text{ quando } n \rightarrow \infty.$$

Então podemos escrever

$$\log(D_{j,n}) = -(1 + \epsilon_{1,j,n})(1 + \epsilon_{2,j,n}) \sum_{k=0}^{j-1} \frac{j}{n}.$$

Simplificando

$$\log(D_{j,n}) = -(1 + \epsilon_{3,j,n}) \sum_{k=0}^{j-1} \frac{j}{n} = -(1 + \epsilon_{3,j,n}) \frac{j^2}{n}$$

onde $\epsilon_{3,j,n} = \epsilon_{1,j,n} + \epsilon_{2,j,n} + \epsilon_{1,j,n} \cdot \epsilon_{2,j,n}$. Portanto $\epsilon_{3,j,n} \rightarrow 0$ quando $n \rightarrow \infty$ uniformemente em j . Exponenciando

$$D_{j,n} = e^{-j^2/n} (1 + \Delta_{j,n})$$

onde $\Delta_{j,n} \rightarrow 0$ quando $n \rightarrow \infty$ uniformemente em j .

Usando a fórmula de Stirling

$$P_n = 2^{-2n} \frac{(2n)!}{N! \cdot n!} = \frac{1}{\sqrt{n\pi}} (1 + \delta_n).$$

Resumindo

$$\begin{aligned}
 \mathbf{P}_{2n} [|S_{2n} - n| < x\sqrt{2n}/2] &= \sum_{|j| < x\sqrt{n}/2} 2^{-2n} \binom{2n}{n+j} \\
 &= \sum_{|j| < x\sqrt{n}/2} P_n \cdot D_{j,n} \\
 &= \sum_{|j| < x\sqrt{n}/2} P_n \cdot e^{-j^2/n} (1 + \Delta_{j,n}) \\
 &= (1 + \delta_n) \sum_{|j| < x\sqrt{n}/2} \frac{1}{\sqrt{2\pi}} \cdot e^{-j^2/n} \sqrt{\frac{2}{n}}
 \end{aligned}$$

Faça a mudança de variáveis $t_j = j\sqrt{2/n}$, $\Delta t = t_{j+1} - t_j = \sqrt{2/n}$ então a soma está acima do intervalo $-x < t_j < x$. Então

$$\mathbf{P}_{2n} [|S_{2n} - n| < x\sqrt{2n}/2] = (1 + \delta_n) \sum_{-x < t_j < x} \frac{1}{\sqrt{2\pi}} e^{-t_j^2/2} \Delta t.$$

O fator à direita é a soma aproximada para a integral de a densidade normal padrão ao longo do intervalo $[-x, x]$. Assim sendo,

$$\lim_{n \rightarrow \infty} \mathbf{P}_{2n} [|S_{2n} - n| < x\sqrt{2n}/2] = \frac{1}{\sqrt{2\pi}} \int_{-x}^x e^{-t^2/2} dt.$$

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

$$[X]^M := \begin{cases} X, & \text{se } |X| \leq M \\ 0, & \text{se } |X| > M \end{cases} = X \cdot \mathbf{1}_{\{|X| \leq M\}}.$$

Então

- $|[X]^M| \leq M$ pontualmente, e $|[X]^M| \leq |X|$
- $[X]^M \rightarrow X$ pontualmente quando $M \rightarrow \infty$,
- A variável $[X]^M$ possui todos os momentos finitos.
- Pelo Teorema da Convergência Dominada $\mathbf{E}[[X]^M] \rightarrow \mathbf{E}[X]$ quando $M \rightarrow \infty$.

Ex. 10.2 — Sejam $X_1, X_2, \dots$ variáveis aleatórias independentes com distribuição comum $\text{No}(0,1)$.

Prove que

$$\frac{X_1^2 + \dots + X_n^2}{(X_1 - 1)^2 + \dots + (X_n - 1)^2}$$

converge quase certamente e determine para qual valor.

Ex. 10.3 — Sejam $X_1, X_2, \dots$ variáveis aleatórias independentes e identicamente distribuídas, com $X_1 \sim \text{Uni}[0,1]$. 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 $X_1, X_2, \dots, X_n$ ser i.i.d. variáveis aleatórias com valor esperado

$$E[X_i] = \mu < \infty$$

e funções geradoras de momento $M_X(s)$ que é finito em algum intervalo $[-c, c]$ onde $c > 0$ é uma constante. Seja

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}. \quad (10.14)$$

Prove

$$\lim_{n \rightarrow \infty} M_{\bar{X}}(s) = e^{s\mu}, \quad \text{para todo } s \in [-c, c]. \quad (10.15)$$

Uma vez que esta é a função geradora de momento da variável aleatória constante μ , concluímos que a distribuição de $\bar{X}$ converge para μ .

Dica: use que para uma variável aleatória X com funções geradoras de momento bem definida, $M_X(s)$, temos

$$\lim_{n \rightarrow \infty} \left[M_X\left(\frac{s}{n}\right) \right]^n = e^{sE[X]}.$$

Ex. 10.5 — Sejam $X_1, X_2, \dots$ variáveis aleatórias independentes com $E[X_i] = 0$ e $\sigma_i = \text{Var}(X_i) < \infty$ 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 $X_1, X_2, \dots$ 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 de Cantor.

Ex. 10.7 — Integração de Monte Carlo

Sejam f uma função contínua com imagem no intervalo $[0,1]$ e $X_1, c_1, X_2, c_2, \dots$ uma sequência de variáveis aleatórias independentes uniformes no $[0,1]$. Tomamos

$$Y_i = \begin{cases} 1, & \text{se } f(X_i) > c_i \\ 0, & \text{se } f(X_i) \leq c_i. \end{cases}$$

Prove que:

$$\frac{1}{n} \sum_{i=1}^n Y_i \rightarrow \int_0^1 f(x) dx, \quad \text{quase certamente}$$

Teorema do Limite Central

11.1 Funções Características

Nesta capítulo, estudaremos a convergência fraca de vetores 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 7.8 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 uma 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

$$f(t) = \int e^{itX} d\mathbf{P}$$

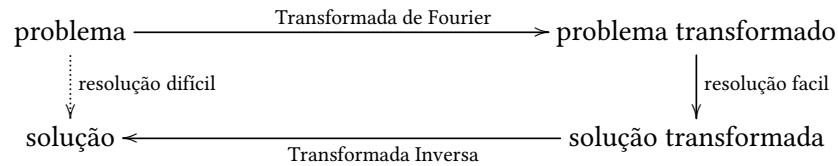
ou seja a *Transformada de Fourier* da distribuição de probabilidade e dessa forma, como veremos, conseguimos um objeto que determina exclusivamente a distribuição.

11.1 Definição (Função Característica)

Dado uma variável aleatória X definimos a **função característica** de X , denotada por φ_X , como:

$$\begin{cases} \varphi_X: \mathbb{R} \rightarrow \mathbb{C} \\ \varphi_X(t) = \mathbf{E}[e^{itX}] = \int_{\mathbb{R}} e^{itx} dF_X(x) = \int_{\mathbb{R}} e^{itx} f_X(x) d\mu \end{cases}$$

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.



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 $f: (\Omega, \mathcal{F}, \mu) \rightarrow \mathbb{C}$ é mensurável se e somente se $f = g + ih$ onde $g, h: (\Omega, \mathcal{F}, \mu) \rightarrow \mathbb{R}$ são mensuráveis.

Partindo da medida em Ω podemos construir uma integral para as funções $f: (\Omega, \mathcal{F}, \mu) \rightarrow \mathbb{C}$ 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 $f: (\Omega, \mathcal{F}, \mu) \rightarrow \mathbb{C}$ uma função mensurável, então $f = g + ih$ onde $g, h: (\Omega, \mathcal{F}, \mu) \rightarrow \mathbb{R}$ são mensuráveis. Definimos a integral de f como:

$$\int f \, d\mu = \int g \, d\mu + i \int h \, d\mu$$

11.4 Proposição

□ Sejam f, g funções integráveis e $a, b \in \mathbb{C}$ então $af + bg$ é integrável e

$$\int (af + bg) \, d\mu = a \int f \, d\mu + b \int g \, d\mu$$

□ Se f for integrável, então $\overline{\int f \, d\mu} = \int \bar{f} \, d\mu$

□ Se $\mu(A) < \infty$, então $\left| \int f \, d\mu \right| \leq \int |f| \, d\mu$.

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 $\left| \int f \, d\mu \right| \leq \int |f| \, d\mu$. $\triangleleft$

11.5 Definição

Seja μ uma medida de probabilidade em $\mathbb{R}^n$. A **transformada de Fourier**, denotada por $\hat{\mu}$, é a função complexa em $\mathbb{R}^n$ definida por

$$\hat{\mu}(u) = \int e^{i\langle u, x \rangle} \, d\mu(x) = \int \cos(\langle u, x \rangle) \, d\mu(x) + i \int \sin(\langle u, x \rangle) \, d\mu(x)$$

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.

11.6 Teorema

Seja μ uma medida de probabilidade, então $\hat{\mu}$ existe e é uma função limitada uniformemente contínua com $\hat{\mu}(0) = 1$.

Demonstração. Para ver que $\hat{\mu}$ existe, usaremos representação

$$\hat{\mu}(u) = \int \cos(\langle u, x \rangle) \, d\mu(x) + i \int \sin(\langle u, x \rangle) \, d\mu(x)$$

e que $|\cos \theta| \leq 1$ e $|\sin \theta| \leq 1$ para concluir que ambas as integrais são limitadas.

