PROBABILIDADE

Tópicos para um Curso Avançado

Pedro A. Morettin

Christophe F. Gallesco

Edição dos Autores, 2026

PROBABILIDADE

Tópicos para um Curso Avançado

PEDRO A. MORETTIN

Universidade de São Paulo

CHRISTOPHE F. GALLESCO

Universidade Estadual de Campinas

Dados Internacionais de Catalogação na Publicação (CIP)
(Câmara Brasileira do Livro, SP, Brasil)

Morettin, Pedro A.
Probabilidade [livro eletrônico] : tópicos para um curso avançado / Pedro A. Morettin, Christophe F. Gallesco. -- São Paulo : Ed. dos Autores, 2026.
PDF

Bibliografia.
ISBN 978-65-01-91821-1

1. Estatística 2. Probabilidades
3. Probabilidades - Estudo e ensino I. Gallesco, Christophe F. II. Título.

26-332987.0 CDD-519.5

Índices para catálogo sistemático:

1. Probabilidades e estatística : Matemática 519.5

Eliane de Freitas Leite - Bibliotecária - CRB 8/8415

Sumário

Figura da capa: The infinite monkey theorem

The infinite monkey theorem states that a monkey hitting keys independently and at random on a typewriter keyboard for an infinite amount of time will almost surely type any given text, including the complete works of William Shakespeare.

Prefácio

> An old joke says that "if you copy from one book that is plagiarism, but if you copy from ten books, that is scholarship."From that viewpoint this is a scholarly book. If you bought all these books you would spend more than a thousand dollars but for a fraction of that cost you can have this book, the intellectual equivalent of a ginsu knife.
>
> Richard Durrett, from the preface of his 1996 book.

Partes destas notas foram organizadas ao longo dos anos em que o primeiro autor ministrou a disciplina Probabilidade Avançada, para alunos de doutorado do Programa de Pós Graduação em Estatística do IME-USP e, o segundo autor, no Programa de Pós Graduação em Estatística do IMECC-UNICAMP. Outros tópicos foram acrescentados para seminários realizados com o Grupo de Séries Temporais do IME-USP.

Há vários excelentes livros sobre probabilidade em nível avançado. Alguns, clássicos, como Loève (1963), Chung (1968) e Breiman (1968). Outros, mais contemporâneos, como Ash e Doléans-Dade (2000), Billingsley (1995) e Durrett (2019). Então, por que mais um livro? Basicamente, porque quase não há literautura em Português sobre o assunto e o primeiro autor de há muito se preocupa com isso.

Os livros citados acima, é claro, diferem entre si sob vários aspectos, mas em comum trazem parte substancial da teoria da medida, ou na forma de capítulo ou apêndice. Além disso, alguns deles, apresentam números de páginas acima de quinhentas. O presente livro é mais modesto, nesse último sentido, pois um objetivo foi colocar material que pudesse ser ministrado em dois semestres.

O leitor deste livro deverá ter tido uma disciplina de probabilidade em nível de mestrado, como em James (1981) e, idealmente, uma disciplina sobre medida e integração, ou uma disciplina avançada de análise real. Anexamos, no final do livro, um apêndice contendo conceitos básicos dessas disciplinas necessários para o melhor entendimento do texto. Respostas a problemas selecionados também são adicionados após as Referências.

Ao prepararmos nossas aulas, nos baseamos em livros existentes sobre o assunto, citados nas Referências, em especial Chung (1968) e Breiman (1968), bem como as notas de aulas e dos seminários mencionados acima. Nesse sentido, parafraseando Durrett, na citação acima, afirmamos que nenhum resultado apresentado no texto é de nossa autoria.

Finalmente, o material incluído e a ordem de apresentação são resultados da influência de três disciplinas cursadas pelo primeiro autor na University of California, Berkeley, ministradas por Warry Millar, recentemente falecido. Naqueles dias, suas aulas e os livros de Chung e Breiman foram as referências que guiaram os estudantes. Em seu obituário (Institute of Mathematical Statistics, April 2025), lemos que "Warry was outstanding as a lecturer ... and had an almost magical ability to pace his lectures so that students could take effective notes, while he kept the dialogue flowing naturally".

Este livro pode ser usado para um curso de um ou dois semestres, sendo que, em cada caso, o professor deverá escolher os capítulos mais convenientes. Para um curso semestral recomendamos os capítulos 1, 2, 3, 4, 6, 7 e 8. Para dois cursos semestrais, recomendamos, no segundo, os capítulos 5, 9, 10, 11 e 12.

Pedro A. Morettin, São Paulo,
Christophe F. Gallesco, Campinas,
dezembro de 2025

Este livro é dedicado à memória de
Pressley Warwick (Warry) Millar (1939-2024).

Capítulo 1

Preliminares

Neste capítulo, introduzimos espaços de probabilidade e variáveis aleatórias. Para melhor entendimento dos conceitos e propriedades apresentados, introduzimos noções básicas sobre medidas, integrais e alguns espaços de funções no Apêndice. As referências pertinentes a esse capítulo são Kolmogorov (1933, 1956), Chung (1968, 1974), Breiman (1968, 1992) e Billingsley (1995).

1.1 Fundamentos

As origens da teoria de probabilidades podem ser encontradas em trabalhos de Fermat, Pascal, DeMoivre, Laplace e outros em meados do século 17. Huygens escreveu o primeiro livro sobre o asunto em 1657. O interesse na época residia em jogos de dados, baralhos etc. A primeira versão do teorema de Bayes foi publicada em 1763.

Há pelo menos quatro abordagens para a definição de probabilidade. A primeira, chamada de definição clássica, devida a Laplace (Laplace, 1795), que trata de espaços amostrais finitos, no qual os eventos elementares têm a mesma probabilidade de ocorrer. A segunda, usa o conceito de frequência relativa de ocorrência de determinado evento em um grande número de repetições. Veja von Mises (1931, 1939). A terceira, usa o conceito de probabilidade subjetiva, que Gnedenko (1989) chama de de uma medida do "grau de certeza"do observador. Referências para esse assunto são de Finetti (1974) e Jeffreys (1939). Finalmente, a quarta abordagem refere-se à construção axiomática da teoria de probabilidades, devida a Kolmogorov (1933), cuja tradução em inglês está em Kolmogorov (1956).

A construção axiomática de Kolmogorov começa com um conjunto arbitrário não vazio Ω e $\mathcal{F}$ um conjunto de subconjuntos de Ω.

Axioma 1. $\mathcal{F}$ é uma álgebra (veja o Apêndice A.1);

Axioma 2. A cada evento A de $\mathcal{F}$ está associada um número real não negativo,

$P(A)$, chamado a *probabilidade do evento* A;

Axioma 3. $P(\Omega) = 1$;

Axioma 4. Se A e B são eventos disjuntos (ou mutuamente exclusivos), então

$$P(A \cup B) = P(A) + P(B). \tag{1.1}$$

Desses axiomas podemos deduzir várias propriedades e resultados que podem ser encontrados em livros como Gnedenko (1989), mas não serão considerados aqui.

A seguir, considera-se $\mathcal{F}$ uma σ-álgebra sobre Ω (veja o Apêndice A.1) e adiciona-se o seguinte axioma:

Axioma 5. Para uma sequência decrescente de eventos $A_n \in \mathcal{F}$, com $A_n \downarrow \varnothing$, a seguinte relação vale

$$\lim_{n\to\infty} P(A_n) = 0. \tag{1.2}$$

Este axioma é chamado de **Axioma da Continuidade**. A partir desse axioma pode-se provar que (Kolmogorov, 1956):

(a) $P(\bigcup_n A_n) = \sum_n P(A_n)$, se $A_n \in \mathcal{F}$ disjuntos;

(b) $P(\bigcup_n A_n) \leq \sum_n P(A_n)$, se $A_n \in \mathcal{F}$.

Uma maneira equivalente de formular o conceito de probabilidade, sem recorrer ao Axioma 5, é a seguinte, usando o conceito de medida (veja o Apêndice A.2):

Definição 1.1. Dado o espaço mensurável $(\Omega, \mathcal{F})$, uma *probabilidade* P sobre este espaço é uma medida tal que $P(\Omega) = 1$. A tripla $(\Omega, \mathcal{F}, P)$ é chamado um *espaço de probabilidade* e Ω é o *espaço amostral*. Os elementos de $\mathcal{F}$ são chamados de *eventos*.

Segue que as relações (a) e (b) decorrem naturalmente.

Com esta formulação, o Axioma 5 pode, na realidade, ser provado. O seguinte resultado (a prova a seguir é baseada em Chung (1968)) dá uma condição necessária e suficiente para que uma função de conjunto seja uma medida de probabilidade.

Teorema 1.1. *Seja P uma função de conjunto não negativa, finitamente aditiva sobre o espaço mensurável $(\Omega, \mathcal{F})$, com $P(\Omega) = 1$. Então P é uma medida de probabilidade se e somente se o Axioma da Continuidade valer.*

Prova: (a) Suponha que P seja uma medida de probabilidade, isto é, é enumeravelmente aditiva. Então temos que

$$A_n = [(A_n - A_{n+1}) \cup (A_{n+1} - A_{n+2}) \cup \cdots] \cup [\cap_{n=1}^{\infty} A_n].$$

Se os A_n são disjuntos, o último termo é vazio. Logo,

$$P(A_n) = \sum_{k=n}^{\infty} P(A_k - A_{k+1}).$$

Como a série $\sum_{k=1}^{\infty} P(A_k - A_{k+1})$ é convergente, temos que $\lim_{n\to\infty} P(A_n) = 0$, logo (1.2) é verdadeira.

(b) Suponha, agora, que (1.2) seja verdadeira e sejam $A_n, n \geq 1$, disjuntos. Então, $\bigcup_{k=n+1}^{\infty} A_k \downarrow \varnothing$ e por (1.2) $\lim_{n\to\infty} P(\bigcup_{k=n+1}^{\infty} A_k) = 0$. A aditividade finita implica que

$$P(\cup_{k=1}^{\infty} A_k) = P(\cup_{k=1}^{n} A_k) + P(\cup_{k=n+1}^{\infty} A_k) = \sum_{k=1}^{n} P(A_k) + P(\cup_{k=n+1}^{\infty} A_k),$$

e fazendo $n \to \infty$, obtemos

$$P(\cup_{k=1}^{\infty} A_k) = \lim_{n\to\infty} \sum_{k=1}^{n} P(A_k) + \lim_{n\to\infty} P(\cup_{k=n+1}^{\infty} A_k) = \sum_{k=1}^{\infty} P(A_k). \quad \square$$

Corolário 1.1. *Seja $A_n \in \mathcal{F}$, $n \geq 1$. Se $A_n \uparrow A$, então $\lim_n P(A_n) = P(A)$ e se $A_n \downarrow A$, então $\lim_n P(A_n) = P(A)$.*

Exemplo 1.1. (a) Seja $\Omega = \{\omega_1, \omega_2, \ldots\}$, $\mathcal{F} = 2^{\Omega}$ e P definida por $P(\{\omega_i\}) = p_i$, $p_i \geq 0$, $\sum_i p_i = 1$. Então, $(\Omega, \mathcal{F}, P)$ é um espaço de probabilidade *discreto* e $\{p_i, i \geq 1\}$ é chamada de *função probabilidade*.

(b) Seja $\Omega = \mathbb{R}$, $\mathcal{F} = \mathcal{B}$ e $f \geq 0$ uma função mensurável tal que $\int_{\mathbb{R}} f(x)dx = 1$. Para cada $A \in \mathcal{F}$, define $P(A) = \int_A f(x)dx$ (veja o Apêndice A.2 para a definição da integral). Então, $(\Omega, \mathcal{F}, P)$ é um espaço de probabilidade *contínuo* e f é chamada de *função densidade de probabilidade*.

Definição 1.2. Seja $(\Omega, \mathcal{F}, P)$ um espaço de probabilidade. Um conjunto $\Lambda \in \mathcal{F}$ é chamado um *conjunto nulo* se $P(\Lambda) = 0$. Uma propriedade que vale para todo $\omega \in \Omega$ exceto para ω em um conjunto nulo é dita valer *quase certamente* (q.c), *quase em toda parte* (q.t.p) ou *com probabilidade 1* (c.p 1). O espaço de probabilidade $(\Omega, \mathcal{F}, P)$ é chamado *completo* se, sempre que $A \subset B$, com $B \in \mathcal{F}$ tal que $P(B) = 0$, então $A \in \mathcal{F}$.

Teorema 1.2 *Se $(\Omega, \mathcal{F}, P)$ for um espaço de probabilidade qualquer, então existe um espaço de probabilidade $(\Omega, \overline{\mathcal{F}}, \overline{P})$, tal que $\mathcal{F} \subset \overline{\mathcal{F}}$, $\overline{P}(A) = P(A)$, se $A \in \mathcal{F}$ e $(\Omega, \overline{\mathcal{F}}, \overline{P})$ é completo.*

Prova: Considere $\overline{\mathcal{F}} = \{A \cup N : A \in \mathcal{F}, N \subset \Delta,\ \Delta$ é conjunto nulo$\}$. Mostre que $\overline{\mathcal{F}}$ é uma σ-álgebra. Para cada $B = A \cup N \in \overline{\mathcal{F}}$ defina $\overline{P}(B) = P(A)$. Mostre que esta definição não depende da escolha de $A \in \mathcal{F}$. $\square$

Exemplo 1.2. Seja $\Omega = [0,1]$, $\mathcal{F} = \mathcal{B}([0,1])$ a σ-álgebra dos conjuntos de Borel em $[0,1]$ e P a medida de Lebesgue. Esta tripla determina um espaço de probabilidade não completo. Completando este espaço obtemos *os conjuntos mensuráveis de Lebesgue*.

Teorema 1.3. *Seja $\mathcal{F}_0$ uma álgebra e $\mathcal{F}$ a σ-álgebra gerada por $\mathcal{F}_0$. Sejam P_1 e P_2 duas probabilidades definidas sobre $\mathcal{F}$, tais que $P_1(A) = P_2(A)$, para todo $A \in \mathcal{F}_0$. Então, $P_1(A) = P_2(A)$, para todo $A \in \mathcal{F}$.*

Prova: Seja $\mathcal{C}$ a classe de conjuntos para os quais $P_1(A) = P_2(A)$. Então, $\mathcal{C} \supset \mathcal{F}_0$, por hipótese. Sejam E_n conjuntos de $\mathcal{F}$ e suponha que $E_n \uparrow E$. Então, $P_1(E) = \lim_n P_1(E_n) = \lim_n P_2(E_n) = P_2(E)$. Ou seja, se $E_n \in \mathcal{C}$, com $E_n \uparrow E$, segue-se que $E \in \mathcal{C}$. De modo análogo, se $E_n \downarrow E$, $E_n \in \mathcal{C}$, então $E \in \mathcal{C}$. Logo $\mathcal{C}$ é uma classe monotônica e portanto $\mathcal{C} \supset \mathcal{F}$ (veja o Apêndice A.1). $\square$

Teorema 1.4. *Seja $\mathcal{F}_0$ uma álgebra. Seja P uma função de conjuntos não negativa sobe $\mathcal{F}_0$, finitamente aditiva, com $P(\Omega) = 1$. Suponha que, se $A_n \in \mathcal{F}_0$, com $A_n \downarrow \emptyset$, então $\lim_n P(A_n) = 0$. Seja $\mathcal{F}$ a σ-álgebra gerada por $\mathcal{F}_0$. Então, existe uma probabilidade $P^{'}$ sobre $(\Omega, \mathcal{F})$, tal que $P^{'}(A) = P(A)$, se $A \in \mathcal{F}_0$.*

Esta é uma versão do Teorema da Extensão de Carathéodory. Veja Loève (1963) para detalhes. Uma consequência desse resultado é a seguinte. Seja $\mathcal{F}_0$ a álgebra sobre $\mathbb{R}$ consistindo de reuniões finitas de intervalos disjuntos, semi-abertos à esquerda. Seja F uma função crescente, contínua à direita tal que $\lim_{x\to\infty} F(x) = 1$, $\lim_{x\to-\infty} F(x) = 0$. Defina P como no Exemplo A.1.4. do Apêndice. Então, existe uma medida de probabilidade P' sobre $\mathcal{B}$, a σ-álgebra de Borel sobre $\mathbb{R}$, que coincide com P sobre $\mathcal{F}_0$. Tal medida é denotada por dF ou $F(dx)$. Esse argumento pode ser estendido para o $\mathbb{R}^n$.

1.2 Variável aleatória

Iniciamos com a seguinte definição.

Definição 1.3. Seja $(\Omega, \mathcal{F})$ um espaço mensurável. Uma *variável aleatória* (v.a) X sobre esse espaço é uma função definida em Ω com valores em $\mathbb{R}$ tal que $X^{-1}(B) \in \mathcal{F}$, se $B \in \mathcal{B}$. Em outras palavras, X é uma função mensurável com respeito a $\mathcal{F}$.

Para qualquer função $X : \Omega \to \mathbb{R}$, não necessariamente uma v.a, a função inversa X^{-1} tem as propriedades:

$$
\begin{aligned}
X^{-1}(A^c) &= [X^{-1}(A)]^c, \\
X^{-1}(\bigcup_\alpha A_\alpha) &= \bigcup_\alpha X^{-1}(A_\alpha), \\
X^{-1}(\bigcap_\alpha A_\alpha) &= \bigcap_\alpha X^{-1}(A_\alpha).
\end{aligned}
\tag{1.3}
$$

onde α pertence a um conjunto arbitrário de índices. Veja o Problema 2.

Teorema 1.5. *X é uma v.a se e somente se $\{\omega : X(\omega) \leq x\} \in \mathcal{F}$, para todo $x \in \mathbb{R}$.*

Prova: (a) Suponha que X seja uma v.a; então, $\{\omega : X(\omega) \leq x\} \in \mathcal{F}$, pois $\{\omega : X(\omega) \leq x\} = X^{-1}((-\infty, x])$ e esse último conjunto pertence a $\mathcal{B}$, e pela definição de v.a, para qualquer conjunto de Borel B, $X^{-1}(B) \in \mathcal{F}$.

(b) Suponha, agora, que $\{\omega : X(\omega) \leq x\} \in \mathcal{F}$, isto é, para todo x, $X^{-1}((-\infty, x]) \in \mathcal{F}$. Seja $\mathcal{C}$ a coleção dos conjuntos B tais que $X^{-1}(B) \in \mathcal{F}$. Então, $\mathcal{C}$ contém conjuntos da forma $(-\infty, a]$, por hipótese. $\mathcal{C}$ é uma σ-álgebra, pois se $B \in \mathcal{C}$, então $X^{-1}(B^c) = (X^{-1}(B))^c \in \mathcal{F}$, e se $B_j \in \mathcal{C}$, para todo j, então $X^{-1}(\cup_j B_j) = \cup_j X^{-1}(B_j) \in \mathcal{F}$, usando o Lema 1.1. Segue-se que $\mathcal{C}$ contem conjuntos da forma $(-\infty, a]$, que geram $\mathcal{B}$, logo $\mathcal{C} \supset \mathcal{B}$ (pois $\mathcal{B}$ é a menor σ-álgebra contendo conjuntos da forma $(-\infty, a]$). Isso significa que $X^{-1}(B) \in \mathcal{F}$, para cada $B \in \mathcal{B}$, logo X é uma v.a. $\square$

Exemplo 1.3. Seja um espaço mensurável $(\Omega, \mathcal{F})$ e $\Lambda \in \mathcal{F}$. Defina I_Λ como segue:

$$
I_\Lambda(\omega) = \begin{cases} 1, & \text{se } \omega \in \Lambda, \\ 0, & \text{se } \omega \notin \Lambda. \end{cases}
$$

Então, I_Λ é uma v.a, chamada a *função indicadora* de Λ. Se $c_1, \ldots, c_m$ são números reais e se $X : \Omega \leftarrow \mathbb{R}$ é definida por

$$
X(\omega) = \sum_{i=1}^{m} c_i I_{\Lambda_i}(\omega), \quad \Lambda_i \in \mathcal{F},
$$

então dizemos que X é uma *variável aleatória simples.*

Exemplo 1.4. Seja $\Omega = [0,1]$, $\mathcal{F} = \mathcal{B}([0,1])$. Nesse caso, uma v.a sobre $(\Omega, \mathcal{F})$ é, por definição, uma *função de Borel* (uma função f definida em Ω é uma função de Borel se $f^{-1}(B) \in \mathcal{B}$, para todo $B \in \mathcal{B}$).

O resultado seguinte é um teorema bem conhecido na Teoria da Medida.

Teorema 1.6. *Seja X uma v.a. Então, X é um limite de v.a's simples. Se $X \geq 0$, então X é o limite de uma sequência crescente de v.a's simples.*

Definição 1.4. Se X é uma v.a, então $\mathcal{F}\{X\}$ é a menor σ-álgebra sobre Ω com respeito à qual X é mensurável.

Proposição 1.1. $\mathcal{F}\{X\} = \{\Lambda : \ \Lambda = X^{-1}(B), \ B \in \mathcal{B}\}$.

Prova: Basta usar (1.3). $\square$

Lema 1.1. *Se X é uma v.a e f é uma função mensurável de Borel sobre $(\mathbb{R}, \mathcal{B})$, então $f(X)$ é uma v.a.*

Prova: Basta encarar $f(X)$ como a aplicação composta $f \circ X : \ \omega \ \rightarrow \ f(X(\omega))$. Veja o Problema 3. $\square$

Teorema 1.7. *Uma v.a Y é $\mathcal{F}\{X\}$-mensurável se, e somente se, Y for da forma $Y = g(X)$, sendo g uma função de Borel.*

Prova: (a) Suponha que $Y = g(X)$. Queremos provar que $Y^{-1}(B) \in \mathcal{F}\{X\}$, para todo $B \in \mathcal{B}$. Mas isso é verdade pelo Lema 1.1.

(b) Suponha, agora, que Y seja $\mathcal{F}\{X\}$-mensurável. É suficiente provar o resultado para $Y \geq 0$.

(i) Primeiramente, o resultado é verdadeiro se $Y = I_A, \ A \in \mathcal{F}\{X\}$. De fato, se $A \in \mathcal{F}\{X\}$, então $A = X^{-1}(B)$, para algum $B \in \mathcal{B}$, pela Proposição 1.1. Assim, se tomarmos $g = I_B$, teremos

$$Y(u) = I_A(u) = I_{X^{-1}(B)}(u) = I_B(X(u)),$$

ou seja, $Y = g(X)$.

(ii) A seguir, o resultado é verdadeiro se Y for da forma $Y = \sum_{i=1}^{m} c_i I_{A_i}$, onde $A_i \in \mathcal{F}\{X\}$. Então $A_i = X^{-1}(B_i), \ \ B_i \in \mathcal{B}$. Se definirmos $g = \sum_{i=1}^{m} c_i I_{B_i}$, então teremos $Y = g(X)$.

(iii) Finalmente, o caso geral. Existe uma sequência de funções simples Y_n, que tende a Y, quando $n \to \infty$. Por (ii), $Y_n = g_n(X)$, para alguma função de Borel g_n. Logo, $Y = \lim_{n\to\infty} g_n(X)$. Como $g_n(X) \to Y$, segue-se que g_n converge na imagem de X. Defina g como segue:

$$g(u) = \begin{cases} \lim_{n\to\infty} g_n(u), & \text{se o limite existir,} \\ 0, & \text{caso contrário.} \end{cases}$$

Segue-se que g é uma função de Borel e $Y = g(X)$. $\square$

Passemos, agora, a estudar os conceitos de distribuição e função de distribuição de uma v.a.

Definição 1.5. Seja X uma v.a sobre $(\Omega, \mathcal{F}, P)$. A *distribuição de* X é a medida de probabilidade P_X definida sobre $(\mathbb{R}, \mathcal{B})$ de tal sorte que, para $B \in \mathcal{B}$, $P_X(B) = P\{X^{-1}(B)\} = P\{X \in B\} = P\{\omega : \ X(\omega) \in B\}$.

Observação 1. Devemos verificar que P_X assim definida é um probabilidade. Claramente, $P_X(B) \geq 0$. Se os B_n são conjuntos disjuntos em $\mathcal{B}$, então $X^{-1}(B_n)$ são disjuntos por (1.3) e

$$P_X(\cup_n B_n) = P\{X^{-1}(\cup_n B_n)\} = P\{\cup_n X^{-1}(B_n)\} = \sum_n P\{X^{-1}(B_n)\} = \sum_n P_X(B_n).$$

Finalmente, $X^{-1}(\mathbb{R}) = \Omega$, logo $P_X(\mathbb{R}) = P(\Omega) = 1$.

O espaço de probabilidade $(\mathbb{R}, \mathcal{B}, P_X)$ é chamado o *espaço de probabilidade induzido* pela v.a X. No caso em que a imagem de X é um subconjunto enumerável de $\mathbb{R}$, X é chamada de v.a *discreta* e a sua distribuição é caracterizada pela *função de probabilidade* $p_X(x) := P\{X = x\}$, para $x \in \mathbb{R}$. No caso em que existe uma função não negativa e mensurável f tal que para todo boreliano A, $P_X(A) = \int_{\mathbb{R}} f(x)dx$, X é dita *absolutamente contínua* e f é chamada de *função densidade de probabilidade* de X.

Observação 2. A v.a X determina univocamente sua distribuição P_X, mas duas v.a's distintas podem ter a mesma distribuição. Por exemplo, considere $\Omega = [0, 1]$, $\mathcal{F} = \mathcal{B}([0, 1])$, e P é a medida de Lebesgue. Considere as v.a's X e Y definidas sobre esse espaço de probabilidade por meio de

$$X(\omega) = \omega, \quad Y(\omega) = 1 - \omega.$$

Então, $P_X = P_Y = P$ (use o fato que P é invariante por translações).

Definição 1.6. A *função de distribuição* F_X de uma v.a X é a função definida por

$$F_X(x) = P\{\omega : X(\omega) \leq x\} = P_X((-\infty, x]), \quad \text{para todo real } x,$$

e que escreveremos, simplesmente, $P\{X \leq x\}$.

As seguintes propriedades de F_X são válidas.

Teorema 1.8. *Seja F_X a função de distribuição (f.d) da v.a X. Então:*

(i) F_X *é não decrescente;*

(ii) $\lim_{x \to -\infty} F_X(x) = 0, \quad \lim_{x \to \infty} F_X(x) = 1$;

(iii) F_X *é contínua à direita.*

Prova: Veja o Problema 4.

Observação 3. A função F_X pode ser definida por $F_X(x) = P\{\omega : X(\omega) < x\}$. Nesse caso, F_X é contínua à esquerda. Veja o Problema 13.

Teorema 1.9. *Sejam X e Y duas v.a's. Então $P_X = P_Y$ se, e somente se, $F_X = F_Y$.*

Prova: (a) Suponha $P_X = P_Y$; então, $P_X(B) = P_Y(B)$, para todo $B \in \mathcal{B}$. Em particular, se $B = (-\infty, x]$, obtemos $F_X(x) = F_Y(x)$.

(b) Considere a classe $\mathcal{C}$ de todos os conjuntos B, tais que $P_X(B) = P_Y(B)$. Essa classe contem conjuntos da forma $B = (-\infty, b]$, porque $F_X(x) = F_Y(x)$, para todo real x. Logo, $\mathcal{C}$ contem conjuntos da forma $(a, b]$, do que segue que $\mathcal{C}$ contem reuniões finitas de intervalos fechados à direita, disjuntos. Logo, P_X e P_Y coincidem numa álgebra que gera $\mathcal{B}$, portanto P_X e P_Y coincidem sobre $\mathcal{B}$, pelo Teorema 1.3. $\square$

Uma definição equivalente de f.d é dada a seguir.

Definição 1.7. Uma *função de distribuição* é qualquer função F não decrescente, contínua à direita, tal que $\lim_{x\to-\infty} F(x) = 0$ e $\lim_{x\to\infty} F(x) = 1$.

Um problema que se coloca é o seguinte:

Suponha que seja dada uma f.d F. Existe um espaço de probabilidade $(\Omega, \mathcal{F}, P)$ e uma v.a sobre esse espaço, tendo F como sua f.d?

A resposta é afirmativa e uma construção é a seguinte. Construa $(\Omega, \mathcal{F}, P)$ como:

$$\Omega = \mathbb{R}, \quad \mathcal{F} = \mathcal{B}, \quad P = dF.$$

Além disso, defina a v.a X por $X(\omega) = \omega$. Então, a f.d de X é F. De fato,

$$P\{\omega : X(\omega) \leq x\} = P\{\omega : \omega \leq x\} = F(x) - F(-\infty) = F(x).$$

1.3 Vetor aleatório

Nessa seção iremos generalizar os conceitos da seção anterior para o caso de termos mais de uma v.a.

Definição 1.8. Considere $X_1, \ldots, X_n$ v.a's sobre $(\Omega, \mathcal{F}, P)$. Então $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ é um *vetor aleatório.*

Observe que $\mathbf{X} : \Omega \to \mathbb{R}^n$.

Proposição 1.2. *Se $\mathcal{B}^n$ é a σ-álgebra de conjuntos de Borel do $\mathbb{R}^n$, então $\mathbf{X}^{-1}(B) \in \mathcal{F}$, para todo $B \in \mathcal{B}^n$.*

Prova: Para simplificar as notações, suponha $\mathbf{X} = (X_1, X_2)$. Considere a classe $\mathcal{C}$ de conjuntos B para os quais $\mathbf{X}^{-1}(B) \in \mathcal{F}$. Esta classe contem retângulos. De fato,

$$\begin{aligned}\mathbf{X}^{-1}(A \times B) &= \{\omega : X_1(\omega) \in A, X_2(\omega) \in B\} = \\ &= \{\omega : X_1(\omega) \in A\} \cap \{\omega : X_2(\omega) \in B\} = M \cap N.\end{aligned}$$

Logo, $\mathbf{X}^{-1}(A \times B) = M \cap N$, tal que $M \in \mathcal{F}$, $N \in \mathcal{F}$, do que segue que $\mathbf{X}^{-1}(A \times B) \in \mathcal{F}$. Além disso, essa classe $\mathcal{C}$ é uma σ-álgebra, e como a classe dos conjuntos de Borel de $\mathcal{B}^2$ é a menor σ-álgebra contendo todos os retângulos, $\mathcal{C} \supset \mathcal{B}^2$. $\square$

Definição 1.9. Seja $\mathbf{X} = (X_1, X_2, \ldots, X_n)$. Então, $\mathcal{F}\{X_1, \ldots, X_n\}$ é a menor σ-álgebra com respeito à qual todas as v.a's X_i, $i = 1, \ldots, n$, são mensuráveis.

Teorema 1.10. *Uma v.a Y é $\mathcal{F}\{X_1, \ldots, X_n\}$-mensurável se, e somente se, $Y = g(X_1, \ldots, X_n)$, onde g é uma função de Borel de $\mathbb{R}^n$ em $\mathbb{R}$.*

Prova: Como no Teorema 1.7, observando que, por exemplo, se X e Y são v.a's e f é uma função de Borel mensurável de duas variáveis, então $f(X, Y)$ é uma v.a. $\square$

Definição 1.10. A *distribuição* de $\mathbf{X}$ é a probabilidade sobre $(\mathbb{R}^n, \mathcal{B}^n)$ definida por

$$P_{\mathbf{X}}(B) = P\{\omega \in \Omega : (X_1(\omega), \ldots, X_n(\omega) \in B\}, \quad B \in \mathcal{B}^n.$$

Dois casos particulares importantes de distribuição de vetores aleatórios são os vetores aleatórios *discretos* e *absolutamente contínuos*. As definições são obtidas adaptando de maneira natural as definições dadas no caso de variáveis aleatórias.

Definição 1.11. A *função de distribuição* de $\mathbf{X}$ é a função

$$F_{\mathbf{X}}(x_1, \ldots, x_n) = P\{\omega : X_1(\omega) \leq x_1, \ldots, X_n(\omega) \leq x_n\}.$$

Como antes, podemos provar os seguintes resultados.

Teorema 1.11. *Se $\mathbf{X}$ e $\mathbf{Y}$ são dois vetores aleatórios, então $P_{\mathbf{X}} = P_{\mathbf{Y}}$ se, e somente se, $F_{\mathbf{X}} = F_{\mathbf{Y}}$.*

Teorema 1.12. *A f.d. $F_{\mathbf{X}}$ de $\mathbf{X}$ tem as seguintes propriedades*:

(i) *$F_{\mathbf{X}}$ é não decrescente em cada variável;*

(ii) *$F_{\mathbf{X}}$ é contínua à direita em cada variável;*

(iii) $\lim_{x_i \to -\infty} F_{\mathbf{X}}(x_1, \ldots, x_n) = 0$ para todo $1 \leq i \leq n$;

(iv) $\lim_{x_1, \ldots, x_n \to \infty} F_{\mathbf{X}}(x_1, \ldots, x_n) = 1$;

(v) *para* $(a_1, \ldots, a_n)$, $(b_1, \ldots, b_n) \in \mathbb{R}^n$ *tais que* $a_i \leq b_i$ *para todo* $1 \leq i \leq n$, *temos que* $\sum_{\mathbf{x}} sgn(\mathbf{x}) F_{\mathbf{X}}(\mathbf{x}) \geq 0$, *onde a soma é sobre todos os vetores* $\mathbf{x} = (x_1, \ldots, x_n)$ *tais que para todo* i, $x_i \in \{a_i, b_i\}$ *e* $sgn(\mathbf{x})$ *é igual a 1 (resp.* -1*) se o número de coordenadas do tipo* a_i *no vetor* $\mathbf{x}$ *é par (resp. ímpar).*

Como no caso de uma v.a, podemos definir uma f.d multidimensional como sendo qualquer função F satisfazendo (i)-(v) do Teorema 1.12. Note que, ao contrário do caso unidimensional, para $n \geq 2$ as propriedades (i)-(iv) não são suficientes para que uma função F seja uma f.d.

1.4 Processo estocástico

Nesta seção discutiremos um conceito mais geral do que variável aleatória ou vetor aleatório, pois teremos uma função, que além de ser indexada por um elemento de Ω, também será indexada por um elemento pertencente a um conjunto T, que usualmente será um conjunto de instantes de tempo, mas não necessariamente. Antes de definir o que seja um processo estocástico (ou função aleatória), alguns conceitos são necessários.

Definição 1.12. Seja T um conjunto arbitrário. Para cada $t \in T$, seja Ω_t um conjunto. Então, $\prod_{t \in T} \Omega_t$ é o conjunto de todas as funções $\omega : T \to \bigcup_{t \in T} \Omega_t$, tal que $\omega(t) \in \Omega_t$.

Exemplo 1.5. (a) Se $T = \{1, 2, \ldots, n\}$, então $\prod_{t \in T} \Omega_t = \Omega_1 \times \Omega_2 \times \cdots \times \Omega_n$.

(b) Se $T = \{1, 2, \ldots\}$, $\Omega_t = \mathbb{R}$, para cada $t \in T$, então $\prod_{t \in T} \Omega_t$ é o conjunto de todas as sequências de números reais.

(c) Se $T = (a, b]$, $\Omega_t = \mathbb{R}$, então $\prod_{t \in T} \Omega_t$ é o conjunto de todas as funções de $(a, b]$ em $\mathbb{R}$.

Se $\Omega_t = \Omega$, para todo t, então iremos escrever $\prod_{t \in T} \Omega_t = \Omega^T$.

Definição 1.13. (σ-álgebra produto) Seja T um conjunto de índices; para cada $t \in T$, seja $(\Omega_t, \mathcal{F}_t)$ um espaço mesurável. Um *retângulo* A é qualquer conjunto da forma $A = \prod_{t \in T} A_t$, onde $A_t = \Omega_t$, para todo t, exceto por um número finito deles. A *σ-álgebra produto* é a menor σ-álgebra sobre $\prod_{t \in T} \Omega_t$ que contém esses retângulos.

Usaremos a notação $\bigotimes_{t \in T} \mathcal{F}_t$ para denotar essa σ-álgebra produto. Se $\Omega_t = \mathbb{R}$ e $\mathcal{F}_t = \mathcal{B}$, para todo $t \in T$, então chamamos o espaço resultante de σ-álgebra de Borel sobre $\mathbb{R}^T$ e a denotamos por $\mathcal{B}^T$.

Definição 1.14. Seja $(\Omega, \mathcal{F}, P)$ um e.p e T um subconjunto de $\mathbb{R}$. Um *processo estocástico* é uma coleção de v.a's $X = \{X_t,\ t \in T\}$ definidas sobre $(\Omega, \mathcal{F}, P)$.

Dizemos que T é o conjunto paramétrico do processo.

Note que $X : \Omega \to \mathbb{R}^T$. Alguns casos especiais são:

(a) $T = \{1, 2, \ldots, n\}$, neste caso $X = \{X_k,\ 1 \leq k \leq n\}$ é um vetor aleatório de dimensão n.

(b) $T = \mathbb{N}$, neste caso $X = \{X_n,\ n \geq 1\}$ é uma sequência de variáveis aleatórias. X é um processo estocástico com parâmetro discreto.

(c) $T = [a, b]$, neste caso $X = \{X_t\ , a \leq t \leq b\}$ é um processo estocástico com parâmetro contínuo.

Dado o processo estocástico (p.e) $X :\ \Omega \to \mathbb{R}^T$, a questão que surge é: X é mensurável? A resposta é dada pelo

Teorema 1.13. *Seja $B \in \mathcal{B}^T$. Então, $X^{-1}(B) \in \mathcal{F}$.*

Prova: Similar à da Proposição 1.2. $\square$

Definição 1.15. Defina uma probabilidade P_X sobre $(\mathbb{R}^T, \mathcal{B}^T)$ por $P_X(B) = P\{X^{-1}(B)\}$. P_X é chamada de distribuição de X.

Teorema 1.14. *Seja P^* uma probabilidade sobre $(\mathbb{R}^T, \mathcal{B}^T)$. Então, existe um processo estocástico $X = \{X_t, t \in T\}$ tal que P^* é a distribuição de X e esse processo está definido sobre $(\mathbb{R}^T, \mathcal{B}^T, P^*)$.*

Prova: Se $\omega \in \mathbb{R}^T$, defina $X_t(\omega) = \omega(t)$. Lembremos que $\omega \in \mathbb{R}^T$ significa que $\omega : T \to \mathbb{R}$, isto é, $\omega(t) \in \mathbb{R}$. $\square$

Como exemplo, seja $T = \mathbb{N}$; aqui $\mathbb{R}^T = \mathbb{R} \times \mathbb{R} \times \cdots$ é o conjunto de todas as sequências de números reais, isto é, $\omega \in \mathbb{R}^T$ se $\omega = \{x_1, x_2, \ldots\}$. Então, $X_n(\omega) = x_n$.

Contudo, usualmente a situação do Teorema 1.14 não aparece. A situação comum é a seguinte. Seja dado um conjunto de índices $T \subset \mathbb{R}$; para uma sequência finita de elementos distintos $t_1, \ldots, t_n \in T$, é dada uma probabilidade $P_{t_1,\ldots,t_n}$ sobre $(\mathbb{R}^n, \mathcal{B}^n)$. Pede-se para construir um processo estocástico $\{X_t, t \in T\}$, tal que a distribuição de $(X_{t_1}, X_{t_2}, \ldots, X_{t_n})$ seja $P_{t_1,\ldots,t_n}$.

Por exemplo, frequentemente temos a seguinte situação: $X_1, X_2, \ldots$ são v.a's independentes, com f.d comum F. A questão então é: Tal sequência existe? Ou seja, existe um e.p $(\Omega, \mathcal{F}, P)$, tal que um processo estocástico $\{X_n,\ n \geq 1\}$ esteja definido sobre o mesmo?

Definição 1.16. Dado um processo estocástico $X = \{X_t,\ t \in T\}$, as probabilidades $P_{t_1,\cdots,t_n}$ são chamadas as *distribuições finito-dimensionais* do processo estocástico X.

Portanto, podemos refrasear o nosso problema da seguinte forma: seja dada uma coleção de probabilidades, que são supostas serem as distribuições finito-dimensionais de algum processo estocástico. Podemos construir um processo estocástico com essas distribuições? A resposta é afirmativa, desde que essas probabilidades sejam “consistentes”. Eis um exemplo do que entendemos por consistência. Se $P\{X_1 \in A, X_2 \in B, X_3 \in \mathbb{R}\} = P_{1,2,3}(A \times B \times \mathbb{R})$, então o primeiro membro dessa igualdade é igual a $P\{X_1 \in A, X_2 \in B\} = P_{1,2}(A \times B)$.

[C] Condições de Consistência. Para toda sequência finita $t_1, \ldots, t_n$ de elementos distintos de T, seja $P_{t_1,\cdots,t_n}$ uma probabilidade. A primeira condição de consistência estabelece que para toda permutação $\pi = (\pi(1), \ldots, \pi(n))$ de $\{1, \ldots, n\}$ e $A_1, \ldots, A_n \in \mathcal{B}$,

$$P_{t_1,\ldots,t_n}(A_1 \times \cdots \times A_n) = P_{t_{\pi(1)},\ldots,t_{\pi(n)}}(A_{\pi(1)} \times \cdots \times A_{\pi(n)}).$$

A segunda condição é

$$P_{t_1,\ldots,t_{n-1}}(A_1 \times \cdots \times A_{n-1}) = P_{t_1,\ldots,t_n}(A_1 \times \cdots \times A_{n-1} \times \mathbb{R}).$$

O Teorema de Consistência de Kolmogorov ou Teorema de Extensão de Kolmogorov-Daniell, pode ser enunciado como segue.

Teorema 1.15. *Seja T um conjunto de índices. Suponha que para cada sequência finita $t_1, \ldots, t_n$ de elementos distintos de T tenhamos uma probabilidade $P_{t_1,\ldots,t_n}$ sobre $(\mathbb{R}^{\{t_1,\ldots,t_n\}}, \mathcal{B}^{\{t_1,\ldots,t_n\}})$. Suponha que essas probabilidades satisfaçam [C]. Então, existe uma única probabilidade P sobre $(\mathbb{R}^T, \mathcal{B}^T)$, tal que P restrita a $(\mathbb{R}^{\{t_1,\ldots,t_n\}}, \mathcal{B}^{\{t_1,\ldots,t_n\}})$ seja $P_{t_1,\ldots,t_n}$.*

Para uma prova veja Billingsley (1995).

Corolário 1.2. *Sejam $X = \{X_t, t \in T\}$ e $Y = \{Y_t, t \in T\}$ dois processos estocásticos tendo as mesmas distribuições finito-dimensionais. Então, X e Y têm a mesma distribuição. Ou seja, as distribuições finito-dimensionais determinam a distribuição de um processo.*

Uma formulação alternativa do Teorema 1.15 pode ser dada em termos de f.d’s. Para cada $t_1, \ldots, t_n$ distintos contidos em T seja dada uma f.d $F_{t_1,\ldots,t_n}$. Então, existe um processo estocástico $X = \{X_t, t \in T\}$ tal que $\{X_{t_1}, \ldots, X_{t_n}\}$ tenha f.d $F_{t_1,\ldots,t_n}$, desde que essa seja consistente. Consistência, agora, significa:

(a) $\lim_{x_n \to \infty} F_{t_1,\ldots,t_n}(x_1, \ldots, x_n) = F_{t_1,\ldots,t_{n-1}}(x_1, \ldots, x_{n-1})$,
(b) para toda permutação $\pi = (\pi(1), \ldots, \pi(n))$ de $\{1, \ldots, n\}$,

$$F_{t_{\pi(1)},\ldots,t_{\pi(n)}}(x_{\pi(1)}, \ldots, x_{\pi(n)}) = F_{t_1,\ldots,t_n}(x_1, \ldots, x_n).$$

Teorema 1.16. *Seja $X = \{X_t, t \in T\}$ um processo estocástico e Y uma v.a $\mathcal{F}\{X\}$-mensurável. Suponha que T seja não enumerável. Então, existe uma sequência $\{t_n,\ n \geq 1\}$ de T, tal que Y seja $\mathcal{F}\{X_{t_1}, X_{t_2}, \ldots\}$-mensurável.*

Prova: (i) Suponha $Y = I_A$, $A \in \mathcal{F}\{X\}$. Considere a classe $\mathcal{A}$ de conjuntos A tais que I_A seja determinada por um número enumerável de coordenadas. Então, $\mathcal{A}$ é uma σ-álgebra que contem retângulos (porque esses são determinados por um número finito de coordenadas). Então, $\mathcal{A} \supset \mathcal{F}\{X\}$.

(ii) O resultado é válido se $Y = \sum_{i=1}^{m} c_i I_{A_i}$, $A_i \in \mathcal{F}\{X\}$.

(iii) Logo, é valido para todo Y que seja $\mathcal{F}\{X\}$-mensurável, pois cada tal Y é um limite de funções simples. $\square$

Teorema 1.17. *Seja $X = \{X_t, t \in T\}$ um processo estocástico e T não enumerável. A classe de funções Y que são $\mathcal{F}\{X\}$-mensuráveis é a menor classe Γ tal que:*

(i) *Se $t_1, \ldots, t_n$ pertencem a T e se g é uma função de Borel n-dimensional, então $g(X_{t_1}, \ldots, X_{t_n}) \in \Gamma$;*

(ii) *Se $Y_1, Y_2, \cdots \in \Gamma$ e se $Y = \lim_{n\to\infty} Y_n$, então $Y \in \Gamma$.*

Prova: Veja o Problema 6.

1.5 Esperança

O conceito de esperança matemática (ou valor esperado, ou simplesmente esperança) de uma v.a X é equivalente ao conceito de integral de uma função mensurável sobre um espaço de probabilidade.

Definição 1.17. Seja $(\Omega, \mathcal{F}, P)$ um e.p e X uma v.a sobre esse espaço. A *esperança* de X, quando existe, é definida por

$$E(X) = \int_{\Omega} X(\omega) dP(\omega).$$

Para cada $\Lambda \in \mathcal{F}$, definimos

$$\int_{\Lambda} X(\omega) dP(\omega) = E(X I_{\Lambda}).$$

Como uma integral, a esperança de uma v.a tem as propriedades familiares de uma integral (veja o Apêndice A.2), como as que seguem.

(a) $E(aX + bY) = aE(X) + bE(Y)$, desde que o lado direito tenha sentido (ou seja, não pode ser $\infty - \infty$ ou $-\infty + \infty$);

(b) se $X_n \geq 0$ e $X = \lim_n X_n$ q.c, então $E(X) \leq \liminf_n E(X_n)$ (lema de Fatou);

(c) se $X_n \geq 0$ e se $X_n \uparrow X$ q.c então $\lim_n E(X_n) = E(X)$, desde que $+\infty$ seja permitido como um valor de cada lado (teorema da convergência monótona);

(d) se $X_n \to X$, em probabilidade ou q.c e $|X_n| \leq Y$, para todo n, com $E(Y) < \infty$, então $\lim_n E(X_n) = E(X)$ (teorema da convergência dominada).

Teorema 1.18. *A seguinte desigualdade é válida:*

$$\sum_{n=1}^{\infty} P\{|X| \geq n\} \leq E(|X|) \leq 1 + \sum_{n=1}^{\infty} P\{|X| \geq n\}, \tag{1.4}$$

de modo que $E(|X|) < \infty$ se, e somente se, a série em (1.4) convergir.

Prova: A prova segue Chung (1968). Se $\{\Lambda_n,\ n \geq 1\}$ são conjuntos disjuntos, então temos

$$\int_{\cup_n \Lambda_n} X dP = \sum_n \int_{\Lambda_n} X dP.$$

Se tomarmos $\Lambda_n = \{n \leq |X| < n+1\}$, então

$$E(|X|) = \sum_{n=0}^{\infty} \int_{\Lambda_n} |X| dP.$$

Também, temos que, se $a \leq X \leq b$ sobre Λ, então $aP(\Lambda) \leq \int_\Lambda X dP \leq bP(\Lambda)$, de modo que, para cada conjunto Λ_n,

$$\sum_{n=0}^{\infty} nP(\Lambda_n) \leq E(|X|) \leq (n+1)P(\Lambda_n) = 1 + \sum_{n=0}^{\infty} nP(\Lambda_n). \tag{1.5}$$

Resta provar que

$$\sum_{n=0}^{\infty} nP(\Lambda_n) = \sum_{n=1}^{\infty} P\{|X| \geq n\}, \tag{1.6}$$

onde as somas podem ser finitas ou infinitas.

Agora, as somas parciais do lado esquerdo de (1.6) podem ser rearranjadas (método de Abel) de modo que, para $N \geq 1$,

$$\begin{aligned}
\sum_{n=0}^{N} nP(\Lambda_n) &= \sum_{n=0}^{N} n(P\{|X| \geq n\} - P\{|X| \geq n+1\}) \\
&= \sum_{n=1}^{N} (n-(n-1))P\{|X| \geq n\} - NP\{|X| \geq N+1\} \\
&= \sum_{n=1}^{N} P\{|X| \geq n\} - NP\{|X| \geq N+1\}. \qquad (1.7)
\end{aligned}$$

Temos que

$$N \cdot P\{|X| \geq N+1\} \leq \int_{\{|X|\geq N+1\}} |X| dP. \tag{1.8}$$

Se $E(|X|) < \infty$, deduzimos que o lado direito de (1.8) tende a zero. Portanto $\sum_{n=0}^{\infty} nP(\Lambda_n) = \sum_{n=1}^{\infty} P\{|X| \geq n\}$. No caso em que $E(|X|) = \infty$, deduzimos de (1.5) que $\sum_{n=0}^{\infty} nP(\Lambda_n) = \infty$ e portanto usando (1.7) deduzimos que $\sum_{n=1}^{\infty} P\{|X| \geq n\} = \infty$. $\square$

Existe uma relação básica entre a integral abstrata com respeito a P, sobre conjuntos de $\mathcal{F}$, de um lado, e a integral de Lebesgue-Stieltjes com respeito a P_X, sobre conjuntos de $\mathcal{B}$, induzida pela v.a X, de outro lado.