Para ver que $\hat{\mu}(0) = 1$, basta calcular

$$\hat{\mu}(0) = \int \cos(\langle 0, x \rangle) \, d\mu(x) + i \int \sin(\langle 0, x \rangle) \, d\mu(x) = \int d\mu(x) = 1$$

A limitação é um cálculo simples

$$|\widehat{\mu}(u)| \leq \int |e^{i\langle u, x \rangle}| \, d\mu(x) = \int d\mu(x) = 1$$

Por último, para provar a continuidade uniforme, primeiro observe que para qualquer $u, v \in \mathbb{R}^n$, temos

$$\begin{aligned} |e^{i\langle u, x \rangle} - e^{i\langle v, x \rangle}|^2 &= |e^{i\langle u-v, x \rangle} - 1|^2 \\ &= (\cos(\langle u-v, x \rangle) - 1)^2 + \sin^2(\langle u-v, x \rangle) \\ &= 2(1 - \cos(\langle u-v, x \rangle)) \\ &\leq \langle u-v, x \rangle^2 \\ &\leq \|u-v\|^2 \|x\|^2 \end{aligned} \quad \text{por Cauchy Schwartz}$$

Por outro lado, é claro a partir da desigualdade triangular que

$$|e^{i\langle u, x \rangle} - e^{i\langle v, x \rangle}| \leq |e^{i\langle u, x \rangle}| + |e^{i\langle v, x \rangle}| \leq 2$$

e, portanto, temos o seguinte limitante $|e^{i\langle u, x \rangle} - e^{i\langle v, x \rangle}| \leq \|u-v\| \|x\| \wedge 2$. Observe que, em um ponto, $x \in \mathbb{R}^n$, $\lim_{n \rightarrow \infty} \frac{1}{n} \|x\| \wedge 2 = 0$ e trivialmente $\frac{1}{n} \|x\| \wedge 2 \leq 2$ então o Teorema da Convergência Dominada mostra que $\lim_{n \rightarrow \infty} \int \frac{1}{n} \|x\| \wedge 2 \, d\mu(x) = 0$. Dado um $\epsilon > 0$, escolha $N > 0$ tal que $\int \frac{1}{N} \|x\| \wedge 2 \, d\mu(x) < \epsilon$ então para $\|u-v\| \leq \frac{1}{N}$,

$$\begin{aligned} |\widehat{\mu}(u) - \widehat{\mu}(v)| &\leq \int |e^{i\langle u, x \rangle} - e^{i\langle v, x \rangle}| \, d\mu(x) \\ &\leq \int \|u-v\| \|x\| \wedge 2 \, d\mu(x) \\ &\leq \int \frac{1}{N} \|x\| \wedge 2 \, d\mu(x) < \epsilon \end{aligned}$$

provando continuidade uniforme. ◁

11.7 Definição

Seja X um vetor aleatório tomando valores em $\mathbb{R}^n$. A **função característica**, denotada φ_X , é a função complexa tomando valores em $\mathbb{R}^n$ definida por

$$\begin{aligned} \varphi_X(u) &= \mathbf{E}[e^{i\langle u, X \rangle}] \\ &= \int e^{i\langle u, x \rangle} \mathbf{P}^X(dx) = \widehat{\mathbf{P}^X}(u) \end{aligned}$$

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.

Para mostrar o que devemos esperar, vamos supor que podemos escrever uma série de potência:

$$e^{itX} = \frac{i^n X}{n!} t^n$$

e desta forma

$$\hat{\mu}(t) = \int e^{itX} d\mu = \sum_{n=0}^{\infty} \frac{i^n M_n}{n!} t^n$$

Ainda trabalhando formalmente, vemos que podemos diferenciar as séries com respeito a t para isolar cada momento individual M_n

$$\frac{d^n}{dt^n} \hat{\mu}(0) = i^n M_n$$

Claramente, isso motiva o seguinte teorema:

11.8 Teorema (Momentos e Derivadas)

Seja μ uma medida de probabilidade em $\mathbb{R}^n$ tal que $f(x) = |x|^m$ é integrável em relação a μ . Então $\hat{\mu}$ tem derivadas parciais contínuas até a ordem m e

$$\frac{\partial^m \hat{\mu}}{\partial x_{j_1} \dots \partial x_{j_m}}(u) = i^m \int x_{j_1} \dots x_{j_m} e^{i\langle u, x \rangle} \mu(dx)$$

Demonstração. Primeiro, demonstraremos para $m = 1$. Escolha $1 \leq j \leq n$ e tome $v \in \mathbb{R}^n$ o vetor da base canônica, i.e., $e_j = 1$ e $e_i = 0$ para $i \neq j$. Então por $u \in \mathbb{R}^n$ e $t > 0$,

$$\begin{aligned} \frac{\hat{\mu}(u + te_j) - \hat{\mu}(u)}{t} &= \frac{1}{t} \int e^{i\langle u + te_j, x \rangle} - e^{i\langle u, x \rangle} d\mu(x) \\ &= \frac{1}{t} \int e^{i\langle u, x \rangle} (e^{itx_j} - 1) d\mu(x) \end{aligned}$$

Agora note que

$$\begin{aligned} \left| \frac{1}{t} e^{i\langle u, x \rangle} (e^{itx_j} - 1) \right|^2 &= \left| \frac{e^{itx_j} - 1}{t} \right|^2 \\ &= \frac{\cos^2(tx_j) - 2\cos(tx_j) + 1 + \sin^2(tx_j)}{t^2} \\ &= 2 \left(\frac{1 - \cos(tx_j)}{t^2} \right) \\ &\leq x_j^2 \end{aligned}$$

Mas assumimos que $|x_j|$ é integrável, portanto, podemos aplicar o Teorema da Convergência Dominada para ver

$$\begin{aligned} \frac{\partial}{\partial x_j} \int e^{i\langle u, x \rangle} d\mu(x) &= \lim_{t \rightarrow 0} \frac{1}{t} \int e^{i\langle u + te_j, x \rangle} - e^{i\langle u, x \rangle} d\mu(x) \\ &= \int \lim_{t \rightarrow 0} \frac{e^{i\langle u + te_j, x \rangle} - e^{i\langle u, x \rangle}}{t} d\mu(x) \\ &= i \int x_j e^{i\langle u, x \rangle} d\mu(x) \end{aligned}$$

A continuidade da derivada segue da fórmula que acabamos de provar. Suponha que $u_n \rightarrow u \in \mathbb{R}^n$. Então

$$\frac{\partial}{\partial x_j} \widehat{\mu}(u_n) = i \int x_j e^{i\langle u_n, x \rangle} d\mu(x)$$

e assim temos o seguinte limitante $|x_j e^{i\langle u_n, x \rangle}| < |x_j|$ com $|x_j|$ integrável por hipótese. Aplicando o Teorema da Convergência Dominada vemos que

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{\partial}{\partial x_j} \widehat{\mu}(u_n) &= i \int \lim_{n \rightarrow \infty} x_j e^{i\langle u_n, x \rangle} d\mu(x) \\ &= i \int x_j e^{i\langle u, x \rangle} d\mu(x) \\ &= \frac{\partial}{\partial x_j} \widehat{\mu}(u) \end{aligned}$$

◁

A chave para a compreensão da relação entre convergência fraca e funções características é uma propriedade básica de Transformadas de Fourier denominada de Teorema de Plancherel. Em nosso caso particular O Teorema de Plancherel mostra que podemos calcular as integrais de funções contínuas em respeito as medidas de probabilidade bem como em respeito as Transformadas de Fourier.

11.9 Teorema (Teorema de Plancherel)

Seja $\rho_\epsilon(x)$ a densidade gaussiana com variância ϵ^2 :

$$\rho_\epsilon(x) = \frac{1}{\epsilon\sqrt{2\pi}} e^{-\frac{x^2}{2\epsilon^2}}.$$

Dado o espaço de probabilidade $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu)$ e uma função integrável $f : \mathbb{R} \rightarrow \mathbb{R}$, então para qualquer $\epsilon > 0$,

$$\int_{-\infty}^{\infty} f * \rho_\epsilon(x) d\mu(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-\frac{\epsilon^2 u^2}{2}} \widehat{f}(u) \overline{\widehat{\mu}(u)} du$$

Se além disso, $f \in C_b(\mathbb{R})$ e $\widehat{f}(u)$ é integrável então

$$\int_{-\infty}^{\infty} f d\mu = \frac{1}{2\pi} \int_{-\infty}^{\infty} \widehat{f}(u) \overline{\widehat{\mu}(u)} du$$

Demonstração. Faremos a demonstração utilizando o Teorema de Fubini (Teorema 8.9) para o cálculo da integral tripla

$$\int \int \int e^{-\frac{\epsilon^2 u^2}{2}} f(x) e^{iux} e^{-iuy} d\mu(y) dx du$$

Observamos que pelo teorema de Tonelli,

$$\begin{aligned} \int \int \int \left| e^{-\frac{\epsilon^2 u^2}{2}} f(x) e^{iux} e^{-iuy} \right| d\mu(y) dx du &= \int \int \int e^{-\frac{\epsilon^2 u^2}{2}} |f(x)| d\mu(y) dx du \\ &= \int |f(x)| dx \int e^{-\frac{\epsilon^2 u^2}{2}} du < \infty \end{aligned}$$

e, portanto, podemos usar o Teorema de Fubini para calcular através de integrais iteradas

$$\begin{aligned} \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-\frac{\epsilon^2 u^2}{2}} \widehat{f}(u) \overline{\widehat{\mu}(u)} du &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-\frac{\epsilon^2 u^2}{2}} \left(\int_{-\infty}^{\infty} f(x) e^{iux} dx \right) \left(\int_{-\infty}^{\infty} e^{-iuy} d\mu(y) \right) du \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} f(x) \left(\int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} e^{iu(x-y)} e^{-\frac{\epsilon^2 u^2}{2}} du \right) d\mu(y) \right) dx \end{aligned}$$

Agora, a integral interna é apenas a Transformada de Fourier de uma Gaussiana com média 0 e variância $\frac{1}{\epsilon^2}$. Então, por outra aplicação do Teorema de Fubini e a definição de convolução,

$$\begin{aligned} &= \frac{1}{2\pi} \int_{-\infty}^{\infty} f(x) \left(\int_{-\infty}^{\infty} \frac{\sqrt{2\pi}}{\epsilon} e^{-(x-y)^2/2\epsilon^2} d\mu(y) \right) dx \\ &= \int_{-\infty}^{\infty} f(x) \left(\int_{-\infty}^{\infty} \rho_{\epsilon}(x-y) d\mu(y) \right) dx \\ &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} f(x) \rho_{\epsilon}(x-y) dx \right) d\mu(y) \\ &= \int_{-\infty}^{\infty} f * \rho_{\epsilon}(y) d\mu(y) \end{aligned}$$

A segunda parte do teorema é apenas uma aplicação de lema ?? e a primeira parte do Teorema. Pelo lema, sabemos que por qualquer $f \in C_c(\mathbb{R}; \mathbb{R})$, temos $\lim_{\epsilon \rightarrow 0} \sup_x |f * \rho_{\epsilon}(x) - f(x)| = 0$. Então temos,

$$\lim_{\epsilon \rightarrow 0} \left| \int_{-\infty}^{\infty} f - f * \rho_{\epsilon} d\mu \right| \leq \lim_{\epsilon \rightarrow 0} \int_{-\infty}^{\infty} |f - f * \rho_{\epsilon}| d\mu \leq \lim_{\epsilon \rightarrow 0} \sup_x |f - f * \rho_{\epsilon}| = 0$$

e pela integrabilidade de $\widehat{f}(u)$, o fato de que $|\widehat{\mu}| \leq 1$ podemos usar o Teorema da Convergência Dominada para ver que

$$\begin{aligned} \lim_{\epsilon \rightarrow 0} \int_{-\infty}^{\infty} f * \rho_{\epsilon} d\mu &= \frac{1}{2\pi} \lim_{\epsilon \rightarrow 0} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\epsilon^2 u^2} \widehat{f}(u) \overline{\widehat{\mu}(u)} du \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \lim_{\epsilon \rightarrow 0} e^{-\frac{1}{2}\epsilon^2 u^2} \widehat{f}(u) \overline{\widehat{\mu}(u)} du \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \widehat{f}(u) \overline{\widehat{\mu}(u)} du \end{aligned}$$

e, portanto, temos o resultado. $\triangleleft$

Agora demonstraremos que a função característica caracteriza completamente as medidas de probabilidade.

11.10 Teorema (Teorema da Unicidade de Fourier)

Sejam μ e ν medida de probabilidade em $\mathbb{R}^n$ tais que $\widehat{\mu} = \widehat{\nu}$, então $\mu = \nu$.

Demonstração. Considere $f \in C_c(\mathbb{R})$, então sabemos pelo Lema 11.17 que $\lim_{\epsilon \rightarrow 0} \|\rho_\epsilon * f - f\|_\infty = 0$. Então, para cada $\epsilon > 0$, e usando o Teorema de Plancherel

$$\begin{aligned} \left| \int f \, d\mu - \int f \, d\nu \right| &\leq \left| \int \rho_\epsilon * f \, d\mu - \int \rho_\epsilon * f \, d\nu \right| + \int |\rho_\epsilon * f - f| \, d\mu + \int |\rho_\epsilon * f - f| \, d\nu \\ &\leq \left| \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-\frac{\epsilon^2 u^2}{2}} \widehat{f}(u) (\widehat{\mu}(u) - \widehat{\nu}(u)) \, du \right| + 2\|\rho_\epsilon * f - f\|_\infty \\ &= 2\|\rho_\epsilon * f - f\|_\infty \end{aligned}$$

Tomando o limite quando ϵ vai para 0, vemos que $\int f \, d\mu = \int f \, d\nu$ para toda função $f \in C_c(\mathbb{R})$.

Agora, considere um intervalo finito $[a, b]$ e aproximando a $\mathbb{1}_{[a, b]}$

$$f_n(x) = \begin{cases} 1 & \text{para } a \leq x \leq b \\ 0 & \text{for } x < a - \frac{1}{n} \text{ ou } x > b + \frac{1}{n} \\ n(x - a) + 1 & \text{para } a - \frac{1}{n} \leq x < a \\ 1 - n(x - b) & \text{para } b < x \leq b + \frac{1}{n} \end{cases}$$

É claro que $f_n(x)$ é decrescente em n e que $\lim_{n \rightarrow \infty} f_n(x) = \mathbb{1}_{[a, b]}$ assim pelo Teorema da Convergência Monótona temos

$$\mu([a, b]) = \lim_{n \rightarrow \infty} \int f_n \, d\mu = \lim_{n \rightarrow \infty} \int f_n \, d\nu = \nu([a, b])$$

Como a σ -álgebra de Borel é gerada pelos intervalos fechados, vemos que $\mu = \nu$. $\triangleleft$

11.11 Teorema

Seja $X = (X_1, \dots, X_n)$ um vetor aleatório com valores em $\mathbb{R}^n$. Então as variáveis aleatórias X_i são independentes se e somente se

$$\varphi_X(u_1, \dots, u_n) = \prod_{j=1}^n \varphi_{X_j}(u_j)$$

Demonstração.

Este é um corolário que permite calcular a função característica do produto e depois usar o fato de que a função característica define de forma exclusiva a distribuição. Primeiro suponha que as variáveis X_i sejam independentes. Então,

$$\varphi_X(u) = \mathbf{E}[e^{i\langle u, X \rangle}] = \mathbf{E}\left[\prod_{k=1}^n e^{iu_k X_k}\right] = \prod_{k=1}^n \mathbf{E}[e^{iu_k X_k}] = \prod_{k=1}^n \varphi_{X_k}(u_k)$$

Por outro lado, se assumirmos que $\varphi_X(u_1, \dots, u_n) = \prod_{j=1}^n \varphi_{X_j}(u_j)$, então sabemos que se escolhermos variáveis aleatórias independentes Y_j onde cada Y_j tem a mesma distribuição que X_j então pelo cálculo acima $\varphi_X(u) = \varphi_Y(u)$. Pelo Teorema 11.10 sabemos que X e Y têm a mesma distribuição. Assim, X_j são independentes e temos a igualdade das distribuições de cada X_j e Y_j . $\triangleleft$

11.12 Lema

Sejam X e Y vetores aleatórios independentes em $\mathbb{R}^n$. Então $\varphi_{X+Y}(u) = \varphi_X(u)\varphi_Y(u)$.

Demonstração. Isso decorre do cálculo

$$\begin{aligned}\varphi_{X+Y}(u) &= \mathbf{E}[e^{i\langle u, X+Y \rangle}] = \mathbf{E}[e^{i\langle u, X \rangle} e^{i\langle u, Y \rangle}] \\ &= \mathbf{E}[e^{i\langle u, X \rangle}] \mathbf{E}[e^{i\langle u, Y \rangle}] = \varphi_X(u)\varphi_Y(u)\end{aligned}$$

$\triangleleft$

11.13 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 $\sin ux$ é uma função ímpar temos que

$$\begin{aligned}\varphi_X(u) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{iux} e^{-\frac{x^2}{2}} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} \cos ux dx + \frac{i}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} \sin ux dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} \cos ux dx\end{aligned}$$

Por outro lado, pelo Lema 11.8 e pelo fato que $x \cos ux$ é uma função par que temos

$$\begin{aligned}\frac{d\varphi_X(u)}{du} &= \frac{i}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{iux} e^{-\frac{x^2}{2}} dx \\ &= \frac{i}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-\frac{x^2}{2}} \cos ux dx - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-\frac{x^2}{2}} \sin ux dx \\ &= \frac{-1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-\frac{x^2}{2}} \sin ux dx\end{aligned}$$

Esta última integral pode ser integrada por partes (faça $df = x e^{-\frac{x^2}{2}} dx$ e $g = \sin ux$, portanto $f = -e^{-\frac{x^2}{2}}$ e $dg = u \cos ux$) para obter

$$\frac{d\varphi_X(u)}{du} = \frac{-u}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} \cos ux dx$$

e, portanto, mostramos que a função característica satisfaz a equação diferencial de primeira ordem simples $\frac{d\varphi_X(u)}{du} = -u\varphi_X(u)$ que tem a solução geral $\varphi_X(u) = C e^{-\frac{u^2}{2}}$ para alguma constante C .

Deixamos a determinação dessa constante como exercício.

<

11.14 Lema

Seja X um vetor aleatório em $\mathbb{R}^d$ então para $a \in \mathbb{R}^r$ e A uma matriz $r \times d$, temos

$$\varphi_{a+AX}(u) = e^{i\langle a, u \rangle} \varphi_X(A^*u)$$

onde A^* denota a transposição de A .

Demonstração. Este é um cálculo simples

$$\varphi_{a+AX}(u) = \mathbf{E}[e^{i\langle u, a+AX \rangle}] = \mathbf{E}[e^{i\langle u, a \rangle} e^{i\langle u, AX \rangle}] = e^{i\langle u, a \rangle} \mathbf{E}[e^{i\langle A^*u, X \rangle}] = e^{i\langle a, u \rangle} \varphi_X(A^*u)$$

onde usamos o fato elementar da álgebra linear que

$$\langle u, Av \rangle = u^*Av = (u^*A)^*v = v^*A^*u = \langle A^*u, v \rangle$$

<

11.15 Exemplo

Seja X uma variável aleatória. Então $\varphi_X(u) = e^{iu\mu - \frac{1}{2}u^2\sigma^2}$. Sabemos que se Y é uma variável aleatória $N(0,1)$, então $\mu + \sigma Y$ é $N(\mu, \sigma^2)$, então pelo lema anterior 11.14 e Exemplo 11.13

$$\varphi_X(u) = e^{iu\mu} \varphi_Y(\sigma u) = e^{iu\mu - \frac{1}{2}u^2\sigma^2}$$

<

O último pedaço do quebra-cabeça que precisamos montar antes de demonstrar o teorema do limite central é um resultado que mostra que podemos testar convergência em distribuição, observando a convergência funções características associadas.