Teorema 1.19. *Seja X sobre $(\Omega, \mathcal{F}, P)$, induzindo o e.p $(\mathbb{R}, \mathcal{B}, P_X)$ e seja f uma função mensurável de Borel. Então, temos*

$$\int_{\Omega} f(X(\omega)) dP(\omega) = \int_{\mathbb{R}} f(y) dP_X(y), \tag{1.9}$$

desde que pelo menos uma das duas integrais exista.

Prova. (a) Seja $B \in \mathcal{B}$ e tome $f = I_B$. Então, o lado esquerdo de (1.9) fica

$$\int_{\Omega} I_B(X(\omega)) dP(\omega) = P(X \in B),$$

e, o lado direito, $P_X(B)$. Há, então, igualdade, pela definição de P_X.

(b) Em seguida, (1.9) vale para f simples, ou seja, da forma $f = \sum_j b_j I_{B_j}$.

(c) Se $f \geq 0$, existe uma sequência $\{f_n,\ n \geq 1\}$ de funções simples, tal que $f_n \uparrow f$. Para cada f_n temos $\int_{\Omega} f_n(X(\omega)) dP(\omega) = \int_{\mathbb{R}} f_n(y) dP_X(y)$. Quando $n \to \infty$ e usando o teorema da convergência monótona, obtemos que (1.9) é válida.

(d) No caso geral, tome $f = f_+ - f_-$. $\square$

Vejamos, agora, algumas desigualdades importantes.

Proposição 1.3. (desigualdade de Chebyshev) *Se φ é uma função par, estritamente crescente e positiva sobre $[0, \infty)$, e X uma v.a, então, para cada $\lambda > 0$, temos*

$$P\{|X| \geq \lambda\} \leq \frac{E(\varphi(X))}{\varphi(\lambda)}.$$

Prova. Observamos que $P\{|X| \geq \lambda\} = P\{\varphi(|X|) \geq \varphi(\lambda)\}$. Depois, observamos que $\varphi(|X|) \geq \varphi(\lambda)\mathbf{1}_{\{\varphi(|X|)>\varphi(\lambda)\}}$ e tomamos a esperança dos dois lados. Enfim, como φ é par, temos que $\varphi(|X|) = \varphi(X)$. $\square$

Exemplos especiais são:

(a) Se $\varphi(x) = x^2$, a desigualdade fica

$$P\{|X| \geq \lambda\} \leq \frac{E(X^2)}{\lambda^2}.$$

(b) Se $\varphi(x) = |x|^p, \ \ 0 < p < \infty$, então

$$P\{|X| \geq \lambda\} \leq \frac{E(|X|^p)}{\lambda^p}.$$

(c) Se $\varphi(x) = \exp(t|x|), \ \ t > 0$, então

$$P\{|X| \geq \lambda\} \leq \frac{E(e^{t|X|})}{e^{t\lambda}}.$$

Recordemos que $L_p = L_p(\Omega, \mathcal{F}, P)$ é a coleção de todas as v.a's X sobre $(\Omega, \mathcal{F}, P)$ tais que $E(|X|^p) < \infty$ (identificamos neste caso variáveis aleatórias iguais q.c, veja o Apêndice A.3). Lembramos também que a norma L_p de X é definida por

$$||X||_p = \left[\int |X|^p dP\right]^{1/p} = [E(|X|^p]^{1/p}.$$

Proposição 1.4. (desigualdade de Hölder) *Se $X \in L_p$ e $Y \in L_q$, com $p \geq 1, \ q \geq 1$ e $\frac{1}{p} + \frac{1}{q} = 1$, então*

$$||XY||_1 \leq ||X||_p \cdot ||Y||_q, \tag{1.10}$$

ou, de modo equivalente,

$$E(|XY|) \leq [E(|X|^p)]^{1/p} \cdot [E(|Y|^q)]^{1/q}.$$

Prova. Veja o Problema 7.

Se $Y = 1$ em (1.10), obtemos $E(|X|) \leq [E(|X|^p)]^{1/p}$. Se $p = 2$ em (1.10) temos a desigualdade de Cauchy-Schwarz.

Proposição 1.5. (desigualdade de Minkowski) *Se X e Y estão em L_p, temos*

$$||X + Y||_p \leq ||X||_p + ||Y||_p,$$

ou seja,

$$[E(|X + Y|^p]^{1/p} \leq [E(|X|^p)]^{1/p} + [E(|Y|^p)]^{1/p}.$$

Prova. Veja o Problema 8.

Proposição 1.6. (desigualdade de Jensen) *Seja φ uma função mensurável convexa e suponha que $E(X)$ e $E(\varphi(X))$ existam. Então,*

$$E(\varphi(X)) \geq \varphi(E(X)).$$

Igualdade ocorre se, e somente se, φ for linear.

Prova. Dado ξ, existe uma reta passando por $\varphi(\xi)$, que está totalmente abaixo da curva φ. Tal reta é dada por

$$y - \varphi(\xi) = \lambda(x - \xi),$$

para algum λ. Então, $\varphi(x) \geq \lambda(x - \xi) + \varphi(\xi)$, para todo x. Segue-se que

$$E(\varphi(X)) \geq \lambda E(X - \xi) + \varphi(\xi).$$

Escolha $\xi = E(X)$ e o resultado esperado é obtido. O caso linear segue da prova anterior e é deixado para o leitor. $\square$

Se $\varphi(x) = x^2$, obtemos $E(X^2) \geq [E(X)]^2$. Se $\varphi(x) = |x|^p$, temos $E(|X|^p) \geq |E(X)|^p$, para $p \geq 1$.

1.6 Convergência

Nesta seção apresentamos os conceitos dos diversos modos de convergência de sequências de v.a's, que são idênticos aos conceitos correspondentes sobre sequências de funções mensuráveis em um espaço de medida.

Consideremos um e.p $(\Omega, \mathcal{F}, P)$ e $\{X_n\}$ uma sequência de v.a's definidas sobre esse espaço.

Definição 1.18. Dizemos que X_n *converge para X quase certamente*, se existe um conjunto nulo N tal que $\lim_{n\to\infty} X_n(\omega) = X(\omega)$, sempre que $\omega \in N^c$.

Dizemos, também, que X_n converge para X *com probabilidade um* e usamos a notação $X_n \overset{q.c.}{\to} X$, ou $X_n \to X$ q.c.

Teorema 1.20. *A sequência $\{X_n\}$ converge para X q.c. se, e somente se, tivermospara todo $\varepsilon > 0$*

$$\lim_{m\to\infty} P\{|X_n - X| \leq \varepsilon, \text{ para todo } n \geq m\} = 1,$$

ou

$$\lim_{m\to\infty} P\{\cap_{n=m}^{\infty}\{|X_n - X| \leq \varepsilon\}\} = 1. \tag{1.11}$$

A relação (1.11) é equivalente a

$$\lim_{m\to\infty} P\{|X_n - X| > \varepsilon, \text{ para algum } n \geq m\} = 0,$$

ou

$$\lim_{m\to\infty} P\{\cup_{n=m}^{\infty}\{|X_n - X| > \varepsilon\}\} = 0.$$

Prova. (a) Suponha que $X_n \overset{q.c.}{\to} X$. Para $m \geq 1$, seja A_m o evento em (1.11), ou seja, $A_m = \cap_{n=m}^{\infty}\{|X_n - X| \leq \varepsilon\}$. Então, $\{A_m\}$ é uma sequência crescente. Para cada $\omega \in N^c$, a convergência de $\{X_n(\omega)\}$ para $X(\omega)$ implica que, dado $\varepsilon > 0$, existe um $m(\omega, \varepsilon)$ tal que se $n \geq m(\omega, \varepsilon)$, então $|X_n(\omega) - X(\omega)| \leq \varepsilon$.

Logo, cada tal ω pertence a algum A_m, e portanto $N^c \subset \cup_{m=1}^{\infty} A_m$. Logo, $P(N^c) \leq P(\cup_{m=1}^{\infty} A_m) = \lim_m P(A_m)$, pois a sequência é crescente. Portanto, $\lim_m P(A_m) = 1$.

(b) Suponha que X_n não convirja para X sobre um conjunto Λ, com $P(\Lambda) > 0$. Considere a v.a Z definida por $Z(\omega) = \lim_n \sup |X_n(\omega) - X(\omega)|$, que pode não ser finita. Observe que

$$\{Z > 0\} = \bigcup_{n=1}^{\infty} \left\{Z > \frac{1}{n}\right\}. \tag{1.12}$$

Para cada $\omega \in \Lambda$, temos que $Z(\omega) > 0$ e, portanto, $\Lambda \subset \{Z > 0\}$. Segue-se que, para algum n, um membro da reunião do lado direito de (1.12) deve ter probabilidade estritamente positiva, e portanto, para algum $\varepsilon > 0$, o conjunto $\{Z > \varepsilon\}$ tem probabilidade estritamente positiva. Pela definição de Z, este último conjunto está contido no conjunto A_m^c, para todo m, logo $P(A_m^c) \geq P(Z > \varepsilon)$, e portanto (1.11) não pode ser verdadeira. □

Definimos o $\limsup_n A_n$ como o conjunto de todos os elementos de Ω que pertencem a um número infinito de conjuntos A_n, e o $\liminf_n A_n$ como o conjunto dos elementos de Ω que pertencem a todos os conjuntos A_n com exceção de um número finito deles (veja o Apêndice A.1). Também dizemos que o evento $\limsup_n A_n$ ocorre se, e somente se, os eventos A_n ocorrem *infinitas vezes* (*infinitely often*) e escrevemos

$$P\Big(\limsup_n A_n\Big) = P(A_n \text{ i.v}).$$

Proposição 1.7. *Para cada $A_n \in \mathcal{F}$, temos que:*

$$P\Big(\limsup_n A_n\Big) = \lim_{m\to\infty} P(\cup_{n=m}^{\infty} A_n),$$
$$P\Big(\liminf_n A_n\Big) = \lim_{m\to\infty} P(\cap_{n=m}^{\infty} A_n).$$

Prova: Chamemos $F_m = \cup_{n=m}^{\infty} A_n$, para todo $m \geq 1$. Então, F_m decresce, quando m cresce. Pela monotonicidade de P,

$$P(\cap_{m=1}^{\infty} F_m) = \lim_{m\to\infty} P(F_m). \quad \square$$

Teorema 1.21. *X_n converge para X q.c se, e somente se, para todo $\varepsilon > 0$, tivermos $P\{|X_n - X| > \varepsilon \text{ i.v }\} = 0$.*

Prova: Seja $A_m = \cap_{n=m}^{\infty}\{|X_n - X| \leq \varepsilon\}$. Segue-se que

$$\{|X_n - X| > \varepsilon \text{ i.v }\} = \cap_{m=1}^{\infty} \cup_{n=m}^{\infty} \{|X_n - X| > \varepsilon\} = \cap_{m=1}^{\infty} A_m^c.$$

De acordo com o Teorema 1.21, $X_n \to X$ q.c, se e somente se, para todo $\varepsilon > 0$, tivermos $P(A_m^c) \to 0$, quando $m \to \infty$. Como a sequência A_m^c é decrescente, isso é equivalente a $P\{|X_n - X| > \varepsilon \text{ i.v }\} = 0$. $\square$

Um conceito mais fraco do que convergência q.c é o de convergência em probabilidade.

Definição 1.19. Dizemos que a sequência $\{X_n\}$ converge para X *em probabilidade* se, e somente se, para todo $\varepsilon > 0$, tivermos $\lim_{n\to\infty} P\{|X_n - X| > \varepsilon\} = 0$.

Usaremos a notação $X_n \stackrel{P}{\to} X$. Note que $P\{|X_n - X| > \varepsilon\}$ significa $P\{\omega : |X_n(\omega) - X(\omega)| > \varepsilon\}$.

Teorema 1.22. *Se X_n convergir para X q.c., então X_n converge para X em probabilidade.*

Prova: Se $X_n \stackrel{q.c}{\to} X$, então $P\{\cup_{n=m}^{\infty}(|X_n - X| > \varepsilon\} \to 0$, quando $m \to \infty$. Mas isso implica $P\{|X_n - X| > \varepsilon\} \to 0$, quando $n \to \infty$, logo $X_n \stackrel{P}{\to} X$. $\square$

A recíproca do teorema não vale. O que se verifica é o seguinte resultado.

Teorema 1.23. *Se $X_n \stackrel{P}{\to} X$, existe uma subsequência $\{n_k\}$, tal que $X_{n_k} \stackrel{q.c}{\to} X$. Ou seja, convergência em probabilidade implica em convergência quase certa ao longo de uma subsequência.*

Prova: Como $X_n \stackrel{P}{\to} X$ se, e somente se $X_n - X \stackrel{P}{\to} 0$, podemos supor $X = 0$. Então, por hipótese, para todo $k > 0$, $P\{|X_n| > 1/2^k\} \to 0$, quando $n \to \infty$. Segue-se que, para cada $k > 0$, existe um n_k tal que

$$P\{|X_{n_k}| > 1/2^k\} \leq \frac{1}{2^k}.$$

Tendo escolhido tal sequência $\{n_k\}$, seja $E_k = \{|X_{n_k}| > 1/2^k\}$. Então, $P\{E_k \text{ i.v}\} = 0$, logo pelo Teorema 1.22, $X_{n_k} \stackrel{q.c}{\to} 0$. $\square$

Um tipo de convergência importante em Probabilidade e Estatística é a convergência *em média quadrática*. Essa é um caso particular do seguinte modo de convergência.

Definição 1.20. A sequência $\{X_n\}$ converge para X em L_p, se e somente se,

$$\lim_{n\to\infty} E(|X_n - X|^p) = 0.$$

Usaremos a notação $X_n \stackrel{L_p}{\to} X$.

Teorema 1.24. *Se X_n converge para zero em L_p, então X_n converge para zero em probabilidade. A recíproca é verdadeira desde que $|X_n| \leq Y$ q.c, com $Y \in L_p$.*

Prova: (a) Se $X_n \stackrel{L_p}{\to} 0$, então $E(|X_n|^p) \to 0$, quando $n \to \infty$. Pela desigualdade de Chebyshev, com $\varphi(x) = |x|^p$, temos

$$P\{|X_n| \geq \varepsilon\} \leq \frac{E(|X_n|^p)}{\varepsilon^p}.$$

Quando $n \to \infty$, o lado direito tende a zero, logo $X_n \stackrel{P}{\to} 0$.

(b) Suponha $|X_n| \leq Y$ q.c, com $E(|Y|^p) < \infty$. Temos que

$$E(|X_n|^p) = \int_{\{|X_n|<\varepsilon\}} |X_n|^p dP + \int_{\{|X_n|\geq\varepsilon\}} |X_n|^p dP \leq \varepsilon^p + \int_{\{|X_n|\geq\varepsilon\}} Y^p dP.$$

Por hipótese, a última integral tende a zero (veja o Problema 9). Assim, faça $n \to \infty$ e depois $\varepsilon \to 0$ para obter $E(|X_n|^p) \to 0$, ou seja, $X_n \stackrel{L_p}{\to} 0$. $\square$

Dizemos que a sequência $\{X_n\}$ é *uniformemente limitada* se, e somente se, $|X_n| \leq M$ q.c, com M constante. Como um corolário do Teorema 1.24, se $\{X_n\}$ for uniformemente limitada, convergência em probabilidade e convergência em L_p são equivalentes. O resultado geral segue.

Teorema 1.25. *$X_n \stackrel{P}{\to} 0$ se, e somente se, $E\left(\frac{|X_n|}{1+|X_n|}\right) \to 0$. Além disso, a aplicação $d(\cdot,\cdot)$ definida por $d(X,Y) = E\left(\frac{|X-Y|}{1+|X-Y|}\right)$ é uma métrica no espaço das v.a's, desde que identifiquemos v.a's que sejam iguais q.c.*

Prova: Veja Chung (1974).

Algumas vezes temos que tratar com problemas de convergência de v.a's quando nenhum limite esteja evidenciado. Se $X_n - X_m \to 0$, quando $m, n \to \infty$, segundo qualquer um dos modos anteriores, então existe uma v.a. X tal que $X_n \to X$ segundo o mesmo modo. Esse é o critério de Cauchy.

Problemas

1. Prove completamente o Teorema 1.2.

2. Prove as relações (1.3).

3. Prove completamente o Lema 1.1.

4. Prove o Teorema 1.8.

5. Dada uma f.d. $F(x, y)$, encontre um vetor aleatório $\mathbf{X}$ tendo F como sua f.d. [Sugestão: proceda extamente como no caso de uma v.a.]

6. Prove o Teorema 1.17.

7. Prove a Proposição 1.4.

8. Prove a Proposição 1.5.

9. Se X é uma v.a com $E(|X|) < \infty$, então $A_n \in \mathcal{F},\ A_n \downarrow \emptyset$ implica que $\lim_n \int_{A_n} X dP = 0$.

10. Seja $\mathcal{F}_0$ uma álgebra e $\mathcal{F}$ a σ-álgebra gerada por $\mathcal{F}_0$. Seja P uma probabilidade sobre $\mathcal{F}$. Mostre, diretamente do Teorema da Convergência Monotônica, que dado $\varepsilon > 0$ e $A \in \mathcal{F}$, existe $A_\varepsilon \in \mathcal{F}_0$, tal que $P(A \triangle A_\varepsilon) < \varepsilon$ ($A \triangle B$ significa a diferença simétrica entre A e B).

11. Mostre que, para qualquer sequência de v.a's limitadas, existe uma sequência de constantes $\{b_n\}$, tal que X_n/b_n converge para zero q.c.

12. Seja F contínua, estritamente crescente, $\lim_{x\to\infty} F(x) = 1,\ \lim_{x\to-\infty} F(x) = 0$, e seja λ a medida de Lebesgue.

    (a) Mostre que $P(A) = \lambda(F(A))$ é uma medida de probabilidade sobre $\mathcal{B}((0, 1))$;

    (b) Mostre que P pode ser obtida via uma construção similar à construção do Exemplo A.1.4.

13. Mostre que a f.d de uma v.a X como definida no texto é contínua à direita, mas como definida na Observação 3 é contínua à esquerda.

14. Seja $X = \{X_t, t \in T\}$ um processo estocástico, T não enumerável. Mostre que a classe de funções $\mathcal{F}\{X\}$-mensuráveis é a menor classe Γ tal que:

    (a) se f for uma função de Borel sobre $\mathbb{R}^n$, então $f(X_{t_1}, \ldots, X_{t_n}) \in \Gamma$, para $t_1, \ldots, t_n \in T$;

    (b) se $Y_1, Y_2, \ldots$ estão em Γ e se $\lim_n Y_n = Y$ existir, então $Y \in \Gamma$.

15. Apresente uma sequência de eventos $\{A_n, n \geq 1\}$, de um mesmo e.p, tais que:

    (a) $\sum_{n=1}^{\infty} P(A_n) = \infty$ e $P(A_n \text{ i.v}) < 1$;

    (b) $P(A_n \text{ i.v}) = 1$ e $P(A_n^c \text{ i.v}) = 0$.

16. Considere $\Omega = [0,1]$, $\mathcal{F} = \mathcal{B}([0,1])$ e P a medida de Lebesgue. Seja $A_0 = \Omega$, $A_1 = A_0 \cap (1/3, 2/3)^c = [0,1/3] \cup [2/3,1]$, $A_2 = [0,1/9] \cup [2/9,3/9] \cup [6/9,7/9] \cup [8/9,1]$ (isto é, retiramos a terceira parte central de cada um dos intervalos de A_1). Prosseguindo desse modo, obteremos uma sequência de eventos $\{A_n, n \geq 1\}$, monotônica não crescente ($A_n \supset A_{n+1}$), com A_n sendo a reunião de 2^n intervalos fechados.

    O conjunto $C = \cap_{n=1}^{\infty}$ é chamado *conjunto de Cantor*. Prove que $P(C) = 0$ (esse é um exemplo de um conjunto infinito não enumerável com probabilidade (comprimento) zero).

17. Se $(\Omega, \mathcal{F}, P)$ for um e.p., $A \in \mathcal{F}, P(A) > 0$, defina $P(B|A) = P(A \cap B)/P(A)$, para todo $B \in \mathcal{F}$. Prove que $(\Omega, \mathcal{F}, P(\cdot|A))$ é um e.p.

18. Seja $\Omega = \{(x,y) : \ 0 < x, y \leq 1\}$, $\mathcal{F}$ a classe dos conjuntos da forma $\{(x,y) : x \in A \cap (0,1], 0 < y \leq 1\}$, onde $A \in \mathcal{B}$, e seja P dada pela medida de Lebesgue nesse conjunto. Prove que $(\Omega, \mathcal{F}, P)$ é um espaço de probabilidade.

Capítulo 2

Independência

Neste capítulo estudaremos o importante conceito de independência, juntamente com resultados relacionados: leis zero-um, leis dos grandes números, séries aleatórias e aplicações. Algumas referências para esse capítulo são Lamperti (1966), Chung (1968, 1974), Breiman (1968, 1992) e Billingsley (1995).

2.1 Fatos básicos

Definição 2.1. (i) Seja $(\Omega, \mathcal{F}, P)$ um e.p e sejam $\mathcal{F}_1, \mathcal{F}_2, \ldots, \mathcal{F}_n$ σ-álgebras (ou álgebras) contidas em $\mathcal{F}$. Dizemos que essas σ-álgebras são *independentes* se para quaisquer $A_1 \in \mathcal{F}_1, A_2 \in \mathcal{F}_2, \ldots, A_n \in \mathcal{F}_n$, tivermos

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1)P(A_2) \cdots P(A_n).$$

(ii) As v.a'a $X_1, \ldots, X_n$ definidas sobre $(\Omega, \mathcal{F}, P)$ são *independentes* se $\mathcal{F}\{X_1\}, \ldots, \mathcal{F}\{X_n\}$ são σ-álgebras independentes.

(iii) As σ-álgebras $\mathcal{F}_1, \mathcal{F}_2, \ldots$, contidas em $\mathcal{F}$, são *independentes* se para cada n, $\mathcal{F}_1, \ldots, \mathcal{F}_n$ forem independentes.

(iv) As v.a's $X_1, X_2, \ldots$ sobre $(\Omega, \mathcal{F}, P)$ são *independentes* se para cada n, $X_1, \ldots, X_n$ são independentes.

Lema 2.1 *Sejam $\mathcal{F}_0$ e $\mathcal{G}_0$ álgebras independentes. Sejam $\mathcal{F}$ e $\mathcal{G}$ as σ-álgebras geradas por $\mathcal{F}_0$ e $\mathcal{G}_0$, respectivamente. Então, $\mathcal{F}$ e $\mathcal{G}$ são independentes.*

Prova: Fixemos $A \in \mathcal{F}_0$. Se $P(A) > 0$, defina a probabilidade P_A por

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

P_A é uma probabilidade sobre $\mathcal{G}_0$, tal que, para todo $B \in \mathcal{G}_0$,

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

pois $\mathcal{F}_0$ e $\mathcal{G}_0$ são independentes. Logo, $P_A(B) = P(B)$, para todo $B \in \mathcal{G}$, pelo Teorema 1.3. Logo, para cada $A \in \mathcal{F}_0$ e cada $B \in \mathcal{G}$, teremos $P(A \cap B) = P(A)P(B)$ (esta última igualdade é trivial se $P(A) = 0$). Para terminar a prova, fixe $B \in \mathcal{G}$ e repita o argumento. $\square$

Definição 2.2. Os vetores aleatórios $\mathbf{X} = (X_1, \ldots, X_n)$ e $\mathbf{Y} = (Y_1, \ldots, Y_n)$ são *independentes* se $\mathcal{F}\{\mathbf{X}\}$ é independente de $\mathcal{F}\{\mathbf{Y}\}$.

Definição 2.3. Os processos estocásticos $X = \{X_n, n \geq 1\}$ e $Y = \{Y_n, n \geq 1\}$ são *independentes* se $(X_1, \ldots, X_n)$ e $(Y_1, \ldots, Y_n)$ são independentes, para todo n.

Proposição 2.1. *As v.a's $X_1, X_2, \ldots, X_n$ são independentes se, e somente se, para toda coleção de conjuntos de Borel $A_1, \ldots, A_n$ tivermos*

$$P\{X_1 \in A_1, \ldots, X_n \in A_n\} = P\{X_1 \in A_1\} \cdots P\{X_n \in A_n\}.$$

Prova: Segue imediatamente da definição de v.a's independentes e do fato que todo conjunto em $\mathcal{F}\{X_i\}$ é da forma $X_i^{-1}(B),\ B \in \mathcal{B}$. $\square$

O mesmo resultado vale para uma sequência infinita de v.a's independentes. Veja, por exemplo, Breiman (1968).

Definição 2.4. Sejam $A_1, \ldots, A_n$ eventos. Esses eventos são *independentes* se as respectivas σ-álgebras geradas por eles são independentes.

Teorema 2.1. *Sejam $X = \{X_n, n \geq 1\}$ e $Y = \{Y_n, n \geq 1\}$ dois processos estocásticos independentes sobre $(\Omega, \mathcal{F}, P)$. Então, toda função $\mathcal{F}\{X\}$-mensurável é independente de toda função $\mathcal{F}\{Y\}$-mensurável.*

Prova: Como X e Y são independentes, $(X_1, \ldots, X_n)$ é independente de $(Y_1, \ldots, Y_n)$, para cada n. Ou seja, $\mathcal{F}\{X_1, \ldots, X_n\}$ é independente de $\mathcal{F}\{Y_1, \ldots, Y_n\}$, pela definição de vetores independentes. Isso implica que $\cup_n \mathcal{F}\{X_1, \ldots, X_n\}$ é independente de $\cup_n \mathcal{F}\{Y_1, \ldots, Y_n\}$. De fato, se $A \in \cup_n \mathcal{F}\{X_1, \ldots, X_n\}$ e $B \in \cup_n \mathcal{F}\{Y_1, \ldots, Y_n\}$, então $A \in \mathcal{F}\{X_1, \ldots, X_n\}$, para algum n e $B \in \mathcal{F}\{Y_1, \ldots, Y_m\}$, para algum m. Portanto, $\vee_n \mathcal{F}\{X_1, \ldots, X_n\}$ é independente de $\vee_n \mathcal{F}\{Y_1, \ldots, Y_n\}$, devido à independência das álgebras $\cup_n \mathcal{F}\{X_1, \ldots, X_n\}$ e $\cup_n \mathcal{F}\{Y_1, \ldots, Y_n\}$.

Seja, agora, Z $\mathcal{F}\{X\}$-mensurável e W $\mathcal{F}\{Y\}$-mensurável. Então $\mathcal{F}\{Z\} \subset \mathcal{F}\{X\}$ e $\mathcal{F}\{W\} \subset \mathcal{F}\{Y\}$. Logo, $\mathcal{F}\{Z\}$ e $\mathcal{F}\{W\}$ são independentes, isto é, Z e W são independentes. $\square$

Alguns casos especiais desse teorema são:

(a) $\limsup_n X_n$ é independente de $\limsup_n Y_n$;

(b) Se f e g são funções de Borel sobre $\mathbb{R}^n$, então $f(X_1,\ldots,X_n)$ é independente de $g(Y_1,\ldots,Y_n)$.

Teorema 2.2. *Se $X_1,\ldots,X_n$ são v.a's sobre $(\Omega,\mathcal{F},P)$ e $\mathbf{X}=(X_1,\ldots,X_n)$, então $X_1,\ldots,X_n$ são independentes se, e somente se, $F_{\mathbf{X}}(x_1,\ldots,x_n)=F_{X_1}(x_1)\ldots F_{X_n}(x_n)$ para todo $x_1,\ldots,x_n$.*

Prova: Para simplificar as notações consideramos somente o caso $n=2$.
(a) Se X_1 e X_2 são independentes, $P(X_1\in A, X_2\in B)=P(X_1\in A)P(X_2\in B)$, logo basta tomar $A=(-\infty,x_1]$ e $B=(-\infty,x_2]$.

(b) Fixe b, um número real, e defina as medidas finitas μ_1 e μ_2 como segue:

$$\begin{aligned}\mu_1(A) &= P(X_1\le b, X_2\in A),\\ \mu_2(A) &= P(X_1\le b)P(X_2\in A).\end{aligned}$$

Agora, mostremos que μ_1 e μ_2 são iguais, para todo conjunto de Borel A. Elas são iguais sobre conjuntos da forma $A=(-\infty,c]$, pela definição de f.d e por hipótese. Logo, elas são iguais também para todos os conjuntos A da forma $(c,d]$ ou (c,∞). Deduzimos que μ_1 e μ_2 são iguais para conjuntos de uma álgebra que gera os conjuntos de Borel, logo elas concordam sobre todos os conjuntos de Borel. Ou seja, provamos que: se $F_{\mathbf{X}}=F_{X_1}F_{X_2}$, então para todo real b e todo conjunto de Borel A, temos $P(X_1\le b, X_2\in A)=P(X_1\le b)P(X_2\in A)$. Queremos provar que $P(X_1\in B, X_2\in A)=P(X_1\in B)P(X_2\in A)$. Basta fixar A, conjunto de Borel, e proceder como antes. $\square$

Uma consequência desse teorema é o seguinte resultado. Veja, também, Lamperti (1966) para uma construção de v.a's sobre um espaço de probabilidade.

Teorema 2.3. *Sejam $F_1,F_2,\ldots$ f.d's. Então, existe um e.p $(\Omega,\mathcal{F},P)$ e v.a's $X_1,X_2,\ldots$ sobre esse espaço, tais que:*

(i) *As v.a's X_i são independentes;*

(ii) *A f.d de X_i é F_i, $i=1,2,\ldots$.*

Prova: Forme o conjunto consistente de f.d's $F_1(x_1),\ldots,F_n(x_n)$ e use o teorema da extensão de Kolmogorov. $\square$

Teorema 2.4. *Sejam $\mathcal{F}_0,\mathcal{F}_1,\mathcal{F}_2,\ldots$ σ-álgebras independentes sobre $(\Omega,\mathcal{F},P)$. Então, $\mathcal{F}_0$ é independente da σ-álgebra gerada por $\mathcal{F}_1,\mathcal{F}_2,\ldots$.*

Prova: Considere a álgebra $\hat{\mathcal{F}}$ definida como sendo a classe de reuniões finitas de conjuntos disjuntos da forma $A_1\cap\cdots\cap A_n$, onde $A_i\in\mathcal{F}_i,\ i\ge 1$. Essa álgebra gera a σ-álgebra $\mathcal{F}\{\mathcal{F}_1, \mathcal{F}_2, \ldots\}$, e $\mathcal{F}_0$ e $\hat{\mathcal{F}}$ são independentes. De fato, se provarmos essa afirmação, o teorema fica provado, pois álgebras independentes geram σ-álgebras independentes.

Sejam A e B pertencentes a $\mathcal{F}_0$ e $\hat{\mathcal{F}}$, respectivamente, $B = A_1 \cap \cdots \cap A_n$. Então,

$$P(A \cap B) = P(A \cap A_1 \cap \cdots \cap A_n) = P(A)P(A_1 \cap \cdots \cap A_n),$$

pois $\mathcal{F}_0, \mathcal{F}_1, \mathcal{F}_2, \ldots \mathcal{F}_n$ são independentes, logo $P(A \cap B) = P(A)P(B)$.

Se $B \in \hat{\mathcal{F}}$ e $B = D_1 \cup D_2$, $D_1 \cap D_2 = \emptyset$, $D_1 = A_1 \cap \cdots \cap A_n$, $D_2 = C_1 \cap \cdots \cap C_n$, então

$$\begin{aligned} P(A \cap B) &= P(A \cap D_1) + P(A \cap D_2) = P(A)P(D_1) + P(A)P(D_2) \\ &= P(A)[P(D_1) + P(D_2)] = P(A)P(D_1 \cup D_2) = P(A)P(B). \quad \square \end{aligned}$$

Corolário 2.1. *Sejam $X_1, X_2, \ldots$ v.a's independentes. Sejam $\{i_1, i_2, \ldots\}$ e $\{j_1, j_2, \ldots\}$ conjuntos disjuntos de inteiros. Então, $\mathcal{F}\{X_{i_1}, X_{i_2}, \ldots\}$ é independente de $\mathcal{F}\{X_{j_1}, X_{j_2}, \ldots\}$.*

Prova: Considere i_1; então, $\mathcal{F}\{X_{i_1}\}$ é independente de $\mathcal{F}\{X_{j_1}\}$, $\mathcal{F}\{X_{j_2}\}, \ldots$, por hipótese. Segue-se que $\mathcal{F}\{X_{i_1}\}$ é independente de $\bigvee_{k \geq 1} \mathcal{F}\{X_{j_k}\} = \mathcal{F}\{X_{j_1}, X_{j_2}, \ldots\}$. De modo análogo, $\mathcal{F}\{X_{i_k}\}$ é independente de $\mathcal{F}\{X_{j_1}, X_{j_2}, \ldots\}$, logo $\bigvee_{k \geq 1} \mathcal{F}\{X_{i_k}\}$ é independente de $\mathcal{F}\{X_{j_1}, X_{j_2}, \ldots\}$. $\square$

Esse corolário implica, por exemplo, que $\mathcal{F}\{X_1, \ldots, X_n\}$ é independente de $\mathcal{F}\{X_{n+1}, X_{n+2}, \ldots\}$, se $X_1, X_2, \ldots$ são independentes. Também, se $\varphi_1, \varphi_2, \ldots$ são mensuráveis sobre $(\mathbb{R}^n, \mathcal{B}^n)$, então as v.a's $Z_1 = \varphi_1(X_1, \ldots, X_n)$, $Z_2 = \varphi_2(X_{n+1}, \ldots, X_{2n}), Z_3 = \varphi_3(X_{2n+1}, \ldots, X_{3n}), \ldots$ são independentes.

Teorema 2.5. *Sejam X e Y v.a's independentes. Suponha que ou $E(|X|) < \infty$, $E(|Y|) < \infty$. Então, $E(|XY|) < \infty$ e*

$$E(XY) = E(X)E(Y).$$

Prova: Pelo teorema de Fubini e a independência de X e Y temos

$$\begin{aligned} E(XY) &= \int_\Omega X(\omega)Y(\omega)dP(\omega) \\ &= \int_{\mathbb{R}\times\mathbb{R}} xydF_{X,Y}(x,y) \\ &= \int_{\mathbb{R}} \left[\int_{\mathbb{R}} xydF_X(x)\right] dF_Y(y) \\ &= \int_{\mathbb{R}} xdF_X(x) \int_{\mathbb{R}} ydF_Y(y) = E(X)E(Y). \quad \square \end{aligned}$$

Como, se X e Y são independentes, também o serão $f(X)$ e $g(Y)$, com f e g funções de Borel, o teorema implica que $E[f(X)g(Y)] = E[f(X)][g(Y)]$. Também, o resultado pode ser generalizado para um número finito $X_1, \ldots, X_n$ de v.a's independentes.

Definição 2.5. Seja X uma v.a tal que $E(X^2) < \infty$ e $\mu = E(X)$. Então a *variância* de X é definida por

$$\mathrm{Var}(X) = E((X-\mu)^2).$$

Proposition 2.2. *(a) Se c constante,* $\mathrm{Var}(cX) = c^2\,\mathrm{Var}(X)$ *e* $\mathrm{Var}(X+c) = \mathrm{Var}(X)$.
(b) $\mathrm{Var}(X) = E(X^2) - \mu^2$.
(c) Se X *e* Y *são independentes,* $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$.

Prova: Veja Problema 1. □

Exemplo 2.1. (Funções de Rademacher) Seja $\Omega = [0,1)$, $\mathcal{F} = \mathcal{B} \cap [0,1)$ e P a medida de Lebesgue. Defina a sequência $\{X_n, n \geq 1\}$ de v.a's sobre $(\Omega, \mathcal{F}, P)$ como segue:

$$X_n(\omega) = \begin{cases} 1, & \text{se } \omega \in [2j/(2^n), (2j+1)/(2^n)), \;\; j \in \{0, 1, \ldots, 2^{n-1} - 1\}, \\ -1, & \text{caso contrário.} \end{cases}$$

Então as v.a's X_n são independentes. Para provar isso, temos que verificar que

$$P(X_1 = e_1, \ldots, X_k = e_k) = P(X_1 = e_1) \cdots P(X_k = e_k),$$

para todas as escolhas $e_1, \ldots, e_k \in \{-1, 1\}$.

Exemplo 2.2. Seja Ω, $\mathcal{F}$ e P como no exemplo anterior. Defina $\{X_n, n \geq 1\}$ v.a's sobre $(\Omega, \mathcal{F}, P)$ por $X_n(\omega) = a_n$, se $\omega \in \Omega$ e $\omega = 0, a_1 a_2 a_3 \cdots$ (no caso em que ω tem duas expansões decimais escolhemos a expansão infinita). Então essas v.a's são independentes. Veja o Problema 2.

2.2 Leis Zero-Um

Nesta seção investigaremos eventos cujas probabilidades são iguais a zero ou um.

2.2.1 Lema de Borel-Cantelli

Lembremos que, se $\{E_n\}$ são eventos, então $\limsup_n E_n = \cap_{m=1}^{\infty} \cup_{n=m}^{\infty} E_n$ e escrevemos $P(\limsup_n E_n) = P(E_n \text{ i.v.})$.

Teorema 2.6. *Seja* $\{E_k, k \geq 1\}$ *uma sequência de eventos de* $\mathcal{F}$. *Os seguintes resultados são válidos:*

(a) Se $\sum_{k=1}^{\infty} P(E_k) < \infty$, então $P(E_k \text{ i.v}) = 0$;

(b) Se os E_k's são independentes e $\sum_{k=1}^{\infty} P(E_k) = \infty$, então $P(E_k \text{ i.v}) = 1$.

Prova: (a) Seja $F_n = \cup_{k=n}^{\infty} E_k$, então $F_n \downarrow \limsup_n E_n$, quando $n \to \infty$. Logo,

$$P(E_k \text{ i.v}) = P(\limsup_n E_n) = \lim_{n\to\infty} P(F_n).$$

Mas $P(F_n) \leq \sum_{k=n}^{\infty} P(E_k) \to 0$, quando $n \to \infty$, porque essa é a cauda de uma série convergente e portanto $P(E_k \text{ i.v}) = 0$.

(b) $1 - P(F_n) \leq 1 - P(\cup_{k=n}^{n+p} E_k) = P((\cup_{k=n}^{n+p} E_k)^c) = P(\cap_{k=n}^{n+p} E_k^c)$. Usando a independência dos E_k, temos que $1 - P(F_n) \leq \prod_{k=n}^{n+p} P(E_k^c) = \prod_{k=n}^{n+p} [1 - P(E_k)]$. Mas $1 - x \leq e^{-x}$ para $x \geq 0$, logo $1 - P(F_n) \leq \prod_{k=n}^{n+p} \exp\{-P(E_k)\} = \exp\{-\sum_{k=n}^{n+p} P(E_k)\} \to 0$, quando $p \to \infty$. Logo, $1 - P(E_k \text{ i.v}) \to 0$, pois $F_n \downarrow \limsup_n E_n$. $\square$

Observações: (1) A independência é necessária em (b); de fato, seja Λ um conjunto, com $0 < P(\Lambda) < 1$ e seja $E_k = \Lambda$, para todo k. Então, $\sum P(E_k) = \sum P(\Lambda) = \infty$, mas $P(\limsup_k E_k) = P(\Lambda) < 1$.

(2) A parte (b) continua válida se os eventos E_k's são independentes dois a dois. Veja Chung (1968).

Aplicações: (1) Lembremos que $E(|X|) < \infty$ se, e somente se, $\sum_{n=0}^{\infty} P(|X| > n) < \infty$. Sejam $X_1, X_2, \ldots$ v.a's independentes, com a mesma distribuição (i.i.d). Então,

$$P\{\omega : |X_n(\omega)| > n \text{ i.v}\} = 0, \text{ se } E(|X_1|) < \infty.$$

(2) Sejam $X_1, X_2, \ldots$ v.a's i.i.d e suponha que $(X_1 + X_2 + \ldots + X_n)/n$ convirja q.c. Então $E(|X_1|) < \infty$. De fato, temos que $|X_n|/n \to 0$ q.c, em particular $P(|X_n|/n > 1 \text{ i.v}) = 0$, ou seja $\sum P\{|X_n|/n > 1\} < \infty$ ou $\sum P\{|X_1| > n\} < \infty$ e portanto $E(|X_1|) < \infty$.

2.2.2 Lei Zero-Um de Kolmogorov

Essa lei depende do conceito de σ-álgebra caudal (*tail σ-field* em inglês) que passamos a definir.

Definição 2.6. *Sejam $\{X_n, n \geq 1\}$ v.a's sobre $(\Omega, \mathcal{F}, P)$ e seja $\mathcal{F}_n = \mathcal{F}\{X_n, X_{n+1}, \ldots\}$. Então, $\mathcal{F}_\infty = \cap_{n=1}^{\infty} \mathcal{F}_n$ é chamada a σ-álgebra caudal e qualquer conjunto $\Lambda \in \mathcal{F}_\infty$ é chamado evento caudal.*

Isso significa que um evento caudal não depende de qualquer número finito de coordenadas. Por exemplo, considere um evento Λ para o qual $S_n/n = (X_1 + \ldots + X_n)/n \nrightarrow 1/2$, quando se lança uma moeda “honesta”. Ou seja,

$$\Lambda = \Big\{\omega : \frac{X_1(\omega) + \ldots + X_n(\omega)}{n} \nrightarrow 1/2\Big\}.$$

Esse conjunto tem a propriedade que, se $\omega \in \Lambda$ ou não, isso não depende das primeiras n coordenadas de ω, não importando o quão grande n seja. Ou seja, Λ é um evento caudal. Formalmente, como para todo $k \geq 1$,

$$\Lambda = \Big\{\omega : \frac{X_k(\omega) + \ldots + X_n(\omega)}{n} \nrightarrow 1/2\Big\},$$

então $\Lambda \in \mathcal{F}\{X_k, X_{k+1}, \ldots\}$, para todo $k \geq 1$, logo $\Lambda \in \mathcal{F}_\infty$.

Teorema 2.7. *Sejam $\{X_n, n \geq 1\}$ v.a's independentes sobre $(\Omega, \mathcal{F}, P)$. Se $\Lambda \in \mathcal{F}_\infty$, então $P(\Lambda) = 0$ ou $P(\Lambda) = 1$.*

Prova: Seja n um inteiro; então X_n é independente de $\mathcal{F}\{X_m\},\ m > n$, e consequentemente, X_n é independente de $\mathcal{F}\{X_m, X_{m+1}, \ldots\}$. Isso implica que X_n é independente de $\mathcal{F}_\infty$. Mas n é arbitrário, logo todo X_n é independente de $\mathcal{F}_\infty$, ou seja, $\mathcal{F}\{X_1, X_2, \ldots\}$ é independente de $\mathcal{F}_\infty$. Mas $\mathcal{F}_\infty \subset \mathcal{F}\{X_1, X_2, \ldots\}$, logo $\mathcal{F}_\infty$ é independente de si mesmo. Tome $\Lambda \in \mathcal{F}_\infty$. Então, $P(\Lambda \cap \Lambda) = P(\Lambda)P(\Lambda)$, ou seja $P(\Lambda) = [P(\Lambda)]^2$, isto é, $P(\Lambda) = 0$ ou $P(\Lambda) = 1$. $\square$

Aplicações: 1) Sejam $\{X_n, n \geq 1\}$ independentes, $S_n = \sum_{i=1}^n X_i$. Se $\Lambda = \{\omega : S_n(\omega) \text{ converge}\}$, então $P(\Lambda) = 0$ ou $P(\Lambda) = 1$. Informalmente, se S_n converge ou não, isso depende somente das somas parciais $(X_n + X_{n+1} + \ldots)$, isto é, o conjunto de convergência é um conjunto de $\mathcal{F}\{X_n, X_{n+1} \ldots\}$, ainda um conjunto de $\cap_{n=1}^\infty \mathcal{F}\{X_n, X_{n+1} \ldots\}$. Portanto, $\Lambda \in \mathcal{F}_\infty$, logo $P(\Lambda) = 0$ ou $P(\Lambda) = 1$. Veja Problema 3.

2) Sejam $\{X_n, n \geq 1\}$ independentes e $\Lambda = \{\omega : S_n(\omega)/n \text{ converge}\}$. Então $P(\Lambda) = 0$ ou $P\Lambda) = 1$.

3) Sejam $\{X_n, n \geq 1\}$ independentes. Seja Y qualquer v.a que seja $\mathcal{F}_\infty$-mensurável. Então, Y é constante q.c.

2.2.3 Lei Zero-Um de Hewitt-Savage

Essa lei vale para conjuntos simétricos, cuja definição é dada a seguir.

Definição 2.7. *Sejam $X_1, X_2, \ldots$ v.a's e $\Lambda \in \mathcal{F}\{X_1, X_2, \ldots\}$. Então Λ é simétrico se existe um conjunto de Borel B em $\mathbb{R}^\infty$ tal que*

$$\Lambda = \{\omega : (X_1(\omega), X_2(\omega), \ldots) \in B\} = \{\omega : (X_{\sigma_1}(\omega), X_{\sigma_2}(\omega), \ldots) \in B\},$$

onde σ é uma permutação de um número finito de inteiros.

Exemplo 2.3 (i) Sejam $\{X_n, n \geq 1\}$ v.a's. Então, $\{\omega : X_n(\omega) \text{ converge}\}$ é um conjunto simétrico. Note que esse conjunto é também um evento caudal.

(ii) Se $S_n = X_1 + \ldots + X_n$, então $\{\omega : S_n(\omega) \text{ converge}\}$ é um conjunto simétrico. De fato, tome $B \in \mathcal{B}^\infty$ como segue: B é o conjunto de todas as sequências $(a_1, a_2, \ldots)$ tais que $a_1 + \ldots + a_n$ converge. Então $\Lambda = X^{-1}(B)$ é simétrico, onde $X = (X_1, X_2, \ldots)$. De fato, seja σ uma permutação de $1, 2, \ldots, N$ e seja $\hat{X} = (X_{\sigma_1}, X_{\sigma_2}, \ldots)$, $S_{\sigma_n} = X_{\sigma_1} + \ldots + X_{\sigma_n}$, para todo n. Note que $S_{\sigma_n} = S_n$, para todo $n \geq N$. Agora,

$$
\begin{aligned}
X^{-1}(B) = \Lambda &= \bigcap_{m=1}^{\infty} \bigcup_{k \geq 1} \Big\{\omega : \sup_{j \geq k} |S_j(\omega) - S_k(\omega)| \leq 1/m\Big\} \\
&= \bigcap_{m=1}^{\infty} \bigcup_{k \geq n_0} \Big\{\omega : \sup_{j \geq k} |S_j(\omega) - S_k(\omega)| \leq 1/m\Big\}, \quad \forall n_0 \geq 1 \\
&= \bigcap_{m=1}^{\infty} \bigcup_{k \geq N} \Big\{\omega : \sup_{j \geq k} |S_{\sigma_j}(\omega) - S_{\sigma_k}(\omega)| \leq 1/m\Big\} \\
&= \bigcap_{m=1}^{\infty} \bigcup_{k \geq 1} \Big\{\omega : \sup_{j \geq k} |S_{\sigma_j}(\omega) - S_{\sigma_k}(\omega)| \leq 1/m\Big\} \\
&= \hat{X}^{-1}(B),
\end{aligned}
$$

ou seja, $X^{-1}(B) = \hat{X}^{-1}(B)$, portanto Λ é simétrico.

(iii) Vejamos um exemplo de um evento simétrico que não seja um evento caudal. Seja $\{B_n\}$ uma sequência de conjuntos de Borel e $S_n = X_1 + \ldots + X_n$. Então, $\Lambda = \{\omega : S_n(\omega) \in B_n \text{ i.v}\}$ é simétrico (veja o Problema 8). Mas esse evento não necessita ser um evento caudal. Tome $X_1, X_2, \ldots$ independentes com $X_1 = 1$, com probabilidade $1/2$ e $X_1 = 0$, com probabilidade $1/2$ e $X_2 = X_3 = \ldots = 0$. Então, $\{S_n = 0 \text{ i.v}\}$ não é caudal, pois $P\{S_n = 0 \text{ i.v}\} = 1/2$.

Para provar a lei de Hewitt-Savage, necessitamos dos dois lemas a seguir.

Lema 2.1. *Sejam A e B conjuntos de $\mathcal{F}$, para algum e.p $(\Omega, \mathcal{F}, P)$. Defina $d(A, B) = P(A \Delta B)$. Então, d é uma pseudo-métrica sobre $\mathcal{F}$, e se $A_n \to A$, $B_n \to B$ nessa pseudo-métrica, então $A_n \cap B_n \to A \cap B$, $A_n \cup B_n \to A \cup B$ e $A_n^c \to A^c$. Também, se $A_n \to A$ nessa pseudo-métrica, então $P(A_n) \to P(A)$.*

Lemma 2.2. *Seja $\mathcal{F}_0$ uma álgebra e $\mathcal{F}$ a σ-álgebra gerada por $\mathcal{F}_0$. Então, se $A \in \mathcal{F}$, existe uma sequência $A_n \in \mathcal{F}_0$, tal que $A_n \to A$ na pseudo-métrica d.*

Veja o Problema 21 deste capítulo e o Problema 10 do Capítulo 1.

Teorema 2.8. (Hewitt and Savage, 1955) *Sejam $\{X_i, i \geq 1\}$ v.a's independentes e identicamente distribuídas sobre $(\Omega, \mathcal{F}, P)$. Se Λ for um conjunto simétrico em*

$\mathcal{F}\{X_1, X_2, \ldots\}$, então $P(\Lambda) = 0$ ou $P(\Lambda) = 1$.

Prova: Suponha Λ simétrico. Pelo Lema 2.2, existe $\Lambda_n \in \mathcal{F}\{X_1, \ldots, X_n\}$ tal que $\Lambda_n \to \Lambda$, de acordo com $d(\Lambda_n, \Lambda) = P(\Lambda_n \triangle \Lambda)$. Ou seja, $P(\Lambda_n \triangle \Lambda) \to 0$. Então, $\Lambda = \{\omega : (X_1(\omega), X_2(\omega), \ldots) \in B\}$, para algum $B \in \mathcal{B}^\infty$ e $\Lambda_n = \{\omega : (X_1(\omega), \ldots, X_n(\omega)) \in B_n\}$, para algum $B_n \in \mathcal{B}^n$. Defina uma permutação σ_n como:

$$\sigma_n = \begin{pmatrix} 1 & 2 & \cdots & n & n+1 & n+2 & \cdots & 2n \\ n+1 & n+2 & \cdots & 2n & 1 & 2 & \cdots & n \end{pmatrix}.$$

Defina $M_n = \{\omega : (X_{n+1}(\omega), \ldots, X_{2n}(\omega)) \in B_n\}$, ou seja $M_n = \sigma_n \Lambda_n$. Então, $P(M_n \triangle \sigma_n \Lambda) = P(\Lambda_n \triangle \Lambda)$, pois $(X_1, \ldots X_n)$ tem a mesma distribuição que $(X_{n+1}, \ldots, X_{2n})$. Segue-se que $P(M_n \triangle \Lambda) = P(\Lambda_n \triangle \Lambda)$, porque Λ é simétrico. Temos, então:
(i) $\Lambda_n \to \Lambda$, ou $P(\Lambda_n \triangle \Lambda) \to 0$;
(ii) $M_n \to \Lambda$, pois $P(M_n \triangle \Lambda) = P(\Lambda_n \triangle \Lambda) \to 0$;
(iii) M_n e Λ_n são independentes.

De acordo com o Lema 2.1, $\Lambda_n \cap M_n \to \Lambda \cap \Lambda$, e portanto $P(\Lambda_n \cap M_n) \to P(\Lambda)$ ou $P(\Lambda_n)P(M_n) \to P(\Lambda \cap \Lambda)$, ainda pelo Lema 2.1. Como o lado esquerdo converge para $P(\Lambda)P(\Lambda)$, temos $P(\Lambda) = [P(\Lambda)]^2$, ou seja, $P(\Lambda) = 0$ ou $P(\Lambda) = 1$. $\square$

2.3 Leis dos grandes números

Sejam $X_1, X_2, \ldots$ v.a's. Uma lei dos grandes números (LGN) é qualquer teorema relacionado com a convergência de

$$\frac{X_1 + X_2 + \ldots + X_n - a_n}{b_n},$$

onde $\{a_n\}, \{b_n\}$ são sequências de constantes, $b_n \uparrow +\infty$. Como caso especial temos $(X_1 + \ldots + X_n)/n$.

Uma LGN é chamada uma *lei fraca* (LFrGN) se a convergência for em probabilidade e *lei forte* (LFGN) se a convergência for q.c.

Teorema 2.9. (Desigualdade de Kolmogorov) *Sejam $\{X_k, k \geq 1\}$ v.a's independentes, $E(X_k) = 0$, para todo k, $\mathrm{Var}(X_k) = \sigma_k^2 < \infty$. Se $S_k = X_1 + \ldots + X_k$, então para todo $\lambda > 0$,*

$$P\Big\{\max_{1 \leq k \leq n} |S_k| > \lambda\Big\} \leq \frac{1}{\lambda^2} \sum_{k=1}^{n} \sigma_k^2. \tag{2.1}$$

Prova: Seja $A = \{\omega : \max_{1 \leq k \leq n} |S_k| > \lambda\}$ e $A_i = \{\omega : |S_k| \leq \lambda,\ k < i,\ |S_i| > \lambda\}$. Então, os A_i's são disjuntos e $\cup_{i=1}^n A_i = A$. Temos que

$$\mathrm{Var}(S_n) = E(S_n^2) \geq \int_A S_n^2 dP = \sum_{i=1}^{n} \int_{A_i} S_n^2 dP.$$

Mas,

$$\int_{A_i} S_n^2 dP = \int_{A_i} (S_n - S_i + S_i)^2 dP \geq \int_{A_i} S_i^2 dP + 2 \int_{A_i} S_i(S_n - S_i) dP,$$

e

$$\int_{A_i} S_i(S_n - S_i) dP = \int I_{A_i} S_i (S_n - S_i) dP = \int I_{A_i} S_i dP \int (S_n - S_i) dP,$$

pois $I_{A_i} S_i$ e $S_n - S_i$ são independentes. Este último termo é nulo pois $\int (S_n - S_i) dP = 0$, logo

$$\int_{A_i} S_n^2 dP \geq \int_{A_i} S_i^2 dP.$$

Segue-se que

$$\mathrm{Var}(S_n) \geq \sum_{i=1}^{n} \int_{A_i} S_i^2 dP \geq \sum_{i=1}^{n} \lambda^2 P(A_i),$$

pois em A_i, $|S_i| > \lambda$. Como $\sum_{i=1}^{n} P(A_i) = P(A)$, temos que $\mathrm{Var}(S_n) \geq \lambda^2 P(A)$. Portanto, $\sum_{i=1}^{n} \sigma_i^2 = \mathrm{Var}(S_n) \geq \lambda^2 P(A)$, ou $P(A) \leq \frac{1}{\lambda^2} \sum_{i=1}^{n} \sigma_i^2$. $\square$

Teorema 2.10. *Sejam $\{X_k, k \geq 1\}$ independentes e $E(X_k) = 0$, para todo k. Suponha que $\sum_{k=1}^{\infty} \mathrm{Var}(X_k) = \sum_{k=1}^{\infty} E(X_k^2) < \infty$. Então, $\sum_k X_k$ converge q.c.*

Prova: Pela desigualdade de Kolmogorov,

$$P\Big\{\omega : \sup_{0 \leq i \leq n} |S_{m+i} - S_m| > \varepsilon\Big\} \leq \frac{1}{\varepsilon^2} \sum_{k=m+1}^{m+n} \sigma_k^2.$$

Então,

$$P\Big\{\omega : \sup_{i \geq 0} |S_{m+i} - S_m| > \varepsilon\Big\} \leq \frac{1}{\varepsilon^2} \sum_{k=m+1}^{\infty} \sigma_k^2,$$

porque os conjuntos decrescem. Logo, $\lim_{m\to\infty} P\{\omega : \sup_{i\geq 0} |S_{m+i} - S_m| > \varepsilon\} = 0$, porque temos a cauda da série convergente $\sum_k \mathrm{Var}(X_k)$. Isso implica que $\{S_n\}$ é uma sequência de Cauchy, para quase todo ω, logo S_n converge q.c. $\square$

No Teorema 2.10, dizer que $\sum_{k=1}^{\infty} \mathrm{Var}(X_k) < \infty$ é equivalente a dizer que $\sum_k X_k$ converge em L_2, logo o teorema pode ser enunciado como: Se $\sum_k X_k$ converge em L_2, então essa soma converge q.c.

Exemplo 2.4. Como uma aplicação do Teorema 2.10, vamos mostrar que a série $\sum_{k=1}^{\infty}\left(\frac{\pm 1}{k}\right)$ converge para quase todas as escolhas de $\pm$.

Sejam $\{r_k\}$ as funções de Rademaker sobre $[0,1)$ e considere a série $\sum_{k=1}^{\infty}\frac{r_k(t)}{k}$. Lembremos que essas funções são independentes. Segue-se que $X_k = r_k/k$ é uma sequência de v.a.'s, de média zero, $\sum_k \mathrm{Var}(X_k) = \sum_k k^{-2} < \infty$. Logo, pelo teorema anterior, $\sum_k r_k(t)/k$ converge q.c, ou seja, converge para quase todo $t \in [0,1)$. Mas $r_k(t) = \pm 1$, e o resultado segue.

Para provarmos uma versão da LFGM necessitamos do resultado seguinte (veja por exemplo Breiman (1968)).

Lema 2.3. (Lema de Kronecker) *Seja $\{x_k, k \geq 1\}$ uma sequência de números reais tais que $\sum_k x_k/a_k$ converge, onde $\{a_k, k \geq 1\}$ é uma sequência de números reais positivos, tais que $a_k \uparrow \infty$. Então, $\sum_{k=1}^{n} x_k/a_n \to 0$.*

Teorema 2.11. (Uma versão da LFGN) *Sejam $\{X_k, k \geq 1\}$ v.a's independentes, $E(X_k) = \mu_k$ e suponha que $\sum_{k=1}^{\infty}\sigma_k^2/k^2 < \infty$. Então,*

$$\frac{X_1 + X_2 + \ldots + X_n}{n} - \frac{\mu_1 + \mu_2 + \ldots + \mu_n}{n} \to 0 \;\; q.c. \tag{2.2}$$

Prova: Considere $Y_k = (X_k - \mu_k)/k$; então $\{Y_k, k \geq 1\}$ são independentes, têm média zero e $\sum_k \mathrm{Var}(Y_k) = \sum_k \sigma_k^2/k^2 < \infty$. Pelo Teorema 2.10, $\sum_k Y_k$ converge q.c. Pelo Lema de Kronecker, $\sum_{k=1}^{n} kY_k/n \to 0$ q.c, isto é,

$$\frac{1}{n}\left((X_1 - \mu_1) + (X_2 - \mu_2) + \ldots + (X_n - \mu_n)\right) \to \; 0 \;\; \text{q.c},$$

ou seja, obtemos (2.2). $\square$

Para provarmos a LFGN de Kolmogorov, precisamos do seguinte resultado.

Lema 2.4. *Se X é uma v.a com f.d F, tal que $E(|X|) < \infty$, então*

$$\sum_{n=1}^{\infty}\frac{1}{n^2}\int_{-n}^{n} x^2 dF(x) < \infty.$$

Prova: Temos que $\sum_{n=1}^{\infty}\frac{1}{n^2}\int_{-n}^{n} x^2 dF(x) = \sum_{n=1}^{\infty}\sum_{k=1}^{n}\frac{1}{n^2}\int_{\{k-1<|x|\leq k\}} x^2 dF(x)$. Mudemos a ordem de integração e usemos $\sum_{n>k} 1/n^2 < 2/k,\; n \geq 1$, para obter

$$\sum_{n=1}^{\infty}\sum_{k=1}^{n}\frac{1}{n^2}\int_{\{k-1<|x|\leq k\}} x^2 dF(x) = \sum_{k=1}^{\infty}\int_{\{k-1<|x|\leq k\}} x^2 dF(x)\sum_{n=k}^{\infty}\frac{1}{n^2}$$

$$\leq \sum_{k=1}^{\infty}\frac{2}{k}\int_{\{k-1<|x|\leq k\}} x^2 dF(x) \leq \sum_{k=1}^{\infty}\frac{2}{k}\int_{\{k-1<|x|\leq k\}} k|x| dF(x) = 2E(|X|) < \infty. \quad \square$$

Teorema 2.12. (LFGN de Kolmogorov) *Sejam* $\{X_n, n \geq 1\}$ *v.a's i.i.d. Se* $\mu = E(X_1)$ *e* $E(|X_1|) < \infty$, *então* $(X_1 + \ldots + X_n)/n \to \ \mu$ *q.c.*

Prova: Defina v.a's truncadas $\{Y_n, n \geq 1\}$ como segue:

$$Y_n = \begin{cases} X_n, & \text{se } |X_n| \leq n, \\ 0, & \text{se } |X_n| > n. \end{cases}$$

Defina $Z_n = X_n - Y_n$. Então,

$$\frac{X_1 + \ldots + X_n}{n} = \frac{Y_1 + \ldots + Y_n}{n} + \frac{Z_1 + \ldots + Z_n}{n}.$$

Sem perda de generalidade, podemos supor $E(X_1) = 0$, pois se não, considere $X_n - E(X_n)$, no lugar de X_n. A prova consiste de duas partes:

(a) Mostrar que $(Z_1 + \ldots + Z_n)/n \to \ 0$ q.c. Para isso, mostramos que $P\{Z_n \neq 0 \text{ i.v}\} = 0$. De fato, $P\{Z_n \neq 0\} = P\{|X_n| > n\}$. Mas, usando o fato de que as variáveis são i.i.d,

$$\sum_n P(|X_n| > n\} = \sum_n P\{|X_1| > n\} \leq E(|X_1|) < \infty,$$

usando o Teorema 1.18, logo pelo Lema de Borel-Cantelli, o resultado segue.

(b) Mostrar que $(Y_1 + \ldots + Y_n)/n \to \ 0$ q.c. Para isso, aplicamos o Teorema 2.11. Temos que

$$\sum_{k=1}^{\infty} \frac{\text{Var}(Y_k)}{k^2} \leq \sum_{k=1}^{\infty} \frac{E(Y_k^2)}{k^2} = \sum_{k=1}^{\infty} \frac{1}{k^2} \int_{-k}^{k} x^2 dF(x) < \infty,$$

pelo Lema 2.4. Então, usando o Teorema 2.11, $(Y_1 + \ldots + Y_n)/n - (E(Y_1) + \ldots + E(Y_n))/n \to \ 0$, q.c. Por hipótese, $E(X_1) = 0$, de modo que $E(Y_k) = \int_{-k}^{k} x dF(x) \to 0$, q.c, quando $k \to \infty$. Portanto $(E(Y_1) + \ldots + E(Y_n))/n \to 0$, q.c, do que segue o resultado. $\square$

Teorema 2.13. (Recíprocas à LFGN) *Sejam* $\{X_n, n \geq 1\}$ *v.a's i.i.d.*

(a) *Se* $(X_1 + \ldots + X_n)/n$ *converge q.c para um limite finito, então* $E(|X_1|) < \infty$;

(b) *Se* $E(|X_1|) = \infty$, *então* $\limsup_{n\to\infty} |(X_1 + \ldots + X_n)/n| = +\infty$, *q.c.*

Prova: (a) Como $(X_1 + \ldots + X_n)/n$ converge q.c, $|X_n|/n \to 0$ q.c, pois $X_n/n = S_n/n - S_{n-1}/n$, e a diferença converge para zero. Logo, $P(|X_n|/n > 1 \text{ i.v}\} = 0$ e pelo Lema de Borel-Cantelli, $\sum_n P(|X_n| > n) < \infty$ e como as v.a.'s são i.i.d, $E(|X_1|) < \infty$.

(b) Se $E(|X_1|) = \infty$, então $E(|cX_1|) = \infty$, para todo $c > 0$. Logo, $\sum P(|X_1| > cn) = \infty$, ou ainda $\sum P(|X_n| > cn) = \infty$. Por Borel-Cantelli, $P(|X_n|/n > c \text{ i.v}) = 1$. Agora, se $|X_n|/n = |S_n/n - S_{n-1}/n| > c$ i.v, então ou $|S_n/n| > c/2$ i.v, ou $|S_{n-1}/n| > c/2$ i.v. Como c é arbitráriamente grande, segue-se que $\limsup_n |S_n/n| = +\infty$. □

Seja $X_1, X_2, \ldots$ uma sequência de v.a's i.i.d, com f.d comum F, suposta desconhecida. Considere n valores observados de $X_1, \ldots, X_n$ e defina a *função de distribuição empírica* (f.d.e) como

$$F_n(\lambda, \omega) = \frac{\text{número dos } X_i(\omega), i \leq n, \text{ que são } \leq \lambda}{n}.$$

Então temos o seguinte importante resultado.

Teorema 2.14. (Glivenko-Cantelli) *Para quase todo ω, $F_n(\lambda, \omega)$ converge para $F(\lambda)$, uniformemente em λ.*

Prova: (a) Em primeiro lugar, verificamos que para cada λ, existe um conjunto nulo A_λ, tal que se $\omega \notin A_\lambda$, então $F_n(\lambda, \omega) \to F(\lambda)$.

De fato, defina

$$Y_n(\omega) = \begin{cases} 1, & \text{se } X_n(\omega) \leq \lambda, \\ 0, & \text{se } X_n(\omega) > \lambda. \end{cases}$$

Então, $Y_1, Y_2, \ldots$ são v.a's i.i.d. e $F_n(\lambda, \omega) = (Y_1 + \ldots + Y_n)/n$. Essa v.a converge para $E(Y_1)$, pela LFGN e $E(Y_1) = P(X_n \leq \lambda) = F(\lambda)$.
Deduzimos que existe um conjunto nulo N, tal que se $\omega \notin N$, então $\lim_{n\to\infty} F_n(\lambda, \omega) = F(\lambda)$, sempre que λ seja um número racional (Seja $\lambda_1, \lambda_2, \ldots$ uma enumeração dos racionais e tome $N = \cup_{k=1}^{\infty} A_{\lambda_k}$).

(b) F tem um número enumerável de descontinuidades. Seja $a_1, a_2, \ldots$ uma enumeração de tais descontinuidades. Vamos mostrar que para todo a_k, $F_n(a_k, \omega) - F_n(a_k-, \omega)$ converge para $F(a_k, \omega) - F(a_k-, \omega)$ q.c. Ou seja, para $\omega \notin A_{a_k}$, um conjunto nulo, temos essa convergência. De fato, seja

$$Z_k(\omega) = \begin{cases} 1, & \text{se } X_n(\omega) = a_k, \\ 0, & \text{caso contrário.} \end{cases}$$

Então, usando o mesmo argumento que em (a), obtemos o resultado. Além disso, existe um conjunto nulo M tal que, se $\omega \notin M$, teremos

$$\lim_{n\to\infty} \{F_n(\lambda, \omega) - F_n(\lambda-, \omega)\} = F(\lambda, \omega) - F(\lambda-, \omega),$$

sempre que λ seja um ponto no qual F tenha um salto.

(c) Tome qualquer $\omega \notin M \cup N$. Para tal ω, temos $F_n(\lambda, \omega) \to F(\lambda, \omega)$, para qualquer λ racional, e $\{F_n(\lambda, \omega) - F_n(\lambda-, \omega)\} \to F(\lambda, \omega) - F(\lambda-, \omega)$ para qualquer ponto de salto λ de F. Isso implica a convergência uniforme de F_n para F para o ω escolhido. $\square$

2.4 Séries aleatórias

Nessa seção provamos o teorema das três séries de Kolmogorov, que dá uma condição necessária e suficiente para que a série $\sum_k X_k$, de v.a's independentes, convirja q.c. Primeiramente, obtemos uma cota inferior na desigualdade de Kolmogorov.

Proposição 2.1. *Sejam $X_1, X_2, \ldots$ v.a's independentes, $E(X_k) = 0$, para todo k. Suponha que $|X_k| \leq c$, $\forall k \geq 1$. Se Var$(X_k) = \sigma_k^2$, então para todo $a > 0$*

$$P\Big\{\max_{0 \leq k \leq n} |S_k| > a\Big\} \geq 1 - \frac{(a+c)^2}{\sum_{k=1}^n \sigma_k^2}.$$

Prova: Seja $A = \{\max_{0 \leq k \leq n} |S_k| > a\}$, $A_i = \{\omega : |S_j(\omega)| \leq a,\ j < i,\ |S_i| > a\}$. Então, os conjuntos A_i são disjuntos e $\cup_{i=1}^n A_i = A$. Portanto,

$$E[I_A S_n^2] = \sum_{i=1}^n E(I_{A_i} S_i^2) + \sum_{i=1}^n E(I_{A_i}(S_n - S_i)^2).$$

Agora, $E(I_{A_i} S_i^2) \leq (a+c)^2 P(A_i)$, pois $S_{i-1} \leq a,\ X_i \leq c$. Por outro lado, $E(I_{A_i}(S_n - S_i)^2) = P(A_i) E(S_n - S_i)^2 \leq P(A_i) \sum_{k=1}^n \sigma_k^2$, usando a independência de I_{A_i} e $(S_n - S_i)^2$. Segue-se que temos

$$E(I_A S_n^2) \leq (a+c)^2 P(A) + P(A) \sum_{k=1}^n \sigma_k^2 = \Big[(a+c)^2 + \sum_{k=1}^n \sigma_k^2\Big] P(A), \tag{2.3}$$

e

$$E(I_A S_n^2) = E(S_n^2) - E(S_n^2 I_{A^c}) \geq \sum_{k=1}^n \sigma_k^2 - a^2[1 - P(A)], \tag{2.4}$$

de modo que combinando (2.3) e (2.4) temos

$$[(a+c)^2 + \sum_{k=1}^n \sigma_k^2] P(A) \geq \sum_{k=1}^n \sigma_k^2 - a^2[1 - P(A)],$$

logo

$$P(A) \geq \frac{\sum_{k=1}^n \sigma_k^2 - a^2}{(a+c)^2 + \sum_{k=1}^n \sigma_k^2 - a^2} = 1 - \frac{(a+c)^2}{(a+c)^2 + \sum_{k=1}^n \sigma_k^2 - a^2} \geq 1 - \frac{(a+c)^2}{\sum_{k=1}^n \sigma_k^2}.$$

□

Provamos antes que, se $X_1, X_2, \ldots$ são independentes, de média zero, e se $\sum_k \sigma_k^2 < \infty$, então $\sum_k X_k$ converge q.c. A recíproca é verdadeira se adicionarmos uma outra condição.

Teorema 2.15 *Sejam $X_1, X_2, \ldots$ independentes, de média zero, $|X_k| \leq c$, para todo k e algum c. Suponha que $\sum_k X_k$ convirja q.c. Então $\sum_k \sigma_k^2 < \infty$, onde $\sigma_k^2 =$Var(X_k).*

Prova: Suponha que $S_n = \sum_{k=1}^n X_k$ converge. Então $\sup_{n\geq 1} |S_{n+N} - S_N| \to 0$ q.c, quando $N \to \infty$. Portanto, $\lim_{N\to\infty} P\{\sup_{n\geq 1} |S_{n+N} - S_N| > \varepsilon\} = 0$. Tome N grande de modo que $P\{\sup_{n\geq 1} |S_{n+N} - S_N| > \varepsilon\} \leq 1/2$. Suponha que $\sum_{k=1}^{\infty} \sigma_k^2 = \infty$. Então, teremos

$$\frac{1}{2} \geq P\Big\{\sup_{n\geq 1} |S_{n+N} - S_N| > \varepsilon\Big\} \geq P\Big\{\sup_{M\geq n\geq 1} |S_{n+N} - S_N| > \varepsilon\Big\} \geq 1 - \frac{(\varepsilon + c)^2}{\sum_{k=N+1}^{N+M} \sigma_k^2},$$

usando a Proposição 2.1. Como $\sum_{k=N+1}^{N+M} \sigma_k^2 \to \infty$, quando $M \to \infty$, obtemos $1/2 \geq 1$, uma contradição. Logo devemos ter $\sum_{k=1}^{\infty} \sigma_k^2 < \infty$. □

Teorema 2.16. (Teorema das três séries de Kolmogorov) *Sejam $X_1, X_2, \ldots$ v.a's independentes. A série $\sum_n X_n$ converge q.c se e somente se para algum $c > 0$, as seguintes três séries convergem:*

(i) $\sum_n P\{|X_n| > c\}$; *(ii)* $\sum_n$ Var(X_n^c); *(iii)* $\sum_n E(X_n^c)$,

onde $X_n^c = X_n$, se $|X_n| \leq c$ e $X_n^c = 0$, se $|X_n| > c$.

Prova: (a) Suponha que as três séries dadas em (i)-(iii) convergem. Para mostrar que $\sum_n X_n$ converge q.c, basta mostrar que $\sum_n X_n^c$ converge q.c. pois (i) implica que $X_n^c = X_n$, com exceção de um número finito de índices n, pelo Lema de Borel-Cantelli. Mas, por (ii), temos que $\sum_{n=1}^{\infty} [X_n^c - E(X_n^c)]$ converge q.c, pelo Teorema 2.10. Também, por (iii) obtemos que $\sum_n X_n^c$ converge q.c.

(b) Suponha, agora, que $\sum_n X_n$ converge q.c. Então $X_n \to 0$ q.c, o que implica $P\{|X_n| > c \text{ i.v}\} = 0$, para todo $c > 0$, logo por Borel-Cantelli, $\sum_n P(|X_n| > c) < \infty$, e (i) segue. Também segue que $\sum_n X_n^c$ converge q.c, pois as caudas de ambas as séries são as mesmas.

Sejam $Y_1, Y_2, \ldots$ v.a's independentes tais que, para todo k, Y_k tenha a mesma distribuição que X_k^c e $\mathcal{F}\{Y_1, Y_2, \ldots\}$ seja independente de $\mathcal{F}\{X_1^c, X_2^c, \ldots\}$. Então, $\sum_n (X_n^c - Y_n)$ converge q.c. Os termos dessa soma têm média zero, e são limitados em valor absoluto por $2c$. Portanto, pelo Teorema 2.15, $\sum_n \text{Var}(X_n^c - Y_n) < \infty$. Mas $\text{Var}(X_n^c - Y_n) = 2\text{Var}(X_n^c)$, portanto $\sum_n \text{Var}(X_n^c) < \infty$, provando (ii).

Novamente, usando o Teorema 2.10, segue-se que $\sum_n [X_n^c - E(X_n^c)] < \infty$ e como $\sum_n X_n^c < \infty$, obtemos que $\sum_n E(X_n^c)$ converge q.c, e (iii) fica provada. □

Até agora usamos três métodos importantes:

(i) Truncamento: substituímos X_k por X_k^c;

(ii) Centralização com respeito a médias: substitiuímos X_k por $X_k - E(X_k)$;

(iii) Simetrização: substituímos X_k por $X_k - Y_k$, onde Y_k é independente de X_k e tem a mesma distribuição que X_k.

Uma outra possibilidade: centrar com respeito a medianas. Lembremos que a mediana de uma v.a. X é um número m tal que $P(X \geq m) \geq 1/2$ e $P(X \leq m) \geq 1/2$. Usaremos a notação $m(X)$.

Teorema 2.17 (Desigualdade de Lévy) *Sejam $X_1, X_2, \ldots$ independentes e $S_n = \sum_{k=1}^n X_k$. Então,*

$$P\Big\{ \max_{1\leq k\leq n} |S_k - m(S_n - S_k)| \geq \lambda \Big\} \leq 2P\{|S_n| \geq \lambda\}.$$

Prova: (a) Provamos primeiro que $P\{\max_{1\leq k\leq n}(S_k - m(S_n - S_k)) \geq \lambda\} \leq 2P\{S_n \geq \lambda\}$. Chamemos $m_{k,n} = m(S_n - S_k)$. Então, temos:

(i) $P\{\max_{1\leq k\leq n}(S_k - m_{k,n}) \geq \lambda, S_n \geq \lambda\} \leq P\{S_n \geq \lambda)$,

(ii) $P\{\max_{1\leq k\leq n}(S_k - m_{k,n}) \geq \lambda, S_n < \lambda\} = \sum_{k=1}^{n-1} P(\tau = k, S_n < \lambda)$,

onde τ é o primeiro inteiro k tal que $S_k - m_{k,n} \geq \lambda$. Segue-se que a última soma é $\leq \sum_{k=1}^{n-1} P(\tau = k, S_n < S_k - m_{k,n}) = \sum_{k=1}^{n-1} P(\tau = k)P(S_n - S_k < -m_{k,n}) = \sum_{k=1}^{n-1} P(\tau = k)P(m_{k,n} < S_k - S_n))$, devido à independência entre $\{\tau = k\}$ e $\{S_n < S_k - m_{k,n}\}$.

Pela definição de mediana, $P(m_{k,n} < S_k - S_n) \leq P(S_k - S_n \leq m_{k,n})$, logo $\sum_{k=1}^{n-1} P(\tau = k)P(m_{k,n} < S_k - S_n)) \leq \sum_{k=1}^{n-1} P(\tau = k)P(S_n \geq S_k - m_{k,n}) = \sum_{k=1}^{n-1} P(\tau = k, S_n \geq S_k - m_{k,n}) \leq \sum_{k=1}^{n-1} P(\tau = k, S_n \geq \lambda) = P(S_n \geq \lambda)$.

Portanto, $P\{\max(S_n - m_{k,n}) \geq \lambda, S_n < \lambda\} \leq P(S_n \geq \lambda\}$. Adicione (i) e (ii) para obter o desejado.

(b) Para o caso geral, na parte (a) substitua X_n por $-X_n$ na prova. Obtenha

$$P\Big\{ \max_{1\leq k\leq n} (-S_k + m_{k,n}) \geq \lambda \Big\} \leq 2P\{-S_n \geq \lambda\},$$

portanto

$$P\Big\{ \max_{1\leq k\leq n} |S_k - m_{k,n}| \geq \lambda \Big\} \leq P\Big\{ \max_{1\leq k\leq n} (S_k - m_{k,n}) \geq \lambda \Big\} + P\Big\{ \max_{1\leq k\leq n} (-S_k + m_{k,n}) \geq \lambda \Big\}$$

$$\leq 2P\{S_n \geq \lambda) + 2P\{-S_n \geq \lambda\} = 2P\{|S_n| \geq \lambda\}. \quad \square$$

Esse resultado pode ser usado para provar o teorema a seguir.

Teorema 2.18. *Sejam $\{X_n, n \geq 1\}$ v.a's independentes. Então $\sum_n X_n$ converge em probabilidade se e somente se $\sum_n X_n$ converge q.c.*

Prova: ($\Leftarrow$) Trivial
($\Rightarrow$) Seja $S_n = \sum_{k=1}^{n} X_k$. Suponha que $S_k \ \to \ S$ em probabilidade. Então, existe uma subsequência $\{n_k\}$ tal que $S_{n_k} \to S$ q.c e $\sum_{k=1}^{\infty} P\{|S_{n_k} - S_{n_{k+1}}| > 1/2^k\} < \infty$. Defina

$$M_k = \max_{n_k \leq n \leq n_{k+1}} |S_n - S_{n_k} - m(S_n - S_{n_{k+1}})|.$$

Então, pela desigualdade de Lévy,

$$P(M_k \geq 1/2^k) \leq 2P\{|S_{n_k} - S_{n_{k+1}}| > 1/2^k\}.$$

Logo $\sum_k P(M_k \geq 1/2^k) < \infty$, implicando que $M_k \to 0$ q.c. Ou seja, para $n_k \leq n \leq n_{k+1}$,

$$|S_n - m(S_{n_{k+1}} - S_n) - S| \leq |S - S_{n_k}| + |S_n - S_{n_k} - m(S_{n_{k+1}} - S_n)|,$$

e como $S - S_{n_k} \to 0$, e o segundo termo é menor ou igual a M_k, que tende a zero q.c, segue-se que $S_n - m(S_{n_{k+1}} - S_n) \to S$ q.c. Mas $S_n \to S$ em probabilidade, portanto $m(S_{n_{k+1}} - S_n) \to 0$, isto é, $S_n \to S$ q.c. $\square$

Problemas

1. Prove a Proposição 2.2.
2. Prove que as v.a's definidas no Exemplo 2.2 são independentes.
3. Prove a Aplicação 1, logo após o Teorema 2.6.
4. Prove formalmente a Aplicação 1 (b), após o Teorema 2.7.
5. Idem, Aplicação 3.
6. Sejam $X_1, X_2, \ldots$ independentes. Prove que se S_n/n converge a um limite finito Y, então Y é necessariamente constante.
7. Prove (i) do Exemplo 2.3.
8. Prove que o evento Λ de (iii) do Exemplo 2.3 é um evento simétrico.
9. Prove que a classe dos eventos simétricos é uma σ-álgebra.
10. Sejam $\{X_i, i \geq 1\}$ v.a's i.i.d. Mostre que $P\{\omega : X_n(\omega) \text{ converge}\} = 0$, supondo que a distribuição de X_1 não está concentrada num único ponto.

11. Sejam $\{X_i, i \geq 1\}$ v.a's i.i.d., com f.d F definida por $F(x) = 1 - e^{-x},\ x \geq 0,\ F(x) = 0,\ x < 0$. Prove que:

    (a) $P\{(X_n/\log n) > 2 \text{ i.v}\} = 0$, mas que

    (b) $P\{(X_n/\log n) > 1 \text{ i.v}\} = 1$.

12. Sejam $X_1, X_2, \ldots$ v.a's i.i.d., $P(X_1 = 1) = p,\ P(X_1 = -1) = q,\ p > q,\ p+q = 1$. Seja $S_n = X_1 + \ldots + X_n$. S_n é um passeio aleatório. Então, prove que $P(S_n = 0 \text{ i.v}) = 0$.

13. Prove que $P\{\omega : \liminf_{n\to\infty} \sum_{k=1}^n X_k(\omega) > -\infty\} = 0$ ou 1, onde $\{X_n, n \geq 1\}$ é uma sequência de v.a's independentes, cada uma q.c finita.

14. Seja $\{X_n, n \geq 1\}$ uma sequência de v.a's i.i.d, com $E(X_n) = 0$ e seja $\{c_n, n \geq 1\}$ uma sequência limitada de constantes. Prove que $\sum_{k=1}^n c_k X_k/n \to 0$ q.c.

15. A afirmação: $\limsup_n (X_1+\ldots+X_n)/n$ é mensurável relativamente à σ-álgebra caudal, é falsa ou verdadeira? Justifique.

16. Sejam X e Y v.a's independentes e suponha que $E(|X+Y|) < \infty$. Prove que $E(|X|) < \infty$.

17. Sejam $\{X_i, i \geq 1\}$ v.a's i.i.d, $E(|X_1|) < \infty$. Prove que $(X_1 + \ldots + X_n)/n$ converge para $E(X_1)$ em L_1.

18. Sejam $\{X_i, i \geq 1\}$ v.a's i.i.d, cada uma $N(0,1)$. Mostre que $\sum_n (X_n/n^\alpha)$ converge q.c se $\alpha > 1/2$ e diverge se $\alpha \leq 1/2$.

19. Suponha $\{X_i, i \geq 1\}$ v.a's com médias μ_i e variâncias σ_i^2, não necessariamente independentes. Suponha X_i não correlacionadas.

    (a) Prove que $\mathrm{Var}(\sum_{i=1}^n X_i) = \sum_{i=1}^n \mathrm{Var}(X_i)$;

    (b) Prove que, se $\sum_{i=1}^n \sigma_i^2/n^2 \to 0$, para $n \to \infty$, então $(X_1 + \ldots + X_n)/n - (\mu_1 + \ldots + \mu_n)/n \to 0$, em L_1 e em probabilidade.

20. Seja $(\Omega, \mathcal{F}, P)$ um e.p. e os eventos A, B de $\mathcal{F}$. Prove que, se $P(A) = 0$ ou 1, então A e B são independentes.

21. Prove o Lema 2.1.

22. Seja $\Omega = \{1,2,3,4,5,6,7,8\}$ e $\mathcal{F} = 2^\Omega$, Suponha que $P(\{1\}) = \cdots = P(\{8\}) = 1/8$. Sejam $\mathcal{F}_1$ e $\mathcal{F}_2$ σ-álgebras sobre Ω geradas por $\{1,3,5,7\}$ e $\{1,2,3,4\}$, respectivamente. Verifique se $\mathcal{F}_1$ e $\mathcal{F}_2$ são independentes.

23. Prove que as v.a's do Exemplo 2.1 são independentes, mostrando que ambos os lados da igualdade são iguais a 2^{-k}.

24. (Doukhan, 2015) (a) Sejam X, Y v.a's reais e independentes com X simétrica (isso significa que $-X$ tem a mesma distribuição de X), $E(X^2) < \infty$ e $P(Y = \pm 1) = 1/2$. Considere $Z = XY$. Prove que $\mathrm{Cov}(X, Z) = 0$ e, além disso, se $|X|$ não for constante q.c, então X, Z não são independentes.

    (b) Se as v.a's X e Y têm valores em $\{0, 1\}$ e satisfazem $\mathrm{Cov}(X, Y) = 0$, prove que X e Y são independentes.

Capítulo 3

Esperança Condicional

Neste capítulo iremos tratar do importante conceito de esperança condicional com respeito à uma σ-álgebra. Estudaremos as suas propriedades mais importantes e terminaremos com o conceito de probabilidade condicional regular. As referências principais para este capítulo são Chung (1968, 1974), Breiman (1968, 1992) e Billingley (1995).

3.1 Definições e fatos básicos

No Problema 17 do Capítulo 1, para $A \in \mathcal{F}, P(A) > 0$, definimos a probabilidade condicional $P(B|A) = P(A \cap B)/P(A)$, para todo $B \in \mathcal{F}$. Segue-se que $(\Omega, \mathcal{F}, P(\cdot|A))$ é um e.p. Dessa definição seguem resultados importantes, como a lei da probabilidade total, $P(B \cap A) = P(A)P(B|A)$, e o Teorema de Bayes,

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

que nos diz que a probabilidade *a posteriori* de B, dado que A ocorreu, é obtida, essencialmente, pelo produto da probabilidade *a priori de* B, $P(B)$, pela *verossimilhança* $P(A|B)$. Veja os Problemas 21 e 22.

Se X for uma v.a definida neste e.p, com valores $\{x_k, k \geq 1\}$, podemos também definir a probabilidade condicional

$$P(A|X = x_k) = \frac{P(A, X = x_k)}{P(X = x_k)},$$

se $P(X = x_k) > 0$, e definida arbitrariamente como sendo zero, se a probabilidade do denominador for zero. No caso geral, podemos considerar $A \in \mathcal{F}$ e $B \in \mathcal{B}$, com $P(X \in B) > 0$ e definir

$$P(A|X \in B) = \frac{P(A, X \in B)}{P(X \in B)}.$$

Se quisermos dar um significado preciso para $P(A|X = x)$ teremos que recorrer ao conceito de derivada de Radon-Nikodym, o que será feito a seguir, quando definirmos o conceito mais geral de esperança condicional. Uma maneira equivalente é definir a probabilidade condicional de A, dada $X(\omega)$, como qualquer v.a sobre Ω, $\mathcal{F}\{X\}$-mensurável, satisfazendo

$$P(A, X \in B) = \int_{\{X \in B\}} P(A|X)dP, \quad \text{para todo } B \in \mathcal{B}.$$

Quaisquer duas versões de $P(A|X)$ diferem num conjunto de probabilidade nula. Ver Breiman (1968) para detalhes.

De modo análogo, podemos considerar a esperança condicional $E(Y|X = x)$, dadas duas v.a's X e Y sobre $(\Omega, \mathcal{F}, P)$. Se $E(|Y|) < \infty$, então $E(Y|X)$ é qualquer função $\mathcal{F}\{X\}$-mensurável satisfazendo

$$\int_A E(Y|X)dP = \int_A Y dP, \quad \text{para todo } A \in \mathcal{F}\{X\}.$$

Ou seja, tanto $P(A|X)$ como $E(Y|X)$ dependem somente de $\mathcal{F}\{X\}$.

A seguir definimos uma esperança condicional mais geral, ou seja, a esperança condicional de uma v.a com respeito a uma σ-álgebra.

Definição 3.1. *Seja $(\Omega, \mathcal{G}, P)$ um e.p e $\mathcal{F}$ uma σ-álgebra contida em $\mathcal{G}$. Seja X uma v.a integrável sobre $(\Omega, \mathcal{G}, P)$. A esperança condicional de X com respeito a $\mathcal{F}$, denotada por $E(X|\mathcal{F})$, é qualquer v.a satisfazendo:*

(i) *$E(X|\mathcal{F})$ é $\mathcal{F}$-mensurável;*

(ii) *Se Λ é qualquer conjunto em $\mathcal{F}$, então*

$$\int_\Lambda E(X|\mathcal{F})dP = \int_\Lambda X dP.$$

Note que $E(X|\mathcal{F})$ não é definida univocamente, mas quaisquer duas v.a's que satisfazem (i) e (ii) serão iguais q.c. Assim, $E(X|\mathcal{F})$ é qualquer uma das classes de equivalência de v.a's sobre Ω satisfazendo (i) e (ii).

Seja $(\Omega, \mathcal{G})$ qualquer espaço mensurável, μ uma medida sobre esse espaço e ν uma medida sinalizada sobre o mesmo espaço. Dizemos que ν é *absolutamente contínua* com respeito a μ se $\nu(A) = 0$ sempre que $\mu(A) = 0$, para todo $A \in \mathcal{G}$. Escrevemos $\nu << \mu$. O seguinte resultado é fundamental (veja Halmos (1976)).

Teorema 3.1. (Radon-Nikodym) *Seja $(\Omega, \mathcal{G})$ um espaço mensurável e μ uma medida finita sobre o mesmo. Suponha $\nu << \mu$. Então, existe uma função $\mathcal{G}$-mensurável X tal que, para todo $A \in \mathcal{G}$,*

$$\nu(A) = \int_A X d\mu.$$

A v.a X é única a menos de conjuntos de medida μ-nula; dizemos que X é a derivada de Radon-Nikodym de ν com respeito a μ e escrevemos $X = \frac{d\nu}{d\mu}$. Usamos esse fato para provar o seguinte resultado.

Teorema 3.2. *A esperança condicional como definida acima existe.*

Prova: Considere a função de conjunto ν sobre $\mathcal{F}$ definida por

$$\nu(A) = \int_A X dP, \quad \forall A \in \mathcal{F}.$$

Esta função tem valores finitos e é enumeravelmente aditiva, logo é uma medida sinalizada. Se $P(A) = 0$, então $\nu(A) = 0$, logo $\nu << P$. Pelo Teorema 3.1, existe uma função $\mathcal{F}$-mensurável Y tal que $\nu(A) = \int_A Y dP$. Segue-se que Y satisfaz a definição de esperança condicional, e $d\nu/dP = E(X|\mathcal{F})$. $\square$

Para a σ-álgebra $\mathcal{F}\{X\}$ gerada pela v.a X, escrevemos $E(Y|X)$ para $E(Y|\mathcal{F}\{X\})$. De modo similar, $E(Y|X_1, \ldots, X_n)$ é definida como $E(Y|\mathcal{F}\{X_1, \ldots, X_n\})$. Para $\Lambda \in \mathcal{G}$, defina $P(\Lambda|\mathcal{F}) = E(I_\Lambda|\mathcal{F})$, como sendo a probabilidade condicional de Λ com respeito a $\mathcal{F}$. Especificamente, $P(\Lambda|\mathcal{F})$ é qualquer uma das classes de equivalência de v.a's $\mathcal{F}$-mensuráveis satisfazendo

$$P(\Lambda \cap B) = \int_B P(\Lambda|\mathcal{F}) dP, \text{ para todo } B \in \mathcal{F}.$$

Considere $X_1, X_2, \ldots$ v.a's sobre $(\Omega, \mathcal{G}, P)$ e $E(Y|X_1, \ldots, X_n)$. Cada versão da esperança condicional de Y, dadas $X_1, \ldots, X_n$ é $\mathcal{F}\{X_1, \ldots, X_n\}$-mensurável. Tome qualquer uma dessas versões. Então existe uma função mensurável de Borel $\varphi : \mathbb{R}^n \to \mathbb{R}$, tal que

$$E(Y|X_1, \ldots, X_n) = \varphi(X_1, \ldots, X_n) \text{ q.c},$$

pelo Teorema 1.10. Como consequência desse fato, a função $E(Y|X)$ (ou $E(Y|X_1, \ldots, X_n)$), como função de ω, é constante q.c em cada conjunto sobre o qual $X(\omega)$ seja constante (ou sobre o qual $(X_1, \ldots, X_n)$ seja constante). Frequentemente, usamos a notação $E(Y|X = x) = \varphi(x)$ ou $E(Y|X_1 = x_1, \ldots, X_n = x_n) = \varphi(x_1, \ldots, x_n)$.

Exemplo 3.1. (a) Se $\mathcal{F} = \{\emptyset, \Omega\}$, então $E(X|\mathcal{F}) = E(X)$.

(b) Se X é $\mathcal{F}$-mensurável, então $E(X|\mathcal{F}) = X$.

Exemplo 3.2. (a) Sejam $\Lambda_1, \ldots, \Lambda_n$ eventos disjuntos em $\mathcal{G}$ tais que e $\cup_i \Lambda_i = \Omega$ e $P(\Lambda_i) > 0$ para todo $i \in \{1, \ldots, n\}$. Seja $\mathcal{F}$ a σ-álgebra gerada pelos $\{\Lambda_i\}$. Então,

$$E(X|\mathcal{F})(\omega) = \frac{1}{P(\Lambda_i)}\int_{\Lambda_i} XdP, \quad \text{para todo } \omega \in \Lambda_i,\ i \in \{1,\ldots,n\}. \tag{3.1}$$

De fato, (i) da definição está satisfeita, pois (3.1) é constante nos Λ_i. Quanto a (ii), devemos mostrar que, se $\Lambda \in \mathcal{F}$, então $\int_\Lambda E(X|\mathcal{F})dP = \int_\Lambda XdP$. É suficiente verificar a igualdade para cada Λ_i, pois $\mathcal{F}$ é composta por reuniões dos Λ_i. Então, para cada i,

$$\int_{\Lambda_i} E(X|\mathcal{F})dP = \int_{\Lambda_i}\left[\frac{1}{P(\Lambda_i)}\int_{\Lambda_i} XdP\right]dP = \frac{1}{P(\Lambda_i)}\int_{\Lambda_i} dP\int_{\Lambda_i} XdP = \int_{\Lambda_i} XdP.$$

Note que $E(X|\mathcal{F})$ é constante q.c sobre os átomos de $\mathcal{F}$ (dada uma σ-álgebra $\mathcal{F}$, um átomo de $\mathcal{F}$ é qualquer conjunto $\Lambda \in \mathcal{F}$ tal que, se $A \subset \Lambda$ e se $A \in \mathcal{F}$, então $P(A) = 0$ ou $P(A) = P(\Lambda)$).

(b) Dados os Λ_i de (a), defina $Y = \sum_{i=1}^n c_i I_{\Lambda_i}$, sendo os c_i distintos. Então, $E(X|Y) = E(X|\mathcal{F})$, onde $\mathcal{F}$ é a σ-álgebra gerada pelos Λ_i.

(c) Seja $\mathcal{F}$ gerada por um conjunto Λ, isto é, $\mathcal{F} = \{\emptyset, \Omega, \Lambda, \Lambda^c\}$. Se $0 < P(\Lambda) < 1$, então para $A \in \mathcal{G}$,

$$P(A|\mathcal{F}) = \begin{cases} P(A|\Lambda), & \text{se } \omega \in \Lambda, \\ P(A|\Lambda^c), & \text{se } \omega \in \Lambda^c. \end{cases}$$

Exemplo 3.3. Seja $\Omega = [-1,1]$, $\mathcal{F} = \mathcal{B}([-1,1])$ e P=(medida de Lebesgue)/2. Defina uma v.a Y sobre $(\Omega, \mathcal{F}, P)$ por $Y(\omega) = \omega$. Então, Y gera $\mathcal{F}$. Seja X uma v.a sobre o mesmo e.p, integrável.

(a) $E(X|Y) = X$, pois X é $\mathcal{F}$-mensurável e $\mathcal{F}\{Y\} = \mathcal{F}$.

(b) $E(X|Y^3) = X$, pela mesma razão.

(c) $E(X|Y^2) = [X(\omega) + X(-\omega)]/2$. Note que, agora, a σ-álgebra gerada por Y^2 consiste de todos os conjuntos de Borel M, tais que $M = -M$. Mostre que (i) e (ii) da definição estão satisfeitas.

Exemplo 3.4. Seja $\Omega = \mathbb{R}^2, \mathcal{G}$ a σ-álgebra de Borel sobre $\mathbb{R}^2$; seja uma função não negativa, $f : \mathbb{R}^2 \to \mathbb{R}$, com $\int_{\mathbb{R}^2} f(x,y)dxdy = 1$. Defina P sobre $(\Omega, \mathcal{G})$ por $P(A) = \int_A fdxdy$, para todo $A \in \mathcal{G}$. Defina v.a's X e Y sobre $(\Omega, \mathcal{G}, P)$ por:

$$\text{se } \omega \in \mathbb{R}^2, \omega = (\omega_1, \omega_2), \ \ X(\omega) = \omega_1,\ Y(\omega) = \omega_2.$$

Então, a f.d conjunta de (X, Y) é dada por

$$F(x,y) = P\{\omega : X(\omega) \leq x,\ Y(\omega) \leq y\} = \int_{-\infty}^{x}\int_{-\infty}^{y} f(x,y)dxdy.$$

Seja $f_1(x)$ a densidade marginal de X. Então, afirmamos que uma versão da $E[g(Y)|X]$, para alguma função de Borel g, é dada por

$$E[g(Y)|X] = \frac{\int_{-\infty}^{\infty} g(y)f(X,y)dy}{f_1(X)} =: h(X).$$

Para justificar tal afirmação, temos que provar que $h(X)$ é $\mathcal{F}\{X\}$-mensurável e satisfaz a propriedade (ii).

(i) $f_1(x)$ é $\mathcal{F}\{X\}$-mensurável, o mesmo valendo para $\int_{-\infty}^{\infty} g(y)f(x,y)dy$, pelo teorema de Fubini. Em seguida, vamos mostrar que $P\{f_1(X) = 0\} = 0$. Logo, o quociente $h(X)$ em (3.8) é bem definido q.c. Definindo $h(x) = 0$ nos pontos tais que $f_1(x) = 0$, obtemos que h é $\mathcal{F}\{X\}$-mensurável. Seja $\Lambda = \{x : f_1(x) = 0\}$. Então,

$$P(\Lambda) = \int_{\Lambda}\int_{-\infty}^{\infty} f(x,y)dxdy = \int_{\Lambda} f_1(x)dx = 0.$$

(ii) Tome $\Lambda \in \mathcal{F}\{X\}$. Devemos provar que

$$\int_{\Lambda} g(Y)dP = \int_{\Lambda} h(X)dP.$$

Mas qualquer tal Λ é da forma $\Lambda = A_1 \times \mathbb{R}$, onde $A_1 \in \mathcal{B}$, logo

$$\int_{A_1\times\mathbb{R}} g(Y)dP = \int_{A_1}\int_{\mathbb{R}} g(y)f(x,y)dxdy = \int_{A_1}\Big(\int_{\mathbb{R}} \frac{g(y)f(x,y)}{f_1(x)} f_1(x)dy\Big)dx$$

$$= \int_{A_1} h(x)f_1(x)dx = \int_{A_1} h(x)\Big(\int_{\mathbb{R}} f(x,y)dy\Big)dx = \int_{A_1\times\mathbb{R}} h(x)f(x,y)dxdy$$

$$= \int_{A_1\times\mathbb{R}} h(x)dP.$$

Note que, na primeira e última igualdades, usamos os fato que $dP = fdxdy$.