11.16 Teorema (Teorema de Continuidade Glivenko-Levy)

Se $\mu, \mu_1, \mu_2, \dots$ são medidas de probabilidade em $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$, então μ_n converge fracamente para μ se e somente se $\hat{\mu}_n(u)$ converge para $\hat{\mu}(u)$ pontualmente.

Para a demonstração desse Teorema usaremos o seguinte lema que nos diz que podemos aproximar uniformemente funções contínuas de suporte compacto por convoluções com Gaussianas.

11.17 Lema

Defina $\rho(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$ e seja $\rho_n(x) = n\rho(nx)$. Seja $f \in C_c(\mathbb{R})$ então defina $f_n(x) = (f * \rho_n)(x)$. Então $f_n(x) \in C_c^\infty(\mathbb{R})$ e f_n converge para f uniformemente.

Demonstração. Pelo Teorema 8.32 basta demonstrar para $f \in C_c^\infty(\mathbb{R}^n, \mathbb{R})$, temos $\lim_{n \rightarrow \infty} \int f \, d\mu_n = \int f \, d\mu$. Por 11.17 sabemos que $\lim_{\epsilon \rightarrow 0} \|\rho_\epsilon * f - f\|_\infty = 0$. Escolha $\delta > 0$ e encontre $\epsilon > 0$ tal que

$\|\rho_\epsilon * f - f\|_\infty < \delta$. Agora,

$$\begin{aligned} & \left| \int f \, d\mu_n - \int f \, d\mu \right| \\ & \leq \left| \int (f - \rho_\epsilon * f) \, d\mu_n \right| + \left| \int \rho_\epsilon * f \, d\mu_n - \int \rho_\epsilon * f \, d\mu \right| + \left| \int (\rho_\epsilon * f - f) \, d\mu \right| \\ & \leq \delta + \frac{1}{2\pi} \left| \int \widehat{f}(t) e^{-\frac{1}{2}\epsilon^2 t^2} (\widehat{\mu}_n(t) - \widehat{\mu}(t)) \, dt \right| + \delta \end{aligned}$$

onde usamos o Teorema de Plancherel (11.9) e a aproximação uniforme de f por $\rho_\epsilon * f$.

Como f é compacta, sabemos que $\widehat{f}(t) \leq \|f\|_\infty$ e junto com Lema 11.6 vemos que

$$\left| \widehat{f}(t) e^{-\frac{1}{2}\epsilon^2 t^2} (\widehat{\mu}_n(t) - \widehat{\mu}(t)) \right| \leq 2\|f\|_\infty e^{-\frac{1}{2}\epsilon^2 t^2}$$

onde o limite superior é uma função integrável de t . Portanto, pelo Teorema da Convergência Dominada vemos que $\limsup_{n \rightarrow \infty} \left| \int f \, d\mu_n - \int f \, d\mu \right| \leq 2\delta$. Como $\delta > 0$ é arbitrário, temos $\int f \, d\mu_n = \int f \, d\mu$. $\triangleleft$

11.2 Teorema do Limite Central

11.18 Teorema (Teorema do Limite Central)

Sejam $X, X_1, X_2, \dots$ variáveis aleatórias i.i.d com $\mu = \mathbf{E}[X]$ e $\sigma = \text{Var}[X_n] < \infty$, 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 o teorema em $\frac{X_i - \mu}{\sigma}$, basta 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 11.16 é suficiente mostrar que

$$\lim_{n \rightarrow \infty} \mathbf{E} \left[e^{itS_n/\sqrt{n}} \right] = e^{t^2/2}$$

Para calcular o limite, observamos que, por independência e i.i.d. temos

$$\mathbf{E} \left[e^{itS_n/\sqrt{n}} \right] = \prod_{k=1}^n \mathbf{E} \left[e^{itX_k/\sqrt{n}} \right] = \left[\mathbf{E} \left[e^{itX/\sqrt{n}} \right] \right]^n$$

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 $\left| \frac{d}{dx} e^{ix} \right| \leq 1$, temos que $|R(x)| \leq \frac{1}{6} |x|^3$. Observe que essa estimativa não é muito boa para valores grandes de $|x|$, mas é fácil de torná-la melhor para $|x| > 1$ usando a desigualdade triangular

$$\left| e^{ix} - 1 - ix + \frac{1}{2}x^2 \right| \leq 2 + |x| + \frac{1}{2}x^2 \leq \frac{7}{2}x^2$$

Portanto, temos o limite $|R(x)| \leq \frac{7}{2}(|x|^3 \wedge x^2)$. Aplicando a expansão de Taylor e usando a hipótese de média zero e variância um, 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

$$\begin{aligned} n \left| \mathbf{E}\left[R\left(\frac{tX}{\sqrt{n}}\right)\right] \right| &\leq \frac{7}{2} \mathbf{E}\left[\frac{t^3 |X|^3}{\sqrt{n}} \wedge t^2 X^2\right] \\ &\leq \frac{7}{2} \mathbf{E}[t^2 X^2] = \frac{7t^2}{2} \end{aligned}$$

Pela desigualdade acima mencionada e pelo Teorema da Convergência Dominada, podemos concluir que

$$\lim_{n \rightarrow \infty} n \left| \mathbf{E}\left[R\left(\frac{tX}{\sqrt{n}}\right)\right] \right| = 0$$

então se definimos $\epsilon_n = \frac{2n}{t^2} \left| \mathbf{E}\left[R\left(\frac{tX}{\sqrt{n}}\right)\right] \right|$ então temos $\lim_{n \rightarrow \infty} \epsilon_n = 0$ e

$$\lim_{n \rightarrow \infty} \mathbf{E}\left[e^{itS_n/\sqrt{n}}\right] e = \lim_{n \rightarrow \infty} \left(1 - \frac{t^2}{2n}(1 + \epsilon_n)\right)^n = \lim_{n \rightarrow \infty} e^{n \log(1 - \frac{t^2}{2n}(1 + \epsilon_n))} = e^{-t^2/2}$$

◁

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	$e^{i\mu t - \frac{(\sigma t)^2}{2}}$

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 $X_1, X_2, \dots, X_n$ variáveis aleatórias independentes e identicamente distribuídas, de Poisson de parâmetro λ . Calcule a distribuição de $X_1 + X_2 + \dots + X_n$

Ex. 11.5 — O objetivo neste problema é provar o Teorema do Limite Central usando funções Características.

Seja $\{X_i\}_{i \geq 1}$ sequência de variáveis aleatórias i.i.d com $\mathbf{E}[X_1] < \infty$, $S_n = X_1 + \dots + X_n$ e $\mathbf{E}[X_1] = \mu$. Então $\frac{S_n}{n} \xrightarrow{P} \mu$ tal que para todo $\varepsilon > 0$

$$\mathbf{P}\left(\left|\frac{S_n}{n} - \mu\right| \geq \varepsilon\right) \xrightarrow{n \rightarrow \infty} 0$$

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:

$$\varphi(t) = 1 + it\mu + r(t), \quad \text{quando } t \rightarrow 0$$

$$\text{Com } \frac{r(t)}{t} \xrightarrow{t \rightarrow 0} 0$$

Dica: Polinômio de Taylor

3. Conclua que

$$\varphi_{S_n/n}(t) = \left[1 + i\frac{t}{n}\mu + r\left(\frac{1}{n}\right)\right]^n \rightarrow e^{it\mu}$$

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

$$\lim_{n \rightarrow \infty} \sum_{k=0}^n e^{-n} \frac{n^k}{k!} = 1/2$$

Ex. 11.8 — Seja $M_X(s)$ finita para todo $s \in [-c, c]$. Prove que

$$\lim_{n \rightarrow \infty} \left[M_X\left(\frac{s}{n}\right)\right]^n = e^{s\mathbf{E}[X]}.$$

Dica: Calcule

$$\lim_{n \rightarrow \infty} n \ln \left(M_X\left(\frac{s}{n}\right)\right) = s\mathbf{E}[X].$$

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 $X_1, X_2, \dots, X_n$ variáveis aleatórias i.i.d com valor esperado $E[X_i] = \mu < \infty$, $\text{Var}(X_i) = \sigma^2 < \infty$, e funções geradoras de momento $M_X(s)$ que é finito em algum intervalo $[-c, c]$, onde $c > 0$ é uma constante. Considere

$$Z_n = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{X_1 + X_2 + \dots + X_n - n\mu}{\sqrt{n}\sigma}. \quad (11.1)$$

Prove

$$\lim_{n \rightarrow \infty} M_{Z_n}(s) = e^{\frac{s^2}{2}}, \quad \text{para todo } s \in [-c, c]. \quad (11.2)$$

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.

Medidas com Sinal e Teorema de Radon-Nikodym

12.1 Medidas com Sinal

12.1 Definição

Seja $(\Omega, \mathcal{F})$ um espaço mensurável. Uma função $\nu : \mathcal{F} \rightarrow [-\infty, \infty]$ é uma medida com sinal se

- $\nu(\emptyset) = 0$
- ν só toma um dos valores $\{-\infty, \infty\}$
- se $\{E_i\}_{i=1}^{\infty}$ são uma coleção de conjuntos mensuráveis disjuntos, então

$$\nu\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} \nu(E_i)$$

onde a soma converge absolutamente se $\sum_{i=1}^{\infty} \nu(E_i) < \infty$.