3.2 Propriedades da esperança condicional

As propriedades da esperança condicional são de três tipos: aquelas análogas a propriedades das integrais, aquelas denominadas de suavização e uma propriedade relacionada a espaços lineares. Lembramos aqui que todas as igualdades (e desigualdades) a seguir envolvendo esperanças condicionais valem q.c.

[A] Propriedades de Integrais

Seja X, Y v.a's sobre $(\Omega, \mathcal{G}, P)$, integráveis e $\mathcal{F} \subset \mathcal{G}$.

P1. $E(\alpha X + \beta Y|\mathcal{F}) = \alpha E(X|\mathcal{F}) + \beta E(Y|\mathcal{F})$.

Veja o Problema 4.

P2. Se $X \leq Y$ q.c, então $E(X|\mathcal{F}) \leq E(Y|\mathcal{F})$

De fato, para cada $\Lambda \in \mathcal{F}$, temos

$$\int_\Lambda E(X|\mathcal{F})dP = \int_\Lambda X dP \leq \int_\Lambda Y dP = \int_\Lambda E(Y|\mathcal{F})dP.$$

P3. (Teorema da Convergência Dominada). Seja $\{X_n\}$ uma sequência de v.a's integráveis, $X_n \to X$ q.c. Suponha $\sup_n |X_n|$ integrável. Então,

$$\lim_{n\to\infty} E(X_n|\mathcal{F}) = E(X|\mathcal{F}).$$

(a) Suponha $X_n \geq 0$, $X_n \downarrow 0$ q.c. Então, $X_1 \geq X_2 \geq \cdots$ e por P2,

$$E(X_1|\mathcal{F}) \geq E(X_2|\mathcal{F}) \geq \cdots, \tag{3.2}$$

logo $E(X_n|\mathcal{F}) \downarrow W \geq 0$. Agora, $E(X_n) \to 0$ pelo TCD e $E[E(Y|\mathcal{F})] = E(Y)$, pela definição de esperança condicional, com $\Lambda = \Omega$. Por (3.2), $E(X_n) \geq E(W)$, portanto $E(W) = 0$. Segue-se que $\lim_{n\to\infty} E(X_n|\mathcal{F}) = 0$.

(b) Para o caso geral, temos que

$$|E(X_n|\mathcal{F}) - E(X|\mathcal{F})| = |E(X_n - X)|\mathcal{F})| \leq E(|X_n - X| \; |\mathcal{F}) \leq E\Big(\sup_{k\geq n} |X_k - X| \; |\mathcal{F}\Big),$$

e o último termo tende a zero pela parte (a). A primeira desigualdade na expressão acima, isto é, $|E(X|\mathcal{F})| \leq E(|X||\mathcal{F})$ pode ser obtida usando $-|X| \leq X \leq |X|$.

P4. (Teorema da Convergência Monotônica) Suponha que $X_n \geq 0$, $X_n \uparrow X$, X integrável. Então, $\lim_{n\to\infty} E(X_n|\mathcal{F}) = E(X|\mathcal{F})$ q.c.

Chamemos de Z o limite de $E(X_n|\mathcal{F})$, que existe q.c, pela propriedade anterior. Então, para cada $\Lambda \in \mathcal{F}$, temos

$$\int_\Lambda Z dP = \lim_n \int_\Lambda E(X_n|\mathcal{F})dP = \lim_n \int_\Lambda X_n dP = \int_\Lambda X dP,$$

sendo que na última igualdade usamos o TCM. Segue-se que Z satisfaz a definição de esperança condicional e é $\mathcal{F}$-mensurável, logo $Z = E(X|\mathcal{F})$.

P5. (Lema de Fatou). Se $X_n \geq 0$ e X_n, $\liminf_n X_n$ são integráveis, então

$$E(\liminf_n X_n|\mathcal{F}) \leq \liminf_n E(X_n|\mathcal{F}).$$

P6. (Desigualdade de Jensen). Se φ é uma função convexa sobre $\mathbb{R}$ e X e $\varphi(X)$ são integráveis, então

$$\varphi\left(E(X|\mathcal{F})\right) \leq E\left(\varphi(X)|\mathcal{F}\right).$$

[B] Propriedades de Suavização

P7. Se X for integrável e $\mathcal{F}_1 \subset \mathcal{F}_2$, então

$$E\left[E(X|\mathcal{F}_1)|\mathcal{F}_2\right] = E\left[E(X|\mathcal{F}_2)|\mathcal{F}_1\right] = E(X|\mathcal{F}_1).$$

$E(X|\mathcal{F}_1)$ é $\mathcal{F}_1$-mensurável e como $\mathcal{F}_1 \subset \mathcal{F}_2$, é também $\mathcal{F}_2$-mensurável, logo $E[E(X|\mathcal{F}_1)|\mathcal{F}_2] = E(X|\mathcal{F}_1)$, pelo Exemplo 3.1 (b). Seja, agora, $W = E[E(X|\mathcal{F}_2)|\mathcal{F}_1]$. Tome $\Lambda \in \mathcal{F}_1$. Devemos mostrar que $\int_\Lambda W dP = \int_\Lambda X dP$.
Mas $\int_\Lambda W dP = \int_\Lambda E(X|\mathcal{F}_2) dP = \int_\Lambda X dP$, pois se $\Lambda \in \mathcal{F}_1$, então $\Lambda \in \mathcal{F}_2$.

Como casos especiais desse resultado, temos:

(i) $E[E(X|\mathcal{F}_1)] = E(X)$.
(b) Se $\mathcal{F}_1 = \mathcal{F}\{Y_1\}$, $\mathcal{F}_2 = \mathcal{F}\{Y_2\}$, temos

$$E\left[E(X|Y_1)|Y_1, Y_2\right] = E\left[E(X|Y_1, Y_2)|Y_1\right] = E(X|Y_1).$$

P8. Suponha que X e XY sejam integráveis e X seja $\mathcal{F}$-mensurável. Então,

$$E(XY|\mathcal{F}) = XE(Y|\mathcal{F}). \tag{3.3}$$

É fácil ver que (3.3) vale para $X = I_B$, $B \in \mathcal{F}$. A seguir, é válida para uma função simples e logo se $X \geq 0$, $\mathcal{F}$-mensurável. Para o caso geral, considere $X = X^+ - X^-$.

P9. Seja $(\Omega, \mathcal{G}, P)$ um e.p, X uma v.a sobre esse espaço, integrável, e $\mathcal{F}_1$, $\mathcal{F}_2$ duas sub-σ-álgebras de $\mathcal{G}$. Suponha que $\mathcal{F}\{X\} \vee \mathcal{F}_1$ seja independente de $\mathcal{F}_2$. Então, $E(X|\mathcal{F}_1 \vee \mathcal{F}_2) = E(X|\mathcal{F}_1)$.

Temos que $E(X|\mathcal{F}_1)$ é mensurável relativamente a $\mathcal{F}_1 \vee \mathcal{F}_2$. É suficiente mostrar que se $\Lambda \in \mathcal{F}_1 \vee \mathcal{F}_2$, então

$$\int_\Lambda E(X|\mathcal{F}_1) dP = \int_\Lambda X dP. \tag{3.4}$$

Primeiramente, suponha que $\Lambda = A_1 \cap A_2$, $A_i \in \mathcal{F}_i$, $i = 1, 2$. Então,

$$\int_{A_1\cap A_2} E(X|\mathcal{F}_1)dP = \int I_{A_1}I_{A_2}E(X|\mathcal{F}_1)dP = \int I_{A_2}dP \int I_{A_1}E(X|\mathcal{F}_1)dP$$

$$= P(A_2)\int_{A_1} E(X|\mathcal{F}_1)dP = P(A_2)\int_{A_1} XdP = \int_{A_1\cap A_2} XdP.$$

a segunda igualdade por independência e por definição de esperança condicional, a última igualdade novamente usando a independência.

Logo, (3.4) vale para esse caso. Também, (3.4) vale se Λ for uma reunião de conjuntos disjuntos da forma $A_1 \cap A_2,\ A_i \in \mathcal{F}_i,\ i = 1, 2$. Mas a coleção de todos os conjuntos dessa forma é uma álgebra, de modo que (3.4) é verdadeira para todos os conjuntos em uma álgebra, e a σ-álgebra gerada por essa álgebra é $\mathcal{F}_1 \vee \mathcal{F}_2$. Concluimos a prova observando que os eventos satisfazendo (3.4) formam uma classe monotônica. Obtemos finalmente que (3.4) é satisfeita para todo $\Lambda \in \mathcal{F}_1 \vee \mathcal{F}_2$ (veja o Apêndice A.1).

Um caso especial importante desse resultado é que, se X e $\mathcal{F}$ são independentes,, então

$$E(X|\mathcal{F}) = E(X).$$

[C] Propriedade de Espaços Lineares

Seja $(\Omega, \mathcal{G}, P)$ um e.p e $\mathcal{F} \subset \mathcal{G}$. Então, $E(\cdot|\mathcal{F})$ pode ser considerada como um operador linear no espaço $L_2(\Omega, \mathcal{G}, P)$ (veja o Apêndice A.3). Como tal, $E(\cdot|\mathcal{F})$ é uma projeção ortogonal de $L_2(\Omega, \mathcal{G}, P)$ sobre o subespaco $L_2(\Omega, \mathcal{F}, P)$. Para ver isto, lembramos que o espaço $L_2(\Omega, \mathcal{G}, P)$ é um espaço de Hilbert com norma dada por

$$||X||_2 := \sqrt{\langle X, X\rangle} = \sqrt{E(|X|^2)}. \tag{3.5}$$

Dada X uma v.a qualquer sobre $L_2(\Omega, \mathcal{G}, P)$, $E(X|\mathcal{F})$ é a função $\mathcal{F}$-mensurável que é a "mais próxima"de X em termos da norma definida em (3.5). De fato, seja Y uma v.a $\mathcal{F}$-mensurável. Temos que

$$||X - Y||_2^2 = E(|X - Y|^2) = E(|X - E(X|\mathcal{F}) + E(X|\mathcal{F}) - Y|^2)$$

$$= E(|X - E(X|\mathcal{F})|^2) + E(|E(X|\mathcal{F}) - Y|^2) + 2E[(X - E(X|\mathcal{F}))(E(X|\mathcal{F}) - Y)].$$

Chamando o último termo dentro da esperança de H, temos que $E(H) = E[E(H|\mathcal{F})]$ $= 0$, pois $E(H|\mathcal{F}) = 0$. Logo, basta tomar $Y = E(X|\mathcal{F})$.

3.3 Probabilidade condicional regular

Até agora temos definidas a esperança condicional $E(X|\mathcal{F})$ e a probabilidade condicional $P(\Lambda|\mathcal{F})$. Para a esperança de uma v.a X temos

$$E(X) = \int X(\omega)dP(\omega).$$

Uma questão que surge é: podemos escrever a esperança condicional de modo similar, isto é,

$$E(X|\mathcal{F})(\omega) = \int X(\alpha)P(\alpha|\mathcal{F})(\omega)?$$

Parte do problema é: será que $P(\cdot|\mathcal{F})$ é uma probabilidade sobre alguma σ-álgebra, para cada ω?

Definição 3.2. *Seja $(\Omega, \mathcal{G}, P)$ um e.p e $\mathcal{A}$, $\mathcal{B}$ sub-σ-álgebras de $\mathcal{G}$. Uma probabilidade condicional regular sobre $\mathcal{A}$, relativamente a $\mathcal{B}$, é uma função $\nu(\cdot,\cdot)$ definida em $\mathcal{A}\times\Omega$ com valores em $\mathbb{R}$ tal que:*

(i) *Para cada ω, $\nu(\cdot,\omega)$ é uma medida de probabilidade sobre $\mathcal{A}$;*

(ii) *Para cada A fixo em $\mathcal{A}$, $\nu(A,\cdot)$ é uma versão de $P(A|\mathcal{B})$.*

Teorema 3.3. *Seja $(\Omega, \mathcal{G}, P)$ um e.p, $\mathcal{A}$, $\mathcal{B}$ sub-σ-álgebras de $\mathcal{G}$. Suponha que exista uma probabilidade condicional regular $\nu(\cdot,\cdot)$ sobre $\mathcal{A}$, relativamente a $\mathcal{B}$. Então, se X é $\mathcal{A}$-mensurável, uma versão de $E(X|\mathcal{B})$ é $\int_\Omega X(\alpha)\nu(d\alpha,\omega)$, ou seja,*

$$\int_\Omega X(\alpha)\nu(d\alpha,\omega) = E(X|\mathcal{B})(\omega) \quad q.c. \tag{3.6}$$

Prova: Considere a classe das funções $\mathcal{A}$-mensuráveis para as quais (3.6) vale.

(i) Essa classe contém funções indicadoras de conjuntos de $\mathcal{A}$, pois se $A \in \mathcal{A}$, temos $E(I_A|\mathcal{B}) = P(A|\mathcal{B})$ por definição, e

$$\int_\Omega I_A(\alpha)\nu(d\alpha,\omega) = \int_A \nu(d\alpha,\omega) = \nu(A,\omega) = P(A|\mathcal{B}) \text{ q.c,}$$

onde na penúltima igualdade usamos (i) e na última usamos (ii) da definição. Logo, (3.6) é válida se $X = I_A$, para algum $A \in \mathcal{A}$.

(ii) Depois, (3.6) vale se $X = \sum c_i I_{A_i}$, $A_i \in \mathcal{A}$, ou seja, se X é uma função simples.

(iii) A seguir,(3.6) vale para $X \geq 0$, integrável, $\mathcal{A}$-mensurável. Basta tomar uma sequência $X_n \uparrow X$, de funções simples e usar o TCM.

(iv) Portanto, (3.6) vale para qualqer X que seja $\mathcal{A}$-mensurável, usando $X = X^+ - X^-$. $\square$

Um caso importante na prática é o caso em que $\mathcal{A} = \mathcal{F}\{X\}$ e $\mathcal{B} = \mathcal{F}\{Y\}$ onde X e Y são dois vetores aleatórios. Neste caso é possível mostrar que a probabilidade condicional sobre $\mathcal{A}$ relativamente a $\mathcal{B}$ sempre existe (Veja Kallenberg (2002) para um teorema mais geral). Neste contexto falamos de *lei condicional* de X dado Y. Na maioria das vezes, uma construção explicita da lei condicional nos permite evitar recorrer ao teorema de existência. Em relação à unicidade, temos que se ν e $\tilde{\nu}$ são duas leis condicionais, então para todo $A \in \mathcal{F}\{X\}$,

$$\nu(A,\omega) = \tilde{\nu}(A,\omega), \text{ q.c.}$$

Como $\mathcal{F}\{X\}$ é contavelmente gerada, obtemos que q.c,

$$\nu(A,\omega) = \tilde{\nu}(A,\omega), \ \ \forall A \in \mathcal{F}\{X\}.$$

É neste sentido que temos unicidade e claramente por (3.6) não podemos esperar mais do que isto. Por abuso de linguagem, falamos *da* lei condicional de X dado Y.

Terminamos esta seção com uns exemplos de leis condicionais.

Exemplo 3.5. Sejam X e Y vetores aleatórios com dimensões m e n respectivamente.
(a) Supomos que Y é um vetor discreto. Seja $g : \mathcal{B}_m \times \mathbb{R}^n$ tal que

$$g(A,y) = \begin{cases} P(X \in A \mid Y = y), & \text{se } P(Y=y) > 0; \\ \delta_{x_0}(A), & \text{se } P(Y=y) = 0, \end{cases}$$

onde x_0 é um ponto arbitrário de $\mathbb{R}^m$. Seja $\nu(A,\omega) := g(A, Y(\omega))$. Usando a propriedade característica da esperança condicional podemos verificar que ν satisfaz a Definição 3.2. e portanto ν é a lei condicional de X dado Y.

(b) Assumimos agora que o vetor (X,Y) é absolutamente contínuo com densidade $f_{X,Y}$. A densidade de Y é dada por

$$f_Y(y) = \int_{\mathbb{R}^m} f_{X,Y}(x,y)dx.$$

Note que é possível que o termo da direita seja igual a $+\infty$ num conjunto de medida de Lebesgue igual a 0. Neste caso, para que f_Y tenha valores reais, redefinimos f_Y igual 0 neste conjunto. Agora, definimos $g : \mathcal{B}_m \times \mathbb{R}^n$ por

$$g(A,y) = \begin{cases} \frac{1}{f_Y(y)} \int_A f_{X,Y}(x,y)dx, & \text{se } f_Y(y) > 0; \\ \delta_{x_0}(A), & \text{se } f_Y(y) = 0. \end{cases}$$

Seja $\nu(A,\omega) := g(A, Y(\omega))$. Como na parte (a), podemos mostrar que ν é a lei condicional de X dado Y.

Problemas

1. Prove (a) e (b) do Exemplo 3.1.
2. Prove (b) do Exemplo 3.2.
3. Prove (c) do Exemplo 3.3.
4. Prove a Propriedade P1.
5. Prove as Propriedades P5 e P6.
6. Seja X integrável, Y limitada. Prove que
$$E\left[E(X|\mathcal{F})Y\right] = E\left[XE(Y|\mathcal{F})\right].$$
7. Prove que se $X \geq 0$, então $E(X|\mathcal{F}) \geq 0$ q.c.
8. Prove que, se $\mathcal{F}\{X\}$ for independente de $\mathcal{F}$, então $E(X|\mathcal{F}) = E(X)$ q.c.
9. Prove que $\text{Var}[E(Y|\mathcal{F})] \leq \text{Var}(Y)$.
10. Dê um exemplo onde $E[E(Y|X_1) \mid X_2] \neq E[E(Y|X_2) \mid X_1]$.
11. Seja Y uma v.a com f.d $F(x) = 1 - e^{-x}$, se $x \geq 0$ e $F(x) = 0$, se $x < 0$. Calcule:

    (a) $E(Y|Y \vee t)$; (b) $E(Y|Y \wedge t),\ t > 0$.
12. Sejam X e Y independentes e B um conjunto de Borel. Prove que $P\{(X+Y) \in B|X\} = P_Y\{B - X\}$ q.c.
13. Sejam $X_1, \ldots, X_n$ independentes e $S_n = X_1 + \ldots + X_n$. Prove que
$$P\{S_n \in B|S_1, \ldots, S_{n-1}\} = P\{S_n \in B|S_{n-1}\}.$$
14. Seja $\Omega = [-\pi, \pi]$, $\mathcal{F}$ a σ-álgebra de Borel e $P =$ (medida de Lebesgue sobre $[-\pi, \pi]$)$/2\pi$. Calcule $E(X|Y)$, se X integrável sobre $(\Omega, \mathcal{F}, P)$ e $Y(\omega) = \text{sen}(n\omega)$, n um inteiro positivo fixo.
15. Sejam $X_1, X_2, \ldots$ v.a's i.i.d e $S_n = X_1 + X_2 + \ldots + X_n$. Prove que $E(X_1|S_n, S_{n+1}, \ldots) = S_n/n$.
16. Se X e Y estão em L_2, prove que $E[E(X|\mathcal{F})Y] = E[XE(Y|\mathcal{F})]$.
17. Seja $\Omega = [-1,1]^2$, $\mathcal{F} = \mathcal{B}([-1,1]^2)$, $P =$ (medida de Lebesgue sobre $[-1,1]^2$)$/4$. Se $\omega = (\omega_1, \omega_2) \in \Omega$, seja $X(\omega) = \omega_1,\ Y(\omega) = \omega_2$. Calcule $E[X|(X+Y)^2]$.
18. Seja X uma v.a, $\mathcal{F}_1$, $\mathcal{F}_2$ duas σ-álgebras. Prove que, se $\mathcal{F}\{X\} \vee \mathcal{F}_1$ for independente de $\mathcal{F}_2$, então $E(X|\mathcal{F}_1 \vee \mathcal{F}_2) = E(X|\mathcal{F}_1)$ q.c.
19. Suponha que X seja uma v.a com variância finita e $\mathcal{F}$ uma sub-σ-álgebra de $\mathcal{G}$. Prove que
$$E(X - E(X|\mathcal{F}))^2 = E(X^2) - E(E(X|\mathcal{F}))^2.$$

20. Defina a *variância condicional* como $\text{Var}(X|\mathcal{F}) = E((X - E(X|\mathcal{F}))^2|\mathcal{F})$. Sejam X e Y duas v.a's com variâncias finitas, $\mathcal{F} \subset \mathcal{G}$, e seja g uma função com valores reais, tal que $E(g(X)^2) < \infty$. Prove que

$$E[(Y - g(X))^2] = E[\text{Var}(Y|\mathcal{F})] + E[E(Y|\mathcal{F}) - g(X)]^2 \geq E[\text{Var}(Y|\mathcal{F})],$$

com igualdade se $g(X) = E(Y|\mathcal{F})$.

21. Considere $\{C_1, \ldots, C_n\}$ uma partição de Ω (ou seja, uma coleção de eventos mutuamente exclusivos cuja reunião é Ω) tal que $P(C_k) > 0$ para todo k. Prove que, para todo evento $A \subset \Omega$, $P(A) = \sum_{k=1}^n P(A|C_k)P(C_k)$.

22. Com a mesma partição do problema anterior, prove que, para todo evento $A \subset \Omega$,

$$P(C_k|A) = \frac{P(A|C_k)P(C_k)}{\sum_{j=1}^n P(A|C_j)P(C_j)}, \quad k = 1, 2, \ldots, n.$$

Capítulo 4

Martingales

Neste capítulo, vamos nos restringir ao estudo de martingales com tempo discreto. No capítulo seguinte trataremos de processo com tempo contínuo e, em particular, martingales com tempo contínuo. Os conceitos de tempo de parada e integrabilidade uniforme serão estudados antes de definir martingales.

Martingales são generalizações de somas de v.a's independentes com média zero. Veja o Exemplo 4.3 (a). Aparentemente, foram definidos pela primeira vez por Ville (1939), sendo que os resultados mais inovadores aparecem em Doob (1953). A teoria de martingales tem aplicações em diversas áreas, em particular em confiabilidade e finanças. Referências importantes são Neveu (1975) e Williams (1991). Notamos por fim que na literatura em Português, é comum usar também os termos martingal e martingais.

4.1 Tempos de parada

Seja $(\Omega, \mathcal{F}, P)$ um e.p e $\{\mathcal{F}_n, n \geq 1\}$ uma sequência de sub-σ-álgebras de $\mathcal{F}$. Dizemos que $\{\mathcal{F}_n\}$ é *crescente* se $\mathcal{F}_n \subset \mathcal{F}_{n+1}$, para todo $n \geq 1$.

Seja $X = \{X_n, n \geq 1\}$ um processo estocástico sobre $(\Omega, \mathcal{F}, P)$; dizemos que X *é adaptado a* $\{\mathcal{F}_n, n \geq 1\}$ se X_n é $\mathcal{F}_n$-mensurável, para cada $n \geq 1$.

Na seção 4.3 estaremos interessados na seguinte questão: o que acontece quando paramos um martingale num instante de tempo aleatório? Para responder a essa questão, temos que ter uma regra para parar um processo estocástico num dado instante, de modo que essa regra não dependa do futuro. Isso nos leva à definição de tempo de parada.

Definição 4.1. *Um tempo de parada* τ *(ou v.a opcional) relativo a uma família crescente* $\{\mathcal{F}_n, n \geq 1\}$, *é uma v.a sobre* $(\Omega, \mathcal{F}, P)$ *com valores em* $\overline{\mathbb{N}} = \mathbb{N} \cup \{+\infty\}$, $\mathbb{N} = \{1, 2, 3, \ldots\}$, *tal que*

$$\{\omega \ : \ \tau(\omega) = n\} \in \mathcal{F}_n, \quad \text{para cada } \ n \in \mathbb{N}. \tag{4.1}$$

Quando dizemos que τ é um tempo de parada para um processo X, isso significa que τ é um tempo de parada relativo à sequência crescente de σ-álgebras $\mathcal{F}_n = \mathcal{F}\{X_1, \ldots, X_n\}$. Intuitivamente, τ é uma v.a com valores inteiros positivos (possivelmente ∞) que fornece uma regra para parar um processo estocástico. A equação (4.1) nos diz que a decisão de parar ou não o processo no instante n depende somente da informação disponível no instante n (ou seja, a história do proceso até e incluindo o instante n). Nenhum conhecimento do futuro é necessário.

Definição 4.2. *Seja $X = \{X_n, n \geq 1\}$ um processo estocástico e τ um tempo de parada para X. Suponha $\tau < \infty$ q.c, isto é, $P\{\tau < \infty\} = 1$. Considere v.a X_τ tomando o valor $X_n(\omega)$ sobre o conjunto $\{\omega : \tau(\omega) = n\}$. Ou seja, se $\tau(\omega) = n$, então $X_{\tau(\omega)}(\omega) = X_n(\omega)$. A σ-álgebra gerada por $X_\tau, X_{\tau+1}, \ldots$ é chamada σ-álgebra pós-τ, e indicada $\mathcal{F}_{\tau+}$.*

Definição 4.3. *Seja $\{\mathcal{F}_n, n \geq 1\}$ uma família crescente de σ-álgebras e τ um tempo de parada relativo a essa família. A σ-álgebra pré-τ, denotada $\mathcal{F}_{\tau-}$, é a σ-álgebra consistindo de todos os conjuntos Λ de $\mathcal{F}_\infty := \bigvee_{n\geq 1} \mathcal{F}_n$ tal que $\Lambda \cap \{\omega : \tau(\omega) = n\} \in \mathcal{F}_n$, para todo $n \geq 1$.*

Segue-se que os eventos de $\mathcal{F}_{\tau-}$ são aqueles ocorrendo antes do tempo de parada τ, enquanto que os eventos de $\mathcal{F}_{\tau+}$ são aqueles eventos ocorrendo depois de τ.

Exemplo 4.1. (a) Seja $\{\mathcal{F}_n, n \geq 1\}$ e defina $\tau(\omega) = p$, constante, $p \in \overline{\mathbb{N}}$. Então, τ é um tempo de parada e $\mathcal{F}_{\tau-} = \mathcal{F}_p$.

(b) Seja X um processo estocástico e B um conjunto de Borel. Defina τ por: $\tau(\omega) = \inf\{n \geq 1 : X_n(\omega) \in B\}$, e $\tau(\omega) = \infty$, se o conjunto anterior for vazio. Segue-se que τ é o primeiro instante de tempo que X entra em B, e é um tempo de parada. De fato, $\{\omega : \tau(\omega) = n\} = \{\omega : X_j(\omega) \notin B, j < n, X_n(\omega) \in B\}$, que pertence a $\mathcal{F}_n = \mathcal{F}\{X_1, \ldots, X_n\}$.

(c) Sejam $Y_1, Y_2, \ldots$ v.a's i.i.d, $P(Y_1 = 1) = p$, $P(Y_1 = -1) = q$, $p > q$. Defina $X_n = 1 + Y_1 + Y_2 + \ldots + Y_n$ e seja $\tau = \inf\{n \geq 1 : X_n(\omega) = 0\}$, e $\tau = \infty$, se esse conjunto for vazio. Então, τ é um tempo de parada.

(d) Considere a situação de (c) mas desta vez $X_n = -1 + Y_1 + \cdots + Y_n$. Sabemos que com probabilidade 1, X_n atingirá o valor zero somente um número finito de vezes, ou seja, $P\{X_n = 0 \text{ i.v}\} = 0$. Defina τ como o último tempo em que $X_n = 0$. Então, $\tau < \infty$ q.c, está bem definido, mas não é um tempo de parada.

4.1.1 Propriedades dos tempos de parada

Algumas propriedades dos tempos de paradas são enunciadas a seguir. Algumas serão demonstradas, as demais ficam como exercícios.

[1] Para cada n, os seguintes conjuntos estão em $\mathcal{F}_n$, onde τ é um tempo de parada relativo a $\{\mathcal{F}_n, n \geq 1\}$:

$$\{\tau \leq n\}, \quad \{\tau = n\}, \quad \{\tau < n\}, \quad \{\tau > n\}, \quad \{\tau \geq n\}.$$

De fato, $\{\tau \leq n\} = \cup_{m=0}^{n}\{\tau = m\}$ e $\{\tau = m\} \in \mathcal{F}_m \subset \mathcal{F}_n$, se $m \leq n$. O conjunto $\{\tau > n\} \in \mathcal{F}_n$, tomando o complementar.

[2] Se σ e τ são tempos de parada, então $\sigma \wedge \tau$ e $\sigma \vee \tau$ são tempos de parada.

De fato, $\{\sigma \wedge \tau = n\} = \{\sigma = n, \tau > n\} \cup \{\sigma > n, \tau = n\} \cup \{\sigma = n, \tau = n\}$.

[3] Seja $k \in \mathbb{N}$ e τ um tempo de parada. Então, $\tau + k$ é também um tempo de parada (Note que $\tau - k$ não necessita ser um tempo de parada).

[4] Sejam $\tau_1, \tau_2, \ldots$ tempos de parada; então $\sup_n \tau_n$ é também um tempo de parada.

[5] τ é $\mathcal{F}_{\tau-}$-mensurável.

[6] Seja τ um tempo de parada finito relativo a $X = \{X_n, n \geq 1\}$. Se X é adaptado à sequência $\{\mathcal{F}_n, n \geq 1\}$, então X_τ é $\mathcal{F}_{\tau-}$-mensurável.

[7] Sejam τ_1, τ_2 tempos de parada, $\tau_1 \leq \tau_2$. Então, $\mathcal{F}_{\tau_1-} \subset \mathcal{F}_{\tau_2-}$.

De fato, se $B \in \mathcal{F}_{\tau_1-}$, temos que $B \cap \{\tau_2 \leq n\} = B \cap \{\tau_1 \leq n\} \cap \{\tau_2 \leq n\} \in \mathcal{F}_n$, para todo $n \geq 1$, ou seja $B \in \mathcal{F}_{\tau_2-}$.

[8] Seja $X = \{X_n, n \geq 1\}$ um processo estocástico e τ um tempo de parada. Para qualquer inteiro positivo n e qualquer $\omega \in \Omega$, defina $\tau \wedge n(\omega) = \min\{\tau(\omega), n\}$. Definimos, então, o *processo parado* $X^\tau = \{X_n^\tau, n \geq 1\}$ por $X_n^\tau(\omega) = X_{\tau \wedge n(\omega)}(\omega)$.

Na seção 4.3 veremos que se X é um martingale, então $X^\tau = \{X_{\tau \wedge n}\}$ é também um martingale.

Exemplo 4.2. Sejam $X_1, X_2, \ldots$ v.a's i.i.d e seja τ um tempo de parada finito para $X = \{X_n, n \geq 1\}$. Então, $\mathcal{F}_{\tau-}$ e $\mathcal{F}_{\tau+}$ são independentes e $(X_1, X_2, \ldots)$ tem a mesma distribuição que $(X_{\tau+1}, X_{\tau+2}, \ldots)$ (esse é um caso especial da chamada propriedade forte de Markov).

De fato, seja $\Lambda \in \mathcal{F}_{\tau-}$. É suficiente provar que

$$P(\Lambda, X_{\tau+1} \in B_1, \ldots, X_{\tau+n} \in B_n) = P(\Lambda)P(X_1 \in B_1, \ldots, X_n \in B_n),$$

onde B_i são conjuntos de Borel.

Temos que

$$P(\Lambda, \{\tau = k\}, X_{k+1} \in B_1, \ldots, X_{k+n} \in B_n) = P(\Lambda, \{\tau = k\})P(X_{k+1} \in B_1, \ldots, X_{k+n} \in B_n)$$

$$= P(\Lambda, \{\tau = k\})P(X_1 \in B_1, \ldots, X_n \in B_n),$$

onde notamos que, $\Lambda \cap \{\tau = k\} \in \mathcal{F}_k$, a primeira igualdade decorre do fato de $\mathcal{F}_k$ ser independente de $\{X_{k+1}, \ldots, X_{k+n}\}$ e a última igualdade vale pois os X_i são i.i.d. Agora, basta somar sobre k para obter o resultado.

4.2 Integrabilidade uniforme

Nesta secção apresentamos o importante conceito de integrabilidade uniforme que usaremos no decorrer do capítulo.

Definição 4.4. *Seja I um conjunto de índices. A família de v.a's $\{X_i, i \in I\}$ é uniformemente integrável (u.i) se*

$$\sup_{i \in I} \int_{\{|X_i| > \lambda\}} |X_i| dP \to 0, \quad \textit{quando } \lambda \to \infty. \tag{4.2}$$

Teorema 4.1. *Seja X uma v.a integrável e suponha que $|X_i| \leq X$, para todo $i \in I$. Então, a família $\{X_i, i \in I\}$ é uniformemente integrável*

Prova: Temos que $\int_{\{|X_i|>\lambda\}} |X_i| dP \leq \int_{\{|X_i|>\lambda\}} |X| dP \leq \int_{\{|X|>\lambda\}} |X| dP$. Como isso é verdade para todo $i \in I$, temos que

$$\sup_{i \in I} \int_{\{|X_i|>\lambda\}} |X_i| dP \leq \int_{\{|X|>\lambda\}} |X| dP \ \to \ 0, \quad \lambda \to \infty,$$

pelo Teorema de Convergencia Dominada. $\square$

Teorema 4.2. *A família $\{X_i, i \in I\}$ é u.i se, e somente se:*

(a) $\sup_{i \in I} E(|X_i|) < \infty$*;*

(b) Dado $\varepsilon > 0$, existe $\delta = \delta(\varepsilon) > 0$, tal que se $P(A) \leq \delta$, então $\int_A |X_i| dP \leq \varepsilon$, para todo $i \in I$.

Prova: ($\Rightarrow$) Suponha que $\{X_i, i \in I\}$ seja u.i. Então, (a) vale, pois, tomando-se λ_0 tão grande tal que $\int_{\{|X_i|>\lambda_0\}} |X_i| dP \leq 1$, para todo $i \in I$, por i.u, temos que

$$\int_{\Omega} |X_i| dP = \int_{\{|X_i|>\lambda_0\}} |X_i| dP + \int_{\{|X_i| \leq \lambda_0\}} dP \leq 1 + \lambda_0.$$

Também, seja $\varepsilon > 0$; então,

$$\int_A |X_i|dP = \int_{A\cap\{|X_i|>\lambda\}} |X_i|dP + \int_{A\cap\{|X_i|\le\lambda\}} |X_i|dP \le \int_{\{|X_i|>\lambda\}} |X_i|dP + \lambda P(A).$$

Tome λ grande de modo que a primeira integral seja menor que $\varepsilon/2$, para todo i, por i.u. Com λ assim escolhido (e igual a λ_1, digamos), tome $\delta = \varepsilon/(2\lambda_1)$. Então, se $P(A) < \delta$, o segundo termo é menor que $\lambda_1\varepsilon/(2\lambda_1) = \varepsilon/2$, portanto $\int_A |X_i|dP < \varepsilon$, e (b) vale.

($\Leftarrow$) Pela desigualdade de Chebyshev, temos que

$$P\{|X_i| > \lambda\} \le E(|X_i|)/\lambda \le \frac{\sup_i E(|X_i|)}{\lambda} \le \frac{M}{\lambda},$$

onde $M < \infty$, usando (a). Então, dado $\varepsilon > 0$, $\int_{\{|X_i|>\lambda\}} |X_i|dP \le \varepsilon$, sempre que $P\{|X_i| > \lambda\} \le \delta(\varepsilon)$, por (b), isto é, se $\lambda \ge M/\delta$. $\square$

Teorema 4.3. *Suponha que $X_n \xrightarrow{P} X$. Então, $X_n \xrightarrow{L_1} X$ se, e somente se, $\{X_n, n \ge 1\}$ for u.i.*

Prova: ($\Rightarrow$) Suponha que $X_n \xrightarrow{L_1} X$. Mostraremos que as condições (a) e (b) do Teorema 4.2 valem.
(a) tome N tão grande de modo que $E(|X_n - X|) \le 1$, para $n \ge N$, pois X_n converge para X em L_1. Então,

$$E(|X_n|) \le E(|X_n - X|) + E(|X|) \le 1 + E(|X|), \quad \text{para todo } \ n \ge N.$$

Portanto,

$$\sup_n E(|X_n|) \le \max\{1 + E(|X|), E(|X_1|), \ldots, E(|X_{N-1}|)\} < \infty.$$

(b) Temos que $\int_A |X_n|dP \le \int_A |X_n - X|dP + \int_A |X|dP$. Seja $\varepsilon > 0$ e seja N grande de tal sorte que, para $n \ge N$, tenhamos $\int_\Omega |X_n - X|dP < \varepsilon/2$, por hipótese. Logo,

$$\int_A |X_n|dP \le \int_A |X_n - X|dP + \int_A |X|dP \le \varepsilon/2 + \int_A |X|dP, \ \ n \ge N.$$

Seja δ_1 pequeno, tal que se $P(A) < \delta_1$, então $\int_A |X|dP < \varepsilon/2$, pois X é integrável. Seja δ_2 tão pequeno, de modo que, se $P(A) < \delta_2$, então $\int_A |X_n|dP < \varepsilon$, para $n = 1, 2, \ldots, N$. Segue-se que, para todo n, se $P(A) \le \min\{\delta_1, \delta_2\}$, temos que $\int_A |X_n|dP \le \varepsilon$.

($\Leftarrow$) Suponha que $X_n \stackrel{P}{\to} X$ e $\{X_n, n \geq 1\}$ seja u.i. Primeiro, X é integrável, porque se n_k é uma subsequência tal que $X_{n_k} \stackrel{q.c}{\to} X$, pelo Lema de Fatou, teremos

$$E(|X|) \leq \liminf_k E(|X_{n_k}|) \leq \sup_n E(|X_n|) < \infty.$$

Agora,

$$\int_\Omega |X_n - X| dP = \int_{\{|X_n - X| > \varepsilon\}} |X_n - X| dP + \int_{\{|X_n - X| \leq \varepsilon\}} |X_n - X| dP$$
$$\leq \int_{\{|X_n - X| > \varepsilon\}} |X_n| dP + \int_{\{|X_n - X| > \varepsilon\}} |X| dP + \varepsilon.$$

Se $n \to \infty$, então $\limsup_n \int_\Omega |X_n - X| dP \leq 0 + 0 + \varepsilon$, logo $E(|X_n - X|) \to 0$. $\square$

Teorema 4.4 *Suponha que* $X_n \stackrel{P}{\to} X$ *e* $X_n \geq 0$, *para todo* $n \geq 1$. *Então,* $\{X_n, n \geq 1\}$ *é u.i se, e somente se,* $\lim_n E(X_n) = E(X) < \infty$.

Prova: (a) Se $\{X_n, n \geq 1\}$ é u.i, pelo Teorema 4.3, $X_n \stackrel{L_1}{\to} X$, logo $E(X_n) \to E(X)$ e X é integrável.

(b) Pelo teorema anterior, X_n é u.i se, e somente se, $X_n \stackrel{L_1}{\to} X$. Temos que $E(|X_n - X|) = E\{2(X - X_n) \vee 0 - (X - X_n)\}$. Mas $\{2(X - X_n) \vee 0\} \to 0$ em probabilidade, logo $E\{2(X - X_n) \vee 0\} \to 0$, pelo TCD, pois $0 \leq \{2(X - X_n) \vee 0\} \leq 2X$. Também, $E(X_n - X) \to 0$, por hipótese. Segue-se que $E(|X_n - X|) \to 0$ e X_n é u.i. $\square$

Teorema 4.5. *Seja* $f : [0, \infty) \to [0, \infty)$, *tal que* $\lim_{x \to \infty} f(x)/x = +\infty$. *Se* $\sup_i E(f(|X_i|)) < \infty$, *então* $\{X_i, i \in I\}$ *é u.i.*

Prova: Seja $\varepsilon > 0$ e seja $M = \sup_i E(f(|X_i|))$. Tome x_0 tal que se $x \geq x_0$, $f(x)/x \geq M/\varepsilon$, pois $f(x)/x \to \infty$. Logo, $f(x)\varepsilon/M \geq x$ e portanto para todo $i \in I$,

$$\int_{\{|X_i| > \lambda\}} |X_i| dP \leq \frac{\varepsilon}{M} \int_\Omega f(|X_i|) dP \leq \frac{\varepsilon}{M} M = \varepsilon,$$

sempre que $\lambda \geq x_0$. $\square$

Como casos especiais importantes do teorema, temos $f(x) = x^p$, para $p > 1$ e $f(x) = x \log^+ x$.

4.3 Martingales

Como dissemos na introdução desse capítulo, trataremos aqui o caso de martingales com tempo discreto. No entanto, daremos a definição de martingale para o caso em que o conjunto paramétrico é um subconjunto dos números reais. Um resultado importante que será provado nessa seção é o Teorema da Amostragem Opcional.

Definição 4.5. *Seja $(\Omega, \mathcal{F}, P)$ um e.p e T um subconjunto de $\mathbb{R}$. Seja $\{\mathcal{F}_t, t \in T\}$ uma família crescente de sub-σ-álgebras de $\mathcal{F}$, ou seja, $\mathcal{F}_s \subset \mathcal{F}_t$, se $s \leq t$. Seja $\{X_t, t \in T\}$ um processo estocástico adaptado a $\{\mathcal{F}_t, t \in T\}$. Um processo $X = \{X_t, \mathcal{F}_t, t \in T\}$ é um martingale se:*

(a) X_t é integrável, para cada $t \in T$;

(b) Se $s \leq t$, então $E(X_t|\mathcal{F}_s) = X_s$.

Definição 4.6. *Um submartingale tem as mesmas características da definição anterior, exceto que (b) é substituída por:*

(b$'$) *Se $s \leq t$, então $E(X_t \mid \mathcal{F}_s) \geq X_s$.*

Um supermartingale substitui (b) por

(b$''$) Se $s \leq t$, então $E(X_t \mid \mathcal{F}_s) \leq X_s$.

Uma interpretação da definição em termos de jogos é a seguinte. Se X_n representa a fortuna de um jogador após o jogo n e $\mathcal{F}_n$ representa a sua história até (incluindo) o instante n, então $E(X_{n+1}|\mathcal{F}_n) = X_n$ significa que o ganho esperado do jogador no instante $n+1$, dado todo o conhecimento passado, é igual à sua fortuna presente. Teremos um *jogo justo*. Vale uma interpretação similar para (sub)supermartingale.

Observações: (a) Quando dizemos que $\{X_t, t \in T\}$ é um martingale, queremos dizer que as σ-álgebras da definição são $\mathcal{F}_t = \mathcal{F}\{X_s, s \leq t\}$.

(b) Um martingale *com parâmetro discreto* ou com *tempo discreto* é aquele para o qual T é uma coleção de números inteiros. Usualmente consideramos $T = \{1, 2, 3, \ldots\}$ ou $T = \mathbb{Z}$.

(c) Se $\{X_t, t \in T\}$ é um submartingale, então $\{-X_t, t \in T\}$ é um supermartingale.

(d) Se $\{X_t, t \in T\}$ e $\{Y_t, t \in T\}$ são martingales, então não é necessário que $\{X_t + Y_t, t \in T\}$ seja um martingale. É verdade se X e Y são martingales com repeito à mesma sequência de σ-álgebras.

(e) Para verificar que $\{X_n, \mathcal{F}_n, n \geq 1\}$ é um martingale, é suficiente provar que $E(X_n|\mathcal{F}_{n-1}) = X_{n-1}$.

(f) Seja $\{X_n, n \geq 1\}$ um submartingale. Então $\{E(X_n), n \geq 1\}$ é uma sequência crescente.

De fato, $E(X_n|\mathcal{F}_{n-1}) \geq X_{n-1}$, bastando tomar a esperança de ambos os membros.

(g) Se $\{X_n, n \geq 1\}$ é um martingale, então $E(X_n) = E(X_1)$, para todo n.

De fato, de $E(X_n|\mathcal{F}_{n-1}) = X_{n-1}$, obtemos $E(X_n) = E(X_{n-1}) = \ldots = E(X_1)$, para todo n.

(h) Se $\{X_n, n \geq 1\}$ é um submartingale ou supermartingale, e se $E(X_n) = E(X_1)$ para todo n, então, $\{X_n, n \geq 1\}$ é um martingale.

Exemplo 4.3. (a) Sejam $\{X_n, n \geq 1\}$ v.a's independentes, com média zero. Então, $Y_n = X_1 + \ldots + X_n$ é um martingale.

De fato, temos que

$$E(Y_{n+1}|\mathcal{F}_n) = E(Y_{n+1}|X_1, \ldots, X_n) = E(\sum_{i=1}^{n+1} X_i|X_1, \ldots, X_n)$$

$$= X_1 + \ldots + X_n + E(X_{n+1}|X_1, \ldots, X_n) = X_1 + \ldots + X_n + E(X_{n+1}) = Y_n,$$

sendo que a penúltima igualdade vale pela independência e a última porque a média é zero.

(b) Se $\{X_n, n \geq 1\}$ são v.a's independentes, $E(X_i) = \mu_i \geq 0$, então $Y_n = \sum_{i=1}^n X_i$ é um submartingale e $\sum_{i=1}^n (X_i - \mu_i) = Z_n$ é um martingale.

(c) Seja X integrável, $\{\mathcal{F}_n, n \geq 1\}$ uma família crescente de σ-álgebras. Então, $X_n = E(X|\mathcal{F}_n)$ é um martingale.

De fato,

$$E(X_n|\mathcal{F}_{n-1}) = E[E(X|\mathcal{F}_n)|\mathcal{F}_{n-1}] = E(X|\mathcal{F}_{n-1}) = X_{n-1},$$

pelo fato que $\mathcal{F}_{n-1} \subset \mathcal{F}_n$.

Exemplo 4.4. (a) Seja $\{Y_n\}$ um martingale com tempos $\{\ldots, -3, -2, -1\}$. Então, Y_n deve ter a forma dada no Exemplo 4.3 (c), isto é, $Y_{-n} = E(Y_{-1}|\mathcal{F}_{-n})$. Esse é um exemplo de um *martingale reverso*.

(b) Sejam $\{X_n, n \geq 1\}$ v.a's i.i.d, integráveis. Defina:

$$Y_{-1} = X_1,$$

$$Y_{-2} = (X_1 + X_2)/2;$$

$$\ldots$$

$$Y_{-n} = (X_1 + \ldots + X_n)/n.$$

Então, $\{\dots, Y_{-2}, Y_{-1}\}$ é um martingale com respeito a $\mathcal{F}\{\dots, Y_{-2}, Y_{-1}\} = \mathcal{F}_{-1}$, $\mathcal{F}\{\dots, Y_{-3}, Y_{-2}\} = \mathcal{F}_{-2}$ etc.

Martingales podem não ter a forma dada no Exemplo 4.3 (c). Veja o Problema 6. Em algumas situações, como em econometria, finanças e no estudo de somas de v.a's independentes, é mais conveniente considerar incrementos. Uma sequência $\{U_n, \mathcal{F}_n, n \geq 1\}$ é chamada uma *diferença martingale* se

$$E(U_{n+1}|\mathcal{F}_n) = 0, \quad \text{para todo } n \geq 1.$$

Diferentemente de martingales, diferenças martingales são ortogonais. Veja o Problema 22.

Teorema 4.6. *(a) Seja φ uma função convexa e $\{X_n, \mathcal{F}_n, n \geq 1\}$ um martingale. Suponha que $\varphi(X_n)$ seja integrável, para cada n. Então, $\{\varphi(X_n), \mathcal{F}_n, n \geq 1\}$ é um submartingale.*
(b) Suponha φ convexa crescente e $\{X_n, \mathcal{F}_n, n \geq 1\}$ um submartingale. Se $\varphi(X_n)$ é integrável, então $\{\varphi(X_n), \mathcal{F}_n, n \geq 1\}$ é um submartingale.

Prova: (a) Temos que

$$E\left(\varphi(X_{n+1}|\mathcal{F}_n)\right) \geq \varphi\left(E(X_{n+1}|\mathcal{F}_n)\right) = \varphi(X_n),$$

onde usamos a desigualdade de Jensen e o fato que X_n é um martingale.

(b) De modo análogo,

$$E\left(\varphi(X_{n+1}|\mathcal{F}_n)\right) \geq \varphi\left(E(X_{n+1}|\mathcal{F}_n)\right) \geq \varphi(X_n),$$

novamente usando Jensen, $E(X_{n+1}|\mathcal{F}_n) \geq X_n$ e φ crescente. $\square$

Exemplo 4.5. (a) Se $\{X_n\}$ é um martingale, então $\{X_n^2\}, \{X_n^+\}, \{X_n \vee M,\ M > 0\}$ são submartingales. Também, $\{|X_n|\}$ é um submartingale, pela parte (a) do teorema anterior.

(b) Se $\{X_n\}$ é um submartingale, então $\{X_n^+\}, \{X_n \vee M,\ M > 0\}$ são submartingales. Mas $\{|X_n|\}$ não necessita ser, pois a função $|x|$ não é crescente (parte (b) do teorema).

Consideramos, a seguir, o Teorema da Amostragem Opcional (TAO) de Doob. Veja Doob (1971) e Williams (1991). O teorema diz, sob determinadas suposições, que o valor esperado de um martingale em um tempo de parada é igual ao valor esperado de seu valor inicial. Como vimos, martingales podem ser usados para modelar a fortuna de um jogador, participando de um jogo justo.

Ou seja, se $\{X_n\}$ é um martingale, temos que $E(X_n) = E(X_{n-1}) = \ldots = E(X_1)$, de modo que a fortuna esperada do jogador em qualquer tempo é igual à sua fortuna esperada inicial.

Suponha, agora, que T seja um tempo de parada e X_T é a fortuna do jogador nesse instante; será que $E(X_T) = E(X_1)$? Em geral, a resposta é negativa, como mostrado em Doyle e Snell (1984). O TAO dá condições para que isso seja verdade.

O TAO é importante em muitas aplicações, em particular em finanças, no contexto do teorema fundamental do apreçamento de ativos. O conteúdo essencial desse teorema é que não se pode ganhar (em média) comprando-se e vendendo-se um ativo cujo preço é modelado por um martingale.

Teorema 4.7. (Teorema da Amostragem Opcional). *Seja $\{X_n, \mathcal{F}_n, n \geq 1\}$ um (sub)martingale e $T_1 \leq T_2 \leq \ldots$ tempos de parada finitos relativamente a $\{\mathcal{F}_n, n \geq 1\}$. Suponha que:*

(a) $E\left(|X_{T_n}|\right) < \infty$, *para cada* n.

(b) $\lim_{N\to\infty} \int_{\{T_n > N\}} |X_N| dP = 0$, *para cada* n.

Então, $\{X_{T_1}, X_{T_2}, \ldots\}$ é um (sub)martingale relativo a $\{\mathcal{F}_{T_1}, \mathcal{F}_{T_2}, \ldots\}$.

Antes de demonstrar o teorema, vamos fazer uma observação e apresentar algumas aplicações do TAO.

Observação: Há uma versão “mais simples”' do TAO, dada em Williams (1991, Theorem 10.10). Seja X um martingale e T um tempo de parada. Suponha que qualquer uma das seguintes condições valha:

(i) Existe um inteiro positivo N tal que $T(\omega) \leq N$, q.c.

(ii) Existe um real positivo K tal que $|X_n(\omega)| < K$, q.c, para todo n, e T é finito q.c.

(iii) $E(T) < \infty$ e existe um real positivo K tal que $|X_n(\omega) - X_{n-1}(\omega)| < K$, q.c, para todo n.

Então X_T é integrável e $E(X_T) = E(X_1)$.

Aplicações do TAO

[1] Seja $M > 0$, inteiro, e T um tempo de parada tal que $T \leq M$ q.c. Então, se $\{X_n, n \geq 1\}$ é um (sub)martingale, também o será $\{X_1, X_T, X_M\}$.

De fato, (b) está satisfeita e (a) também, pois $E(|X_T|) \leq \sum_{k=1}^{M} E(|X_k|) < \infty$. Em particular, se $\{X_n\}$ é um martingale, $E(X_1) = E(X_T) = E(X_M)$. Se for um submartingale, substituir os sinais de igualdade por desigualdade.

[2] Se $\{X_n, n \geq 1\}$ é um (sub)martingale, e T um tempo de parada, finito ou não, então $\{X_{T\wedge 1}, X_{T\wedge 2}, \ldots\}$ é um (sub)martingale.

[3] Sejam $T_1 \leq T_2 \leq \ldots$ tempos de parada e $M_1, M_2, \ldots$ constantes, tais que $T_i \leq M_i$, para todo $i \geq 1$. Se $\{X_n, n \geq 1\}$ é um (sub)martingale, então $\{X_{T_1}, X_{T_2}, \ldots\}$ também o será.

[4] Seja $\{X_n, n \geq 1\}$ um martingale uniformemente integrável. Sejam $T_1 \leq T_2 \leq \ldots$ tempos de parada finitos. Então, $\{X_{T_1}, X_{T_2}, \ldots\}$ é um martingale.

De fato, a condição (b) do TAO vale, pois X é u.i. Por outro lado, $\{|X_n|\}$ é um submartingale pelo Teorema 4.6, logo para todo $k \geq 1$, $\{X_{T_k \wedge n}, n \geq 1\}$ é um submartingale e portanto $E(|X_{T_k \wedge n}|) \leq E(|X_n|)$. Segue-se que $E(|X_{T_k \wedge n}|) \leq \sup_n E(|X_n|) < \infty$, pois temos i.u. Agora, pela condição b) do TAO e o fato que T_k é finito q.c, temos pelo TCM que $E(|X_{T_k}|) = \lim_{n \to \infty} E(|X_{T_k \wedge n}|)$. Logo $E(|X_{T_k}|) \leq \sup_n E(|X_n|) < \infty$.

Prova do TAO: É suficiente provar que, dadas as hipóteses, se $S \leq T$ são dois tempos de parada, então $E(X_T|\mathcal{F}_S) = X_S$, no caso de um martingale e com sinal de desigualdade $\geq$, se submartingale. Ou seja, se $\Lambda \in \mathcal{F}_S$, provar que $\int_\Lambda X_T dP = (\geq) \int_\Lambda X_S dP$. Mas, para provar essa relação é suficiente provar a igualdade (desigualdade se submartingale) mais forte

$$\int_{\Lambda \cap \{S=j\}} X_T dP = (\geq) \int_{\Lambda \cap \{S=j\}} X_S dP.$$

No entanto,

$$\int_{\Lambda \cap \{S=j\}} X_S dP = \int_{\Lambda \cap \{S=j\}} X_j dP = \int_{\Lambda \cap \{S=j, T=j\}} X_j dP + \int_{\Lambda \cap \{S=j, T>j\}} X_j dP.$$

Como o conjunto $\Lambda \cap \{S = j, T > j\} \in \mathcal{F}_j$, obtemos que o último termo é

$$= (\leq) \int_{\Lambda \cap \{S=j, T=j+1\}} X_{j+1} dP + \int_{\Lambda \cap \{S=j, T>j+1\}} X_{j+1} dP,$$

pela definição de (sub)martingale. Fazendo-se $\Lambda \cap \{S = j, T > j\} = (\Lambda \cap \{S = j, T = j+1\}) \cup (\Lambda \cap \{S = j, T > j+1\})$, e assim sucessivamente, obtemos

$$= (\leq) \sum_{k=j}^{N} \int_{\{\Lambda, S=j, T=k\}} X_k dP + \int_{\{\Lambda, S=j, T>N\}} X_N dP$$

$$= \int_{\{\Lambda, S=j,\, j \leq T \leq N\}} X_T dP + \int_{\{\Lambda, S=j,\, T>N\}} X_N dP.$$

Quando $N \to \infty$, o último termo tende a zero, por (b), e pelo TCD a primeira integral tende a $\int_{\{\Lambda, S=j\}} X_T dP$. $\square$

4.4 Convergência de martingales

O objetivo dessa seção é provar alguns teoremas sobre convergência de (sub)martingales. Uma desigualdade devida a Doob (veja Doob, 1953) é essencial nesse contexto.

Seja $x_1, \ldots, x_n$ uma sequência qualquer de números reais e suponha $a < b$. Defina $\mathcal{U}_n(a,b)$ como o número de vezes que a sequência $\{x_k, k \leq n\}$ vai de um valor abaixo de a para um valor acima de b, ou seja, o número de cruzamentos ascendentes (*upcrossings*) de $[a,b]$. A partir de agora usaremos simplesmente a palavra *cruzamento*.

Teorema 4.8. (*Upcrossing inequality* - Doob) *Seja $\{X_1, X_2, \ldots, X_n\}$ um submartingale e $\mathcal{U}_n(a,b)$ o número de cruzamentos de $[a,b]$ pela sequência $\{X_1(\omega), \ldots, X_n(\omega)\}$. Então,*

$$E\left(\mathcal{U}_n(a,b)\right) \leq \frac{E(X_n-a)^+ - E(X_1-a)^+}{b-a} \leq \frac{E(X_n-a)^+}{b-a} \leq \frac{E(|X_n|)+a}{b-a}. \tag{4.3}$$

Prova: Defina $Y_k = (X_k - a)^+$. Segue-se que $\{Y_k, k \leq n\}$ é um submartingale. Ainda, o número de cruzamentos de $[a,b]$ por $X_1, \ldots, X_n$ é o mesmo número de cruzamentos de $[0, b-a]$ por $Y_1, \ldots, Y_k$. Logo, é suficiente calcular $\mathcal{U}_n(0, b-a)$ para a sequência Y.

Defina a sequência de tempos de parada $T_1 \leq T_2 \leq \ldots$ como segue:

$$\begin{aligned}
T_1 &= 1, \\
T_2 &= \inf\{k > 1 : Y_k = 0\}, \\
T_3 &= \inf\{k : k > T_2, Y_k \geq b-a\}, \\
T_4 &= \inf\{k : k > T_3, Y_k = 0\}, \\
T_5 &= \inf\{k : k > T_4, Y_k \geq b-a\} \text{ etc.}
\end{aligned}$$

Defina $T_k = n$ se o conjunto definido for vazio.

Escrevamos:

$$Y_n - Y_1 = (Y_{T_2} - Y_{T_1}) + (Y_{T_3} - Y_{T_2}) + \ldots + (Y_{T_n} - Y_{T_{n-1}}).$$

Observe, por exemplo, que $Y_{T_3} - Y_{T_2} \geq b-a$ e $Y_{T_n} = Y_n$. Portanto,

$$\sum_{k \text{ par}} (Y_{T_{k+1}} - Y_{T_k}) \geq (b-a)\mathcal{U}_n(0, b-a),$$

e, então,

$$Y_n - Y_1 \geq (b-a)\mathcal{U}_n(0,b-a) + \sum_{k \text{ ímpar}} (Y_{T_{k+1}} - Y_{T_k}),$$

do que segue

$$E\left(Y_n - Y_1\right) \geq (b-a)E\left(\mathcal{U}_n(0, b-a)\right) + \sum_{k \text{ ímpar}} E\left(Y_{T_{k+1}} - Y_{T_k}\right),$$

e como o último termo é não negativo (pois temos um submartingale), obtemos

$$\frac{E(X_n - a)^+ - E(X_1 - a)^+}{b-a} \geq E\left(\mathcal{U}_n(a,b)\right). \quad \square$$

Teorema 4.9. (Teorema de Convergência de Submartingales de Doob). *Seja $\{X_n, \mathcal{F}_n, n \geq 1\}$ um submartingale e suponha que X_n seja L_1-limitado (ou seja, $\sup_n(E|X_n|) < \infty$). Então, $\{X_n, n \geq 1\}$ converge q.c para um limite X_∞, que é integrável.*

Prova: Sejam $a < b$ números racionais e defina:

$$M_{ab} = \Big\{\omega : \liminf_n X_n(\omega) \leq a < b \leq \limsup_n X_n(\omega)\Big\}.$$

Pela desigualdade de Doob, $E(\mathcal{U}_n(a,b)) \leq \frac{E(X_n - a)^+}{b-a} \leq \sup_n \frac{E(|X_n|)+a}{b-a} < \infty$.

Seja $\mathcal{U}(a,b) = \lim_{n\to\infty} \mathcal{U}_n(a,b)$, que é o número de cruzamentos da sequência $X_1, X_2, \ldots$. Então, $E(\mathcal{U}(a,b)) < \infty$, de modo que $\mathcal{U}(a,b) < \infty$ q.c. Segue-se que $P(M_{ab}) = 0$, pois o número de cruzamentos é finito q.c. Defina $M = \cup_{a,b} M_{ab}$, onde a união é sobre todos os racionais $a < b$. Então, $P(M) = 0$ e o conjunto no qual $\{X_n\}$ converge é M^c, logo $\{X_n\}$ converge q.c.

A integrabilidade de X_∞ segue do lema de Fatou:

$$E(|X_\infty|) = E(\liminf_n |X_n|) \leq \liminf_n E(|X_n|) \leq \sup_n E(|X_n|) < \infty. \quad \square$$

Observação: A condição $\sup_n E(|X_n|) < \infty$ pode ser substituída pela condição $\sup_n E(X_n^+) < \infty$. Isso porque $E(|X_n|) = 2E(X_n^+) - E(X_n) \leq 2E(X_n^+) - E(X_1)$, pois temos um submartingale e $E(X_n)$ é crescente. Portanto, $\sup_n E(|X_n|) \leq 2\sup_n E(X_n^+) + E(|X_1|) < \infty$.

Exemplo 4.6. (a) Todo submartingale negativo converge.
(b) Todo martingale positivo converge.

Teorema 4.10. *Seja $\{X_n, \mathcal{F}_n, n \geq 1\}$ um (sub)martingale. As seguintes afirmações são equivalentes:*

(a) *$\{X_n, n \geq 1\}$ converge em L_1;*

(b) *$\{X_n, n \geq 1\}$ é uniformemente integrável;*

(c) $X_n \to X_\infty$ *q.c,* $\{X_n,\ 1 \le n \le \infty\}$ *é um (sub)martingale e* $E(X_n) \to E(X_\infty)$.

Prova: $(b) \Rightarrow (a)$ De fato, $\{X_n\}$ u.i implica que $\{X_n\}$ é limitado em L_1 e pelo Teorema 4.9, X_n converge q.c, logo converge em L_1.

$(a) \Rightarrow (c)$ Se X_n converge em L_1, então $\{X_n\}$ é necessariamente limitada em L_1, portanto $X_n \to X_\infty$ q.c, pelo teorema anterior. Resta provar que $\{X_n\}$, $1 \le n \le \infty$, é um (sub)martingale, ou seja, $E(X_\infty|\mathcal{F}_n) = (\ge) X_n$. Mas, se $m > n$ e $\Lambda \in \mathcal{F}_n$, temos $\int_\Lambda X_m dP = (\ge) \int_\Lambda X_n dP$. Se $m \to \infty$, como $X_m \to X_\infty$ em L_1, obtemos $\int_\Lambda X_\infty dP = (\ge) \int_\Lambda X_n dP$.

$(c) \Rightarrow (b)$ Considere $\{X_n^+,\ 1 \le n \le \infty\}$; este é um submartingale, pois $\{X_n\}$ o é. Desta forma, obtemos que $E(X_\infty^+) \ge \lim_{n\to\infty} E(X_n^+)$. Por outro lado, pelo Lema de Fatou temos que $E(X_\infty^+) \le \lim_{n\to\infty} E(X_n^+)$. Obtemos assim que $\lim_{n\to\infty} E(X_n^+) = E(X_\infty^+) < \infty$. Pelo Teorema 4.4, isso implica que $\{X_n^+, n \ge 1\}$ é u.i. Mas, como $X_n^+ \to X_\infty^+$ q.c, obtemos a mesma convergência em L_1. Como $E(X_n) \to E(X_\infty)$ e $E(X_n^+) \to E(X_\infty^+)$, segue-se que $E(X_n^-) \to E(X_\infty^-)$. Como $X_n^- \ge 0$, para todo n, temos que $\{X_n^-\}$ é u.i pelo Teorema 4.4. Finalmente, como $\{X_n^+\}$ e $\{X_n^-\}$ são u.i, segue-se que $\{X_n\}$ é u.i. $\square$

Teorema 4.11. *Suponha que* $\{X_n, \mathcal{F}_n, n \ge 1\}$ *seja um martingale. Então* $\{X_n\}$ *é u.i se, e somente se, existe uma v.a* Y*, integrável, e* $X_n = E(Y|\mathcal{F}_n)$*. Além disso, se* Y *for mensurável relativamente a* $\bigvee_{n=1}^\infty \mathcal{F}_n$*, então* $X_n \to Y$ *q.c.*

Prova: $(\Rightarrow)$ Suponha que $\{X_n\}$ seja u.i. Então, $X_n \to X_\infty$ q.c e em L_1 e $\{X_n,\ 1 \le n \le \infty\}$ é um martingale, pelo resultado anterior. Logo, basta tomar $Y = X_\infty$.

$(\Leftarrow)$ Suponha $X_n = E(Y|\mathcal{F}_n)$, Y integrável. O resultado segue do Problema 13. Para completar, seja $X_n \to X_\infty$ em q.c e L_1. Mostremos que $Y = X_\infty$ q.c. Sabemos que $E(Y|\mathcal{F}_n) = X_n$ e $E(X_\infty|\mathcal{F}_n) = X_n$. Seja $\Lambda \in \mathcal{F}_n$; então, $\int_\Lambda Y dP = \int_\Lambda X_n dP$ e $\int_\Lambda X_\infty dP = \int_\Lambda X_n dP$, do que decorre que $\int_\Lambda Y dP = \int_\Lambda X_\infty dP$. Portanto, a mesma igualdade vale para $\Lambda \in \cup_{n=1}^\infty \mathcal{F}_n$ e essas duas medidas concordam em uma álgebra que gera $\bigvee_{n=1}^\infty \mathcal{F}_n$, e portanto $Y = X_\infty$ q.c. $\square$

Teorema 4.12. *Seja* $\{X_{-n}, \mathcal{F}_{-n}, n \ge 1\}$ *um martingale reverso. Então, esse martingale converge q.c e em* L_1.

Prova: Temos que $X_{-n} = E(X_{-1}|\mathcal{F}_{-n})$, por definição de martingale, logo se X_{-n} convergir, deve fazê-lo em L_1, pois $\{X_{-n}\}$ é u.i. Portanto, é suficiente provar que $\{X_{-n}\}$ converge q.c. Seja $\mathcal{U}_n(a,b)$ o número de cruzamentos de $[a,b]$ pelo martingale $\{X_{-n}, \ldots, X_{-2}, X_{-1}\}$. Pelo Teorema 4.8, $E(\mathcal{U}_n(a,b)) \le \frac{E(|X_{-1}|)+a}{b-a}$, e fazendo $n \to \infty$, obtemos $\mathcal{U}_n(a,b) \to \mathcal{U}(a,b)$, sendo o limite o número de cruzamentos de $[a,b]$ pelo martingale reverso $\{X_{-n}\}$. Pelo TCM, $E(\mathcal{U}(a,b)) \le (E(|X_{-1}|) + a)/(b-a)$. Segue-se que $\{X_{-n}\}$ converge q.c, usando o mesmo argumento usado no caso usual (Teorema 4.9). $\square$

Corolário 4.1. (Teorema da Continuidade de Lévy para a Esperança Condicional) *Suponha que $(\Omega, \mathcal{F}, P)$ seja um espaço de probabilidade. Então:*

(a) *Se $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots$ é uma família crescente de sub-σ-álgebras de $\mathcal{F}$ e se X é uma v.a integrável, então $\lim_{n\to\infty} E(X|\mathcal{F}_n) = E(X|\bigvee_{n\geq 1}\mathcal{F}_n)$ q.c e em L_1.*

(b) *Se $\mathcal{F}_1 \supset \mathcal{F}_2 \supset \cdots$, então $\lim_{n\to\infty} E(X|\mathcal{F}_n) = E(X|\cap_{n\geq 1}\mathcal{F}_n)$ q.c e em L_1.*

Prova: (a) Como $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots$, $E(X|\mathcal{F}_n)$ é um martingale; é u.i, pelo Problema 13. Logo, $\lim_{n\to\infty} E(X|\mathcal{F}_n)$ existe q.c e em L_1. Vamos provar que esse limite é $X_\infty = E(X|\bigvee_{n\geq 1}\mathcal{F}_n)$. É suficiente mostrar que se $\Lambda \in \bigvee_{n\geq 1}\mathcal{F}_n$, então $\int_\Lambda X_\infty dP = \int_\Lambda X dP$. Tome um $\Lambda \in \mathcal{F}_n$. Segue-se que $\int_\Lambda X_m dP = \int_\Lambda X dP$, para $m \geq n$, pois temos um martingale. Faça $m \to \infty$ e obtenha $\int_\Lambda X_\infty dP = \int_\Lambda X dP$, para todo $\Lambda \in \mathcal{F}_n$. Portanto, a classe dos conjuntos Λ para os quais a última igualdade vale contém conjuntos em $\mathcal{F}_n$, para todo n, ou seja, contém conjuntos em $\cup_{n\geq 1}\mathcal{F}_n$, donde também em $\bigvee_{n\geq 1}\mathcal{F}_n$.

(b) Seja $X_{-n} = E(X|\mathcal{F}_n)$, $\mathcal{F}_n \supset \mathcal{F}_{n+1}$. Então, $\{X_{-n}\}$ é um martingale reverso, e pelo teorema anterior, $X_{-n} \to X_{-\infty}$ q.c e em L_1. Resta provar que $X_{-\infty} = E(X|\cap_{n\geq 1}\mathcal{F}_n)$. Mas, se $\Lambda \in \cap_{n\geq 1}\mathcal{F}_n$, então $\int_\Lambda X_{-m} dP = \int_\Lambda X dP$, porque se $\Lambda \in \cap_{n\geq 1}\mathcal{F}_n$, então $\Lambda \in \mathcal{F}_m$. Quando $m \to \infty$, e como $X_{-m} \to X_{-\infty}$ em L_1, temos $\int_\Lambda X_{-\infty} dP = \int_\Lambda X dP$. $\square$

4.5 Aplicações dos martingales

Nesta seção apresentaremos algumas aplicações dos martingales à análise sequencial (equação de Wald), à teoria das v.a's independentes, derivadas, razão de verossimilhanças e divergência de séries.

4.5.1 Igualdade de Wald

Teorema 4.13. (Wald) *Sejam $\{X_n, n \geq 1\}$ v.a's i.i.d, integráveis. Seja T um tempo de parada, $E(T) < \infty$. Se $S_n = X_1 + \ldots + X_n$, então $E(S_T) = E(X_1)E(T)$.*

Prova: (i) Considere, primeiramente, $Y_n = |X_1| + \ldots + |X_n|$. Então, $Z_n = Y_n - nE(|X_1|)$ é um martingale. Logo, $Z_{T\wedge n} = Y_{T\wedge n} - (T\wedge n)E(|X_1|)$ é um martingale e pelo TAO $E(Y_{T\wedge n} - (T\wedge n)E(|X_1|)) = E(Z_{T\wedge n}) = E(Z_1) = 0$, ou seja, $E(Y_{T\wedge n}) = E(T\wedge n)E(|X_1|)$. Para $n \to \infty$, como $E(T) < \infty$, pelo TCM obtemos $E(Y_T) = E(T)E(|X_1|) < \infty$.

(ii) Caso geral: temos que $S_n - nE(X_1)$ é um martingale, logo $S_{T\wedge n} - (T\wedge n)E(X_1)$ também é, pelo TAO. Logo, novamente, $E(S_{T\wedge n}) = E(X_1)E(T\wedge n)$. Fazendo $n \to \infty$, o lado direito da última igualdade converge para $E(X_1)E(T)$, pelo TCM. Para o lado esquerdo, note que $|S_n| \leq |X_1| + \ldots + |X_n|$, portanto $|S_{T\wedge n}| \leq |Y_{T\wedge n}| \leq Y_T$.

Mas Y_T é integrável, pela parte (i), logo podemos fazer $n \to \infty$ no lado esquerdo para obter $E(S_T)$, usando o TCD. $\square$

Exemplo 4.7. (Aplicações da igualdade de Wald).

[1] Sejam $X_1, X_2, \ldots$ v.a's i.i.d., $P(X_1 = 1) = P(X_1 = -1) = 1/2$; sejam a, b inteiros positivos e $S_n = X_1 + \ldots + X_n$. Queremos calcular a probabilidade de atingir b antes de atingir $-a$.

Sejam $T_1 = \inf\{n : S_n = -a\}$ e $T_2 = \inf\{n : S_n = b\}$. Defina $T = T_1 \wedge T_2$. Queremos $P(T = T_2)$. Suponha, por um momento, que $E(T) < \infty$. Então, $E(S_T) = E(X_1)E(T)$, pela igualdade de Wald. Como $E(X_1) = 0$, obtemos $E(S_T) = 0$.

Também, $E(S_T) = bP(T = T_2) - a[1 - P(T = T_2)] = 0$. Resolva e obtenha $P(T = T_2) = a/(a + b)$.

Vamos mostrar que, de fato, $E(T) < \infty$. Seja $c = a + b$. Encontre d tão grande de modo que $P(|X_1 + \ldots + X_d| < c) \leq \delta < 1$. Então, pela independência dos incrementos temos

$$P(T > nd) \leq P\{|X_1 + \ldots + X_d| < c, \ldots, |X_{(n-1)d} + \ldots + X_{nd}| < c\} \leq \delta^n,$$

ou seja, $\sum_n P(T > nd) < \infty$, ou $E(T) < \infty$.

[2] Sejam $X_1, X_2, \ldots$ v.a's i.i.d., de média zero e $S_n = X_1 + \ldots + X_n$. Seja T o primeiro instante de tempo em que $S_n > 0$, ou seja, $T = \inf\{n \geq 1 : S_n > 0\}$. Então, $E(T) = +\infty$. De fato, se $E(T) < \infty$, pela igualdade de Wald devemos ter $E(S_T) = E(T)E(X_1) = 0$, mas $E(S_T) > 0$, uma contradição. Um argumento similar nos mostra que o tempo de espera para que S_n torne-se negativo pela primeira vez é infinito.

Teorema 4.14. *Seja* $\{X_k, \mathcal{F}_k,\ k = 1, 2, \ldots, n\}$ *um submartingale e* $\lambda > 0$. *Então,*

(a) $\lambda P\{\max_{1 \leq k \leq n} X_k \geq \lambda\} \leq \int_{\{\max_k X_k \geq \lambda\}} X_n dP \leq E(X_n^+)$;

(b) $\lambda P\{\min_{1 \leq k \leq n} X_k \leq -\lambda\} \leq E(X_n) - E(X_1) - \int_{\{\min_k X_k \leq -\lambda\}} X_n dP \leq E(X_n^+) - E(X_1)$.

Prova: Defina o tempo de parada limitado T por $T = \inf\{1 \leq k \leq n : X_k \geq \lambda\} \wedge n$ com a convenção $\inf\{\emptyset\} = \infty$. Seja $\Lambda = \{\omega : \max_{1 \leq k \leq n} X_k \geq \lambda\}$. Então, $\Lambda = \{T < n\} \cup \{T = n, X_n \geq \lambda\}$, de modo que $\Lambda \in \mathcal{F}_T$. Segue-se que $\{X_T, X_n\}$ é um submartingale e pelo TAO e $\int_\Lambda X_T dP \leq \int_\Lambda X_n dP$, do que decorre $\lambda P(\Lambda) \leq \int_\Lambda X_T dP \leq \int_\Lambda X_n dP$. O item (b) é provado de maneira similar observando que $-X_n$ é um supermartingal. $\square$

Aplicações

[1] Sejam $\{X_n, n \geq 1\}$ v.a's independentes, média zero e $\mathrm{Var}(X_k) = \sigma_k^2$. Então, $\sum_{i=1}^n X_i = Y_n$ é um martingale e, portanto, $(\sum_{i=1}^n X_i)^2$ é um submartingale, e

portanto

$$\lambda^2 P\{\max_{1\le k\le n} |X_1+\ldots+X_k| \geq \lambda\} = \lambda^2 P\{\max_{1\le k\le n} |X_1+\ldots+X_k|^2 \geq \lambda^2\}$$

$$\leq E(X_1+\ldots+X_n)^2 = \sum_{k=1}^{n} \sigma_k^2,$$

pelo Teorema 4.14. Vemos, pois, que a a desigualdade de Kolmogorov é um caso especial do Teorema 4.14.

[2] Seja $\{X_n, n \geq 1\}$ um submartingale não negativo. Então, para todo $n \geq 1$,

$$E\left(\sup_{1\le k\le n} X_k^p\right) \leq q^p E\left(X_n^p\right), \tag{4.4}$$

para $1/p + 1/q = 1,\ p, q > 1$.

Para a prova, use a fórmula

$$E(Y^p) = p\int_0^\infty \lambda^{p-1} P(Y > \lambda) d\lambda, \quad p > 1,$$

com $Y = \sup_k X_k$, e use o primeiro limite superior para $P(\sup_k X_k > \lambda)$ dado no Teorema 4.14. Como um caso especial temos que, se $\{X_n\}$ é um martingale, então (4.4) é válida para $\{|X_n|, n \geq 1\}$.

4.5.2 Aplicações a variáveis independentes

Nesta seção fazemos duas aplicações: uma, à LFGN e, outra, a funções características.

[1] **LFGN**. Sejam $\{X_n, n \geq 1\}$ v.a's i.i.d, integráveis e $S_n = X_1 + \ldots + X_n$. Então, $E(X_1|\ S_n, S_{n+1}, \ldots)$ é um martingale (reverso) e igual a S_n/n. Portanto, S_n/n converge q.c e em L_1.

Prova: Temos que

$$\begin{aligned} E(X_1|S_n, S_{n+1}, \ldots) &= E(X_1|S_n, X_{n+1}, X_{n+2}, \ldots) \\ &= E(X_1|S_n) = E(X_2|S_n) = \ldots = E(X_n|S_n). \end{aligned} \tag{4.5}$$

A segunda igualdade segue pela independência. Logo,

$$E(X_1|S_n) = \frac{E(X_1|S_n) + \ldots + E(X_n|S_n)}{n} = \frac{E(X_1 + \ldots + X_n|S_n)}{n} = \frac{S_n}{n}.$$

Resta verificar a terceira igualdade de (4.5) e as conseguintes. Vamos mostrar por exemplo que $E(X_1|S_n) = E(X_2|S_n)$. Ou seja, verificar que, se $\Lambda \in \mathcal{F}\{S_n\}$, $\int_\Lambda X_1 dP = \int_\Lambda X_2 dP$. Considere $\Lambda = \{\omega : S_n \leq \lambda\}$. Então,

$$\int_{\{S_n \leq \lambda\}} X_1 dP = \int_{\{x_1+\ldots+x_n \leq \lambda\}} x_1 dF(x_1)\cdots dF(x_n),$$

onde F é a f.d de X_1, e

$$\int_{\{S_n \leq \lambda\}} X_2 dP = \int_{\{x_1+\ldots+x_n \leq \lambda\}} x_2 dF(x_1)\cdots dF(x_n),$$

portanto igualdade vale para conjuntos da forma acima e em consequência para todo $\Lambda \in \mathcal{F}\{S_n\}$. Pelo Teorema 4.12, $S_n/n = E(X_1|S_n, S_{n+1}, \ldots)$ converge q.c e em L_1. $\square$

[2] **Funções características**. Seja X uma v.a. A função $\phi(\lambda) = E(e^{i\lambda X})$ é chamada *função característica* (f.c) de X. No Capítulo 6 iremos estudar essa função com detalhes.

Teorema 4.15. *Sejam $\{X_n, n \geq 1\}$ v.a's independentes. Seja $\phi_n(\lambda)$ a f.c de X_n. Seja $\varphi_n(\lambda) = \prod_{k=1}^n \phi_k(\lambda)$. Se $\varphi_n(\lambda)$ convergir, quando $n \to \infty$, para todo λ que pertence a algum intervalo $[a,b]$, $a < b$, então $\sum_{n=1}^\infty X_n$ converge q.c.*

Prova: Seja $S_n = \sum_{i=1}^n X_i$ e seja $Y_n = e^{i\lambda S_n}/\varphi_n(\lambda)$. Segue-se que $\{Y_n, n \geq 1\}$ é um martingale limitado, logo converge q.c. Como $\varphi_n(\lambda)$ converge, para todo $\lambda \in [a,b]$, obtemos que $\exp\{i\lambda S_n\}$ converge q.c para todo $\lambda \in [a,b]$.

Queremos provar que S_n converge q.c; temos que, para todo $\lambda \in [a,b]$, $\exp\{i\lambda S_n(\omega)\}$ converge, para quase todo ω.

Fato: para quase todo ω, $\exp\{i\lambda S_n(\omega)\}$ converge para quase todo (c.r à medida de Lebesgue) $\lambda \in [a,b]$.

Seja $h_n(\lambda,\omega) = \exp\{i\lambda S_n(\omega)\}$ e $h(\lambda,\omega) = \lim_n h_n(\lambda,\omega)$, sempre que esse limite exista. Temos que h_n é mensurável c.r ao espaço produto $([a,b], \mathcal{B}_{[a,b]}, \mu) \times (\Omega, \mathcal{F}, P)$, onde μ é a medida de Lebesgue em $[a,b]$. Chamemos de $M = \{(\lambda,\omega) : h_n(\lambda,\omega) \text{ converge}\}$. Então, I_M é mensurável relativamente à σ-álgebra produto. Temos que

$$(\mu \times P)(M) = \int_{[a,b]\times\Omega} I_M d\mu \times dP = \int_{[a,b]} \left[\int_\Omega I_M(\lambda,\omega) dP\right] d\mu.$$

Para λ fixo, $\int I_M dP = 1$, logo $(\mu \times P)(M) = \mu[a,b]$. Mas também temos que

$$\int_\Omega \left[\int_{[a,b]} I_M(\lambda,\omega) d\mu\right] dP = \mu[a,b],$$

e a integral interior é menor ou igual a $\mu[a,b]$, portanto devemos ter que essa integral é igual a $\mu[a,b]$, para quase todo ω.

Portanto, para quase todo ω, $I_M(\lambda,\omega) = 1$, para quase todo λ. Agora, seja $[c,d] \subset [a,b]$. Tome qualquer ω_0 tal que $\lim_n \exp\{i\lambda S_n(\omega_0)\}$ exista, para quase todo λ. Mostremos que, para esse ω_0 escolhido, $S_n(\omega_0)$ converge. Pelo TCD,

$$\int_{[c,d]} \lim_n \exp\{i\lambda S_n(\omega_0)\} d\mu = \lim_n \int_{[c,d]} \exp\{i\lambda S_n(\omega_0)\} d\mu. \tag{4.6}$$

Definimos $A_n = \int_{[c,d]} \exp\{i\lambda S_n(\omega_0)\} d\mu$. Por (4.6), $\lim_n A_n$ existe. Observe também que se $S_n(\omega_0) \neq 0$,

$$|A_n| = \Big|\frac{e^{idS_n(\omega_0)} - e^{icS_n(\omega_0)}}{iS_n(\omega_0)}\Big| \leq \frac{2}{|S_n(\omega_0)|}.$$

Segue-se que $\lim_n \sup S_n(\omega_0) = +\infty$ e $\lim_n \inf S_n(\omega_0) = -\infty$ não são possíveis. Porque, se por exemplo, $\lim_n \sup S_n(\omega_0) = +\infty$, então $\lim_n A_n = 0$, mas então $\int_{[c,d]} \lim_n \exp\{i\lambda S_n(\omega_0)\} d\mu = 0$, para todo c,d, ou seja, $\lim_n \exp\{i\lambda S_n(\omega_0)\} = 0$, para quase todo λ, uma contradição, pois $|\exp\{i\lambda S_n(\omega_0)\}| = 1$. Portanto, se $S_n(\omega_0)$ não converge, deve oscilar entre dois valores finitos r e s, $r < s$, ou seja $\lim_n \inf S_n(\omega_0) = r$, $\lim_n \sup S_n(\omega_0) = s$. Mas, se assim for, então $e^{i\lambda r} = e^{i\lambda s}$, para quase todo $\lambda \in [a,b]$. Mas, isso é impossível para dois valores de λ cujo quociente é irracional. Segue-se que $S_n(\omega_0)$ deve convergir. $\square$

4.5.3 Diversas aplicações

Derivadas

Seja $(\Omega, \mathcal{F}, P)$ um e.p. Para cada n, considere $\{\Lambda_{n,1}, \ldots, \Lambda_{n,k}\}$ como uma partição de Ω, ou seja, $\cup_{k=1}^{\infty} \Lambda_{n,k} = \Omega$, $\{\Lambda_{n,k}, k \geq 1\}$ são disjuntos e $P(\Lambda_{n,k}) > 0$.

Suponha que a partição $(n+1)$-ésima seja um refinamento da n-ésima, ou seja para todo $k \geq 1$, $\Lambda_{n+1,k}$ é um subconjunto de $\Lambda_{n,j}$, para algum j.

Seja φ uma função de conjuntos positiva e aditiva e defina

$$X_n(\omega) = \frac{\varphi(\Lambda_{n,k})}{P(\Lambda_{n,k})}, \quad \text{se } \omega \in \Lambda_{n,k}.$$

Então $\{X_n, n \geq 1\}$ é um martingale positivo e, portanto converge q.c.

Exemplo 4.8. Seja $\Omega = [0,1]$, $\mathcal{F}$ a σ-álgebra de Borel e P a medida de Lebesgue sobre $[0,1]$. Considere F uma função crescente definida sobre $[0,1]$. Defina:

$$\begin{aligned}\Lambda_{n,1} &= \left[0, \frac{1}{2^n}\right), \\ \Lambda_{n,2} &= \left[\frac{1}{2^n}, \frac{2}{2^n}\right), \\ \Lambda_{n,3} &= \left[\frac{2}{2^n}, \frac{3}{2^n}\right) \quad \text{etc.}\end{aligned}$$

Seja

$$X_n(\omega) = \frac{F\left(\frac{k+1}{2^n}\right) - F\left(\frac{k}{2^n}\right)}{\frac{k+1}{2^n} - \frac{k}{2^n}}, \quad \text{se } \omega \in \left[\frac{k}{2^n}, \frac{k+1}{2^n}\right).$$

Então, $\{X_n, n \geq 1\}$ é um martingale positivo, portanto converge q.c. Como F é crescente, $F^{'}(\omega)$ existe para quase todo $\omega \in [0,1]$. Assim, obtemos que $X_n \to F'$, q.c.

Razões de verossimilhanças

Sejam $X_1, X_2, \ldots$, v.a's i.i.d, com função densidade de probabilidade $f(x)$ e seja $g(x)$ outra densidade qualquer. Defina

$$L_n = \begin{cases} \frac{g(X_1)g(X_2)\cdots g(X_n)}{f(X_1)f(X_2)\cdots f(X_n)}, & \text{se } f(X_i) > 0, \text{ para todo } i, \\ 0, & \text{se } f(X_i) = 0, \text{ para algum } i. \end{cases} \tag{4.7}$$

Então, $\{L_n, n \geq 1\}$ é um martingale positivo e, portanto, converge q.c, quando $n \to \infty$. Na realidade, se $P(f(X_1) = g(X_1)) < 1$, $L_n \to 0$ q.c. De fato, seja $L = \lim_{n\to\infty} L_n$. Se $\hat{L}$ é qualquer v.a tal que $\hat{L}$ seja independente de L, mas com a mesma distribuição de L, $\hat{L} \sim L$, então $L\hat{L} \sim L$. No entanto, pela lei de Hewitt-Savage, L é uma constante e $L\hat{L} \sim L$ vale somente se $L = 0$ ou $L = 1$. Essa constante não pode ser 1, pois se o quociente em (4.7) for 1, então $\prod_{i=1}^{\infty} \frac{g(X_i)}{f(X_i)} = 1$, logo $g(X_1)/f(X_1) = 1$, q.c, o que contradiz nossa hipótese. Portanto, $L = 0$ é o único limite possível.

A relevância desse fato em Estatística é a seguinte. Suponha $X_1, X_2, \ldots$ i.i.d. A densidade é f ou g, mas não sabemos qual. Para decidir, calcule (4.7); se o limite for 0, é f, se o limite for infinito, é g.

Divergência de séries

Teorema 4.16. *Seja $\{X_n, n \geq 1\}$ um martingale tal que $E(\sup_n |X_{n+1} - X_n|) < \infty$. Sejam $\Omega_1 = \{\omega : X_n(\omega)$ converge $\}$ e
$\Omega_2 = \{\omega : \liminf_n X_n(\omega) = -\infty, \limsup_n X_n(\omega) = +\infty\}$. Então, $\Omega = \Omega_1 \cup \Omega_2$.*

Prova: Seja $M > 0$ e seja $T = \inf\{n \geq 1 : X_n > M\}$, e $T = \infty$, se não existir tal n. Então, T é um tempo de parada e $\{X_{T\wedge 1}, X_{T\wedge 2}, \ldots\}$ é um martingale, pelo TAO. Temos, então,

$$E(X^+_{T\wedge n}) \leq E[M + \sup_n |X_{n+1} - X_n|] \leq M + E(\sup_n |X_{n+1} - X_n|) < \infty,$$

e portanto $\sup_n E(X^+_{T\wedge n}) < \infty$, do que segue que $\{X_{T\wedge n}\}$ converge q.c.

Note que $X_{T\wedge n}(\omega) = X_n(\omega)$, para todo n sobre o conjunto $\{T = +\infty\}$, portanto $X_n(\omega)$ converge q.c sobre $\{T = +\infty\}$, ou seja, $\{X_n\}$ converge no conjunto onde $\sup_n X_n \leq M$, pela definição de T. Para $M \to \infty$, segue-se que $\{X_n\}$ converge sobre o conjunto onde $\limsup_n X_n < \infty$. Proceda do mesmo modo para $\{-X_n\}$ e obtenha que $\{X_n\}$ converge sobre o conjunto onde $\liminf_n X_n > -\infty$. $\square$

Corolário 4.2. *Sejam $\{X_n, n \geq 1\}$ independentes, média zero, $|X_n| \leq M$, para todo $n \geq 1$. Então, para quase todo $\omega \in \Omega$, ou $\sum X_n(\omega)$ converge ou $\liminf_n \sum_{k=1}^n X_k(\omega) = -\infty$ e $\limsup_n \sum_{k=1}^n X_k(\omega) = +\infty$.*

Problemas

1. Prove que τ, como definido no Exemplo 4.1(c), é um tempo de parada.
2. Prove a afirmação do Exemplo 4.1(d).
3. Prove as propriedades [3]-[6] dos tempos de parada.
4. Prove a observação (h) da seção 4.3.
5. Provem o item (b) do Exemplo 4.3.
6. Sejam $\{Z_n, n \geq 1\}$ i.i.d, $P(Z_1 = 1) = P(Z_1 = 0) = 1/2$. Defina $X_n = 2^n \prod_{k=1}^n Z_k$. Prove que $\{X_n\}$ é um martingale. Prove que esse martingale não tem a forma dada no Exemplo 4.3 (c).
7. Prove que $\{Y_{-n}\}$ do Exemplo 4.4 (b) é um martingale.
8. Prove que, se $\{X_n, n \geq 1\}$ é um (sub)martingale, e T um tempo de parada, finito ou não, então $\{X_{T\wedge 1}, X_{T\wedge 2}, \ldots\}$ é um (sub)martingale.
9. Prove (a) e (b) do Exemplo 4.5.
10. Seja $\Omega = \{1, 2, 3, \ldots\}$, $\mathcal{F}$ a classe de todos os subconjuntos de Ω, $P(\{k\}) = 1/k - 1/(k+1)$, se $k \in \Omega$. Suponha que $\{X_n, n \geq 1\}$ seja definido por $X_n(\{k\}) = n$, se $k > n$ e $X_n(\{k\}) = -1$, se $k \leq n$.

    (a) Prove que $\{X_n\}$ é um martingale e encontre seu limite, quando $n \to \infty$.

    (b) Determine se $\{X_n\}$ é ou não u.i.

    (c) Calcule $P\{\sup_n |X_n| > \lambda\}$ exatamente.

11. Produza exemplos de: (a) martingales $\{X_n\}$ e $\{Y_n\}$ tais que $\{X_n + Y_n\}$ não seja um martingale; (b) um submartingale $\{X_n\}$ tal que $\{|X_n|\}$ não seja um submartingale.

12. (Decomposição de Submartingales - Doob) (a) Seja $\{X_n, \mathcal{F}_n, n \geq 1\}$ um submartingale. Mostre que $X_n = M_n + A_n$, univocamente, onde $\{M_n, \mathcal{F}_n, n \geq 1\}$ dé um martingale e $\{A_n\}$ tem as propriedades: $A_1 = 0$, $A_n \leq A_{n+1}$ e A_n é $\mathcal{F}_{n-1}$-mensurável. O processo $\{A_n\}$ é chamado de *compensador*.

    (b) Suponha que se saiba que todo martingale limitado no sentido de L_1 converge q.c. Use esse fato e (a) para provar diretamente que todo submartingale limitado no sentido L_1 converge q.c.

    [Esse resultado é útil em Confiabilidade. Sugestão: $A_n = \sum_{k=1}^{n-1} E(X_{k+1} - X_k|\mathcal{F}_k),\ n \geq 2$.] Veja Gut(2013).

13. Seja X uma v.a. integrável sobre $(\Omega, \mathcal{F}, P)$. Mostre que a classe de funções $\{E(X|\mathcal{G})\}$, onde $\mathcal{G}$ varia sobre todas as sub-σ-álgebras de $\mathcal{F}$, é uniformemente integrável.

14. (a) Seja $\{X_n, \mathcal{F}_n, n \geq 1\}$ um martingale e seja $V = \{V_n, n \geq 1\}$ um processo estocástico tal que $|V_n| \leq 1$, para todo n e V_n sendo $\mathcal{F}_{n-1}$-mensurável (tome $\mathcal{F}_0 = \{\Omega, \emptyset\}$). Prove que $\sum_{k=1}^{n} V_k d_k = Y_n$ é um martingale, onde $d_1 = X_1, d_k = X_k - X_{k-1},\ k > 1$.

    (b) Um modo útil de olhar um tempo de parada é o seguinte. Seja T um tempo de parada para um martingale $\{X_n, \mathcal{F}_n, n \geq 1\}$. Prove que $\{X_{T \wedge n}, n \geq 1\}$ tem a forma dada em (a). Tome $V_k = I\{T \geq k\}$.

15. Sejam $\{Y_n, n \geq 1\}$ independentes, simetricamente distribuídas, mas não necessariamente integráveis. Seja $\mathcal{F}_n = \mathcal{F}\{Y_1, \ldots, Y_n\}$. Seja V_n $\mathcal{F}_{n-1}$-mensurável. Prove que se $\sum_k Y_k$ converge q.c para um limite finito, então $\sum_k V_k Y_k$ converge q.c para um limite finito sobre o conjunto $\{\sup_n |V_n| < \infty\}$. Note que $\sum_k V_k Y_k$ não é uma soma de v.a's independentes.

16. Prove a chamada decomposição de Riesz para supermartingales: se $\{X_n, \mathcal{F}_n, n \geq 1\}$ é um supermartingale u.i, então $X_n = M_n + A_n$, onde $\{M_n, \mathcal{F}_n, n \geq 1\}$ é um martingale u.i. e $\{A_n, \mathcal{F}_n, n \geq 1\}$ é um *potencial*, isto é, um supermartingale não negativo tal que $\lim_{n\to\infty} E(A_n) = 0$.

17. Prove a Decomposição de martingales de Krichberg: se $\{X_n, n \geq 1\}$ é um martingale, então $X_n = X_n^{(1)} - X_n^{(2)}$, onde $\{X_n^{(i)}, n \geq 1\}$ é, para cada $i = 1, 2$, um martingale não negativo.

    [Sugestão: considere a decomposição de Doob do submartingale $X_n^+ = M_n + A_n$ e coloque $X_n^{(1)} = M_n + E(A_\infty|\mathcal{F}_n)$.]

18. Estabeleça a seguinte extensão do TAO. Seja $\{X_n, n \geq 1\}$ um martingale u.i e seja $X_\infty = \lim_n X_n$. Suponha que $S \leq T$ sejam tempos de parada e coloque $X_T = X_\infty$ sobre o conjunto $\{T = \infty\}$. Prove que $E(X_T|\mathcal{F}_S) = X_S$. A diferença para o TAO provado no texto é que S, T podem ser infinitos.

19. Seja $\{X_n, n \geq 1\}$ um supermartingale não negativo. Prove que, para qause todo ω, se $X_n(\omega) = 0$, então $X_{n+k}(\omega) = 0$, para todo $k \geq 0$ (Esse resultado nos diz que sempre

que um supermartingale não negativo atinge zero, ele permance lá permanentemente, o que pode parecer surpreendente).

[Sugestão: considere o tempo de parada $T = \inf\{n : X_n(\omega) = 0\}$ e use o TAO.]

20. Prove o Corolário 4.2.

21. Prove que, de fato, $\mathcal{F}_{\tau-}$ e $\mathcal{F}_{\tau+}$ são σ-álgebras.

22. Se $\{X_n, \mathcal{F}_n, n \geq 1\}$ é um martingale em L_2, com $U_{n+1} = X_{n+1} - X_n$, $n \geq 1$, diferenças martingales, prove que:

    (a) $E(U_n U_m) = E(U_n^2)$, se $n = m$ e igual a zero se $n \neq m$.

    (b) Para $m < n$, $E(U_n X_m) = E(U_n E(X_n|\mathcal{F}_n)) = 0$.

Capítulo 5

Processos Estocásticos com Tempo Contínuo

Neste capítulo, faremos uma introdução aos processos estocásticos com tempo (parâmetro) contínuo. Em particular, veremos as dificuldades encontradas nesse caso. Estenderemos o estudo dos martingales com tempo discreto, estudados no Capítulo 4 para o caso de tempo contínuo e daremos uma introdução aos processos com incrementos independentes. No Capítulo 9 estudaremos com mais detalhes o movimento browniano, ou processo de Wiener. A seção 5.1 traz conceitos mais difíceis e pode ser ignorada numa primeira leitura. Referências para esse capítulo são Doob (1953), Breiman (1968, 1992) e Itô (2006).

5.1 Separabilidade e mensurabilidade

Iniciamos com a definição de processo estocástico com parâmetro contínuo.

Definição 5.1. *Seja* $(\Omega, \mathcal{F}, P)$ *um e.p e* T *um intervalo dos reais. Então,* $X = \{X_t, t \in T\}$ *é um processo estocástico com parâmetro contínuo se cada* X_t *é uma variável aleatória sobre* $(\Omega, \mathcal{F}, P)$*.*

Podemos considerar X como uma aplicação de $T \times \Omega \to \mathbb{R}$, tal que $X : (t, \omega) \to X_t(\omega)$. Frequentemente, escrevemos $X_t(\omega) = X(t, \omega)$.

As funções (uma para cada ω) definidas por $X(\cdot, \omega)$ são chamadas *funções amostrais* ou *trajetórias* de X (ou ainda *realizações*).

Exemplo 5.1. Seja $(\Omega, \mathcal{F}, P)$ um e.p e considere Y uma v.a normal sobre esse e.p. Defina $X_t(\omega) = \text{sen}[at + Y(\omega)]$. Então, para cada ω, uma trajetória de X será uma senóide, como função de t real.

Definição 5.2. *Seja* $(\Omega, \mathcal{F}, P)$ *um e.p,* $X = \{X_t, t \in T\}$ *um processo estocástico.*

(a) *Dizemos que quase todas as trajetórias de X são contínuas se existe um conjunto nulo Λ tal que, se $\omega \notin \Lambda$, então $X(\cdot, \omega)$ é contínua.*

(b) *Dizemos que X é contínuo em t_0 q.c, se existe um conjunto nulo Λ_0 tal que $\lim_{t\to t_0} X_t(\omega) = X_{t_0}(\omega)$, sempre que $\omega \notin \Lambda_0$. X é contínuo q.c se X for contínuo q.c em cada $t \in T$.*

(c) *X é contínuo em probabilidade em t_0 se $\lim_{t\to t_0} X_t = X_{t_0}$, limite este em probabilidade. X é contínuo em probabilidade se for contínuo em probabilidade em todo ponto $t \in T$.*

Observamos que (a) $\Rightarrow$ (b) $\Rightarrow$ (c). Claramente, (c) não necessita implicar (b).