12.2 Exemplo

Seja $g \in L^1(\Omega, \mathcal{F}, \mu)$ onde μ é uma medida e defina

$$\nu(E) = \int_E g d\mu = \int_E g^+ d\mu - \int_E g^- d\mu$$

◁

12.3 Exemplo

Sejam μ_1 e μ_2 medidas finitas em $(\Omega, \mathcal{F})$ e defina

$$\nu(E) = \mu_1(E) - \mu_2(E)$$

<

12.4 Definição

- Um conjunto $E \in \mathcal{F}$ é um conjunto **positivo** para a medida com sinal ν se $\nu(F) \geq 0$ para todos os $F \in \mathcal{F}$ com $F \subset E$.
- Um conjunto $E \in \mathcal{F}$ é um conjunto **negativo** para a medida com sinal ν se $\nu(F) \leq 0$ para todos os $F \in \mathcal{F}$ com $F \subset E$.
- Um conjunto $E \in \mathcal{F}$ é um conjunto **nulo** para a medida com sinal ν se $\nu(F) = 0$ para todos os $F \in \mathcal{F}$ com $F \subset E$.

O conjunto nulo é positivo e negativo. Observamos que a definição de conjunto nulo é mais exigente que a definição de conjuntos de medidas nula.

Se A é positivo então a medida com sinal μ restrita aos subconjuntos de A mensuráveis é uma medida.

12.5 Lema

Suponha que ν seja uma medida com sinal e E é um conjunto positivo. Se $F \subset E$ for mensurável, então F é um conjunto positivo. Além disso, se uma família enumerável $E_i \in \mathcal{F}$ é positiva, então $\bigcup_1^\infty E_i$ também é positiva.

Demonstração.

A primeira afirmação segue imediatamente da definição. Agora suponha que $\{E_i\}_1^\infty$ são conjuntos positivos e escreva $\bigcup_1^\infty E_i$ como uma união disjunta, $\bigcup_1^\infty F_i$ logo $F_i \subset E_i$. Além disso, se $B \subset \bigcup_1^\infty E_i = \bigcup_1^\infty F_i$, então

$$B = \bigcup_1^\infty (B \cap F_i)$$

Então,

$$\nu(B) = \sum_{i=1}^\infty \nu(B \cap F_i) \geq 0$$

como para todo i , $B \cap F_i \subset E_i$.

<

12.6 Lema

Sejam $(\Omega, \mathcal{F})$ um espaço mensurável e ν uma medida com sinal. Suponha que temos um subconjunto $E \in \mathcal{F}$ e $0 < \nu(E) < \infty$. Então, existe um conjunto positivo $P \subset E$ tal que $\nu(P) > 0$.

Demonstração. Se E for positivo, faça $E = P$ e a demonstração terminou.

Portanto, suponha que exista um subconjunto $N \subset E$ tal que $\nu(N) < 0$. Seja $\mathcal{N}(E) = \{N \in \mathcal{F} : N \subset E, \nu(N) < 0\}$. Como E não é positivo, $\mathcal{N}$ é não vazio.

Seja $n \in \mathbb{N}$ o menor número natural, tal que existe $N_1 \in \mathcal{N}(E)$ com $\nu(N_1) < -\frac{1}{n_1}$. Se $E \setminus N_1$ for positivo, estamos finalizados.

Caso contrário, proceda indutivamente. Seja n_2 o menor número inteiro positivo, de modo que existe $N_2 \in \mathcal{N}(E)$ tal que $\nu(N_2) < -\frac{1}{n_2}$. Repita para obter $N_1, \dots, N_k$ tal que $\nu(N_j) < -\frac{1}{n_j}$, $1 \leq j \leq k$ e n_{k+1} é escolhido para ser o menor inteiro positivo, de modo que existe $N_{k+1} \in \mathcal{N}(E \setminus \bigcup_{j=1}^k N_j)$ tal que $\nu(N_{k+1}) < -\frac{1}{n_{k+1}}$.

Observe que $n_{k+1} \geq n_k$ e $\{N_k\}$ é uma família de conjuntos disjuntos.

Seja

$$P = E \setminus \left(\bigcup_{k=1}^{\infty} N_k \right)$$

Nós temos

$$E = P \cup \left(\bigcup_{k=1}^{\infty} N_k \right)$$

De modo a

$$\nu(E) = \nu(P) + \sum_{k=1}^{\infty} \nu(N_k)$$

Como $\nu(E) < \infty$, obtemos que a soma no direito converge absolutamente. Assim,

$$\sum_{k=1}^{\infty} \frac{1}{n_k} < \sum_{k=1}^{\infty} |\nu(N_k)| < \infty$$

Em particular, $n_k \rightarrow \infty$ como $k \rightarrow \infty$. Nós temos

$$\nu\left(\bigcup_{k=1}^{\infty} N_k\right) < 0$$

e

$$\nu(E) > 0$$

então $\nu(P) > 0$. Seja $\varepsilon > 0$ e escolha K tão grande que

$$\frac{1}{n_{k+1} - 1} < \varepsilon, \forall k \geq K$$

Sabemos que

$$P \subset E \setminus \left(\bigcup_{k=1}^K N_k \right),$$

então, por construção, P contém nenhum conjunto mensurável F com

$$\nu(F) < \frac{-1}{n_{k+1} - 1}.$$

Portanto, P não contém subconjuntos $F \in \mathcal{F}$ com $\nu(F) < -\varepsilon$. Mas isso é verdade para todos $\varepsilon > 0$, então P deve ser positivo. $\square$

$\triangleleft$

12.7 Teorema (Decomposição de Hahn)

Seja ν ser uma medida com sinal em $(\Omega, \mathcal{F})$. Então, existe um conjunto positivo $P \subset \Omega$ e um conjunto negativo $N \subset \Omega$, tal que $\Omega = P \cup N$ e $P \cap N = \emptyset$.

Demonstração. A medida com sinal ν omite pelo menos um de $+\infty, -\infty$. Suponha $\nu(E) \neq +\infty$.

Deixe $\alpha = \sup\{\nu(A) : A \in \mathcal{S} \text{ e } A \text{ é um conjunto positivo}\}$. Sejam $\alpha_n \in \mathbb{R}$ com $\alpha_n < \alpha$ e $\alpha_n \rightarrow \alpha$. Sejam $P_n \in \mathcal{F}$ conjuntos positivos de tal forma que $\alpha_n < \nu(P_n) \leq \alpha$ para todos n (por nosso lema anterior, podemos fazer isso). Seja

$$P = \bigcup_{n=1}^{\infty} P_n.$$

Então P é um conjunto positivo assim $\nu(P) \leq \alpha$. Mas, $P \setminus P_n \subset P$, então

$$0 \geq \nu(P \setminus P_n)$$

e

$$\begin{aligned} \alpha &\geq \nu(P) = \nu(P \setminus P_n) + \nu(P_n) \\ &> \nu(P \setminus P_n) + \alpha_n \geq \alpha_n \end{aligned}$$

Por isso, tomando $n \rightarrow \infty$, $\nu(P) = \alpha$. Porque $\nu(P) \neq +\infty$, $\alpha < \infty$.

Seja $N = \Omega \setminus P$.

Suponha que $E \subset N$, $E \in \mathcal{S}$ e $\nu(E) > 0$. Pelo Lema 12.6, existe um conjunto positivo $E' \subset E$, $E' \in \mathcal{S}$. Então, $E' \cup P$ é um conjunto positivo e $E' \cap P = \emptyset$ (da definição de N). Portanto, $\alpha \leq \nu(E' \cup P) = \nu(E') + \nu(P) > \alpha$. Como $\alpha < \infty$, isso diz $\nu(E') = 0$, uma contradição. $\triangleleft$

A decomposição de Ω em P e N é denominada de Decomposição de Hahn Ω . Em geral, esta decomposição não é única. Isto é devido a existência de conjuntos nulos.

Suponha que $\{P, N\}$ seja uma Decomposição de Hahn. Deixe $\nu^+(E) = \nu(E \cap P)$ e $\nu^-(E) = -\nu(E \cap N)$. Então, ν^+ e ν^- são medidas em Ω e $\nu(E) = \nu^+(E) - \nu^-(E)$.

Teorema de Decomposição de Jordan

Não só podemos particionar o espaço Ω em sua parte positiva e negativa, também podemos dividir a própria medida com sinal em duas ou mais peças que, de alguma forma, são disjuntas ou perpendiculares

12.8 Definição (Medidas mutuamente singulares)

Duas medidas com sinais μ e ν são ditas **(mutuamente) singulares**, denotado por $\mu \perp \nu$, se houver uma partição $\{E, F = \Omega \setminus E\}$ de Ω , de modo que:

- F é nulo para μ
- E é nulo para ν .

12.9 Exemplo

A medida de contagem em $\mathbb{N} \subseteq \mathbb{R}$, e a medida de Lebesgue em $\mathbb{R}$ são medidas mutuamente singulares.

<

12.10 Teorema (Decomposição da Jordan)

Uma medida com sinal ν pode ser escrita de forma única como uma diferença $\nu = \nu^+ - \nu^-$ de duas medidas positivas que são mutuamente singulares. **Demonstração.** Seja P ser um conjunto positivo para ν obtido da decomposição de Hahn, e $N = \Omega \setminus P$ ser o conjunto negativo.

Dado E um conjunto mensurável $E \subseteq \Omega$, faça:

$$\nu^+(E) = \nu(E \cap P), \quad \nu^- = -\nu(E \cap N).$$

É claro então que ν^+ e ν^- são medidas positivas, $\nu^+ \perp \nu^-$, e $\nu = \nu^+ - \nu^-$.

Estabelecer a unicidade é direto. Suponha que ν seja uma diferença $\nu = \mu^+ - \mu^-$ de duas outras medidas positivas com $\mu^+ \perp \mu^-$. Por definição, existe um conjunto de P' em que μ^+ vive, e μ^- vive no complemento N . Então, para qualquer E mensurável,

$$\nu(E \cap P') = \mu^+(E \cap P'), \quad \nu(E \cap N') = -\mu^-(E \cap N'),$$

e isso significa que P' é positivo para ν e N' é negativo para ν .