Exemplo 5.2. Vejamos um exemplo de X que satisfaz (b) mas não (a). Considere $\Omega = [0, 1]$, $\mathcal{F}$ σ-álgebra de Borel sobre $[0, 1]$ e P a medida de Lebesgue sobre $[0, 1]$. Considere, também, $T = [0, 1]$. Defina X por

$$X_t(\omega) = \begin{cases} 0, & \text{se } t < \omega, \\ 1, & \text{se } t \geq \omega \end{cases} \tag{5.1}$$

Então, X satisfaz (b), mas não (a).

Vamos discutir algumas dificuldades encontradas no caso de parâmetro contínuo. Seja $\{X_n, n \geq 1\}$ um processo estocástico com parâmetro discreto. Então, operações razoáveis com a sequência $X_1, X_2, \ldots$, sempre produz funções mensuráveis como resultados.

Por exemplo, $\sup_n X_n$ é uma v.a (isto é, mensurável), porque $\{\sup_n X_n < \lambda\} = \cap_n \{X_n < \lambda\}$ é mensurável.

Do mesmo modo, $\lim_n X_n$, $\limsup_n X_n$, $\inf_n X_n$ etc, são mensuráveis.

Seja, agora, um processo estocástico $\{X_t, t \in T\}$ com parâmetro contínuo. Então, $\sup_{t\in T} X_t$ não necessita ser mesurável. De fato, $\{\sup_t X_t < \lambda\} = \cap_{t\in T}\{X_t < \lambda\}$, e este conjunto pode ser não mensurável, porque temos uma intersecção não enumerável. De modo similar, o conjunto $\{\omega : X(\cdot, \omega) \text{ é contínuo}\}$ ou o conjunto $\{\omega : |X(\cdot, \omega)| \text{ permanece limitado}\}$ etc não são necessariamente mensuráveis.

Outra observação: $\{X_t, t \in T\}$ é usualmente construído usando-se somente as distribuições finito-dimensionais (lembre-se do Teorema da Extensão de Kolmogorov). Os conjuntos que temos discutido envolvem todas as distribuições ao mesmo tempo. De fato, dois processos tendo as mesmas distribuições finito-dimensionais podem ter trajetórias radicalmente diferentes.

Exemplo 5.3. Seja $\Omega = [0, 1]$, $\mathcal{F}$ a σ-álgebra de Borel e P a medida de Lebesgue, ambas sobre $[0, 1]$. Defina os processos $\{X_t, t \in T\}$ e $\{Y_t, t \in T\}$ como segue:

$$
\begin{aligned}
X_t(\omega) &= 0, \\
Y_t(\omega) &= \begin{cases} 0, & t \neq \omega, \\ 1, & t = \omega. \end{cases}
\end{aligned}
$$

Note que para todo $t \in [0,1]$, $X_t = Y_t$ q.c, de modo que esses processos têm as mesmas distribuições finito-dimensionais. Mas o comportamento das trajetórias é diferente. Por exemplo, $P(\{\omega : \sup_t X_t(\omega) = 1\}) = 0$, mas $P(\{\omega : \sup_t Y_t(\omega) = 1\}) = 1$. Também, $P(\{\omega : X(\cdot,\omega)\ \text{ é contínuo}\}) = 1$, ao passo que a probabilidade do conjunto similar para Y_t é zero.

Exemplo 5.4. Considere o mesmo e.p do exemplo anterior e seja $A^{'}$ um conjunto não mensurável de $[0,1]$ e defina $A = \{(x,y) : x = y \text{ e } x \in A^{'}\}$. Defina $X_t(\omega) = I_A(t,\omega)$, que define um p.e $\{X_t, t \in T\}$. Então, $\{\omega : \sup_t X_t = 1\} = A^{'}$, que não é mensurável.

Daremos, a seguir três definições de processo separável. Por comodidade, os processos estocásticos envolvidos nestas definições serão considerados com valores em $[-\infty,\infty]$.

Definição 5.4. *Seja $X = \{X_t, t \in T\}$ um p.e com valores em $[-\infty,\infty]$ e $\mathcal{A}$ a classe de todos os conjuntos fechados de $[-\infty,\infty]$. Dizemos que X é separável se a seguinte afirmação é verdadeira: Existe um conjunto enumerável S de pontos em T e um conjunto nulo Λ, tal que se $A \in \mathcal{A}$, então os conjuntos*

$$\Lambda_1 = \{\omega : X_t(\omega) \in A,\ t \in I \cap S\}$$

e

$$\Lambda_2 = \{\omega : X_t(\omega) \in A,\ t \in I \cap T\}$$

diferem no máximo por um subconjunto de Λ. Aqui, I é qualquer subintervalo aberto de T.

O conjunto enumerável envolvido, S, é chamado um *separador*.

Definição 5.5. *X é separável se existe um conjunto nulo Λ tal que, sempre que $\omega \notin \Lambda$, o fecho do gráfico de $X(t,\omega)$, para $t \in S$, contém o gráfico de $X(t,\omega)$, para $t \in T$.*

Definição 5.6. *X é separável se existe um conjunto nulo Λ tal que, se $\omega \notin \Lambda$, então $X_t(\omega) \in \cap_{I:t\in I}\overline{X(I \cap S,\omega)}$, para todo t, onde I é um intervalo aberto da reta e $\overline{X(I \cap S,\omega)} :=$ fecho$\{X(s,\omega) : s \in I \cap S\}$.*

Essas três definições são equivalentes (veja o Problema 1).

Algumas consequências: Seja $\{X_t, t \in T\}$ um processo separável e I qualquer subintervalo de T.

[1] $\sup_{t \in I \cap S} X_t(\omega) = \sup_{t \in I} X_t(\omega)$, se $\omega \notin \Lambda$.

Note que o membro esquerdo é mensurável, pois $I \cap S$ é enumerável. Para a prova, note que o lado esquerdo é menor ou igual ao segundo, pois estamos tomando o supremo sobre um conjunto menor. Cada X_t é um limite de X_s, para $s \in S$, para todo $\omega \notin \Lambda$, logo o segundo membro não pode ser maior do que o primeiro.

De modo similar, temos que

$$\limsup_{t \to t_0,\, t \in S} X_t(\omega) = \limsup_{t \to t_0} X_t(\omega), \quad \omega \notin \Lambda.$$

[2] Seja S' um conjunto enumerável, $S' \supset S$, sendo S um separador. Então S' é também um separador.

[3] S é denso em T.

[4] Se $\{X_t, t \in T\}$ é separável, e se f for contínua, então $Y_t = f(X_t)$ é também separável.

[5] Seja X um processo tal que quase todas as trajetórias de X são contínuas à direita. Então, X é separável.

Exemplo 5.5. Vejamos um exemplo de um processo não separável. Defina $Y_t(\omega)$ como no exemplo 5.2. Então $\{Y_t, t \in T\}$ é não separável. Para provar, use a Definição 5.5.

Definição 5.7. *Sejam $X = \{X_t, t \in T\}$ e $Y = \{Y_t, t \in T\}$ dois processos estocásticos. Dizemos que X e Y são equivalentes se, para cada t, $X_t = Y_t$ q.c.*

Para provar o teorema seguinte, necessitamos de dois lemas.

Lema 5.1. *Seja $\mathcal{A}_0$ o conjunto de todas as reuniões finitas de intervalos abertos ou fechados de $[-\infty, \infty]$, tendo extremos racionais (ou infinitos). Seja $\mathcal{A}$ a classe dos conjuntos que são intersecções de conjuntos de $\mathcal{A}_0$. Então, $\mathcal{A}$ contém todos os conjuntos fechados de $[-\infty, \infty]$.*

Lema 5.2 Seja $\{X_t, t \geq T\}$ um processo estocástico.
(a) Seja A um boreliano de $[-\infty, \infty]$. Existe uma sequência enumerável (finita ou infinita) $\{t_1, t_2, \ldots\}$ dependendo de A tal que para todo $t \in T$,

$$P\{X_{t_n} \in A, n \geq 1, X_t \notin A\} = 0.$$

(b) *Seja $\mathcal{A}_0$ qualquer coleção enumerável de borelianos de $[-\infty, \infty]$. Seja $\mathcal{A}$ a classe de todos os conjuntos que são intersecções de conjuntos de $\mathcal{A}_0$. Então, existe uma sequência enumerável (finita ou infinita) $\{t_1, t_2, \ldots\}$ tal que, para todo $t \in T$, existe um conjunto nulo Λ_t tal que*

$$\{X_{t_k} \in A, k \geq 1, X_t \notin A\} \subset \Lambda_t,$$

para todo $A \in \mathcal{A}$.

Teorema 5.1. *Seja $X = \{X_t, t \in T\}$ um p.e sobre $(\Omega, \mathcal{F}, P)$. Então, existe um p.e $\tilde{X} = \{\tilde{X}_t, t \in T\}$ tal que $\tilde{X}$ é equivalente a X e $\tilde{X}$ é separável. Em geral, $\tilde{X}$ tem valores em $[-\infty, +\infty]$ mesmo que X tenha valores em $\mathbb{R}$.*

Prova: Seja $\mathcal{A}_0$ a coleção de conjuntos que são reuniões finitas de conjuntos abertos ou fechados com extremos racionais (ou infinitos). Seja $\mathcal{A}$ o conjunto que contém todas as intersecções de conjuntos de $\mathcal{A}_0$.

Pelo Lema 5.1, $\mathcal{A}$ contém os conjuntos fechados de $[-\infty, \infty]$. Também, $\mathcal{A}_0$ é uma coleção enumerável de conjuntos.

Seja I um subintervalo com extremos racionais ou infinitos de T. Apliquemos o Lema 5.2 (b) com $\mathcal{A}_0$ e $\mathcal{A}$ descritos aqui e T substituído por $I \cap T$. Temos, então, que para um conjunto enumerável S_I e um conjunto nulo N_t^I, $\{X_s \in A, s \in S_I, X_t \notin A\} \subset N_t^I$. Defina um separador S por $S = \cup_I S_I$ e defina $N_t = \cup_I N_t^I$ e $A(I, \omega) = \overline{\{X_s(\omega) : s \in I \cap S\}}$. Finalmente, seja $A(t, \omega) = \cap_{I: t \in I} A(I, \omega)$.

Observe que $X_t(\omega) \in A(t, \omega)$, se $\omega \notin N_t$. Também, se $\omega \in N_t$, $A(t, \omega)$ é ainda não vazio (pois intersecções finitas dos $A(I, \omega)$'s na definição de $A(t, \omega)$ não são vazias, e cada $A(I, \omega)$ é um conjunto compacto em $[-\infty, \infty]$, portanto vale a propriedade da intersecção finita).

Assim, $A(t, \omega)$ é não vazio, para todo ω. Portanto, definamos $\tilde{X}$ como segue: $\tilde{X}_t(\omega) = X_t(\omega)$, se $\omega \notin N_t$, e igual a qualquer ponto em $A(t, \omega)$, se $\omega \in N_t$.

Então, $\tilde{X}_t(\omega) = X_t(\omega)$ q.c e portanto $\tilde{X}$ é equivalente a X. Também, $\tilde{X}$ é separável, pois por construção

$$\tilde{X}_t(\omega) \in A(t, \omega) = \cap_{I: t \in I} A(I, \omega) = \cap_{I: t \in I} \overline{\tilde{X}(I \cap S, \omega)},$$

para I com extremos racionais (ou infinitos). Pela definição 5.6, devemos mostrar que $\tilde{X}_t(\omega) \in \cap_{I: t \in I} \overline{\tilde{X}(I \cap S, \omega)}$, I aberto. Se I for qualquer intervalo aberto contendo t, segue-se que existe um intervalo aberto racional I', tal que $I' \subset I$ e $t \in I'$. Mas $\overline{X(I' \cap S, \omega)} \subset \overline{X(I \cap S, \omega)}$, portanto $\cap_{I \text{ rac. aberto}: t \in I} \overline{X(I \cap S, \omega)} = \cap_{I \text{ aberto}: t \in I} \overline{X(I \cap S, \omega)}$. $\square$

Definição 5.8. *Seja $X = \{X_t, t \in T\}$ um p.e sobre $(\Omega, \mathcal{F}, P)$. Seja $\mathcal{B}_T \times \mathcal{F}$ a σ-álgebra produto sobre T, onde $\mathcal{B}_T$ é a σ-álgebra de Borel sobre T. Dizemos que o processo X é mensurável se a aplicação $X : (t, \omega) \to \mathbb{R}$ for conjuntamente mensurável em (t, ω).*

Teorema 5.2. *Seja $X = \{X_t, t \in T\}$ um processo mensurável. Então:*

(a) Para todo ω, a trajetória $X(\cdot, \omega)$ é Borel mensurável;

(b) Se X_t for integrável, para cada t, então $E(X_t)$ é uma função Borel mensurável de t.

Prova: (a) Imediata.

(b) Veja o Problema 5. $\square$

Exemplo 5.6. (Um processo mensurável) Se X tem trajetórias contínuas à direita então X é mensurável. De fato: para todo $n \geq 1$, defina $X^{(n)}(t, \omega) = (k+1)/n$, se $\{k/n \leq X(t, \omega) < (k+1)/n\}$, $k \geq 0$. Então, esses procesos são mensuráveis. Pela continuidade à direita, $X^{(n)}(t, \omega) \to X(t, \omega)$, para cada (t, ω) fixado. Deduzimos que X é mensurável como limite simples de processos mensuráveis.

5.2 Martingales com parâmetro contínuo

Nesta seção vamos estender o conceito de (sub/super)martingales para o caso de um processo com tempo contínuo. Em particular, o TAO é estendido para esse caso.

Definição 5.9. *Seja $(\Omega, \mathcal{F}, P)$ um e.p e T um intervalo de $\mathbb{R}$. Para cada $t \in T$, seja $\mathcal{F}_t$ uma σ-álgebra. Suponha que, para $t \leq s$, temos $\mathcal{F}_t \subset \mathcal{F}_s$ (crescente). O processo $X = \{X_t, t \in T\}$ é um martingale com parâmetro contínuo se*

(a) X_t é $\mathcal{F}_t$-mensurável e integrável;

(b) se $s < t$, então $E(X_t|\mathcal{F}_s) = X_s$, q.c.

Submartingales e supermartingales são definidos de maneira similar ao caso discreto.

Teorema 5.3. *Seja $X = \{X_t, \mathcal{F}_t, t \in T\}$ um submartingale separável. Então:*

(a) $\lambda P\{\sup_{t \in T} X_t \geq \lambda\} \leq \sup_{t \in T} E(X_t^+)$;

(b) $E(U(a,b)) \leq \sup_{t \in T} \frac{E(X_t^+)+a}{b-a}$, *onde $U(a,b)(\omega)$ é o número de vezes que a trajetória de $X(\cdot, \omega)$ vai de abaixo de a para cima de b.*

Prova: (a) Seja $S = \{s_1, s_2, \ldots\}$ um conjunto separador para X. Suponha $s_1 < s_2 < \cdots$. Então, $\{X_{s_1}, \ldots, X_{s_n}\}$ é um martingale com parâmetro discreto, logo pelo teorema análogo no caso discreto temos

$$\lambda P\Big\{ \sup_{1 \leq k \leq n} X_{s_k} \geq \lambda \Big\} \leq \sup_{k \leq n} E(X_{s_k}^+) \leq \sup_{t \in T} E(X_t^+).$$

Fazendo $n \to \infty$, obtemos $\lambda P\{\sup_{k\geq 1} X_{s_k} \geq \lambda\} \leq \sup_t E(X_t^+)$. Mas, então, $\sup_k X_{s_k} = \sup_{t\in T} X_t$ q.c, pois X é separável.

(b) Mesmo tipo de prova. Veja o Problema 7. □

Teorema 5.4. *Seja $X = \{X_t, t \in T\}$ um submartingale separável. Então:*

(a) *Quase todas as trajetórias de X são limitadas sobre subintervalos compactos de T.*

(b) *Quase todas as trajetórias de X são livres de descontinuidades oscilatórias.*

Prova: (a) Seja $I = [a, b]$ um subintervalo compacto de T. Pelo teorema anterior,

$$\lambda P\Big\{\sup_{t\in I} X_t \geq \lambda\Big\} \leq \sup_{t\in I} E(X_t^+) = E(X_b^+) < \infty.$$

Para $\lambda \to \infty$, obtemos $P\{\sup_{t\in I} X_t < \infty\} = 1$, portanto quase todas as trajetórias são tais que $\sup_{t\in I} X_t < \infty$. Também, temos que

$$\lambda P\Big\{\inf_{t\in I} X_t \leq -\lambda\Big\} \leq E(X_b^+) - E(X_a).$$

Para $\lambda \to \infty$, temos $P\{\inf_{t\in I} X_t > -\infty\} = 1$, logo para todo $t \in I$ e quase todo ω, $X_t(\omega) > -\infty$. Segue-se que quase todas as trajetórias são limitadas.

(b) Seja I como na parte (a). Sejam $r < s$ números racionais e $U_I(r, s)(\omega)$ o número de cruzamentos de $[r, s]$ pela trajetória $X(t, \omega)$, $a \leq t \leq b$. Então, $E[U_I(r, s)] < \infty$, pelo teorema anterior. Em particular, $U_I(r, s)(\omega) < \infty$, exceto para $\omega \in \Lambda_{r,s}$ de probabilidade nula. Seja $\Lambda = \cup_{r<s}\Lambda_{r,s}$, r, s racionais; Λ é um conjunto nulo. Se $\omega \notin \Lambda$, a trajetória $X(\cdot, \omega)$ não tem descontinuidades oscilatórias sobre I. Pois, suponha que $X(\cdot, \omega)$ não tenha limite à esquerda no ponto $u \in I$. Então, existem números racionais $r < s$ tais que $\limsup_{t\uparrow u} X_t \geq s$ e $\liminf_{t\uparrow u} X_t \leq r$. Mas, então, X_t cruza $[r, s]$ um número infinito de vezes, e isso é impossível, se $\omega \notin \Lambda$. □

Observação: A parte (b) significa que, existe um conjunto nulo Λ tal que, se $\omega \notin \Lambda$, então a trajetória $X(\cdot, \omega)$ tem a seguinte propriedade: se u é um ponto interior de T, então $\lim X_{t\uparrow u}$ e $\lim_{t\downarrow u} X_t$ ambos existem. Deduzimos que trajetórias de submartingales separáveis somente podem ter descontinuidades com saltos.

A seguir, responderemos às seguintes questões:

[1] Suponha que $X = \{X_t, a \leq t < b\}$ seja um submartingale separável. Então

(a) Existe o $\lim_{t\uparrow b} X_t$? Quando?

(b) Se existir o limite em (a) e o chamarmos X_b, quando $X = \{X_t, a \leq t \leq b\}$ será um submartingale?

[2] Suponha $X = \{X_t, a < t \leq b\}$ seja um submartingale separável.

(c) Existe o $\lim_{t\downarrow a} X_t$?

(d) Se existir o limite em (c) e o chamamos X_a, quando $X = \{X_t, a \leq t \leq b\}$ será um submartingale?

Teorema 5.5. *Seja $X = \{X_t, a \leq t < b\}$ um submartingale separável.*

(a) *Se $\sup_{a\leq t<b} E(|X_t|) < \infty$, então $\lim_{t\uparrow b} X_t$ existe q.c.*

(b) Se $X = \{X_t, a \leq t < b\}$ for u.i, então o limite em (a) existe no sentido L_1 e $X = \{X_t, a \leq t \leq b\}$ é um submartingale, se definirmos $X_b = \lim_{t\uparrow b} X_t$.

Prova: (a) Suponha $t_n \uparrow b$, então $\{X_{t_n}, n \geq 1\}$ é submartingale com parâmetro discreto, L_1-limitado, usando a hipótese de (a). Portanto, $\lim_{n\to\infty} X_{t_n}$ existe q.c e denotamos por Y este limite. O limite independe (q.c) da sequência t_n escolhida; de fato, suponha que $s_n \uparrow b$ e tal que $\lim_n X_{s_n} = Y'$. Considere a sequência $\{s_1, t_1, s_2, t_2, \ldots\}$ e seja $\{r_1, r_2, \ldots\}$ essa sequência em ordem crescente. Então, $\{X_{r_n}, n \geq 1\}$ é um submartingale L_1-limitado e portanto $\lim_n X_{r_n}$ existe q.c, e isso é impossível, a menos que $Y = Y'$, q.c.

(b) É suficiente mostrar que $\{X_t, a \leq t \leq b\}$ é um submartingale. Para isso, mostremos que $\int_\Lambda X_t dP \leq \int_\Lambda X_b dP$, sempre que $\Lambda \in \mathcal{F}_t$. Sejam $t < t_n < b$, $t_n \uparrow b$; então, $\int_\Lambda X_t dP \leq \int_\Lambda X_{t_n} dP$, pela definição de submartingale. Para $n \to \infty$, por integrabilidade uniforme, obtemos o desejado, pois $X_{t_n} \to X_b$, por definição. □

Teorema 5.6. *Seja $X = \{X_t, a < t \leq b\}$ um submartingale separável. Então, $\lim_{t\downarrow a} X_t$ existe q.c e em L_1. Além disso, $\{X_t, a \leq t \leq b\}$ é um submartingale, se definirmos $X_a = \lim_{t\downarrow a} X_t$.*

Prova: Suponha que $t_n \downarrow a$, então $\{X_{t_n}, n \geq 1\}$ é um submartingale reverso e portanto $\lim_{n\to\infty} X_{t_n}$ existe q.c e em L_1. Este limite é independente da sequência $\{t_n\}$ escolhida, pelo mesmo argumento feito no teorema anterior. Logo, $\{X_t, a \leq t \leq b\}$ é um submartingale, pelo argumento do teorema anterior. □

Definição 5.10. *Seja T um intervalo real aberto à direita. Seja $\{\mathcal{F}_t, t \in T\}$ uma família crescente de σ-álgebras e $\mathcal{F}_{t+} = \cap_{s>t} \mathcal{F}_s$. Se $\mathcal{F}_{t+} = \mathcal{F}_t$, para todo t, a família $\{\mathcal{F}_t, t \in T\}$ é chamada contínua à direita.*

Exemplo 5.7. Seja $\mathcal{F}_t$ a σ-álgebra gerada pelos conjuntos de Borel de $[0, t)$ e $[t, 1)$. Então, para cada t, o ponto $\{t\}$ satisfaz: $\{t\} \in \mathcal{F}_{t+}$, mas $\{t\} \notin \mathcal{F}_t$.

Definição 5.11. *Sejam $X = \{X_t, t \in T\}$ e $Y = \{Y_t, t \in T\}$ dois p.e. Dizemos que Y é uma modificação càdlàg (do francês "continue à droite, limite à gauche") de X se Y tem as suas trajetórias contínuas à direita, com limites à esquerda e $X_t = Y_t$ q.c, para todo $t \in T$.*

A seguir, apresentamos um resultado que garante a existência de modificações càdlàg de um submartingale. Chamaremos uma família $\{\mathcal{F}_t, t \in T\}$ de completa se todos os subconjuntos de conjuntos nulos de $\mathcal{F}$ pertencem a $\mathcal{F}_t$ para todo $t \in T$. No teorema a seguir, consideraremos o processo $\{X_{t+}, t \in T\}$ definido por $X_{t+} = \lim_{t' \downarrow t} X_{t'}$ se o limite existir e 0 se não. Pela observação seguindo o Teorema 5.4, observe que fora de um conjunto nulo, $\lim_{t' \downarrow t} X_{t'}$ existe para todo $t \in T$. Também observe que por construção, para todo $t \in T$, X_{t+} é $\mathcal{F}_{t+}$ mensurável.

Teorema 5.7. *Seja T um intervalo real aberto à direita. Seja $X = \{X_t, t \in T\}$ um submartingale separável.*

(a) O processo $\{X_{t+}, t \in T\}$ é um submartingale relativamente a $\{\mathcal{F}_{t+}, t \in T\}$ e, é um martingale se X for um martingale.

(b) Seja $\{\mathcal{F}_t, t \in T\}$ uma família completa e contínua à direita. Então $\{X_t, \mathcal{F}_t, t \in T\}$ tem uma modificação càdlàg que também é um submartingale se, e somente se, $E(X_t)$ é contínua à direita como uma função de t. Em particular, todo martingale separável relativo a $\{\mathcal{F}_t, t \in T\}$ tem uma modificação càdlàg que também é um martingale.

Observação: A hipótese de separabilidade não é necessária no Teorema 5.7. Veja Le Gall (2013), Théorème 3.4.

Prova: (a) Devemos provar que se $\Lambda \in \mathcal{F}_{s+}$ e $s < t$, então $\int_\Lambda X_{s+} dP \leq \int_\Lambda X_{t+} dP$. Suponha que $s_n \downarrow s$, $s < s_n < t$ e $t_n \downarrow t$. Então, como $\Lambda \in \mathcal{F}_{s+}$, segue-se que $\Lambda \in \mathcal{F}_{s_n}$, para todo n, pois $\mathcal{F}_{s+} = \cap \mathcal{F}_{s_n}$. Portanto, $\int_\Lambda X_{s_n} dP \leq \int_\Lambda X_{t_n} dP$. Faça $n \to \infty$ para obter o resultado. A segunda afirmação do enunciado decorre imediatamente da prova trocando $\leq$ por $=$.

(b) Primeiramente, pela observação seguindo o Teorema 5.4, sabemos que existe um conjunto nulo N tal que se $\omega \notin N$, as trajetórias $X(\cdot, \omega)$ tem limites à direita e a esquerda. Definimos o processo $Y = \{Y_t, t \in T\}$ por $Y_t(\omega) = X_{t+}(\omega)$ se $\omega \notin N$ e 0 senão. Note que como $\{\mathcal{F}_t\}$ é completa, temos que Y_t é $\mathcal{F}_t$ mensurável para todo $t \in T$ e por (a) obtemos que Y é um submartingale (ou um martingale se X for um martingale). Além disto, por construção, as trajetórias de Y são càdlàg. Agora, devemos mostrar que Y é uma modificação de X se e somente se $t \to E(X_t)$ é contínua à direita. Temos que

$$X_t \leq E(X_{t+} | \mathcal{F}_t) = E(Y_t | \mathcal{F}_{t+}) = Y_t, \tag{5.2}$$

sendo que a primeira igualdade segue da continuidade à direita de $\mathcal{F}_t$ e, a segunda, porque Y_t é $\mathcal{F}_{t+}$-mensurável. Logo, $X_t \leq Y_t$.

Suponha, agora, $t_n \downarrow t$. Se $\Lambda \in \mathcal{F}_t$, temos $\int_\Lambda X_t dP \leq \int_\Lambda X_{t_n} dP$. Faça $n \to \infty$ para obter $\int_\Lambda X_t dP \leq \int_\Lambda Y_t dP$. Como $X_t \leq Y_t$, obtemos $X_t = Y_t$ q.c se e somente se $E(X_t) = E(X_{t+}) = E(\lim_{s \downarrow t} X_t) = \lim_{s \downarrow t} E(X_t)$. $\square$

Passemos, agora, a considerar o Teorema da Amostragem Opcional no caso de martingales com tempo contínuo. Antes, definamos tempo de parada nesse caso.

Definição 5.12. *Seja $\{\mathcal{F}_t, t \in T\}$ uma família crescente de σ-álgebras. Um tempo de parada τ relativamente a $\{\mathcal{F}_t\}$ é uma v.a com valores em $[0, \infty]$ tal que $\{\tau \leq t\} \in \mathcal{F}_t$ para todo $t \in T$.*

Definamos $\mathcal{F}_\tau = \{A \in \bigvee_{t \in T} \mathcal{F}_t : \forall t \in T,\ A \cap \{\tau \leq t\} \in \mathcal{F}_t\}$.

Algumas propriedades:

[1] $\mathcal{F}_\tau$ é uma σ-álgebra.

[2] Os conjuntos $\{\tau < t\}, \{\tau = t\}, \{\tau \leq t\}$ todos pertencem a $\mathcal{F}_t$.

[3] Se τ, ν são tempos de parada, também o serão $\tau \wedge \nu$ e $\tau \vee \nu$.

[4] Se $\tau_1 \leq \tau_2$, então $\mathcal{F}_{\tau_1} \subset \mathcal{F}_{\tau_2}$.

[5] τ é $\mathcal{F}_\tau$ - mensurável.

Proposição 5.1. *Seja $X = \{X_t, t \geq 0\}$ um p.e com trajetórias contínuas à direita e X_t é $\mathcal{F}_t$-mensurável, com $\{\mathcal{F}_t, t \geq 0\}$ uma família crescente de σ-álgebras. Seja τ um tempo de parada finito. Então, X_τ é $\mathcal{F}_\tau$-mensurável.*

Prova: Considere $X(t, \omega) : T \times \Omega \to \mathbb{R}$, com $T = [0, \infty)$. Então, a restrição dessa aplicação a $[0, t] \times \Omega$ é mensurável, considerando $\mathcal{B}_t$ a σ-álgebra de Borel sobre $[0, t]$ e $\mathcal{B}_t \times \mathcal{F}_t$ a σ-álgebra sobre $[0, t] \times \Omega$. A prova é a mesma daquela que mostrou que um processo contínuo à direita é mensurável.

Considere as aplicações

$$\omega \to \tau(\omega) \to (\omega, \tau(\omega)) \to X_{\tau(\omega)}(\omega).$$

Para mostrar que X_τ é $\mathcal{F}_\tau$-mensurável, basta mostrar que, se A for qualquer conjunto de Borel, $\{\omega : X_{\tau(\omega)}(\omega) \in A\} \cap \{\tau \leq t\} \in \mathcal{F}_t$. Note que a aplicação $h : \omega \to (\omega, \tau(\omega))$ é uma aplicação mensurável, por definição de tempo de parada e a aplicação $g : (t, \omega) \to \mathbb{R}$, definida por $g(t, \omega) = X_t(\omega)$ é mensurável, pela observação feita no início da prova. Como $X_{\tau(\omega)}(\omega) = (g \circ h)(\omega)$, obtemos o resultado. $\square$.

Vejamos, agora, o teorema da amostragem opcional de Doob para o caso de submartingales contínuos.

Teorema 5.8. (TAO) *Seja $X = \{X_t, \mathcal{F}_t, t \in T\}$ um submartingale com trajetórias contínuas à direita. Sejam $S \leq R$ tempos de parada finitos. Então. se*

(a) *$R \leq t_0$ q.c, para algum $t_0 \in T$, ou*

(b) *$\{X_t, t \in T\}$ é uniformemente integrável,* *teremos que*

$$E(X_R|\mathcal{F}_S) \geq X_S, \tag{5.3}$$

ou seja, $\{X_S, X_R\}$ *é um submartingale.*

Prova: Suponha $T = [a, b]$ e defina para todo $n \geq 1$, $S^{(n)}$ como:

$$\begin{aligned} S^{(n)}(\omega) &= \frac{k+1}{2^n}, \text{ se } \frac{k}{2^n} < S(\omega) \leq \frac{k+1}{2^n}, k \geq 1 \\ &= \frac{1}{2^n}, \text{ se } 0 \leq S(\omega) \leq \frac{1}{2^n}. \end{aligned}$$

Segue-se que $S(\omega) \leq S^{(n)}(\omega)$, para todo ω. Considere o submartingale $\{X_{k/2^n},\ k \geq 0\}$ e as σ-álgebras $\{\mathcal{F}_{k/2^n}, k \geq 0\}$. Então, $S^{(n)}$ é um tempo de parada para esse submartingale, porque

$$\left\{S^{(n)} = \frac{k+1}{2^n}\right\} = \left\{\frac{k}{2^n} < S \leq \frac{k+1}{2^n}\right\} \in \mathcal{F}_{(k+1)/2^n},$$

pois S é um tempo de parada. Defina $R^{(n)}$ em função de $R(\omega)$ de modo similar ao definido para $S^{(n)}$. Então, $R^{(n)} \geq R$ e $R^{(n)}$ é um tempo de parada para o submartingale $\{X_{k/2^n}, k \geq 0\}$. Se a suposição (a) for válida, então $R^{(n)}$ e $S^{(n)}$ têm somente um número finito de valores (no máximo $2^n t_0$ valores). Logo, pelo TAO no caso discreto,

$$E(X_{R^{(n)}} \mid \mathcal{F}_{S^{(n)}}) \geq X_{S^{(n)}}. \tag{5.4}$$

Se (b) valer, (5.4) é também verdadeira, pois o submartingale $\{X_{k/2^n}, k \geq 0\}$ é u.i. e então podemos aplicar o TAO estendido (veja o Problema 18 do Capítulo 4).

Tome $\Lambda \in \mathcal{F}_S$. Como $S^{(n)} \geq S$, $\Lambda \in \mathcal{F}_{S^{(n)}}$, porque $\mathcal{F}_S \subset \mathcal{F}_{S^{(n)}}$. Logo, (5.4) implica que

$$\int_\Lambda X_{R^{(n)}} dP \geq \int_\Lambda X_{S^{(n)}} dP. \tag{5.5}$$

Faça $n \to \infty$ em (5.5). Como $R^{(n)} \downarrow R$ e $S^{(n)} \downarrow S$, e como $\{X_t\}$ tem todas as trajetórias contínuas à direita, segue-se que $\lim_{n\to\infty} X_{R^{(n)}} = X_R$ e $\lim_{n\to\infty} X_{S^{(n)}} = X_S$, ambos q.c. Esses limites também valem em norma L_1. De fato, note que $S^{(n)} \geq S^{(n+1)}$, para todo n, logo sob a condição (a) ou (b), $E(X_{S^{(n)}}|\mathcal{F}_{S^{(n+1)}}) \geq X_{S^{(n+1)}}$, ou seja, $\{X_{S^{(n)}}, n \geq 1\}$ é um submartingale reverso, que converge em L_1 e q.c.

Logo, podemos tomar o limite em (5.5) para obter $\int_\Lambda X_R dP \geq \int_\Lambda X_S dP$, que vale para todo $\Lambda \in \mathcal{F}_S$. Ou seja, $E(X_R|\mathcal{F}_S) \geq X_S$. $\square$

5.3 Processos com incrementos independentes

Nesta seção estudamos processos importantes, como o Processo de Poisson e o Movimento Brownianno. O primeiro tem aplicações, por exemplo, no modelo de risco adotado em seguros e, o segundo, em modelos de opções financeiras, como a fórmula de Black-Scholes (ver Capítulo 13).

Definição 5.13. *Um p.e $X = \{X_t, t \in T\}$ tem incrementos independentes se, para toda sequência $t_1 < t_2 < \cdots < t_n$ de T, tivermos que $X_{t_2} - X_{t_1}, X_{t_3} - X_{t_2}, \ldots, X_{t_n} - X_{t_{n-1}}$ são v.a's independentes.*

Exemplo 5.8. Seja $c > 0$. Construa um p.e com incrementos independentes como segue:

(a) $X_0 = 0$;

(b) Se $s < t$, suponha $X_t - X_s \sim Poi(c(t-s))$.

Sabemos a distribuição de $X_{t_2} - X_{t_1}, \ldots, X_{t_n} - X_{t_{n-1}}$; suponha que o processo tenha incrementos independentes, portanto teremos a distribuição conjunta de $X_{t_1}, \ldots, X_{t_n}$. Finalmente, teremos que verificar sua consistência, usando o teorema da extensão de Kolmogorov.

Um p.e separável com essa distribuição é chamado um *Processo de Poisson com parâmetro c*. As trajetórias de um processo de Poisson são funções em patamar, não decrescentes, constantes, exceto por saltos de tamanho unitário.

Exemplo 5.9. Construa um processo $\{X_t, t \geq 0\}$ com incrementos independentes como segue:

(a) $X_0 = 0$;

(b) Se $s < t$, suponha $X_t - X_s \sim N(0, t-s)$.

Um p.e. separável com essa distribuição é chamado *Movimento Browniano* (MB) ou *processo de Wiener*. Pode-se provar que quase todas as trajetórias do MB são contínuas, mas não deriváveis q.c. Para detalhes, veja Wiersema (2008), Evans (2013), Dvoretzky et al. (1950, 1954) e a Seção 9.2.

Definição 5.14. *Um processo $X = \{X_t, t \geq 0\}$ com incrementos independentes é estacionário se a distribuição de $X_t - X_s$ somente depender de $t - s,\ t > s$ (e escrevemos $X_t - X_s \sim X_{t-s}$). Também dizemos que o processo tem incrementos estacionários.*

Os processos de Poisson e MB têm incrementos independentes e estacionários.

Proposição 5.2. *Se $X = \{X_t, t \geq 0\}$ é um Movimento Browniano, então X é um martingale.*

Prova: Tomemos $\mathcal{F}_t = \mathcal{F}\{X_s, s \leq t\}$. Devemos mostrar que $E(X_t|\mathcal{F}_s) = X_s$. Temos

$$E(X_t|\mathcal{F}_s) = E(X_t - X_s + X_s|\mathcal{F}_s) = E(X_t - X_s|\mathcal{F}_s) + X_s.$$

Mas, $X_t - X_s$ é independente de $X_u - X_0$, $u \leq s$, logo $E(X_t - X_s|\mathcal{F}_s) + X_s = E(X_t - X_s) + X_s = E(X_{t-s}) + X_s = X_s$, pela estacionariedade e o fato que $X_{t-s} \sim N(0, t-s)$. $\square$.

Proposição 5.3. *Seja $P = \{P_t, t \geq 0\}$ um processo de Poisson com parâmetro $c > 0$. Então, $Y_t = P_t - ct$ é um martingale.*

Prova: Tome $\mathcal{F}_t = \mathcal{F}\{P_s, s \leq t\}$. Então, $E(Y_t|\mathcal{F}_s) = E(P_t - ct|\mathcal{F}_s) = E(P_t - P_s|\mathcal{F}_s) - ct + P_s$, e usando o fato que o processo tem incrementos independentes e estacionários, obtemos que $E(Y_t|\mathcal{F}_s) = E(P_{t-s}) - ct + P_s = c(t-s) - ct + P_s = P_s - cs = Y_s$. $\square$

Observação: Se $X = \{X_t, t \geq 0\}$ é um processo com incrementos independentes e estacionários, e se X_t é integrável, para cada t, então $Y_t = X_t - tE(X_1)$ é um martingale.

Definição 5.15. *Um processo de Lévy é um processo càdlàg com incrementos independentes e estacionários e $P(X_0 = 0) = 1$.*

Os processos de Lévy têm merecido uma grande atenção recentemente, notadamente por suas aplicações em finanças.

A seguir, consideramos um resultado que fornece uma propriedade forte de Markov para processos com incrementos independentes e estacionários.

Teorema 5.9. (Hunt) *Seja $X = \{X_t, t \geq 0\}$ um p.e com incrementos independentes e estacionários e seja $\mathcal{F}_t = \mathcal{F}\{X_s, s \leq t\}$. Suponha que $X_0 = 0$ e que as suas trajetórias sejam contínuas à direita. Seja T um tempo de parada finito. Defina o processo $Y = \{Y_t, t \geq 0\}$ por $Y_t = X_{T+t} - X_T$. Então, Y tem incrementos independentes e estacionários, Y_t tem a mesma distribuição que X_t e Y é independente de $\mathcal{F}_T$.*

Prova: Seja Λ qualquer conjunto em $\mathcal{F}_T$ e $t_1 < t_2 < \cdots < t_n$. Devemos mostrar que

$$P\{\Lambda, Y_{t_1} \in A_1, \ldots, Y_{t_n} \in A_n\} = P(\Lambda)P\{X_{t_1} \in A_1, \ldots, X_{t_n} \in A_n\},$$

onde $A_1, \ldots, A_n$ são conjuntos de Borel arbitrários. Vamos considerar somente o caso $n = 1$ (o caso $n > 1$ é similar). De modo que temos que provar que $P\{\Lambda, Y_t \in A\} = P(\Lambda)P\{X_t \in A\}$.

Ou ainda, temos que provar que

$$E(I_\Lambda f(Y_t)) = P(\Lambda)E(f(X_t)), \tag{5.6}$$

onde $f = I_A$. Para provar (5.6), provamos que essa vale para qualquer função contínua, limitada f. Defina

$$\begin{aligned} T^{(n)}(\omega) &= \frac{k+1}{2^n}, \text{ se } \frac{k}{2^n} < T(\omega) \leq \frac{k+1}{2^n}, k \geq 1, \\ &= \frac{1}{2^n}, \text{ se } 0 \leq T(\omega) \leq \frac{1}{2^n}. \end{aligned}$$

Segue-se que $T^{(n)}$ é tempo de parada e $T^{(n)} \downarrow T$. Agora,

$$f(Y_t) = f(X_{T+t} - X_T) = \lim_{n\to\infty} f(X_{T^{(n)}+t} - X_{T^{(n)}}),$$

pela continuidade à direita de X_t e continuidade de f. Agora provamos que

$$E\left[I_\Lambda f(X_{T^{(n)}+t} - X_{T^{(n)}})\right] = P(\Lambda)E(f(X_t)).$$

Pelo limite acima, o teorema seguirá desse resultado. Agora,

$$\begin{aligned} E\left[I_\Lambda f\left(X_{T^{(n)}+t} - X_{T^{(n)}}\right)\right] &= \sum_{k\geq 1} \int_{\{T^{(n)}=k/2^n\}} I_\Lambda f\left(X_{k/2^n+t} - X_{k/2^n}\right) dP \\ &= \sum_{k\geq 1} \int I_{\Lambda\cap\{T^{(n)}=k/2^n\}} f\left(X_{k/2^n+t} - X_{k/2^n}\right) dP. \end{aligned} \tag{5.7}$$

Mas $\Lambda \in \mathcal{F}_T \subset \mathcal{F}_{T^{(n)}}$, logo $\Lambda \cap \{T^{(n)} = k/2^n\} \in \mathcal{F}_{k/2^n}$, por definição de tempo de parada. Segue-se que $\Lambda \cap \{T^{(n)} = k/2^n\}$ é independente de $X_{k/2^n+t} - X_{k/2^n}$, devido a incrementos independentes. Logo (5.7) torna-se

$$E\Big(f(X_t) \sum_{k\geq 1} P(\Lambda \cap \{T^{(n)} = k/2^n\})\Big) = E(f(X_t))P(\Lambda),$$

usando a estacionariedade. $\square$

Veja Mörters e Peres (2010), Hunt (1956) e Dynkin (1957) para detalhes sobre a propriedade forte de Markov.

Problemas

1. Prove que as definições 5.4, 5.5 e 5.6 são equivalentes.
2. Prove a afirmação do Exemplo 5.2.
3. Prove as Consequências (2)-(5).

4. Prove a afirmação do Exemplo 5.5.

5. Prove (b) do Teorema 5.2.

6. Seja X um processo estocástico com parâmetro contínuo e seja T um intervalo. Suponha que X seja contínuo em probabilidade para cada $t \in T$. Prove que existe um p.e $\tilde{X}$, tal que $\tilde{X}$ seja equivalente a X, separável e mensurável.

7. Prove (b) do Teorema 5.3.

8. Prove o conteúdo da Observação após a Proposição 5.3.

9. Mostre que o processo de Poisson e o Movimento Browniano são processos de Lévy.

10. Mostre que, se $Z \sim N(0,1)$, então para λ real, $E(e^{\lambda Z}) = e^{\lambda^2/2}$.

11. (Movimento Browniano Geométrico). Black and Scholes (1973) e Merton (1973) sugeriram o Movimento Browniano Geométrico para descrever preços num mercado especulativo. Tal processo é dado por

$$X_t = e^{\mu t + \sigma W_t}, \quad t \geq 0,$$

    onde W_t é o Movimento Browniano. Segue-se que $\log X_t$ segue um Movimento Browniano, com *drift* μ real e *volatilidade* $\sigma > 0$. Use o problema anterior para calcular a média e função de autocovariância de X_t. Mostre que esse processo não é gaussiano.

12. (Ponte Browniana) Considere o processo estocástico dado por

$$X_t = W_t - tW_{t-1}, \quad 0 \leq t \leq 1,$$

    onde W_t é Movimento Browniano no intervalo $[0,1]$. Segue-se que $X_0 = X_1 = 0$. Mostre que X_t é um processo gaussiano, com média zero e função de autocovariância dada por $\gamma(t,s) = \min\{t,s\} - ts,\ t,s \in [0,1]$.

13. Encontre exemplos de:

    (a) um processo separável que não seja mensurável;

    (b) um processo mensurável que não seja separável.

14. Prove que, embora tenha trajetórias descontínuas, um processo de Poisson é contínuo em probabilidade.

Capítulo 6

Convergência Fraca

Neste capítulo introduzimos o importante conceito de convergência fraca para uma família de medidas de probabilidade e, depois, para variáveis aleatórias e processos estocásticos. As referências básicas aqui são Billingsley (1999) e Parthasaraty (2005).

6.1 Introdução

Denotaremos por S um espaço métrico e $\mathcal{S}$ a σ-álgebra de conjuntos de Borel sobre S, que coincide com a σ-álgebra gerada pelos conjuntos abertos de S. Usaremos a notação $(S, \mathcal{S})$. Alguns espaços métricos que podemos considerar são:

(a) $(\mathbb{R}, \mathcal{B})$, com métrica $d(x, y) = |x - y|$;

(b) $(\mathbb{R}^k, \mathcal{B}^k)$, com métrica $d(x, y)$ associada a qualquer norma sobre $\mathbb{R}^k$;

(c) $(\mathbb{R}^\infty, \mathcal{B}^\infty)$, sendo $\mathbb{R}^\infty$ o espaço de todas as sequências $(x_1, x_2, \ldots)$ de números reais, e a métrica associada é

$$d(x, y) = \sum_{n \geq 1} 2^{-n} \frac{|x_n - y_n|}{1 + |x_n - y_n|}.$$

(d) $(C([0,1]), \mathcal{C})$, onde $C([0,1])$ é o espaço de todas as funções contínuas sobre $[0,1]$ e $\mathcal{C}$ é a σ-álgebra de Borel sobre este espaço. Se $x = \{x(t),\ t \in [0,1]\}$ e $y = \{y(t),\ t \in [0,1]\}$, então a métrica é definida por

$$d(x, y) = \sup_{0 \leq t \leq 1} |x(t) - y(t)|.$$

Convergência em (c) é ponto a ponto e, em (d), uniforme. Chamemos de $C_b(S)$ o conjuntos de todas as funções contínuas e limitadas sobre S com valores reais.

Definição 6.1. *Seja $\{P_n, n \geq 1\}$ uma família de medidas de probabilidade sobre $(S, \mathcal{S})$ e P uma medida sobre $(S, \mathcal{S})$. Dizemos que P_n converge fracamente para P se*

$$\lim_{n\to\infty} \int_S f(x)dP_n(x) = \int_S f(x)dP(x), \tag{6.1}$$

para toda $f \in C_b(S)$.

Usaremos a notação $P_n \Rightarrow P$. Uma notação padrão bastante usada é a seguinte: se P é uma medida de probabilidade, defina Pf como o valor esperado de f sob P:

$$Pf = \int f(x)dP(x).$$

Então, (6.1) pode ser escrita

$$P_n f \to Pf, \quad \forall f \in C_b(S).$$

Essa notação compacta enfatiza a interpretação de P como um funcional linear.

A medida P em (6.1) é necessariamente uma medida de probabilidade. Tome $f = 1$ e obtemos $\int_S dP_n \to \int_S dP$, ou seja $1 = P_n(S) \to P(S)$, logo $P(S) = 1$.

Exemplo 6.1. (a) Tome $S = \mathbb{R}$ e seja $\{x_n\}$ uma sequência de pontos de $\mathbb{R}$. Considere P_n como massa unitária em x_n. Então, $x_n \to x$ se, e somente se, $P_n \Rightarrow P$, onde P coloca massa unitária em x.

(b) Seja $S = [0, 1]$ e suponha que P_n coloque massa $1/n$ nos pontos $1/n, 2/n, , \ldots, 1$. Então, $P_n \Rightarrow P$, onde P é a medida de Lebesgue sobre $[0, 1]$.

Teorema 6.1. *Sejam P, Q medidas de probabilidade sobre $(S, \mathcal{S})$. Se $\int_S f dP = \int_S f dQ$, para toda $f \in C_b(S)$, então $P = Q$.*

Prova: Seja K um conjunto fechado em S e defina $f_n(x) = e^{-nd(K,x)}$. Então, como K é fechado, temos que $f_n(x) \to I_K(x)$. Também, como $\int_S f dP = \int_S f dQ$, obtemos pelo TCD que $\int_K f dP = \int_K f dQ$, logo $P(K) = Q(K)$ se K for fechado. Também, $P = Q$ sobre todos os conjuntos abertos. Vamos provar que: Para todo $\varepsilon > 0$ e todo conjunto B em $\mathcal{S}$, podemos encontrar um conjunto fechado A e um conjunto aberto C, tal que $A \subset B \subset C$ e $P(C - A) < \varepsilon$. Se isso for verdade, o resultado segue, pois $P = Q$ para conjuntos abertos e fechados.

Seja $\mathcal{H}$ a classe de todos os conjuntos $B \in \mathcal{S}$ tais que, para todo $\varepsilon > 0$, existe A fechado e C aberto, tais que $A \subset B \subset C$ e $P(C - A) < \varepsilon$. Essa classe contém conjuntos fechados. Seja B fechado. Escolha $A = B$. Defina $C_\delta = \{x \in S : d(B, x) < \delta\}$, aberto. Vemos que $C_\delta \downarrow B$, para $\delta \downarrow 0$, pois B é fechado. Tome δ tão pequeno de modo que $P(C_\delta - B) < \varepsilon$. Segue-se que $A \subset B \subset C_\delta$.

Também é fácil ver que $\mathcal{H}$ é fechada sob complementação. Finalmente, é fechada sob reuniões enumeráveis. De fato, sejam $B_1, B_2, \ldots$ conjuntos de $\mathcal{H}$ e $\varepsilon > 0$. Existe uma família $\{A_i\}$ de conjuntos fechados, outra $\{C_i\}$ de conjuntos abertos, tais que $A_i \subset B_i \subset C_i$, tais que $P(C_i - A_i) < \varepsilon/2^{i+1}$, para todo $i \geq 1$. Defina $C = \cup_i C_i,\ A = \cup_{i=1}^N A_i$, onde N é tão grande que $P(\cup_i A_i - \cup_{i=1}^N A_i) < \varepsilon/2$; mais ainda, C é aberto e A é fechado. Então, $A \subset B \subset C$ e $P(C - A) < \varepsilon$. $\square$

Corolário 6.1. *Se $P_n \Rightarrow P$, e $P_n \Rightarrow Q$, então $P = Q$.*

Definição 6.2. *Seja A um conjunto em $(S, \mathcal{S})$. A fronteira de A, denotada ∂A, é o conjunto $\overline{A} - \overset{\circ}{A}$, onde $\overline{A}$ é o fecho de A e $\overset{\circ}{A}$ é o interior de A. Um conjunto A é um conjunto P-contínuo se $P(\partial A) = 0$.*

Exemplo 6.2 (a) Seja $S = \mathbb{R}$ e $A = (a, b]$. Então, $\partial A = \{a\} \cup \{b\}$.

(b) Se $S = \mathbb{R}$ e A é o conjunto dos racionais, $\partial A = \mathbb{R}$.

c) Se $S = \mathbb{R}$ e $A = (a, b]$, então A é um conjunto P-contínuo se P não coloca massa sobre $\{a\} \cup \{b\}$.

O teorema a seguir é chamado "Portmanteau", pois fornece condições úteis que são equivalentes à definição de convergência fraca.

Teorema 6.2. *As seguintes afirmações são equivalentes:*

(a) $P_n \Rightarrow P$;

(b) $\limsup_n P_n(K) \leq P(K)$, *para todo K fechado*;

(c) $\liminf_n P_n(A) \geq P(A)$, *para todo A aberto*;

(d) $\lim_n P_n(A) = P(A)$, *para todo conjunto P-contínuo A.*

Prova: (a) $\Rightarrow$ (b): Suponha que K seja fechado; tome $\varepsilon > 0$ e considere $A_\delta = \{x \in S : d(K, x) < \delta\}$, que é aberto. Tome δ_0 tão pequeno de tal sorte que $P(A_{\delta_0}) \leq P(K) + \varepsilon$, pois $A_\delta \downarrow K$.

Seja f uma função contínua que tome o valor 1 sobre K , o valor 0 fora de A_{δ_0} e $0 \leq f \leq 1$. Para tanto, defina $g(x) = 1$, para $x \leq 0$, $g(x) = 1 - x$, se $0 \leq x \leq 1$ e $g(x) = 0$, se $x \geq 1$; g assim definida é contínua. Agora defina $f(x) = g(d(x, K)/\delta_0)$. Agora,

$$P_n(K) = \int_K f dP_n \leq \int_S f dP_n \ \rightarrow \ \int_S f dP = \int_{A_{\delta_0}} f dP \leq \int_{A_{\delta_0}} 1 \cdot dP = P(A_{\delta_0}) \leq P(K) + \varepsilon,$$

a penúltima igualdade porque f é zero fora de A_{δ_0}. Segue-se que $\limsup_n P_n(K) \leq P(K) + \varepsilon$, mas ε é arbitrário, logo o resultado segue.

(b) $\Leftrightarrow$(c): tome complementos.

(c) $\Rightarrow$ (d): Temos que (c) vale e também (b). Suponha que $P(\overline{A}- \overset{\circ}{A}) = 0$, pois A é um conjunto P-contínuo. Então,

$$\limsup_{n\to\infty} P(A_n)(A) \leq \limsup_{n\to\infty} P_n(\overline{A}) \leq P(\overline{A}) = P(A),$$

sendo que a última desigualdade vale por (b). De modo análogo,

$$\liminf_{n\to\infty} P_n(A) \geq \liminf_{n\to\infty} P_n(\overset{\circ}{A}) \geq P(\overset{\circ}{A}) = P(A),$$

usando (c). Logo, $\lim_{n\to\infty} P_n(A) = P(A)$.

(d) $\Rightarrow$ (a): Seja $f \in C_b(S)$. Devemos provar que $\lim_{n\to\infty} \int f dP_n = \int f dP$.

Defina P_f uma medida na reta por: se B é um conjunto de Borel, $P_f(B) = P\{x \in S : f(x) \in B\}$. Note que, como f é limitada, P_f é concentrada sobre um intervalo limitado, $[a, b)$ digamos. Também, existe no máximo um conjunto enumerável de pontos na reta, sobre os quais P_f coloca massa positiva. Logo, podemos escolher a e b tais que para todo $\varepsilon > 0$, podemos encontrar $a = t_1 < t_2 < \ldots < t_N = b$ tais que $t_{i+1} - t_i < \varepsilon$ e $P_f\{t_i\} = 0$. Seja $A_i = \{x \in S : t_{i-1} \leq f(x) < t_i\}$ para $2 \leq i \leq N$. Então, $\partial A_i \subset \{x : f(x) = t_i \text{ ou } f(x) = t_{i-1}\}$.

Portanto, $P(\partial A_i) \leq P_f\{t_i\} + P_f\{t_{i-1}\} = 0$, logo cada A_i é um conjunto P-contínuo.

Seja $\hat{f}$ a função

$$\hat{f} = \sum_{i=2}^{n} t_{i-1} I_{A_i}. \tag{6.2}$$

Segue-se que $\hat{f}$ é uma função simples e $|\hat{f} - f| < \varepsilon$. Então,

$$\Big|\int f dP_n - \int f dP\Big| \leq \int |f - \hat{f}| dP_n + \Big|\int \hat{f} dP_n - \int \hat{f} dP\Big| + \int |\hat{f} - f| dP$$

$$\leq 2\varepsilon + \Big|\int \hat{f} dP_n - \int \hat{f} dP\Big| \leq 2\varepsilon + \sum_{i=1}^{N} |t_i||P_n(A_i) - P(A_i)|.$$

Para $n \to \infty$ e por (d), $P_n(A_i) - P(A_i) \to 0$, pois A_i é um conjunto P-contínuo. Então,

$$\limsup_{n} \Big|\int f dP_n - \int f dP\Big| \leq 2\varepsilon.$$

Como ε é arbitrario, obtemos o resultado desejado . $\square$

Exemplo 6.3. (i) Suponha que $S = \mathbb{R}$ e que P_n coloque massa unitária em $\{1/n\}$ e P coloque massa unitária em zero. Então, $P_n \Rightarrow P$. Considere $A = [-1, 0]$. Então, $P_n(A) = 0$, mas $P(A) = 1$. Segue-se que $P_n \Rightarrow P$, mas $P_n(A)$ não converge para $P(A)$. Isso ocorre porque $[-1, 0]$ não é um conjunto P-contínuo.

(ii) Suponha $S = [0, 1]$, P_n coloca massa $1/n$ sobre $1/n, 2/n, \ldots, 1$ e P é medida de Lebesgue. Então, $P_n \Rightarrow P$. Considere $A = \mathbb{Q} \cap [0, 1]$, onde $\mathbb{Q}$ é o conjunto dos racionais. Então, $P_n(A) = 1$, mas $P(A) = 0$. Aqui temos $\partial A = [0, 1]$ e $P(\partial A) = P([0, 1]) = 1$.

Corolário 6.2. *Sejam P_n, P medidas de probabilidade sobre $(S, \mathcal{S})$ e $\mathcal{H}$ uma coleção de subconjuntos de S tal que:*

(i) $\mathcal{H}$ é fechado sob intersecções finitas;

(ii) todo conjunto aberto é uma reunião enumerável de conjuntos de $\mathcal{H}$.

Então, se $P_n(A) \to P(A)$, para todo conjunto $A \in \mathcal{H}$, segue-se que $P_n \Rightarrow P$.

Prova: Sejam A, B dois conjuntos em $\mathcal{H}$. Então, $P_n(A \cup B) = P_n(A) + P_n(B) - P_n(A \cap B) \to P(A \cup B)$, para $n \to \infty$, por hipótese. Por indução, $\lim_n P_n(\cup_{i=1}^N A_i) = P(\cup_{i=1}^N A_i)$, para todo N.

Seja B um conjunto aberto. Então, existem $A_1, A_2, \ldots \in \mathcal{H}$, tais que $B = \cup_{i=1}^\infty A_i$. Seja $\varepsilon > 0$ e tome N tão grande de modo que $P(B) - \varepsilon \leq P(\cup_{i=1}^N A_i)$. Então,

$$P(B) - \varepsilon \leq P(\cup_{i=1}^N A_i) = \lim_n P_n(\cup_{i=1}^N A_i) \leq \liminf_n P_n(\cup_i^\infty A_i) = \liminf_n P_n(B).$$

Como $\varepsilon > 0$ é arbitrário, $P_n \Rightarrow P$, pelo critério (c) do teorema anterior. $\square$

Um caso particular importante

Tomemos $S = \mathbb{R}$ e sejam P_n e P probabilidades na reta real. Seja D qualquer conjunto denso de pontos de $\mathbb{R}$. Defina $\mathcal{H}$ como a coleção de todos os intervalos da forma $(a, b]$, com $a, b \in D$. Então, $\mathcal{H}$ satisfaz as condições (i) e (ii) do corolário. Logo, se $P_n((a, b]) \to P((a, b])$, para todos os intervalos $(a, b] \in \mathcal{H}$, segue-se que $P_n \Rightarrow P$.

Se $S = \mathbb{R}$ e P é uma medida de probabilidade sobre $\mathbb{R}$, lembremos que a f.d para P é a função $F(x) = P\{(-\infty, x]\}$, para todo x real. Também, F é não decrescente, contínua à direita, $F(-\infty) = 0$ e $F(+\infty) = 1$.

Teorema 6.3. *Sejam P_n, P probabilidades sobre $\mathbb{R}$, com f.d's F_n, F, respectivamente.*

(a) Se $P_n \Rightarrow P$, então $\lim_{n\to\infty} F_n(x) = F(x)$, para cada ponto x onde F é contínua;

(b) se $F_n(x) \to F(x)$, para todo x em um conjunto denso em $\mathbb{R}$, então $P_n \Rightarrow P$.

Prova: (a) Se F for contínua em x, então o conjunto $(-\infty, x]$ é P-contínuo. Portanto, pela parte (d) do teorema anterior, $P_n((-\infty, x]) \to P((-\infty, x])$, ou seja $F_n(x) \to F(x)$.

(b) Se $F_n(x) \to F(x)$, para todo x num subconjunto denso de $\mathbb{R}$, então $F_n(b) - F_n(a) \to F(b) - F(a)$, para quaisquer a, b nesse conjunto. Logo, $P_n((a, b]) \to P((a, b])$. Logo, (b) segue pelo caso particular discutido acima. $\square$

6.2 Convergência fraca para elementos aleatórios

Lembremos que a distribuição de um elemento aleatório X com valores em S é a probabilidade P_X sobre S definida por $P_X(B) = P\{\omega : X(\omega) \in B\}$, para $B \in \mathcal{S}$. Continuamos supondo que S é um espaço métrico e $\mathcal{S}$ é a sua σ-álgebra de Borel.

Definição 6.3. *Dizemos que X_n converge para X em distribuição se $P_{X_n} \Rightarrow P_X$, ou seja, para toda $f \in C_b(S)$, $E(f(X_n)) \to E(f(X))$.*

Com todo rigor, deveriamos escrever $E_n(f(X_n)) \to E(f(X))$ pois os elementos aleatórios X_n (e X) não precisam ser definidos no mesmo espaço de probabilidade. No entanto, mas para não sobrecarregar as notações não colocaremos explicitamente esta dependência.

Vamos usar a notação $X_n \stackrel{\mathcal{D}}{\to} X$ ou $X_n \stackrel{\mathcal{D}}{\to} P$. Podemos dizer também que X_n converge para X *em lei* e escrevemos $\mathcal{L}(X_n) \to \mathcal{L}(X)$.

Definição 6.4. *Sejam X_n, X elementos aleatórios. Dizemos que A é um conjunto X-contínuo se $P_X(\partial A) = 0$.*

Se X estiver definido sobre $(\Omega, \mathcal{F}, P)$, então A é um conjunto X-contínuo se $P\{\omega : X(\omega) \in \partial A\} = 0$.

Os dois teoremas a seguir podem se provados de modo análogo ao que foi feito com os Teoremas 6.2 e 6.3.

Teorema 6.4. *As afirmações seguintes são equivalentes:*

(a) $X_n \stackrel{\mathcal{D}}{\to} X$;

(b) $\limsup_n P_{X_n}(A) \le P_X(A)$, *$A$ fechado;*

(c) $\liminf_n P_{X_n}(A) \ge P_X(A)$, *$A$ aberto;*

(d) $\lim_n P_{X_n}(A) = P_X(A)$, *para todo conjunto A X-contínuo.*

Teorema 6.5. *Seja $\{X_n, n \ge 1\}$ uma sequência de v.a's com f.d's F_n e X uma v.a com f.d F. Então, temos:*

(a) Se $X_n \xrightarrow{\mathcal{D}} X$, então $\lim_{n\to\infty} F_n(x) = F(x)$, para todos os pontos de continuidade de F;

(b) Se $\lim_{n\to\infty} F_n(x) = F(x)$, para x num conjunto denso de $\mathbb{R}$, então $X_n \xrightarrow{\mathcal{D}} X$.

Para ver a necessidade de F ser uma f.d em (b), basta tomar $F_n(x) = 0$, se $x < n$ e $F(x) = 1$, se $x \geq n$. As $\{F_n\}$ são f.d's e F_n converge, para cada x real, e o limite é zero.

Agora, consideremos as seguintes questões:

[1] Se X_n, X são v.a's e $X_n \xrightarrow{\mathcal{D}} X$, quando $h(X_n) \xrightarrow{\mathcal{D}} h(X)$?

[2] Se $P_n \Rightarrow P$, sabemos que $\int_S f(x)dP_n(x) \to \int_S f(x)dP$, para toda $f \in C_b(S)$. Para quais outras funções f essa implicação vale?

Certamente, [2] não vale para qualquer f. Por exemplo, tome $S = \mathbb{R}$, P_n colocando massa unitária em $1/n$. Então, $P_n \Rightarrow P$, onde P coloca massa unitária no zero. Tome $f(x) = I_{(0,\infty)}(x)$. Então, $\int_S f dP_n = 1$ e $\int_S f dP = 0$.

[3] Sejam P_n, P medidas de probabilidade sobre $(S, \mathcal{S})$. Seja h uma função mensurável de $(S, \mathcal{S})$ em $(S', \mathcal{S}')$, outro espaço métrico. Defina Ph^{-1}, uma medida sobre $(S', \mathcal{S}')$, por meio de:

$$Ph^{-1}(B) = P\{h^{-1}(B)\}, \quad B \in \mathcal{S}'.$$

Suponha que $P_n \Rightarrow P$. Podemos afirmar que $P_n h^{-1} \Rightarrow Ph^{-1}$?

Teorema 6.6. *Sejam P_n, P medidas sobre $(S, \mathcal{S})$ e seja h uma função mensurável de $(S, \mathcal{S})$ em $(S', \mathcal{S}')$. Seja D_h o conjunto dos pontos de S para os quais h é descontínua. Se $P_n \Rightarrow P$ e se $P(D_h) = 0$, então $P_n h^{-1} \Rightarrow Ph^{-1}$.*

Prova: Seja K um conjunto fechado em $(S', \mathcal{S}')$. Então,

$$\limsup_n P_n h^{-1}(K) = \limsup_n P_n[h^{-1}(K)] \leq \limsup_n P_n\left[\overline{h^{-1}(K)}\right] \leq P\left[\overline{h^{-1}(K)}\right],$$

pois $P_n \Rightarrow P$. Como K é fechado $\overline{h^{-1}(K)} \subset h^{-1}(K) \cup D_h$, e $P(D_h) = 0$, logo $P\left[\overline{h^{-1}(K)}\right] \leq P(h^{-1}(K)) = Ph^{-1}(K)$. □

Corolário 6.3. *Seja $S = \mathbb{R}$ e $h : \mathbb{R} \to \mathbb{R}$, mensurável. Suponha $P_n \Rightarrow P$, que h seja limitada e $P(D_h) = 0$. Então, $\int_S h dP_n \to \int_S h dP$.*

Prova: Como h é limitada, $|h(x)| \leq M$, para todo x real, e alguma constante $M > 0$. Defina uma função contínua f por: $f(x) = x$, se $-M \leq x \leq M$, $f(x) = M$, se $x \geq M$, e $f(x) = -M$, se $x \leq -M$. Então, pelo teorema anterior, $\int f dP_n h^{-1} \to \int f dPh^{-1}$, pois $P_n h^{-1} \Rightarrow Ph^{-1}$. Logo, $\int f(h(x))dP_n(x) \to \int f(h(x))dP(x)$, e pela definição de f e pelo fato que h é limitada, temos que$\int h(x)dP_n(x) \to \int h(x)dP(x)$. $\square$

Corolário 6.4. *Sejam X_n, X elementos aleatórios e suponha que $X_n \overset{\mathcal{D}}{\to} X$. Suponha que $h : S \to S'$ seja mensurável e $P_X(D_h) = 0$. Então, $h(X_n) \overset{\mathcal{D}}{\to} h(X)$.*

Prova: Observe que a distribuição de $h(X)$ é $P_X h^{-1}$ e use o Teorema 6.6. $\square$.

Exemplo 6.5. Seja (X_n, Y_n) uma sequência de vetores aleatórios e suponha que $(X_n, Y_n) \overset{\mathcal{D}}{\to} (X, Y)$. Pelo corolário anterior, $X_n + Y_n \overset{\mathcal{D}}{\to} X + Y$.

Note que, se X_n, Y_n são v.a's, com X_n convergindo em lei para X e Y_n convergindo em lei para Y, não é necessariamente verdade que $X_n + Y_n$ convirja em lei para $X + Y$.

Teorema 6.7. *Sejam X_n, X v.a's.*

(a) Se $X_n \overset{\mathcal{D}}{\to} X$, então $E(|X|) \leq \liminf_n E(|X_n|)$.

(b) Se $X_n \overset{\mathcal{D}}{\to} X$, e se $\{X_n\}$ for u.i, então $E(X_n) \to E(X)$.

Prova: (a) Defina $h(x)$ por $h(x) = |x|$, se $|x| < a$ e zero caso contrário. Escolha a como um ponto de continuidade de F, onde F é a f.d correspondente a P. Pelo último corolário, $\int h dP_n \to \int h dP$, donde

$$\liminf_n E(|X_n|) \geq \int_{|x| \leq a} |x| dP. \tag{6.3}$$

Faça $a \to \infty$ pelos pontos de continuidade de F, portanto o limite do termo da direita em (6.3) será $E(|X|)$.

(b) Pelo Corolário 6.4, temos que $X_n^+ \overset{\mathcal{D}}{\to} X^+$ e $X_n^- \overset{\mathcal{D}}{\to} X^-$. Portanto, podemos assumir que $X \geq 0$ e $X_n \geq 0$ para todo n. Novamente, pelo Corolário 6.4, $\int_{|x| \leq a} |x| dP_n \to \int_{|x| \leq a} |x| dP$, se a é ponto de continuidade de F. Tome a tão grande, de modo que $\int_{|x| \leq a} |x| dP$ difira de $E(X)$ de menos que $\varepsilon > 0$ e tão grande de modo que $\int_{|x| \leq a} |x| dP_n$ difira de $E(X_n)$ de menos que ε, uniformemente em n. Segue-se que $E(X_n) \to E(X)$. $\square$

Teorema 6.8. *Se X_n, X são v.a's em $(\Omega, \mathcal{F}, P)$ e X_n converge em probabilidade para X, então X_n converge em lei para X.*

Prova: Se f é qualquer função contínua e se $X_n \to X$ em probabilidade, então $f(X_n) \to f(X)$ em probabilidade. Tome f limitada, pelo TCD $E(f(X_n)) \to E(f(X))$, ou seja $X_n \to X$ em distribuição. $\square$

Note que, se X_n converge em lei para X e se X for uma constante, então, X_n converge em probabilidade para X.

6.3 Convergência fraca sobre $C[0,1]$ e $\mathbb{R}^\infty$

As seguintes questões são de interesse:

(1) Sejam $X^n = (X_1^n, X_2^n, \ldots)$ e $X = (X_1, X_2, \ldots)$ processos estocásticos. Suponha que se saiba que, para cada k, $(X_1^n, X_2^n, \ldots, X_k^n)$ converge fracamente para $(X_1, X_2, \ldots, X_k)$. É verdade que $X^n \xrightarrow{\mathcal{D}} X$?

(2) Sejam $X^n = \{X^n(t), 0 \le t \le 1\}$ e $X = \{X(t), 0 \le t \le 1\}$ dois processos estocásticos. Suponha que, para $0 \le t_1 \le \cdots \le t_k \le 1$, tenhamos $(X^n(t_1), \ldots, X^n(t_k)) \xrightarrow{\mathcal{D}} (X(t_1), \ldots, X(t_k))$. Daqui podemos concluir que $X^n \xrightarrow{\mathcal{D}} X$?

Veremos que a primeira questão tem resposta afirmativa, mas a segunda não.

Teorema 6.9. *Sejam $X^n = (X_1^n, X_2^n, \ldots)$ e $X = (X_1, X_2, \ldots)$ processos estocásticos. Se, para cada k, $(X_1^n, X_2^n, \ldots, X_k^n)$ converge fracamente para $(X_1, X_2, \ldots, X_k)$, então X^n converge fracamente para X.*

Prova: Um conjunto A é um retângulo k-dimensional semi-aberto se A for da forma $A = \{(x_1, x_2, \ldots) \in \mathbb{R}^\infty : a_1 < x_1 \le b_1, \ldots, a_k < x_k \le b_k\}$. Se desprezarmos uma coleção enumerável de a_i's e b_i's, os retângulos remanescentes são conjuntos de continuidade para X. Defina uma coleção $\mathcal{H}$ de conjuntos em $\mathbb{R}^\infty$ como segue: um conjunto está em $\mathcal{H}$ se, para algum k, é um retângulo k-dimensional e também um conjunto X-contínuo. Então, $\mathcal{H}$ é fechada sob intersecções finitas e também todo conjunto aberto em $\mathbb{R}^\infty$ é uma reunião enumerável de conjuntos de $\mathcal{H}$. O resultado, então, segue do Corolário 6.2. $\square$

Uma formulação diferente desse teorema é a seguinte. Defina, para cada k, a função $\pi_k : \mathbb{R}^\infty \to \mathbb{R}^k$ por meio de

$$\pi_k(x_1, x_2, \ldots) = (x_1, \ldots, x_k).$$

Então, o teorema nos diz que, se P_n, P são medidas de probabilidade sobre $\mathbb{R}^\infty$, tais que $P_n\pi_k^{-1} \Rightarrow P\pi_k^{-1}$, para todo k, então $P_n \Rightarrow P$. A recíproca também vale.

Exemplo 6.6. Considere $C[0,1]$ e defina elementos X^n de $C[0,1]$ por meio de:

$$\begin{aligned} X^n &= nt, \;\; 0 \le t \le 1/n, \\ &= 2 - nt, \;\; 1/n \le t \le 2/n, \\ &= 0, \;\; \text{outros casos.} \end{aligned}$$

Além disso, suponha $X \equiv 0$.

Suponha que $P_n = \delta_{X^n}$ e $P = \delta_X$. Então, as distribuições finito-dimensionais convergem fracamente, mas P_n não converge fracamente para P. Uma outra maneira de ver que P_n não converge fracamente para P é a seguinte. Defina a função f sobre $C[0,1]$, com valores reais, como segue: se x é um elemento de $C[0,1]$, $f(x) = \sup_{0 \le t \le 1} |x(t)| \wedge 1$. Então, f é contínua e limitada, contudo $\int f dP_n = 1$ e $\int f dP = 0$, logo P_n não converge fracamente para P.

No espaço $C[0,1]$, além da convergência fraca das distribuições finito-dimensionais é preciso que a sequência $\{P_n\}$ seja *fechada* para ter convergência fraca (Veja o Capítulo 9). Esta condição é o objeto da próxima seção.

6.4 Teoremas de Helly e Prokhorov

Sabemos que F é uma f.d sobre $\mathbb{R}$ se: (i) F for contínua à direita, crescente; (ii) $F(-\infty) = 0$, $F(\infty) = 1$. Dizemos que F é uma f.d *imprópria* se F satisfaz (i) e $0 \le F(x) \le 1$, para todo real x. Dizemos que F_n converge para F se $\lim_{n\to\infty} F_n(x) = F(x)$, sempre que x for um ponto de continuidade de F.

Teorema 6.10. (Teorema da seleção de Helly). *Seja $\{F_n\}$ uma sequência de f.d's sobre $\mathbb{R}$. Essas podem ser impróprias. Então, existe uma subsequência $\{n_k\}$ e uma f.d F (possivelmente imprópria) tal que $F_{n_k}(x)$ converge para $F(x)$, em pontos de continuidade de F.*

Prova: Seja $r_1, r_2, \ldots$ uma enumeração dos racionais $\mathbb{Q}$, então $\{F_n(r_1)\}$ é uma sequência limitada, logo existe uma subsequência $F_{1,k}$ tal que $\{F_{1,k}(r_1), k \ge 1\}$ converge. A seguir, existe uma subsequência da subsequência escolhida, digamos $F_{2,k}$, tal que $\{F_{2,k}(r_2), k \ge 1\}$ converge. Logo, $F_{2,k}(r_1)$ e $F_{2,k}(r_2)$ ambas convergem. Continuando, obtemos

$$
\begin{array}{rc}
F_{1,1}, F_{1,2}, F_{1,3} \cdots & \text{convergem em } r_1 \\
F_{2,1}, F_{2,2}, F_{2,3}, \cdots & \text{convergem em } r_1, r_2 \\
\vdots & \vdots
\end{array}
$$

Agora, observe que a sequência $\{F_{n,n}, n \ge 1\}$ converge para todos os racionais. Seja $\hat{F}$ o limite. A sequência $\{F_{n,n}, n \ge 1\}$ corresponde a alguma subsequência n_k da sequência original. Logo, $\lim_k F_{n_k}(x) = \hat{F}(x)$, para todo x racional. Defina

$$
F(x) = \begin{cases} \hat{F}(x), & \text{se } x \text{ é racional,} \\ \lim_{y \downarrow x, y \in \mathbb{Q}} \hat{F}(y), & \text{se } x \text{ não é racional.} \end{cases}
$$

Então, F é crescente, contínua à direita. Resta provar que $\lim_k F_{n_k}(x) = F(x)$, para todo x no qual F é contínua. Temos:

(a) $\limsup_k F_{n_k}(x) \leq F(x)$, para todo x. De fato, tome $y > x$, y racional, então $\limsup_k F_{n_k}(x) \leq \limsup_k F_{n_k}(y) = F(y)$, pois F_{n_k} é crescente. Faça $y \downarrow x$. Pela continuidade à direita, obtemos o resultado.

(b) $\liminf_k F_{n_k}(x) \geq F(x-)$, para todo x.

Tome $y < x$ racional, então $\liminf_k F_{n_k}(y) \leq \liminf_k F_{n_k}(x)$ e $F(y) = \liminf_k F_{n_k}(y)$; faça $y \uparrow x$ para obter o resultado.

Se x for um ponto de continuidade de F, então $F(x) = F(x-)$, logo por (a) e (b), $\limsup_k F_{n_k}(x) = \liminf_k F_{n_k}(x)$. $\square$

Observe que, mesmo que todas as f.d's F_n sejam próprias, a f.d limite F não é necessariamente própria.

Definição 6.4. *Uma família Π de medidas de probabilidade sobre um espaço métrico $(S, \mathcal{S})$ é chamada fechada (tight) se, para todo $\varepsilon > 0$, existe um conjunto compacto K, tal que $P(K) \geq 1 - \varepsilon$, para toda $P \in \Pi$.*

Exemplo 6.7. (a) Seja $S = \mathbb{R}$, Π é fechada se, para todo $\varepsilon > 0$, existe um intervalo (a, b) tal que $P\{(a, b)^c\} \leq \varepsilon$, para toda $P \in \Pi$. Em termos de f.d's, $F(b) - F(a) \geq 1 - \varepsilon$, para toda F cuja $P \in \Pi$.

(b) Considere P_n uniformemente distribuída sobre $[-n, n]$. Então, $\Pi = \{P_n, n \geq 1\}$ não é fechada.

(c) Suponha que P_n coloque massa unitária em $\{n\}$. Então, $\Pi = \{P_n, n \geq 1\}$ não é fechada.

Lema 6.1. *Sejam $\{P_n, n \geq 1\}$ probabilidades sobre $\mathbb{R}$ com f.d's F_n, respectivamente. Suponha que exista uma f.d F (possivelmente imprópria), tal que $\lim_{n\to\infty} F_n(x) = F(x)$, em pontos de continuidade de F. Se a família $\{P_n\}$ for fechada, então F é uma f.d.*

Prova: Tome a e b tais que $F_n(b) - F_n(a) \geq 1 - \varepsilon$, para todo n, o que é possível, pois P_n é fechada. Suponha, também, que a, b sejam pontos de continuidade de F. Como $F_n(b) \to F(b)$, $F_n(a) \to F(a)$, segue-se que $F(b) - F(a) \geq 1 - \varepsilon$. Portanto, $F(+\infty) - F(-\infty) = 1$, logo F é uma f.d própria. $\square$

Lema 6.2. *Seja $\{P_n, n \geq 1\}$ uma família de probabilidades sobre $(S, \mathcal{S})$. Suponha que exista uma probabilidade P tal que, toda subsequência P_{n_k} possui uma outra subsequência $n_k^{'}$, tal que $P_{n_k^{'}} \Rightarrow P$. Então, $P_n \Rightarrow P$.*

Prova: Suponha que P_n não convirja fracamente para P. Então, existe um $\varepsilon > 0$, uma função contínua e limitada f e uma subsequência n_k tais que $\int f dP_{n_k} \leq \int f dP - \varepsilon$ (ou $\int f dP_{n_k} \geq \int f dP + \varepsilon$), para todo k. Mas existe uma subsequência $n_k^{'}$ dessa sequência tal que $P_{n_k^{'}} \Rightarrow P$, uma contradição. $\square$

Teorema 6.11. *Seja* $\{P_n\}$ *uma família fechada de probabilidades sobre* $\mathbb{R}$, *com f.d's* F_n. *Então, existe uma f.d* F *tal que* $\lim_{n\to\infty} F_n(x) = F(x)$ *em pontos de continuidade de* F *se, e somente se,* $\lim_{n\to\infty} \int f dP_n$ *existe, para toda* f *contínua e limitada.*

Prova: ($\Rightarrow$) já provada, Teorema 6.3; de fato, provamos que $\int f dP_n \to \int f dF$.

($\Leftarrow$) Seja P_{n_k} qualquer subsequência. Vamos provar que existe uma outra subsequência $P_{n'_k}$ e uma medida P, independente dessa subsequência, tal que $P_{n'_k} \Rightarrow P$. Isso será suficiente, pelo Lema 6.2.

Sejam $F_{n'_k}$ as f.d's correspondentes. Pelo Teorema de Helly, existe uma f.d possivelmente imprópria tal que $F_{n'_k}(x) \to F(x)$, para pontos de continuidade x. Pelo fato de a família ser fechada e Lema 6.1, F é, de fato, uma f.d própria. Resta provar que esse limite é independente da subsequência envolvida. Seja $\{F_{m_k}\}$ uma subsequência qualquer e suponha que hajam duas f.d's F e G tais que:

$$\lim_k \int f dF_{n_k} = \int f dF, \quad \forall f \in C_b(\mathbb{R}),$$

$$\lim_k \int f dF_{m_k} = \int f dG, \quad \forall f \in C_b(\mathbb{R}).$$

Como o limite $\lim_{n\to\infty} \int f dF_n$ existe, $\int f dF = \int f dG$, para todo $f \in C_b(\mathbb{R})$, logo $F = G$ pelo Teorema 6.1. $\square$

Definição 6.5. *Seja* Π *uma família de medidas de probabilidade sobre* $(S, \mathcal{S})$. *Dizemos que* Π *é relativamente compacta se toda sequência* $\{P_n, n \geq 1\}$ *de probabilidades de* Π *tem uma subsequência que converge fracamente para alguma medida de probabilidade. A medida de probabilidade limite não necessita estar em* Π.

O teorema a seguir não será provado aqui. Veja Prokhorov (1956), Billingsley (1999) ou Durrett (1996b). A prova pode ser feita, sucessivamente, para $\mathbb{R}^k$, $\mathbb{R}^\infty$, S σ-compacto (uma reunião enumerável de conjuntos compactos) e, finalmente, para S geral.

Teorema 6.12. (Prokhorov). *Seja* Π *uma família de medidas de probabilidade sobre o espaço métrico* $(S, \mathcal{S})$.

(a) Se Π *for fechada, então* Π *é relativamente compacta.*

(b) Se Π *for relativamente compacta e se* S *for completo e separável, então* Π *é fechada.*

A parte (b) nos diz, essencialmente, que para espaços "bem comportados", os dois conceitos (família relativamente compacta e fechada) são equivalentes. Veja o Problema 13.

Problemas

1. Sejam X_n, X, Y_n v.a's, e c uma constante. Prove que:

   (i) Se $X_n \xrightarrow{\mathcal{D}} X$, $Y_n \to c$, em probabilidade, então $X_n + Y_n \xrightarrow{\mathcal{D}} X + c$ e $X_n Y_n \xrightarrow{\mathcal{D}} cX$.

   (ii) Não é verdade, de modo genérico, que se $X_n \xrightarrow{\mathcal{D}} X$ e $Y_n \xrightarrow{\mathcal{D}} Y$, então $X_n + Y_n \xrightarrow{\mathcal{D}} X + Y$.

   (iii) Se $X_n \xrightarrow{\mathcal{D}} X$, então X_n não necessita convergir para X em probabilidade (dê um contra-exemplo). Contudo, $X_n \xrightarrow{\mathcal{D}} X$ implica $X_n \to X$ em probabilidade se X for uma constante.

2. Sejam P_n, P medidas de probabilidade sobre $(S, \mathcal{S})$. Se $P_n(A) \to P(A)$ para todos os abertos A, então $P_n(A) \to P(A)$, para todos os conjuntos de Borel (ou seja, todos os conjuntos de $\mathcal{S}$).

3. Prove que, se F_n, F são f.d's sobre $\mathbb{R}$, e se F for contínua, então $\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| \to 0$, $n \to \infty$.

4. Sejam P, Q probabilidades sobre $\mathbb{R}^k$. Defina a convolução de P e Q como a probabilidade sobre $\mathbb{R}^k$ dada por $P \star Q(A) = \int_{\mathbb{R}^k} P(A-y)Q(dy)$, onde A é um boreliano de $\mathbb{R}^k$.

   (a) Mostre que se $P_n \Rightarrow P$, $Q_n \Rightarrow Q$, então $P_n \star Q_n \Rightarrow P \star Q$.

   (b) Se Π for uma família fechada de medidas de probabilidade, então $\Pi^* = \{P \star Q : P \in \Pi, Q \in \Pi\}$ é fechada.

5. Prove (a)-(c) do Exemplo 6.7.

6. (Métrica de Prokhorov) Se P e Q são medidas de probabilidade, defina $\rho(P, Q) = \inf\{\varepsilon > 0 : Q(A) \leq P(A^\varepsilon) + \varepsilon, \text{ e } P(A) \leq Q(A^\epsilon) + \varepsilon, \forall A \in \mathcal{S}\}$. Mostre que ρ é uma métrica no espaço das medidas de probabilidade sobre $(S, \mathcal{S})$. Aqui, $A^\epsilon = \{x \in S : d(A, x) < \varepsilon\}$, sendo d a métrica sobre S.

7. Seja $(S, \mathcal{S})$ um espaço métrico separável (pode usar $\mathbb{R}^k$). Mostre que $P_n \Rightarrow P$ se, e somente se, $\rho(P_n, P) \to 0$.

8. (Métrica de Lévy) Sejam F, G f.d's sobre $\mathbb{R}$. Defina $\rho_L(F, G) = \inf\{\varepsilon : \forall x \in \mathbb{R}, G(x - \varepsilon) - \varepsilon \leq F(x) \leq G(x + \varepsilon) + \varepsilon\}$.

   (a) Mostre que ρ_L é uma métrica.

   (b) Mostre que $F_n \xrightarrow{\mathcal{D}} F$ se, e somente se $\rho_L(F_n, F) \to 0$.

9. Sejam $\{P_n, n \geq 1\}$ medidas de probabilidade sobre $\mathbb{R}^\infty$, e seja π_k a projeção de $\mathbb{R}^\infty$ sobre $\mathbb{R}^k$, isto é, $\pi_k\{(x_1, x_2, \ldots)\} = (x_1, \ldots, x_k)$. Mostre que, se $\{P_n \pi_k^{-1}, n \geq 1\}$ é uma família fechada, para cada k, então $\{P_n, n \geq 1\}$ é uma família fechada.

10. Seja h uma função mensurável de $\mathbb{R}$ em $\mathbb{R}$, tal que $|h(x)| \to +\infty$, quando $|x| \to +\infty$. Se Π for uma família de medidas de probabilidade e se $\sup_{P\in\Pi} \int |h|dP < \infty$, então Π é fechada. Um caso especial é: se $\{X_n, n \geq 1\}$ são v.a's tais que $\sup_{n\geq 1} E(|X_n|^\delta) < \infty$, para algum $\delta > 0$, então $\{X_n\}$ é fechada.

11. Sejam P_n, P medidas de probabilidade sobre $\mathbb{R}$, cada uma absolutamente contínua com respeito à medida de Lebesgue e tendo densidades g_n, g, respectivamente.

    (a) Se $g_n(x) \to g(x)$ q.c, então $P_n \Rightarrow P$.

    (b) A recíproca de (a) pode não ser verdade; dê um exemplo.

    (c) Suponha $\{P_n\}$ normais e $P_n \Rightarrow P$. Mostre que, nesse caso, P é também normal e a recíproca de (a) vale.

    (d) Mostre que, se cada P_n é normal, então $\{P_n, n \geq 1\}$ é fechada se, e somente se, médias e variâncias são limitadas.

12. Prove que a classe dos conjuntos P-contínuos (P fixa) é uma álgebra. Mostre, por meio de um exemplo, que essa classe não precisa ser uma σ-álgebra.

13. Prove a parte (a) do Teorema de Prokhorov, para $S = \mathbb{R}^k$. Use o Teorema de Helly e o Teorema 6.11.

Capítulo 7

Funções Características

As funções características constituem uma ferramenta importante em diversas áreas da Teoria de Probabilidade e Estatística. Por exemplo, são úteis na demonstração de teoremas limites centrais, do teorema de Bochner para procesos estacionários, no estudo de distribuições estáveis etc. Uma referência adequada aqui é Chung (1974).

7.1 Introdução

Nesta seção definimos a função característica e apresentamos suas propriedades. A seguir, apresentamos dois resultados importantes, o teorema da unicidade e o da continuidade.

Definição 7.1. *Seja P uma medida de probabilidade sobre $\mathbb{R}^k$. A função característica (f.c) de P é a função $\varphi : \mathbb{R}^k \to \mathbb{C}$, definida por*

$$\varphi(\mathbf{t}) = \int e^{i\mathbf{t}\cdot\mathbf{x}} dP(\mathbf{x}), \tag{7.1}$$

onde $\mathbf{t} = (t_1, \ldots, t_k)$, $\mathbf{x} = (x_1, \ldots, x_k)$ *e* $\mathbf{t} \cdot \mathbf{x} = \sum_{i=1}^k t_i x_i$.

Se $\mathbf{X} = (X_1, \ldots, X_k)$ é um vetor aleatório, então a f.c de $\mathbf{X}$ é

$$\varphi(\mathbf{t}) = E\Big(e^{i\mathbf{t}\cdot\mathbf{X}}\Big). \tag{7.2}$$

Algumas propriedades elementares da f.c são:

(1) $|\varphi(\mathbf{t})| \leq 1 = \varphi(0)$; $\quad \overline{\varphi(\mathbf{t})} = \varphi(-\mathbf{t})$.

(2) $\varphi(\mathbf{t})$ é uniformemente contínua. Veja o Problema 1.

(3) Se $\mathbf{X}$ é um vetor aleatório com f.c $\varphi(\mathbf{t})$, então a f.c de $a\mathbf{X}+b$ é $e^{i\mathbf{t}\cdot b}\varphi(a\mathbf{t})$. Em particular a f.c de $-\mathbf{X}$ é $\varphi(-\mathbf{t}) = \overline{\varphi(\mathbf{t})}$. Logo, se φ é uma f.c, também o será $\overline{\varphi}$.

(4) Sejam $\mathbf{X}$ e $\mathbf{Y}$ vetores aleatórios independentes, com f.c's φ_X e φ_Y, respectivamente. Seja φ_{X+Y} a f.c de $\mathbf{X}+\mathbf{Y}$. Então,

$$\varphi_{X+Y}(\mathbf{t}) = \varphi_X(\mathbf{t})\varphi_Y(\mathbf{t}). \tag{7.3}$$

Se essa relação vale, X e Y não precisam ser independentes.

Para provar o teorema da unicidade precisamos dos seguintes lemas.

Lema 7.1. (Uma versão do Teorema de Stone-Weierstrass) *Seja S um espaço de Hausdorff compacto e $C(S)$ a álgebra de todas as funções sobre S, com valores complexos e contínuas. Seja $\mathcal{A}$ uma sub-álgebra de $C(S)$ tal que:*

(a) $\mathcal{A}$ separa pontos (se $x, y \in S$, então existe $f \in \mathcal{A}$ tal que $f(x) \neq f(y)$).

(b) Se $f \in \mathcal{A}$, então $\overline{f} \in \mathcal{A}$.

(c) Para cada ponto $x \in S$, existe $f \in \mathcal{A}$, tal que tal que $f(x) \neq 0$.

Então, $\mathcal{A}$ é densa em $C(S)$, no sentido que, dado $\varepsilon > 0$ e $g \in C(S)$, existe $f \in \mathcal{A}$, tal que $\sup_{x\in S} |f(x) - g(x)| < \varepsilon$.

Para uma prova, veja Simmons (2003).

Lema 7.2. *Sejam P e Q medidas de probabilidade sobre $\mathbb{R}^k$ tais que $\int f dP = \int f dQ$, para toda função real f que seja contínua e se anule fora de um conjunto compacto. Então, $P = Q$.*

Prova: Seja B um conjunto compacto e defina a função g por $g(t) = 1$, para $t \leq 0$, $g(t) = 0$ para $t \geq 1$, e $g(t) = 1 - t$, para $0 \leq t \leq 1$. Seja $\varepsilon > 0$ e defina f_ε por $f_\varepsilon(x) = g(d(x, B)/\varepsilon)$. Como B é fechado, $\lim_{\varepsilon\downarrow 0} f_\varepsilon(x) = I_B(x)$. Por hipótese, $\int f_\varepsilon dP = \int f_\varepsilon dQ$, logo $\int_B dP = \int_B dQ$, pelo TCD. Logo, $P(B) = Q(B)$, para todos os conjuntos compactos e consequentemente pela continuidade de P e Q, $P(F) = Q(F)$, para todos os conjuntos fechados F. Concluimos que $P = Q$. $\square$

Teorema 7.1. (Da Unicidade) *Sejam P e Q medidas de probabilidade sobre $\mathbb{R}^k$, tendo f.c's φ e ψ, respectivamente. Se $\varphi(\mathbf{t}) = \psi(\mathbf{t})$, para todo $\mathbf{t}$, então $P = Q$.*

Prova: Considere $S = [-\pi N, \pi N]^k$ e seja $\mathcal{A}_0$ a classe de todas as funções da forma $f(x) = \exp\{i\mathbf{n}\cdot\mathbf{x}\}$, onde $\mathbf{n} = (n_1/2N, \ldots, n_k/2N)$, $n_j = 0, \pm 1, \ldots$. Seja $\mathcal{A}$ a classe de todas as combinações lineares de funções de $\mathcal{A}_0$. Então, $\mathcal{A}$ é uma álgebra que satisfaz (a)-(c) do Lema 7.1, logo é densa em $C(S)$. Observe que, como $\int e^{it\cdot x} dP = \int e^{it\cdot x} dQ$, para todo t, segue que $\int f dP = \int f dQ$, para toda $f \in \mathcal{A}$.

Tome $f \in C(S)$, seja $\varepsilon > 0$ e tome $g \in \mathcal{A}$ tal que $\|f - g\| = \sup_x |f(x) - g(x)| \le \varepsilon$. Então,

$$\Big|\int f dP - \int f dQ\Big| \le \int |f - g| dP + \int |f - g| dQ + \Big|\int g dP - \int g dQ\Big| \le 2\varepsilon,$$

pois a última integral anula-se. Portanto, temos que $\int f dP = \int f dQ$, para toda f contínua sobre $\mathbb{R}^k$ e que se anula fora de S, e a mesma conclusão vale para f nas mesmas condições que se anula fora de conjuntos compactos. A conclusão do teorema segue do Lema 7.2. $\square$

Teorema 7.2. *Seja $\mathbf{X} = (X_1, \ldots, X_n)$ um vetor aleatório, com f.c φ_X. Seja φ_{X_i} a f.c de X_i, $i = 1, \ldots, n$. Então as v.a's $X_1, \ldots, X_n$ são independentes se, e somente se, $\varphi_X = \prod_{i=1}^n \varphi_{X_i}$.*

Prova: Vamos dar a prova para o caso $n = 2$.

($\Rightarrow$): trivial

($\Leftarrow$) Suponha $\varphi_{(X,Y)} = \varphi_X \cdot \varphi_Y$, isto é, $E\Big(e^{i(tX+sY)}\Big) = E\Big(e^{itX}\Big)E\Big(e^{isY}\Big)$. Então,

$$\int_{\mathbb{R}^2} e^{i(tX+sY)} dP_{(X,Y)}(x, y) = \int_{\mathbb{R}} e^{itX} dP_X(x) \cdot \int_{\mathbb{R}} e^{isY} dP_Y(y) =$$

$$\int_{\mathbb{R}^2} e^{i(tX+sY)} dP_X(x) dP_Y(y),$$

onde a segunda igualdade vale pelo Teorema de Fubini. As medidas $dP_{(X,Y)}$ e $dP_X dP_Y$ têm a mesma f.c., logo pelo teorema da unicidade, elas são iguais, e portanto X e Y são independentes. $\square$

7.2 Funções características e distribuições normais

Nessa seção provamos alguns resultados envolvendo distribuição normal, univariada e multivariada.

Se $X \sim N(\mu, \sigma^2)$, sua função característica é dada por

$$\varphi(t) = e^{it\mu - \sigma^2 t^2/2}. \tag{7.4}$$

Em particular, para uma distribuição normal padrão, $\varphi(t) = e^{-t^2/2}$. Veja o Problema 3.

Um vetor aleatório $\mathbf{X} = (X_1, \ldots, X_k)$ tem distribuição normal multivariada se existem v.a's normais padrões independentes $Z_1, \ldots, Z_k$ e reais μ_i, a_{ij}, $1 \le i, j \le k$, tais que $\mathbf{X}$ tem a mesma distruição que o vetor $(\mu_1 + \sum_{i=1}^k a_{1i} Z_i, \ldots, \mu_k + \sum_{i=1}^k a_{ki} Z_i)$.

Se $\mathbf{A} = [a_{ij}]$, $\mathbf{Z} = (Z_1, \ldots, Z_k)$ e $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_k)$, então podemos escrever $\mathbf{X} \sim \mathbf{AZ}+ \boldsymbol{\mu}$ (o sinal $\sim$ significa "tem a mesma distribuição").

Alguns fatos básicos sobre distribuições normais são dados a seguir.

[1] Suponha que $\mathbf{X} = (X_1, \ldots, X_k)$ tenha distribuição normal multivariada. Então, $\sum_{j=1}^k a_j X_j$ tem distribuição normal univariada.

Devido à caracterização acima de um vetor mutivariado, $\sum_j a_j X_j$ é uma combinação linear de $Z_1, \ldots, Z_k$ e qualquer combinação linear de v.a's normais independentes é normal. Basta calcular a f.c da combinação linear e ver que é dada por (7.4). O resultado segue do teorema da unicidade.

[2] Seja $\mathbf{X} = (X_1, \ldots, X_k)$ um vetor com distribuição normal multivariada. Então, a f.c de $\mathbf{X}$ é dada por

$$\varphi(\mathbf{t}) = \exp\{i\boldsymbol{\mu} \cdot \mathbf{t} - \mathbf{t} \cdot \mathbf{R}\mathbf{t}\}, \tag{7.5}$$

onde $\mathbf{R} = [r_{ij}]$ é a matriz de covariâncias, com $r_{ij} = E[(X_i - \mu_i)(X_j - \mu_j)]$. Note que $\mathbf{R} = \mathbf{A}^2$ com $\mathbf{A}$ dada na definição de X acima. Veja o Problema 4.

[3] (Recíproca de [2]) Dada qualquer matriz $\mathbf{R}$ simétrica semi-definida positiva e qualquer vetor $\boldsymbol{\mu}$, existe um vetor com distribuição normal multivariada com f.c dada por (7.5).

Basta considerar $Z_1, \ldots, Z_k$, independentes, com distribuição normal padrão e definir $\mathbf{X} = \sqrt{\mathbf{R}}\mathbf{Z} + \boldsymbol{\mu}$.

[4] Sejam $\mathbf{X}$ e $\mathbf{Y}$ dois vetores com distribuição normal multivariada, com as mesmas médias e matrizes de covariâncias. Então, $\mathbf{X} \sim \mathbf{Y}$.

De fato, $\mathbf{X}$ e $\mathbf{Y}$ terão a mesma f.c, e o resultado segue do teorema da unicidade.