Mas a decomposição de Hahn é única, então P, P' e N, N' diferem em ν -conjuntos nulos. Portanto, $\nu^+(E) = \nu(E \cap P) = \nu(E \cap P') = \mu^+(E \cap P') = \mu^+(E)$, então $\nu^+ = \mu^+$. E também $\nu^- = \mu^-$. <

12.2 Decomposição de Radon-Nikodym**12.11 Definição (Continuidade absoluta de medidas)**

Uma medida positiva ou com sinal ν em algum espaço mensurável é **absolutamente contínua** com relação a uma medida positiva μ no mesmo espaço mensurável se qualquer conjunto mensurável E que tenha μ -medida zero também tem ν -medida zero.

Nesse caso a medida ν é dita ser **dominada por** μ , fato que indicaremos simbolicamente por $\nu \ll \mu$.

12.12 Observação

A continuidade absoluta é a propriedade oposta a de singularidade. Se $\nu \ll \mu$ e $\nu \perp \mu$ então ν deve ser identicamente zero.

12.13 Teorema (Decomposição de Radon-Nikodym I)

Se ν e μ forem medidas finitas positivas em um espaço mensurável Ω , então ν pode ser decomposto como uma soma $\nu = \nu_a + \nu_s$ de duas medidas positivas, com $\nu_a \ll \mu$ e $\nu_s \perp \mu$.

Além disso, ν_a possui uma função de densidade em relação a μ :

$$\nu_a(A) = \int_A f \, d\mu \text{ para } A \in \mathcal{F}. \quad (12.1)$$

A função f geralmente é denominada de **derivada de Radon-Nikodym** de ν_a em relação a μ , e é denotada por $f = \frac{d\mu}{d\nu_a}$. Podemos ainda pensar em (12.1) como uma versão do Teorema Fundamental do Cálculo.

Demonstração.

Demonstração.

Seja $\mathcal{N}$ a classe das funções não negativas, integráveis com respeito a μ tal que $\int_E f d\mu \leq \nu(E)$ para todo conjunto mensurável E e defina

$$\alpha = \sup\left\{\int f d\mu : f \in \mathcal{N}\right\}.$$

Note que $\mathcal{N} \neq \emptyset$ uma vez que $0 \in \mathcal{N}$. Mais ainda,

$$0 \leq \int f d\mu \leq \nu(X) < \infty \text{ para toda } f \in \mathcal{N}.$$

o que implica que $0 \leq \alpha < \infty$.

Como α é um ponto limite do conjunto $f \in \mathcal{N}$, então existe uma sequência de funções em $\mathcal{N}$ tal que $\alpha = \lim_{n \rightarrow \infty} \int f_n d\mu$. Seja E um conjunto mensurável e n um inteiro positivo. Defina a função $g_n : X \rightarrow [0, \infty]$, por

$$g_n = \max\{f_1, f_2, \dots, f_n\}.$$

Seja

$$A_i = E \cap \left(\bigcap_{k=1, k \neq i}^n (f_i - f_k)^{-1}([0, \infty)) \right)$$

para $i = 1, 2, \dots, n$. Defina

$$E_1 = A_1, E_2 = A_2 - A_1, \dots, E_n = A_n - \bigcup_{i=1}^{n-1} A_i.$$

Assim, $E = E_1 \cup E_2 \cup \dots \cup E_n$ é uma união disjunta tal que $g_n(x) = f_i(x)$ para $x \in E_i$. Note que g_n é uma função integrável visto que $g_n = \sum_{i=1}^n f_i \mathbb{1}_{E_i}$.

Consequentemente

$$\int_E g_n d\mu = \int_E \sum_{i=1}^n f_i \mathbb{1}_{E_i} d\mu = \sum_{i=1}^n \int_E f_i \mathbb{1}_{E_i} d\mu = \sum_{i=1}^n \int_{E_i} f_i d\mu \leq \sum_{i=1}^n \nu(E_i) = \nu(E).$$

Assim, temos que $g_n \in \mathcal{N}$.

Seja $f_0 : X \rightarrow [0, \infty]$ a função definida como

$$f_0(x) = \sup\{f_n(x) : n = 1, 2, \dots\}$$

Temos que $f_0(x) = \lim_{n \rightarrow \infty} g_n(x)$

Como $\{g_n\}$ é uma sequência de funções não decrescente que converge pontualmente para f_0 , pelo teorema da convergência monótona, temos que

$$\int_X f_0 d\mu = \lim_{n \rightarrow \infty} \int_X g_n d\mu$$

Como cada $g_n \in \mathcal{N}$, então

$$\int_X g_n d\mu \leq \alpha$$

para todo n , logo

$$\int_X f_0 d\mu = \lim_{n \rightarrow \infty} \int_X g_n d\mu \leq \alpha.$$

Como $f_n \leq g_n$ para todo n , temos

$$\int_X f_n d\mu \leq \int_X g_n d\mu$$

Assim

$$\alpha = \lim_{n \rightarrow \infty} \int_X f_n d\mu \leq \lim_{n \rightarrow \infty} \int_X g_n d\mu = \int_X f_0 d\mu.$$

Portanto

$$\int_X f_0 d\mu = \alpha.$$

Mais ainda,

$$\int_E f_0 d\mu = \lim_{n \rightarrow \infty} \int_E g_n d\mu \leq \lim_{n \rightarrow \infty} \nu(E) = \nu(E).$$

Então $f_0 \in \mathcal{N}$.

Seja $\mathcal{M} \rightarrow [0, \infty]$ função dada por

$$\nu_0(E) = \nu(E) - \int_E f d\mu.$$

Observação: ν_0 é uma medida finita e que $\nu_0 \ll \mu$.

Vamos demonstrar que ν_0 é identicamente zero. Suponha que não. Então ν_0 satisfaz a hipótese do lema e portanto, existe $\varepsilon > 0$ e um conjunto mensurável A tal que $\mu(A) > 0$ e A é um conjunto positivo para $\nu_0 - \varepsilon \cdot \mu$.

Seja E mensurável, então $E \cap A \subset A$ é mensurável e como A é positivo para $\nu_0 - \varepsilon \cdot \mu$, temos que $\nu_0(E \cap A) - \varepsilon \cdot \mu(E \cap A) \geq 0$, isto é, $\varepsilon \cdot \mu(E \cap A) \leq \nu_0(E \cap A) = \nu(E \cap A) - \int_{E \cap A} f d\mu$

Se $g = f + \mathbb{1}_A$, então g é integrável e

$$\int_E g d\mu = \int_E f d\mu + \varepsilon \cdot \mu(E \cap A) = \int_{E \cap A} f d\mu + \int_{E - A} f d\mu + \varepsilon \cdot \mu(E \cap A)$$

$$\leq \int_{E-A} f d\mu + \nu(E \cap A) \leq \nu(E-A) + \nu(E \cap A) = \nu(E).$$

Então $g \in \mathcal{N}$. Mas, $\int_X g d\mu = \int_X f d\mu + \varepsilon \cdot \mu(A) > \alpha$, temos uma contradição pois α é o supremo do conjunto

$f \in \mathcal{N}$ e $\int_X g d\mu \in \left\{ \int_X f d\mu \right\}$. Então $\nu_0(E) = 0$ para qualquer conjunto mensurável E , isto é

$$\nu(E) = \int_E f d\mu$$

Agora, vamos demonstrar a unicidade de f . Seja g outra função não negativa mensurável satisfazendo 1. Como $\nu(E) < \infty$ para todo $E \in \mathcal{F}$, então

$$0 = \nu(E) - \nu(E) = \int_E g d\mu - \int_E f d\mu = \int_E (g - f) d\mu$$

para todo $E \in \mathcal{F}$. Então, temos que $f = g$ quase certamente. $\triangleleft$

$\triangleleft$

O teorema anterior pode ser generalizado:

12.14 Teorema (Decomposição de Radon-Nikodym)

Seja μ ser uma medida σ -finita em um espaço mensurável Ω . Qualquer outra medida com sinal ou σ -finita medida positiva ν em Ω pode ser decomposto unicamente como $\nu = \nu_a + \nu_s$, onde $\nu_a \ll \mu$ e $\nu_s \perp \mu$. Além disso, ν_a tem uma função de densidade em relação a μ .

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 deixe $X' = X \vee X''$. Como $p' = p + (1 - p)p''$, X' tem a mesma distribuição que a p' -moeda. Mas $X' \geq X$, como queríamos.

13.1 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

$$\|\mu\|_{TV} := 2 \sup_{A \in \mathcal{F}} \mu(A)$$

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:

$$\|\mu - \nu\|_{TV} = 2 \sup_{A \in \mathcal{F}} |\mu(A) - \nu(A)|$$

13.1 Exemplo

Para medidas num espaço finito temos

$$\|\mu - \nu\|_{TV} = \|\mu - \nu\|_1 = \sum_{\omega \in \Omega} |\mu(\omega) - \nu(\omega)|$$

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

$$|\mu - \nu| = \mathbb{1}_{D^+}(\mu - \nu) - \mathbb{1}_{D^-}(\mu - \nu)$$

13.2 Proposição

Sejam μ e ν duas medidas de probabilidade em Ω . Assim,

$$\|\mu - \nu\|_{TV} = \int_{\Omega} d|\mu(x) - \nu(x)|$$

Demonstração. Claramente

$$\|\mu - \nu\|_{TV} = 2 \sup_{A \in \mathcal{F}} |\mu(A) - \nu(A)| = 2(\mu(D^+) - \nu(D^+))$$

Como $\mu(D^+) - \nu(D^+) = \nu(D^-) - \mu(D^-)$

$$\|\mu - \nu\|_{TV} = 2 \sup_{A \in \mathcal{F}} |\mu(A) - \nu(A)| \tag{13.1}$$

$$= (\mu(D^+) - \nu(D^+)) + (\nu(D^-) - \mu(D^-)) \tag{13.2}$$

$$= \int_{\Omega} d|\mu(x) - \nu(x)| \tag{13.3}$$

◁

13.3 Proposição

Seja μ e ν duas medidas de probabilidade em Ω . Então a distância de variação total satisfaz:

$$\|\mu - \nu\|_{TV} = \sup \left\{ \int_{\Omega} f(x) \mu(x) - \int_{\Omega} f(x) \nu(x) \right. \tag{13.4}$$

$$\left. f \text{ satisfazendo } \max_{x \in \Omega} |f(x)| \leq 1 \right\} \tag{13.5}$$

Demonstração. Exercício. Dica: Use a decomposição de Hahn e a proposição anterior. $\triangleleft$

13.4 Observação

Da Proposição e da desigualdade triangular, temos, para distribuições μ , ν e γ , a seguinte desigualdade:

$$\|\mu - \nu\|_{TV} \leq \|\mu - \gamma\|_{TV} + \|\gamma - \nu\|_{TV}$$

13.2 Acoplamento

Um acoplamento de duas distribuições (ou seja, duas medidas de probabilidade) não é mais que uma distribuição conjunta delas. Mais precisamente:

13.5 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.

13.6 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 $X'' = \mathbb{1}_{U \leq q}$ e $Y'' = \mathbb{1}_{U \leq r}$. 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 $\|\mathbf{P} - \mathbf{P}'\|_{TV}$ tão pequeno quanto possível.

13.7 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 $X = X'$.

◁

13.8 Exemplo (Passeio aleatório enviesado em $\mathbb{Z}$)

Para $p \in [0, 1]$, deixe $(S_n(p))$ ser 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 $X_1, X_2, \dots$:

$$S_n(p) = X_1 + \dots + X_n$$

Assuma $0 \leq q < r \leq 1$. 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 $\hat{S}_n(p) = X''_1 + \dots + X''_n$ e $\hat{S}_n(q) = Y''_1 + \dots + Y''_n$, então $(\hat{S}_n(p), \hat{S}_n(q))$ é um acoplamento de $(S_n(p), S_n(q))$ de modo que $\hat{S}_n(q) \leq \hat{S}_n(p)$

◁

13.9 Exemplo (Acoplamento de medidas discretas)

Dadas 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

13.10 Proposição

Seja μ e ν duas medidas de probabilidade em Ω . Então

$$\|\mu - \nu\|_{TV} = \inf\{\mathbf{P}\{X \neq Y\} : (X, Y) \text{ é um acoplamento de } \mu \text{ e } \nu\}$$

Demonstração. Para todo $A \in \mathcal{F}$,

$$\mu(A) - \nu(A) = \mathbf{P}[X \in A] - \mathbf{P}[Y \in A] \quad (13.6)$$

$$= \mathbf{P}[X \in A, X = Y] + \mathbf{P}[X \in A, X \neq Y] \quad (13.7)$$

$$- \mathbf{P}[Y \in A, X = Y] - \mathbf{P}[Y \in A, X \neq Y] \quad (13.8)$$

$$= \mathbf{P}[X \in A, X \neq Y] - \mathbf{P}[Y \in A, X \neq Y] \quad (13.9)$$

$$\leq \mathbf{P}[X \neq Y], \quad (13.10)$$

◁

13.11 Exemplo (Acoplamento de variáveis aleatórias de Poisson)

Dado $X \sim \text{Po}(\lambda_1)$ e $Y \sim \text{Po}(\lambda_2)$ com $\lambda_1 > \lambda_2$. Faremos o seguinte acoplamento: deixe $\hat{Y} \sim \text{Po}(\lambda_2)$, $\hat{Z} \sim \text{Po}(\lambda_1 - \lambda_2)$ independente de Y e deixe $\hat{X} = \hat{Y} + \hat{Z}$. Como a soma de variáveis aleatórias de Poisson é de Poisson, então $(\hat{X}, \hat{Y})$ é um acoplamento e

$$\|\mu_X - \mu_Y\|_{TV} \leq \mathbf{P}[\hat{X} \neq \hat{Y}] = \mathbf{P}[\hat{Z} > 0] = 1 - e^{-(\lambda_1 - \lambda_2)} \leq \lambda_1 - \lambda_2$$

◁

Acoplamentos Maximais

13.12 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

$$\|\mu - \nu\|_{TV} = \mathbf{P}\{X \neq Y\}$$

Acoplamentos maximais existem. Por simplicidade, provamos isso apenas no caso finito.

13.13 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,

$$\|\mu - \nu\|_{TV} = \mathbf{P}\{X \neq Y\} : (X,Y) \text{ acoplamento de } \mu \text{ e } \nu$$

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:

$$\square \quad p = \sum_{x \in \Omega} \mu(x) \wedge \nu(x) = 1 - \|\mu - \nu\|_{TV}$$

$$\square \quad \alpha - \beta = \|\mu - \nu\|_{TV} = 1 - p$$

Começaremos demonstrando o primeiro fato:

$$2\|\mu - \nu\|_{TV} = \sum_{x \in \Omega} |\mu(x) - \nu(x)| \tag{13.11}$$

$$= \sum_{x \in A} [\mu(x) - \nu(x)] + \sum_{x \in B} [\nu(x) - \mu(x)] \tag{13.12}$$

$$= \sum_{x \in A} \mu(x) + \sum_{x \in B} \nu(x) - \sum_{x \in \Omega} \mu(x) \wedge \nu(x) \tag{13.13}$$

$$= 2 - \sum_{x \in B} \mu(x) + \sum_{x \in A} \nu(x) - \sum_{x \in \Omega} \mu(x) \wedge \nu(x) \tag{13.14}$$

$$= 2 - 2 \sum_{x \in \Omega} \mu(x) \wedge \nu(x) \tag{13.15}$$

$$\tag{13.16}$$

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:

$\square$ Com probabilidade p escolha $X = Y$ 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

$$\text{Finalmente } \mathbf{P}[X \neq Y] = 1-p = \|\mu - \nu\|_{TV}$$

◁

13.14 Exemplo (Acoplamento maximal 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 $\mathcal{F} = 2^\Omega$.

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

$$\mathbf{P}[(X', Y') = (i, j)] = \begin{pmatrix} p\gamma_{\min}(0) & (1-p)\gamma_A(0)\gamma_B(1) \\ 0 & p\gamma_{\min}(1) \end{pmatrix} \quad (13.17)$$

$$= \begin{pmatrix} (1-r) & r-q \\ 0 & q \end{pmatrix} \quad (13.18)$$

◁

13.3 Aplicações

Passeio Aleatório

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

$$\int_G X \, d\mathbf{P} \quad \int_G Y \, d\mathbf{P}$$

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 .

$$\text{restrição a } \mathcal{G} : (\Omega, \mathcal{F}, \mathbf{P}) \longrightarrow (\Omega, \mathcal{G}, \mathbf{P}|_{\mathcal{G}}) \quad (14.1)$$

$$\int_G X \, d\mathbf{P} \longmapsto \int_G Y \, d\mathbf{P} \quad (14.2)$$

14.1 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 $\mathbf{E}[X|\mathcal{G}]$ é uma variável aleatória Y tal que:

1 Y é $\mathcal{G}$ -mensurável.

2 Para $G \in \mathcal{G}$,

$$\int_G X d\mathbf{P} = \int_G Y d\mathbf{P}.$$

Vamos começar provando a unicidade.

14.2 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

$$\int_A Y d\mathbf{P} = \int_A Y' d\mathbf{P} \text{ para todo } A \in \mathcal{G}.$$

Então $\int_A (Y - Y') d\mathbf{P} = 0$ para todo $A \in \mathcal{G}$. Considere o conjunto A_ϵ em que $Y - Y' > \epsilon$ para $\epsilon > 0$. Então

$$0 = \int_{A_\epsilon} (Y - Y') d\mathbf{P} \geq \mathbf{P}(A_\epsilon) \cdot \epsilon.$$

Então, $\mathbf{P}(A_\epsilon) = 0$ para todo $\epsilon > 0$. Como $\{Y - Y' > 0\} = \bigcap_{n=1}^{\infty} A_{\frac{1}{n}}$. Então, a propriedade de continuidade das medidas de probabilidade implica em

$$\mathbf{P}(Y - Y' > 0) = 0.$$

Da mesma forma, trocando as funções de Y e Y' , $\mathbf{P}(Y' - Y > 0) = 0$. Logo $Y = Y'$ 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.

14.3 Exemplo

Seja G ser um evento tal que $\mathbf{P}(G) > 0$ e $\mathcal{G} = \sigma(\{G\}) = \{\emptyset, G, G^c, \Omega\}$. Então,

$$\mathbf{E}[X|\mathcal{G}](\omega) = \begin{cases} \mathbf{E}[X|G], & \text{se } \omega \in G \\ \mathbf{E}[X|G^c], & \text{se } \omega \in G^c. \end{cases} = Y(\omega)$$

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:

14.4 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 $\sigma(P)$ seja a sigma-álgebra correspondente.

Então $\mathbf{E}[X|\sigma(P)]$ 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 14.1.

◁

14.5 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 dados para determinar quanto dinheiro temos em cada semana. Nesse caso a esperança condicional é o único modo de aglutinar essas informações preservando as esperanças.

14.1 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

$$\|X\|_2 = (\mathbf{E}[X^2])^{1/2}.$$

Como é um espaço de Hilbert, a norma é realmente dada por um produto interno, que é

$$\langle X, Y \rangle = \mathbf{E}[XY].$$

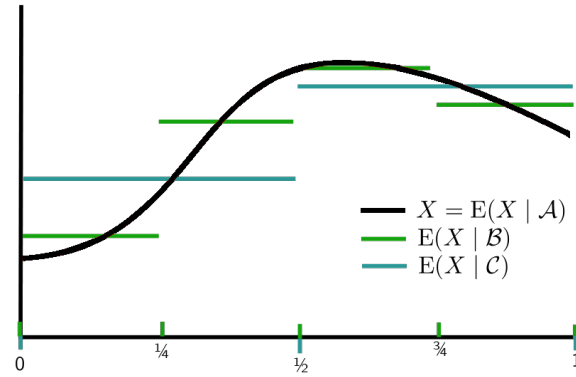
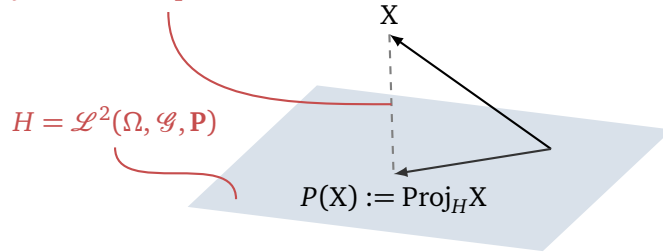


Figura 14.1: A esperança condicional em relação a uma σ -álgebra: neste exemplo, o espaço de probabilidade $(\Omega, \mathcal{G}, \mathbf{P})$ é o intervalo $[0,1]$ com a medida Lebesgue. Definimos as seguintes σ -álgebras: $\mathcal{A} = \mathcal{G}$; $\mathcal{B}$ é a σ -álgebra gerada pelos intervalos 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.

14.6 Lema

$\mathcal{L}^2(\Omega, \mathcal{G}, \mathbf{P})$ é um subspaço fechado de $\mathcal{L}^2(\Omega, \mathcal{F}, \mathbf{P})$.

Projeção usando produto interno



14.7 Teorema

Seja $(\Omega, \mathcal{G}, \mathbf{P})$ um espaço de medida, e $\mathcal{G}$ uma sub- σ -álgebra de $\mathcal{G}$. Seja

$$P : \mathcal{L}^2(\Omega, \mathcal{F}, \mathbf{P}) \rightarrow \mathcal{L}^2(\Omega, \mathcal{G}, \mathbf{P})$$

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.

14.8 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. Concluimos 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:

$$\int_A Y \, d\mathbf{P} - \int_A Y_2 \, d\mathbf{P} - \int_A X^+ \, d\mathbf{P} - \int_A X^- \, d\mathbf{P} = \int_A X \, d\mathbf{P}.$$

<

Alguns exemplos:

14.9 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: $\mathcal{G} = \sigma(\Omega_1, \Omega_2, \dots)$. 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.

◁

14.2 Esperança Condicional em Relação a uma Variável Aleatória

14.10 Definição

Dadas duas variáveis aleatórias definimos

$$\mathbf{E}[X|Y] := \mathbf{E}[X|\sigma(Y)].$$

onde $\sigma(Y) = \{\{\omega : Y(\omega) \in B\}, B \text{ Borel}\}$.

14.11 Exemplo

Por exemplo, se Y é uma variável aleatória que toma um conjunto finito de valores $y_1, \dots, y_n$. Então $\sigma(Y) = \sigma(\Omega_1, \dots, \Omega_n)$, 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 $\omega \in \Omega_k$.

◁

14.3 Propriedades da Esperança Condicional

14.12 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:

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.

2 Para todo $A \in \mathcal{G}$,

$$\begin{aligned} \int_A (a\mathbf{E}[X|\mathcal{G}] + b\mathbf{E}[Y|\mathcal{G}]) d\mathbf{P} &= a \int_A \mathbf{E}[X|\mathcal{G}] d\mathbf{P} + b \int_A \mathbf{E}[Y|\mathcal{G}] d\mathbf{P} \text{ por linearidade} \\ &= a \int_A X d\mathbf{P} + b \int_A Y d\mathbf{P} \text{ por definição} \\ &= \int_A (aX + bY) d\mathbf{P} \text{ por linearidade.} \end{aligned}$$

◁

14.13 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.

14.14 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

$$\begin{aligned} \int_A \mathbf{E}[X|\mathcal{G}] d\mathbf{P} &= \int_A X d\mathbf{P} \\ &\leq \int_A Y d\mathbf{P} \\ &= \int_A \mathbf{E}[Y|\mathcal{G}] d\mathbf{P} \end{aligned}$$

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..

◁

14.15 Corolário

$|\mathbf{E}[X|\mathcal{G}]| \leq \mathbf{E}[|X| | \mathcal{G}]$.

Demonstração. Observe $X \leq |X|$ e $-X \leq |X|$.

◁

Temos também um Teorema de Convergência Dominada para Esperanças Condicionais:

14.16 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,

$$\mathbf{E}[X_n|\mathcal{G}] \rightarrow \mathbf{E}[X|\mathcal{G}] \text{ q.c.}$$

Demonstração.

$$\begin{aligned} |\mathbf{E}[X_n|\mathcal{G}] - \mathbf{E}[X|\mathcal{G}]| &= |\mathbf{E}[X_n - X|\mathcal{G}]| \\ &\leq \mathbf{E}[|X_n - X| \mid \mathcal{G}] \\ &\leq \mathbf{E}[Z_n|\mathcal{G}], \text{ onde } Z_n = \sup_{k \geq n} |X_k - X| \text{ e claramente } Z_n \searrow. \end{aligned}$$

Na verdade, $Z_n \searrow 0$ por hipótese. Queremos mostrar que $\mathbf{E}[Z_n|\mathcal{G}] \searrow 0$ 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:

$$\mathbf{E}[Z_n|\mathcal{G}] \downarrow Z \text{ q.c.}$$

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

$$\begin{aligned} \mathbf{E}Z &= \int_{\Omega} Z \, d\mathbf{P} \\ &\leq \int \mathbf{E}[Z_n|\mathcal{G}] \, d\mathbf{P} \\ &= \mathbf{E}Z_n, \text{ pela definição de esperança condicional.} \end{aligned}$$

Como $Z_n \downarrow 0$, o Teorema de Convergência Monótona fornece

$$\mathbf{E}[Z_n] \rightarrow 0.$$

Portanto, $\mathbf{E}[Z] = 0$. ◁

14.17 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 $\mathcal{G}$.

1 $X \cdot \mathbf{E}[Y|\mathcal{G}]$ é $\mathcal{G}$ -mensurável, porque X e $\mathbf{E}[Y|\mathcal{G}]$ são $\mathcal{G}$ -mensuráveis.

2 Para todo $A \in \mathcal{G}$, precisamos verificar que

$$\int_A X \mathbf{E}[Y|\mathcal{G}] \, d\mathbf{P} = \int_A XY \, d\mathbf{P}. \quad (14.3)$$

a) Se $X = \mathbb{1}_B$ para $B \in \mathcal{G}$, então (14.3) 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 (14.3) 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 (14.3) por X_n e apliquemos o Teorema de Convergência Monótona para completar o argumento e alcançar (14.3) por X .
- d) Para X arbitrário usaremos a decomposição $X = X^+ - X^-$ e a linearidade para provar (14.3).

Isso verifica a propriedade 2..

◁

Essa propriedade se mostrará útil quando lidarmos com martingales.

14.4 Esperança Condicional como Projeção

14.18 Proposição (Suavização/Elevação)

Se $\mathcal{G}_1 \subseteq \mathcal{G}_2$, então

$$\boxed{1} \quad \mathbf{E}[\mathbf{E}[X|\mathcal{G}_1]|\mathcal{G}_2] = \mathbf{E}[X|\mathcal{G}_1].$$

$$\boxed{2} \quad \mathbf{E}[\mathbf{E}[X|\mathcal{G}_2]|\mathcal{G}_1] = \mathbf{E}[X|\mathcal{G}_1].$$

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

$$\int_A \mathbf{E}[X|\mathcal{G}_1] d\mathbf{P} = \int_A \mathbf{E}[X|\mathcal{G}_2] d\mathbf{P}. \quad (14.4)$$

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 (14.4) são iguais à

$$\int_A X d\mathbf{P}.$$

Assim, (1.) e (2.) são verdadeiras.

◁

14.19 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.

14.20 Teorema (A esperança condicional é o melhor estimador)

Para qualquer variável aleatória $\mathcal{G}$ -mensurável Y , temos

$$\mathbf{E}[X - \mathbf{E}[X|\mathcal{G}]]^2 \leq \mathbf{E}[X - Y]^2. \quad (14.5)$$

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 $\mathcal{L}^2(\Omega, \mathcal{G}, \mathbf{P})$.

14.21 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 $= \mathbf{E}ZX - \mathbf{E}ZX = 0$. $\triangleleft$

Como consequência temos uma nova demonstração do seguinte fato:

14.22 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 14.20

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. $\triangleleft$

Acabamos de demonstrar o Teorema de Pitágoras.



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:

$$\int_a^b f(x) dx = \lim_{n \rightarrow \infty} \sum_{k=1}^n f(x_k^*) \Delta x \quad (\text{A.1})$$

onde $\Delta x = \frac{b-a}{n}$ e $x_k^* \in [x_{k-1}, x_k]$ para cada subintervalo $k \in \{1, 2, \dots, n\}$. 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 $P = \{a = x_0, x_1, \dots, x_n = b\} \in \mathcal{P}[a, b]$ e para cada $k \in \{1, 2, \dots, n\}$ 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 $\int_a^b f(x) d\alpha(x) = L$.

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 $\int_a^b f d\alpha = L$ 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 $[a, b]$.

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 sequência 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

Bibliografia

- [1] Mark (<https://mathoverflow.net/users/7392/mark>). *Demystifying the Caratheodory 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.