[5] Suponha $\mathbf{X} = (X_1, \ldots, X_k)$ com distribuição normal multivariada e matriz de covariâncias $\mathbf{R}$. Se todos os elementos de $\mathbf{R}$ são nulos, exceto aqueles sobre a diagonal principal, então $X_1, \ldots, X_k$ são independentes.

É suficiente notar que a f.c de $\mathbf{X}$ é o produto de termos da forma $\exp\{it_j\mu_j - r_{jj}t_j^2/2\}$ e o resultado segue.

Definição 7.2.[Convolução] *Consideremos duas medidas de probabilidades P e Q sobre $\mathbb{R}^k$. Então, $P \star Q$ é a medida de probabilidade sobre $\mathbb{R}^k$ definida por*

$$P \star Q(A) = \int_{\mathbb{R}^k} P(A - y) dQ(y).$$

Note que, se h for uma função integrável, então

$$\int h(x) dP \star Q(x) = \int \int h(x + y) dP(x) dQ(y).$$

Teorema 7.3. *Sejam P e Q medidas de probabilidade sobre $\mathbb{R}^k$ com f.c's φ_1, φ_2, respectivamente. Então temos:*

(a) A f.c de $P \star Q$ é $\varphi_1(\mathbf{t})\varphi_2(\mathbf{t})$.

(b) Sejam $\mathbf{X}$ e $\mathbf{Y}$ dois vetores aleatórios independentes. Então, $P_{\mathbf{X}+\mathbf{Y}} = P_{\mathbf{X}} \star P_{\mathbf{Y}}$, onde $P_{\mathbf{X}+\mathbf{Y}}$ é a distribuição de $\mathbf{X}+\mathbf{Y}$.

Prova: (a) A f.c de $P \star Q$ é

$$\int e^{i\mathbf{t}\cdot\mathbf{x}} dP{\star}Q(\mathbf{x}) = \int\int e^{i\mathbf{t}\cdot(\mathbf{x}+\mathbf{y})} dP(\mathbf{x})dQ(\mathbf{y}) = \int e^{i\mathbf{t}\cdot\mathbf{x}} dP(\mathbf{x}) \int e^{i\mathbf{t}\cdot\mathbf{y}} dQ(\mathbf{y}) = \varphi_1(\mathbf{t})\varphi_2(\mathbf{t}),$$

a primeira igualdade pela nota anterior e a segunda pelo Teorema de Fubini.

(b) A f.c de $P_X \star P_Y$ é $\varphi_X(\mathbf{t})\varphi_Y(\mathbf{t})$, pela parte (a). A f.c de P_{X+Y} é $\varphi_X(\mathbf{t})\varphi_Y(\mathbf{t})$, provada anteriormente. O resultado segue pelo teorema da unicidade. $\square$

A operação de convolução entre duas medidas de probabilidades é uma operação de suavização, no seguinte sentido: sejam P, Q probabilidades sobre $\mathbb{R}^k$.

(i) Se P for absolutamente contínua (com respeito à medida de Lebesgue) e Q for arbitrária, então $P \star Q$ é absolutamente contínua.

(ii) Se P for não atômica, e Q arbitrária, então $P \star Q$ será não atômica. Veja o Problema 6.

7.3 O Teorema da continuidade

O teorema da continuidade para funções características tem sua origem em trabalhos de Lévy (1925), Glivenko (1936) e Cramér (1937). Esse resultado é básico para o estudo do teorema limite central (capítulo seguinte) e, em particular, para caracterizar distribuições infinitamente divisíveis. Para desenvolvimentos recentes, veja Heyer e Kawakami (2005).

Teorema 7.4. (Teorema da continuidade, Lévy-Cramér) *Sejam P_n probabilidades sobre $\mathbb{R}^k$, com f.c's φ_n.*

(a) Se $P_n \Rightarrow P$, então $\lim_{n\to\infty} \varphi_n(\mathbf{t}) = \varphi(\mathbf{t})$, onde φ é a f.c de P. A convergência é uniforme sobre conjuntos compactos e $\{\varphi_n\}$ é uma família uniformemente equicontínua.

(b) Suponha P_n, φ_n como acima. Se:

(i) $\lim_{n\to\infty} \varphi_n(\mathbf{t}) = h(\mathbf{t})$ *existir, para todo* $\mathbf{t}$, *e*

(ii) $h(\mathbf{t})$ *for contínua no zero, então*

existe uma medida de probabilidade P com $P_n \Rightarrow P$, e h é a f.c de P.

Prova: (a) Para cada $\mathbf{t}$, $e^{i\mathbf{t}\cdot\mathbf{x}}$ é uma função limitada e contínua sobre $\mathbb{R}^k$, logo como $P_n \Rightarrow P$, $\int e^{i\mathbf{t}\cdot\mathbf{x}}dP_n \to \int e^{i\mathbf{t}\cdot\mathbf{x}}dP$, para cada $\mathbf{t}$. Para provar a equicontinuidade uniforme, lembremos que uma família de funções $\{f_i, i \in I\}$ é uniformemente equicontínua se para todo $\varepsilon > 0$, existe $\delta > 0$ tal que para todo $|h| \leq \delta$, temos $|f_i(x+h) - f_i(x)| < \varepsilon$, para todo $x \in \mathbb{R}^k$ e $i \in I$. Aqui temos que

$$|\varphi_n(\mathbf{t}+\mathbf{h})-\varphi_n(\mathbf{t})| = \left|\int\{e^{i(\mathbf{t}+\mathbf{h})\cdot\mathbf{x}}-e^{i\mathbf{t}\cdot\mathbf{x}}\}dP_n\right| \leq \int|e^{i\mathbf{h}\cdot\mathbf{x}}-1|dP_n \to \int|e^{i\mathbf{h}\cdot\mathbf{x}}-1|dP,$$

quando $n \to \infty$, uniformemente em $\mathbf{t}$. Como o último termo tende a zero quando $\mathbf{h} \to 0$, obtemos a equicontinuidade de $\{\varphi_n\}$. Convergência uniforme sobre conjuntos compactos segue da convergência ponto a ponto e da equicontinuidade uniforme.

(b) Segue da aplicação dos dois lemas a seguir. □

Lema 7.3. *Se $\lim_{n\to\infty}\varphi_n(\mathbf{t}) = h(\mathbf{t})$ e se $\{P_n\}$ é fechada, então existe uma medida de probabilidade P tal que $P_n \Rightarrow P$.*

Prova: Se $\{n_k\}$ é qualquer subsequência, existe uma subsequência de $\{n_k\}$, $\{n_k^{'}\}$, e uma medida de probabilidade P, tal que $P_{n_k^{'}} \Rightarrow P$, pelo Teorema de Prokhorov. O resultado seguirá se mostrarmos que P é independente da subsequência $\{n_k\}$. Como $P_{n_k^{'}} \Rightarrow P$, segue-se da parte (a) do teorema que $\lim \varphi_{n_k^{'}}(\mathbf{t}) = \varphi(\mathbf{t}) = \lim \varphi_n(\mathbf{t})$, pela hipótese que a sequência original de f.c's converge. Portanto, todas as P's que podem ser candidatas como limites têm a mesma f.c, logo P é única, pelo teorema da unicidade e $P_n \Rightarrow P$. □

Lema 7.4. *Se $\lim_{n\to\infty}\varphi_n(\mathbf{t}) = h(\mathbf{t})$, e $h(\mathbf{t})$ for contínua no zero, então $\{P_n\}$ é fechada.*

Prova: Em primeiro lugar, observamos que

$$P_n\{(X_1^n,\ldots,X_k^n) \notin [-a,a]^k\} \leq \sum_{i=1}^k P_n\{|X_i^n| > a\}.$$

Portanto é suficiente mostrar o resultado para $k = 1$. Temos que

$$\frac{1}{a}\int_{-a}^{a}[1-\varphi_n(t)]dt = \frac{1}{a}\int_{-a}^{a}[1-\int e^{itx}dP_n(x)]dt =$$

$$= \frac{1}{a}\int\int_{-a}^{a}[1-e^{itx}]dtdP_n(x) = 2\int\left[1-\frac{\sin ax}{ax}\right]dP_n(x)$$

$$\geq 2\int_{\{|x|>2/a\}}\left[1-\frac{1}{|ax|}\right]dP_n(x) \geq P_n\{|X_1^n| > 2/a\}.$$

Considere a tão pequeno de modo que $a^{-1}\int_{-a}^{a}[1-h(t)]dt < \varepsilon$, que é possível pois h é contínua no zero e $h(0)=1$. A seguir, tome N tão grande que, se $n \geq N$, então $a^{-1}\int_{-a}^{a}[1-\varphi_n(t)]dt \leq 2\varepsilon$ (pois $\varphi_n(t) \to h(t)$). Logo, para $n \geq N$, $2\varepsilon \geq P_n\{|X_1^n| > 2/a\}$ e tomando ε ainda menor, se necessário, podemos obter $P_j\{|X_1^j| > 2/a\} < 2\varepsilon$, para $j = 1,\ldots,N-1$. Assim, para todo n, com a escolhido dessa forma, temos $2\varepsilon \geq P_n\{|X_1^n| > 2/a\}$, ou seja $\{P_n\}$ é fechada. $\square$

Corolário 7.1. *Sejam $\{X_n, n \geq 1\}$ v.a's independentes com f.c's φ_n. Seja $S_n = \sum_{i=1}^n X_i$. Se S_n converge em distribuição, então S_n converge q.c.*

Prova: Provamos, via martingales, que se $\prod_{k=1}^n \varphi_k(t)$ converge, para todo t num intervalo, então S_n converge q.c. Por hipótese, S_n converge em distribuição, logo a f.c de S_n, o produto em questão, converge para todo t. $\square$

7.4 Funções características sobre $\mathbb{R}$

Nesta seção iremos estudar resultados específicos para f.c's definidas sobre $\mathbb{R}$, em particular a importante fórmula de inversão.

Teorema 7.5. *Seja X uma v.a com f.c φ.*
(a) Se $E(|X|^k) < \infty$, então a k-ésima derivada $\varphi^{(k)}(t)$ existe, é contínua e

$$\varphi^{(k)}(t) = \int (ix)^k e^{itx} dP(x). \tag{7.6}$$

Também, $\varphi^{(k)}(0) = (i)^k E(X^k)$.

(b) Se $\varphi^{(k)}(0)$ existe e se k é par, então $E(|X|^k) < \infty$.

Prova: (a) Vamos dar a prova somente para o caso $k=1$. Temos

$$\frac{\varphi(t+h)-\varphi(t)}{h} = \int \frac{e^{i(t+h)x} - e^{itx}}{h} dP.$$

O módulo do integrando é limitado por $|x|$, que é intergrável por hipótese; faça $h \to \infty$ e use o TCD para obter o resultado.

(b) Para $k=2$,

$$\begin{aligned}\varphi''(0) &= \lim_{h\to 0}\frac{\varphi(h)-2\varphi(0)+\varphi(-h)}{h^2} = \lim_{h\to 0}\frac{1}{h^2}\int [e^{ihx} - 2 + e^{-ihx}]dP\\ &= -2\lim_{h\to 0}\int \frac{1-\cos hx}{h^2} dP.\end{aligned}$$

Então, pelo lema de Fatou,

$$\int x^2 dP = \int \lim_{h\to 0} \frac{1-\cos hx}{h^2} dP \leq \lim_{h\to 0} \int \frac{1-\cos hx}{h^2} dP = -\frac{1}{2}\varphi^{''}(0) < \infty.$$

Para o caso geral k par, suponha o teorema válido para $k-2$ e defina $H(x) = \int_{-\infty}^{x} y^{k-2} dP(y)$. A função H é crescente, logo $H(x)/H(\infty)$ é uma f.d. Seja ψ a f.c. dessa f.d. Então,

$$\psi(t) = \int e^{itx} \frac{dH(x)}{H(\infty)} = \int e^{itx} x^{k-2} \frac{dP(x)}{H(\infty)}.$$

Aplique o caso $k = 2$ a essa f.c. e obtenha

$$\infty > -\frac{1}{2}\psi^{''}(0) \geq \int x^2 \frac{dH(x)}{H(\infty)} = \int x^2 x^{k-2} \frac{dP(x)}{H(\infty)} = \frac{1}{H(\infty)} \int x^k dP(x). \quad \square$$

Observe que o resultado não é em geral válido se k for ímpar.

Exemplo 7.1. $\varphi(t) = e^{-t^4}$ não é uma f.c. A segunda derivada de φ existe e é igual a zero para $t = 0$. Pela parte (b) do teorema, $E(X^2) < \infty$. Por (a) $E(X^2) = 0$, logo $X = 0$, mas φ não é a f.c. de $X = 0$.

Corolário 7.2. (Expansão de Taylor). *Seja X uma v.a com f.c φ e suponha que $E(|X|^n) < \infty$. Então temos para todo $t \in \mathbb{R}$,*

$$\left|\varphi(t) - \sum_{k=0}^{n} \frac{(it)^k E(X^k)}{k!}\right| \leq E\Big(\min\Big\{\frac{|tX|^{n+1}}{(n+1)!}, \frac{2|tX|^n}{n!}\Big\}\Big).$$

Prova: Veja o Problema 12.

Teorema 7.6. (Método dos momentos) *Sejam $\{X_n\}$ v.a's com distribuições $\{P_n\}$. Suponha $E(|X_n|^k) < \infty$, para todo k e n. Suponha que:*

(a) $\lim_{n\to\infty} E(X_n^k) = \mu_k < \infty$;

(b) $\limsup_{n\to\infty} \frac{(\mu_n)^{1/n}}{n} = \lambda < \infty$.

Então, existe uma medida de probabilidade P e $P_n \Rightarrow P$.

Prova: Seja $\{n_k\}$ qualquer subsequência. Vamos provar que podemos extrair uma subsequência $n_k^{'}$ tal que $P_{n_k^{'}} \Rightarrow P$, e P é uma probabilidade que não depende das subsequências consideradas. Note que $\sup_n \int x^2 dP_n < \infty$, por (a). Segue-se que $\{P_n\}$ é uma família fechada (veja o Problema 13). Pelo teorema de Prokhorov, existe uma subsequência $n_k^{'}$ de n_k e uma probabilidade P tal que $P_{n_k^{'}} \Rightarrow P$. Provemos agora que P independe de $\{n_k\}$. Para cada j, temos que $\int x^j dP_{n_k^{'}} \to \int x^j dP$ pois $\{X_n^j\}_n$ é u.i. Logo, como $\int x^j dP_n \to \mu_j$, para todo j, segue-se que todos os P

limites têm os mesmos momentos. Para mostrar que P é única, basta mostrar que é univocamente determinada por seus momentos.

Se φ é a f.c de P, usando a expansão de Taylor dado no Corolário 7.2,

$$\left|\varphi(t+h) - \varphi(t) - \varphi^{'}(t) - \frac{\varphi^{(2)}(t)h^2}{2} - \ldots - \frac{\varphi^{(k)}(t)h^k}{k!}\right| \leq \frac{E(|X|^{k+1})|h|^{k+1}}{(k+1)!},$$

onde X é uma v.a com lei P. Pela parte (b) e usando a fórmula de Stirling, se $|h| < 1/(4\lambda)$, o lado direito converge para zero, quando $k \to \infty$.

Concluímos que φ admite uma expansão de Taylor ao redor de qualquer ponto da reta, ou seja φ é analítica em uma vizinhança da reta, de modo que é univocamente determinada por sua série de potências ao redor do zero. Mas essa é dada por $\sum_k \frac{(it)^k E(X^k)}{k!}$, logo P é univocamente determinada por seus momentos. $\square$

Provaremos, a seguir, a chamada fórmula de inversão para f.c's. Uma motivação para tal fórmula é a seguinte. Para dada f, satisfazendo determinadas condições, a *transformada de Fourier* de f é definida por

$$\hat{f}(\xi) = \int_{-\infty}^{\infty} e^{i\xi x} f(x)dx.$$

Sabe-se, também, que sob certas condições, temos a *transformada inversa de Fourier*

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\xi x}\hat{f}(\xi)d\xi.$$

Se f for uma densidade de probabilidade, com f.d F, a f.c correspondente à F é a transformada de Fourier $\hat{f}$ de f. Então,

$$F(b) - F(a) = \int_a^b f(x)dx = \int_a^b \left[\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\xi x}\varphi(\xi)d\xi\right]dx.$$

O resultado rigoroso é apresentado a seguir.

Teorema 7.7. (Fórmula da inversão) *Seja F uma f.d e φ a f.c correspondente. Então, para $a < b$, temos*

$$\frac{F(b)+F(b-)}{2} - \frac{F(a)+F(a-)}{2} = \frac{1}{2\pi}\lim_{c\to\infty}\int_{-c}^{c} \varphi(t)\frac{e^{-ita} - e^{-itb}}{it}dt. \qquad (7.7)$$

Prova: Os seguintes fatos são necessários:

(i) $\int_0^{\infty} \frac{\sin x}{x}dx = \pi/2$;

(ii) $\lim_{a\to-\infty,b\to\infty}\int_a^b \frac{\sin x}{x}dx = \pi$;

(iii) $\int_{-\infty}^{\infty} \frac{\sin(\alpha x)}{x}dx = \begin{cases} \pi, & \alpha > 0, \\ 0, & \alpha = 0, \\ -\pi, & \alpha < 0; \end{cases}$

(iv) $\int_{-c}^{c} \frac{\sin(\alpha x)}{x}dx$ é limitada como função de c.

A integral em (7.7) é dada por

$$\frac{1}{2\pi}\int_{-c}^{c}\varphi(t)\frac{e^{-ita}-e^{-itb}}{it}dt = \frac{1}{2\pi}\int_{-c}^{c}\left[\frac{e^{-ita}-e^{-itb}}{it}\int_{-\infty}^{\infty}e^{itx}dP(x)\right]dt =$$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty}\left[\int_{-c}^{c}\frac{e^{it(x-a)}-e^{it(x-b)}}{it}dt\right]dP(x).$$

Para obter a última igualdade, o teorema de Fubini foi aplicado, pois

$$\left|\frac{e^{it(x-a)}-e^{it(x-b)}}{it}\right| \leq \left|\frac{e^{-ita}-e^{-itb}}{it}\right| = \left|\int_a^b e^{itx}dx\right| \leq b-a.$$

Seja

$$h_c = \int_{-c}^{c}\frac{e^{it(x-a)}-e^{it(x-b)}}{it}dt = \int_{-c}^{c}\frac{\sin t(x-a)}{t}dt - \int_{-c}^{c}\frac{\sin t(x-b)}{t}dt,$$

que é uma função limitada de c, por (iv) acima. Logo, podemos tomar o limite para $c\to\infty$, sob o sinal da integral, para obter

$$\lim_{c\to\infty}\frac{1}{2\pi}\int_{-c}^{c}\varphi(t)\frac{e^{-ita}-e^{-itb}}{it}dt = \frac{1}{2\pi}\int_{-\infty}^{\infty}\lim_{c\to\infty}h_c dP.$$

Mas, $\lim_{c\to\infty} h_c = \begin{cases} (-\pi)-(-\pi) = 0, & x < a, \\ 0-(-\pi) = \pi, & x = a, \\ \pi-(-\pi) = 2\pi, & a < x < b, \\ \pi, & x = b, \\ 0, & x > b, \end{cases}$ e, portanto, o limite acima fica

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}\lim_{c\to\infty}h_c dP = \frac{1}{2\pi}\left[\int_{\{x=a\}}\pi dF + \int_{\{a<x<b\}}2\pi dF + \int_{\{x=b\}}\pi dF\right]$$

$$= \frac{1}{2}[F(a)-F(a-)]+[F(b)-F(a)]+\frac{1}{2}[F(b)-F(b-)] = \frac{F(b)+F(b-)}{2}-\frac{F(a)+F(a-)}{2}.$$

$\square$

Esse resultado fornece um teorema de unicidade para $\mathbb{R}$.

Corolário 7.3. *Duas medidas de probabilidade sobre $\mathbb{R}$, tendo a mesma f.c., são iguais.*

Teorema 7.8. *Suponha que $\int_{\mathbb{R}} |\varphi(t)|dt < \infty$, onde φ é a f.c da v.a X. Então, X tem uma densidade de probabilidade limitada e contínua.*

Prova: (a) Seja F a f.d de X; então, F é contínua. De fato, para $h > 0$,

$$\frac{F(x+h)+F(x+h-)}{2} - \frac{F(x)+F(x-)}{2} = \frac{1}{2\pi}\int_{-\infty}^{\infty} \varphi(t)\frac{e^{-ixt}-e^{-i(x+h)t}}{it}dt.$$

O integrando é limitado por $|\varphi(t)|h$. Para $h \to 0$ e pelo TCD, o lado esquerdo tende a a zero. Pelo mesmo argumento, $\frac{F(x)+F(x-)}{2} - \frac{F(x-h)+F(x-h-)}{2}$ tende a zero, para $h \to 0$. Deduzimos portanto que F é contínua.

(b) Agora, vamos mostrar que F é derivável. Usando (a), obtemos para $h > 0$

$$\frac{F(x+h)-F(x)}{h} = \frac{1}{2\pi}\int_{-\infty}^{\infty} \varphi(t)\frac{e^{-itx}-e^{-it(x+h)}}{ith}dt.$$

O integrando é limitado por $|\varphi(t)|$; pelo TCD conclua que

$$\lim_{h\to 0}\frac{F(x+h)-F(x)}{h} = \frac{1}{2\pi}\int_{-\infty}^{\infty} \varphi(t)e^{-itx}dx.$$

Podemos mostrar analogamente que $\lim_{h\to 0} \frac{F(x)-F(x-h)}{h}$ tem o mesmo limite. Obtemos assim que F' é derivável. Pela expressão da derivada obtida acima, vemos que F' é limitada e, por uma nova aplicação do TCD, que F' é contínua. $\square$

Corolário 7.4. *Se $\int_{\mathbb{R}} |\varphi(t)|dt < \infty$, então $F'(x)$ existe, é limitada e contínua, e*

$$F'(x) = \frac{1}{2\pi}\int_{\mathbb{R}} \varphi(t)e^{-itx}dt. \tag{7.8}$$

Aplicações

[1] Sabe-se que, se $0 < \alpha \leq 2$, então $\varphi(t) = e^{-|t|^{\alpha}}$ é uma f.c (na realidade, essa é a f.c de uma distribuição estável simétrica, veja o Capítulo 8). Pelo teorema, se X for simétrica e estável, então X tem uma densidade limitada e contínua.

[2] Suponha que $P_n \Rightarrow P$, e P_n, P tenham densidades f_n, f, respectivamente. Sabemos que não é necessariamente verdade que $f_n(x) \to f(x)$ q.c. Contudo, se $\int_{\mathbb{R}} |\varphi_n(t) - \varphi(t)| dt \to 0$, então $f_n(t) \to f(t)$.

O teorema limite central na sua forma mais simples decorre de uma aplicação das f.c's.

Teorema 7.9. *Sejam $X_1, X_2, \ldots$ v.a's i.i.d, $E(X_1) = 0$, $Var(X_1) = 1$. Então, $(X_1 + \ldots + X_n)/\sqrt{n} \stackrel{\mathcal{D}}{\to} N(0,1)$.*

Prova: Seja φ a f.c de X_1 e $S_n = X_1 + \ldots + X_n$. Se ψ_n é a f.c de $S_n/\sqrt{n}$, mostremos que $\psi_n(t) \to e^{-t^2/2}$. Temos que

$$\psi_n(t) = E\left[e^{itS_n/\sqrt{n}}\right] = \left[\varphi\left(\frac{t}{\sqrt{n}}\right)\right]^n =$$

$$= \left[1 + \varphi^{'}(0)\frac{t}{\sqrt{n}} + \frac{\varphi^{''}(0)}{2!}\Big(\frac{t}{\sqrt{n}}\Big)^2 + o\Big(\frac{1}{n}\Big)\right]^n,$$

pela independência dos X_i e usando expansão de Taylor. Mas, $\varphi^{'}(0) = iE(X_1) = 0$, $\varphi^{''}(0) = i^2 E(X_1^2) = -1$, de modo que $\psi_n(t) = [1 - t^2/(2n) + o(1/n)]^n \to e^{-t^2/2}$, quando $n \to \infty$. $\square$

Problemas

1. Prove que $\varphi(\mathbf{t})$ é uniformemente contínua.
2. Prove (7.3).
3. Prove (7.4).
4. Prove (7.5).
5. Prove que a f.c da distribuição de Cauchy padrão (densidade $[\pi(1+x^2)]^{-1}$) é $e^{-|t|}$.
6. Sejam P, Q probabilidades sobre $\mathbb{R}^k$. Prove que: (a) Se P for absolutamente contínua, então $P \star Q$ é absolutamente contínua; (b) Se P for não atômica, $P \star Q$ também não o será.
7. Prove que, se h for uma função integrável, então
$$\int h(x) dP \star Q(x) = \int\int h(x+y) dP(x) dQ(y).$$
8. (a) Se uma família Φ de f.c's sobre $\mathbb{R}$ for equicontínua no zero, então a família correspondente de medidas de probabilidade é fechada.

   (b) Seja $\{Q_n\}$ uma família de f.c's convergindo uniformemente numa vizinhança do zero. Prove que existe uma subsequência convergindo para uma f.c.

9. Suponha que $(X_1, \ldots, X_n)$ seja norma multivariada. Mostre que $E(X_n|X_1, \ldots, X_{n-1}) = \sum_{k=1}^{n-1} a_k X_k$, para constantes a_k.

   [Sugestão: Determine a_k por meio de $E\{(X_n - \sum_{k=1}^{n-1} a_k X_k)X_j\} = 0$.]

10. Seja $\{X_n, n \geq 1\}$ uma sequência de v.a's com f.c's φ_n. Suponha que $|\varphi_n(t)| \to 1$, para todo t, quando $n \to \infty$. Mostre que existem constantes a_n tais que $X_n - a_n$ converge para zero em lei.

    [Sugestão: Simetrize e tome $a_n = \text{mediana}\{X_n\}$.]

11. (a) Sejam $X_1, X_2, \ldots$ v.a's i.i.d, média zero e variância 1. Prove que

$$\frac{\sum_{i=1}^n X_i}{\sqrt{\sum_{i=1}^n X_i^2}} \xrightarrow{\mathcal{D}} N(0,1).$$

    (b) Supomha $X_n \sim$ binomial(n, p_n) e $np_n \to \lambda$. Prove que X_n converge em distribuição para $Y \sim$ Poisson(λ).

    (c) Para o TLC simples (Teorema 7.9), prove que $S_n/\sqrt{n}$ não converge em probabilidade, embora convergindo em distribuição.

    (d) Sejam X e Y independentes, cada uma normal com variância um. Prove que $X + Y$ e $X - Y$ são independentes, usando f.c's.

12. Prove o Corolário 7.2.

13. Prove que a família $\{P_n\}$ do Teorema 7.6, é fechada.

14. Prove a seguinte fórmula de inversão (mesmo método de prova do Teorema 7.7):

$$\frac{1}{2}[F(x) + F(x-)] = \frac{1}{2} + \lim_{c\to\infty, \delta\downarrow 0} \int_\delta^c \frac{e^{itx}\varphi(-t) - e^{-itx}\varphi(t)}{2\pi i t} dt.$$

15. Seja X uma v.a com f.c φ. Prove que, se $\varphi(t) \in L_2$ e se X tem densidade f, então $f \in L_2$ e

$$\int_{-\infty}^{\infty} f^2(x)dx = \frac{1}{2\pi}\int_{-\infty}^{\infty} |\varphi(x)|^2 dx.$$

    [Sugestão: Considere a f.c de $X - X'$, sendo X' independente de X e com a mesma distribuição que X.]

16. Suponha P probabilidade sobre $\mathbb{R}$ com f.c φ.

    (a) Se P for absolutamente contínua com respeito à medida de Lebesgue, então $\lim_{|t|\to\infty} \varphi(t) = 0$;

    (b) Se $|\varphi(t_1)| = 1$, para algum $t_1 \neq 0$, então P é concentrada em um conjunto de pontos da forma $x_n = a + n(2\pi/t_1)$.

    [Sugestão para (a): comece com o caso que a densidade de P é uma função simples. Sugestão para (b): existe θ_1, tal que $1 = e^{-i\theta_1}\varphi(t_1)$. Então, note que $0 = \int[1 - \cos(t_1 x - \theta_1)dP(x)$ e o integrando é não negativo.]

17. (Função de concentração de Lévy) Se P for uma probabilidade sobre $\mathbb{R}$, defina $Q_P(\varepsilon) = \sup_{x\in\mathbb{R}} P([x, x+\varepsilon])$. Prove que:

    (a) o supremo é atingido, $Q_P(\varepsilon)$ é crescente como função de ε e $\lim_{\varepsilon\uparrow\infty} Q_P(\varepsilon) = 1$.

    (b) Se $P = P_1 \star P_2$, então $Q_P(\varepsilon) \leq Q_{P_1}(\varepsilon) \wedge Q_{P_2}(\varepsilon)$.

Capítulo 8

Teoremas Limites Centrais

Um teorema limite central (TLC) é qualquer teorema que trata da convergência fraca de somas de v.a's apropriadamente normalizadas. Os teoremas mais conhecidos tratam da convergência de tais somas de variáveis independentes, satisfazendo certas condições. O caso mais simples, visto no capítulo anterior, trata do caso de v.a's i.i.d com variância finita. Nessas situações, a distribuição limite é a normal (ou gaussiana).

Para v.a's que tenham alguma forma de dependência, podemos ter TLC's sob condições de independência assintótica, também chamadas condições *mixing*. Por exemplo, temos TLC's para processos estacionários satisfazendo condições *mixing*.

Há situações em que a distribuição limite não é a normal. Por exemplo, veremos mais adiante, que uma soma normalizada de v.a's i.i.d converge, em distribuição, para uma v.a estável. Também, o máximo de um número finito de v.a's i.i.d, apropriadamente normalizado, tende para uma distribuição, chamada distribuição generalizada de valores extremos, que pode ser uma de três tipos: Gumbel, Weibull ou Fréchet.

A primeira versão de um TLC foi postulada por de Moivre, em 1733, que usou a distribuição normal como aproximação da distribuição de um número de caras, resultantes de lançamentos de uma moeda. Laplace, em 1812, estendeu o resultado de de Moivre, ao aproximar a distribuição binomial pela normal.

O termo "teorema limite central" foi usado pela primeira vez por Polya, em 1920, e ele se referia ao termo "central" como devido à sua importância em probabilidades. De acordo com L. Le Cam, a escola francesa interpretava o termo no sentido que "descrevia o comportamento do centro da distribuição, em oposição ao comportamento das caudas."

No Capítulo 9 trataremos do teorema de Donsker, que trata do limite de certos processos empíricos, às vezes denominado de TLC funcional.

8.1 Os Teoremas de Lindeberg e Feller

Para provar o Teorema de Lindeberg precisaremos dos seguintes lemas.

Lema 8.1. *Seja C_0 a classe das funções contínuas sobre $\mathbb{R}$ tais que $\lim_{x\to\infty} f(x)$ e $\lim_{x\to-\infty} f(x)$ existem em $\mathbb{R}$. Seja D a classe de todas as funções tendo derivadas contínuas e limitadas de qualquer ordem. Então, qualquer função em C_0 pode ser uniformemente aproximada por uma função de D (isto é, D é densa em C_0 na norma sup).*

Prova: Seja $f \in C_0$ e para $h > 0$ defina $f_h(x) = \int f(t)\phi_h(x-t)dt$, sendo ϕ_h a densidade da $N(0,h)$. Então, f_h tem derivadas contínuas e limitadas de qualquer ordem. Também,

$$|f_h(x) - f(x)| \leq \int |f(t) - f(x)|\phi_h(x-t)dt = \int |f(t-x) - f(x)|\phi_h(t)dt.$$

Tome δ tão pequeno de modo que $|f(t) - f(s)| \leq \varepsilon$, sempre que $|t-s| \leq 2\delta$. Separando a última integral acima em uma integral sobre $[-\delta,\delta]$ e a outra sobre o complementar desse intervalo, obtemos que a integral será menor ou igual a $\varepsilon + 2M[1 - \Phi_h(\delta) + \Phi_h(-\delta)]$, onde M é tal que $|f(x)| \leq M$, para todo x e Φ_h é a f.d da normal. Para $h \to 0$, $1 - \Phi_h(\delta) + \Phi_h(-\delta) \to 0$, logo $\sup_x |f_h(x) - f(x)| \leq \varepsilon$. □

Lema 8.2. *Sejam P_n, P medidas de probabilidade sobre $\mathbb{R}$ e suponha que $\int_{\mathbb{R}} f dP_n \to \int_{\mathbb{R}} f dP$, para toda f tendo derivadas limitas e contínuas de qualquer ordem. Então, $P_n \Rightarrow P$.*

Prova: Sejam F_n, F as f.d's correspondentes a P_n, P e seja x um ponto de continuidade de F. Seja $\delta > 0$ arbitrário e f uma função definida como segue: $f(t) = 1$, para $t \leq x$, f linear entre x e $x+\delta$ e $f(t) = 0$, para $t \geq x+\delta$. Seja f_ε uma função em D, tal que $\sup_x |f_\varepsilon(x) - f(x)| \leq \varepsilon$. Então,

$$\limsup_n F_n(x) \leq \limsup_n \int f dP_n \leq \limsup_n \int (\varepsilon + f_\varepsilon) dP_n$$

$$= \varepsilon + \limsup_n \int f_\varepsilon dP_n = \varepsilon + \int f_\varepsilon dP \leq 2\varepsilon + \int f dP \leq 2\varepsilon + F(x+\delta).$$

Para $\delta \to 0$, $\limsup_n F_n(x) \leq 2\varepsilon + F(x)$, e como $\varepsilon > 0$ arbitrário, obtemos $\limsup_n F_n(x) \leq F(x)$. Por um argumento similar, obtemos $\liminf_n F_n(x) \geq F(x)$, para x ponto de continuidade de F. Basta considerar f como acima, e os pontos $x-\delta$ e x na sua definição, no lugar de x e $x+\delta$. □

Teorema 8.1. (Lindeberg) *Para cada n, sejam $X_{n,1},\ldots,X_{n,k_n}$ v.a's independentes, com média zero, Var$(X_{n,j}) = \sigma^2_{n,j}$. Sejam $S_n = X_{n,1} + \ldots + X_{n,k_n}$ e $s_n^2 = \sum_{j=1}^{k_n} \sigma^2_{n,j}$. Suponha que*

$$\lim_{n\to\infty} \frac{1}{s_n^2} \sum_{j=1}^{k_n} \int_{\{|X_{n,j}|>\varepsilon s_n\}} X_{n,j}^2 \, dP = 0, \ \ \forall \varepsilon > 0. \tag{8.1}$$

Então, $S_n/s_n \xrightarrow{\mathcal{D}} N(0,1)$.

A equação (8.1) é chamada *condição de Lindeberg*.

Prova: Sejam $X_{n,1}, \ldots, X_{n,k_n}$ como no teorema e denotemos por N uma v.a com distribuição $\mathrm{N}(0,1)$. Pelo Lema 8.2, é suficiente provar que

$$E(f(S_n/s_n)) \to E(f(N)), \quad \text{para } f \in D, \tag{8.2}$$

sendo D a classe definida no Lema 8.1. A ideia da prova é: suponha que as v.a's $X_{n,i}$ fossem normais, cada uma $\mathrm{N}(0,\sigma_{n,i}^2)$. Então, S_n/s_n seria $N(0,1)$ e (8.2) valeria nesse caso. Suponha que $Y_{n,1}, \ldots, Y_{n,k_n}$ sejam independentes $N(0,\sigma_{n,i}^2)$, escolhidas de tal maneira que $X_{n1}, \ldots, X_{n,k_n}, Y_{n1}, \ldots, Y_{n,k_n}$ sejam independentes.

Vamos substituir, sucessivamente, em S_n, $X_{n,k_n}, X_{n,k_n-1}, \cdots$ por $Y_{n,k_n}, Y_{n,k_n-1}, \cdots$, de tal sorte que $E(f(S_n/s_n))$ seja substituída por $E((f(Y_{n,1}+\ldots+Y_{n,k_n})/s_n) = E(f(N))$.

Defina $g(t) = \sup_{x\in\mathbb{R}} |f(x+t) - f(x) - f^{'}(x)t - f^{''}(x)t^2/2|$. Então, por Taylor, $|g(t)| \le M_1|t|^3$. Também, $|g(t)| \le M_2 t^2$, pois $g(t) \le \sup_x |f(x+t) - f(x) - f^{'}(x)t| + \sup_x |f^{''}t^2/2|$. Segue que $g(t) \le M(t^2 \wedge |t|^3)$. Note que

$$|f(x+t_1) - f(x+t_2) - f^{'}(x)(t_1-t_2) - f^{''}(x)(t_1^2 - t_2^2)/2| \le g(t_1) + g(t_2). \tag{8.3}$$

Defina

$$Z_{n,k} = \sum_{1\le j<k} X_{n,j} + \sum_{k<j\le k_n} Y_{n,j}.$$

Observe que $Z_{n,k_n} + X_{n,k_n} = S_n$ e $Z_{n,1} + Y_{n,1} \ \sim \ N(0, s_n^2)$.

Considere

$$\left| Ef\left(\frac{S_n}{s_n}\right) - Ef\left(\frac{Y_{n,1}+\ldots+Y_{n,k_n}}{s_n}\right)\right| \le \sum_{k=1}^{k_n} \left| Ef\left(\frac{Z_{n,k}+X_{n,k}}{s_n}\right) - Ef\left(\frac{Z_{n,k}+Y_{n,k}}{s_n}\right)\right|$$

$$\le \sum_{k=1}^{k_n} Eg\left(\frac{X_{n,k}}{s_n}\right) + \sum_{k=1}^{k_n} Eg\left(\frac{Y_{n,k}}{s_n}\right),$$

usando (8.3) e os cálculos seguintes:

$$E(f^{'}(Z_{n,k})(X_{n,k} - Y_{n,k})) = Ef^{'}(Z_{n,k})E(X_{n,k} - Y_{n,k}) = 0,$$

usando a independência de $X_{n,k}, Y_{n,k}$ de $Z_{n,k}$ e $E(X_{n,k} - Y_{n,k}) = 0$. De modo similar, obtemos $E(f^{''}(Z_{n,k})(X_{n,k}^2 - Y_{n,k}^2)) = 0$, notando que $E(X_{n,k}^2 - Y_{n,k}^2) = 0$.

Para terminar a prova, mostraremos que cada uma das somas acima tende a zero. Para a primeira,

$$\sum_{k=1}^{k_n} Eg\left(\frac{X_{n,k}}{s_n}\right) = \sum_{k=1}^{k_n} \int_{\{|X_{n,k}| \le s_n\varepsilon\}} g\left(\frac{X_{n,k}}{s_n}\right)\, dP + \sum_{k=1}^{k_n} \int_{\{|X_{n,k}| > s_n\varepsilon\}} g\left(\frac{X_{n,k}}{s_n}\right)\, dP$$

$$\le M \sum_{k=1}^{k_n} \int_{\{|X_{n,k}| \le s_n\varepsilon\}} \frac{|X_{n,k}|^3}{s_n^3}\, dP + M \sum_{k=1}^{k_n} \int_{\{|X_{n,k}| > s_n\varepsilon\}} \frac{X_{n,k}^2}{s_n^2}\, dP \le$$

$$\le M\varepsilon + M \sum_{k=1}^{k_n} \int_{\{|X_{n,k}| > s_n\varepsilon\}} \frac{X_{n,k}^2}{s_n^2}\, dP,$$

para todo $\varepsilon > 0$. Tomando $n \to \infty$ e em seguida $\varepsilon \to 0$ obtemos que a primeira soma converge para 0.

Para a segunda soma, escrevendo a integral como a soma de duas integrais, como no caso anterior, ou seja uma sobre $\{|Y_{n,k}| \le \varepsilon s_n\}$ e a outra sobre o complementar desse conjunto, obtemos que

$$\sum_{k=1}^{k_n} Eg\left(\frac{Y_{n,k}}{s_n}\right) \le \varepsilon M + M \sum_{k=1}^{k_n} \int_{\{|Y_{n,k}| > \varepsilon s_n\}} \frac{Y_{n,k}^2}{s_n^2}\, dP.$$

Mas,

$$\sum_{k=1}^{k_n} \int_{\{|Y_{n,k}| > \varepsilon s_n\}} \frac{Y_{n,k}^2}{s_n^2}\, dP \le \frac{1}{\varepsilon} \sum_{k=1}^{k_n} \int \frac{|Y_{n,k}|^3}{s_n^3}\, dP = \frac{C}{\varepsilon} \sum_{k=1}^{k_n} \frac{\sigma_{n,k}^3}{s_n^3},$$

pois $E(|Y_{n,k}|^3) = C\sigma_{n,k}^3$, onde C é uma constante absoluta. Logo a última parcela da relação anterior

$$\frac{C}{\varepsilon} \sum_{k=1}^{k_n} \frac{\sigma_{n,k}^3}{s_n^3} \le \frac{C}{\varepsilon} \max_{k \le k_n} \frac{\sigma_{n,k}}{s_n} \sum_{k=1}^{k_n} \frac{\sigma_{n,k}^2}{s_n^2},$$

notando que a soma é igual a um. Logo, é suficiente mostrar que $\max_k \frac{\sigma_{n,k}}{s_n} \to 0$, para $n \to \infty$. Mas, $\frac{\sigma_{n,k}^2}{s_n^2} = \frac{1}{s_n^2} E(X_{n,k}^2)$ e quebrando a integral em duas, uma sobre $\{|X_{n,k}| \le s_n\delta\}$ e outra sobre o complementar, obtemos que

$$\max_{k \le k_n} \frac{\sigma_{n,k}^2}{s_n^2} \le \delta^2 + \max_{k \le k_n} \frac{1}{s_n^2} \int_{\{|X_{n,k}| > s_n\delta\}} |X_{n,k}|^2 dP,$$

sendo que o segundo termo tende a zero por (8.1). Como $\delta > 0$ é arbitrário, o resultado segue. □

Exemplo 8.1. [1] Sejam $\{X_n, n \geq 1\}$ v.a's i.i.d, média zero e variância comum σ^2. Então, $(X_1 + \ldots + X_n)/(\sigma\sqrt{n}) \xrightarrow{\mathcal{D}} N(0,1)$.

De fato, temos que nesse caso, $X_{n,j} = X_j,\ k_n = n,\ s_n^2 = n\sigma^2$ e a condição de Lindeberg fica

$$\frac{1}{\sigma^2}\int_{\{|X_1|>\sigma\sqrt{n}\varepsilon\}} |X_1|^2 dP \to 0.$$

[2] (Teorema de Lyapunov) Com a mesma notação do Teorema 8.1, suponha que

$$\frac{1}{s_n^{2+\delta}}\sum_{j=1}^{k_n} E(|X_{n,j}|^{2+\delta}) \to 0, \text{ para algum } \delta > 0. \tag{8.4}$$

Então, $S_n/s_n \xrightarrow{\mathcal{D}} N(0,1)$.

Basta observar que

$$\frac{1}{s_n^2}\sum_{j=1}^{k_n}\int_{\{|X_{n,j}|>\varepsilon s_n\}} X_{n,j}^2\, dP \leq \frac{1}{s_n^{2+\delta}\varepsilon^{\delta}}\sum_{j=1}^{k_n} E(|X_{n,j}|^{2+\delta}) \to 0,$$

para cada $\varepsilon > 0$, pois $|X_{n,j}|^{\delta}/(\varepsilon^{\delta}s_n^{\delta}) > 1$.

Na condição de Lindeberg (8.1), substitua $X_{n,i}$ por $X_{n,i}/s_n = Y_{n,i},\ i = 1,\ldots,k_n$. Note que $\sum_{i=1}^{k_n}\mathrm{Var}(Y_{n,i}) = 1$. Obtemos, então, a seguinte reformulação do Teorema 8.1. Suponha que

$$\sum_{i=1}^{k_n}\int_{\{|Y_{n,i}|>\varepsilon\}} Y_{n,i}^2\, dP \to 0, \quad n \to \infty. \tag{8.5}$$

Então, $\sum_{i=1}^{k_n} Y_{n,i} \xrightarrow{\mathcal{D}} N(0,1)$.

Note que se (8.5) vale, então

$$\lim_{n\to\infty}\max_{k\leq k_n} P\{|Y_{n,k}| > \varepsilon\} \to 0, \tag{8.6}$$

pois $\max_k P\{|Y_{n,k}| > \varepsilon\} \leq \frac{1}{\varepsilon^2}\sum_{i=1}^{k_n}\int_{\{|Y_{n,i}|>\varepsilon\}} Y_{n,i}^2\, dP \to 0$, por (8.5), sendo que a desigualdade segue de Chebyshev.

Se as v.a's $Y_{n,1}, Y_{n,2}, \ldots$ satisfazem (8.6), elas são chamadas:

(i) u.a.n (*uniformly asymptotically negligible*) (Loève);

(ii) *null array* (Feller);

(iii) *holouspoudic* (Chung).

É natural perguntar se a condição (8.1) do teorema de Lindeberg é também uma condição necessária para ter o TLC. A resposta é em geral negativa mas sob a hipótese adicional (8.6) temos o teorema a seguir.

Teorema 8.2. (Feller) *Sejam $X_{n,1}, \ldots, X_{n,k_n}$ v.a's independentes, de média zero, $Var(X_{n,i}) = \sigma^2_{n,i}$, com $\sum_i \sigma^2_{n,i} = 1$. Seja $S_n = \sum_{i=1}^{k_n} X_{n,i}$. Suponha que:*

(1) $S_n \stackrel{\mathcal{D}}{\to} N(0,1)$;

(2) $\lim_{n\to\infty} \max_{k\le k_n} P\{|X_{n,k}| > \varepsilon\} = 0, \ \ \forall \varepsilon > 0.$

Então, $\sum_{k=1}^{k_n} \int_{\{|X_{n,k}|>\varepsilon\}} X^2_{n,k} dP \to 0, \ \ n \to \infty.$

Prova: A condição (1) implica que

$$\prod_{k=1}^{k_n} \varphi_{n,k}(t) \to e^{-t^2/2}, \tag{8.7}$$

quando $n \to \infty$, sendo $\varphi_{n,k}$ a f.c de $X_{n,k}$.

Provemos, agora, que (2) implica

$$\lim_{n\to\infty} \max_{k\le k_n} |\varphi_{n,k}(t) - 1| \to 0, \quad \text{para cada } t. \tag{8.8}$$

De fato, sendo $P_{n,k}$ a lei de $X_{n,k}$,

$$|\varphi_{n,k}(t) - 1| \le \int |e^{itx} - 1| dP_{n,k}(x) = \int_{\{|x|>\varepsilon\}} |e^{itx} - 1| dP_{n,k}(x) + \int_{\{|x|\le\varepsilon\}} |e^{itx} - 1| dP_{n,k}(x)$$

$$\le 2P\{|X_{n,k}| > \varepsilon\} + \int_{\{|x|\le\varepsilon\}} |tx| dP_{n,k}(x) \le 2P\{|X_{n,k}| > \varepsilon\} + |t|\varepsilon.$$

Logo, temos

$$\limsup_{n\to\infty} \max_{k\le k_n} |\varphi_{n,k}(t) - 1| \le 2 \lim_{n\to\infty} \max_{k\le k_n} P\{|X_{n,k}| > \varepsilon\} + |t|\varepsilon = |t|\varepsilon.$$

Tomando $\varepsilon \to 0$, obtemos (8.8). De (8.8), deduzimos que existe um inteiro $N(t)$, tal que para $n \ge N(t)$, temos $\sup_k |\varphi_{n,k}(t) - 1| \le 1/2$. Logo, podemos escrever, usando (8.7):

$$\sum_{k=1}^{k_n} \log \varphi_{n,k}(t) \to -t^2/2, \tag{8.9}$$

na qual os logaritmos são tomados com argumento em $(-\pi, \pi]$, e

$$\log \varphi_{n,k}(t) = [\varphi_{n,k}(t) - 1] + M|\varphi_{n,k}(t) - 1|^2, \tag{8.10}$$

onde M tem valor complexo e é limitada por 2 em valor absoluto. De fato, pela expansão de Taylor do valor principal de $\log z$, $\log z = \sum_{l\geq 1} \frac{(-1)^{l-1}}{l}(z-1)^l$, que é válida para $|z-1| < 1$. Logo, como $|\varphi_{n,k}(t) - 1| \leq 1/2$, obtemos

$$|\log \varphi_{n,k}(t) - (\varphi_{n,k}(t) - 1)| \leq \sum_{l\geq 2} \frac{1}{k}|\varphi_{n,k}(t) - 1|^l \leq |\varphi_{n,k}(t) - 1|^2 \sum_{l\geq 0} \frac{1}{2^l}.$$

Temos também,

$$\sum_{k=1}^{k_n} |\varphi_{n,k}(t) - 1|^2 \leq \max_{k\leq k_n} |\varphi_{n,k}(t) - 1| \sum_{k=1}^{k_n} |\varphi_{n,k}(t) - 1|. \tag{8.11}$$

Mas

$$\sum_{k=1}^{k_n} |\varphi_{n,k}(t) - 1| \leq \sum_{k=1}^{k_n} \int |e^{itx} - 1| dP_{n,k}(x) \leq \sum_{k=1}^{k_n} \int \frac{t^2x^2}{2} dP_{n,k}(x) = t^2/2,$$

pois $\sum_{k=1}^{k_n} \sigma_{n,k}^2 = 1$. Logo, $\sum_{k=1}^{k_n} |\varphi_{n,k}(t) - 1|^2 \to 0$, para $n \to \infty$, por (8.11) e (8.8).

Usando esse fato, (8.9) e (8.10), obtemos que

$$\sum_{k=1}^{k_n} [\varphi_{n,k}(t) - 1] \to -t^2/2 \tag{8.12}$$

quando $n \to \infty$. Tome a parte real de (8.12) para obter

$$\sum_{k=1}^{k_n} \int [1 - \cos(tx)] dP_{n,k}(x) \to -t^2/2.$$

Portanto,

$$\limsup_n \left| \frac{t^2}{2} - \sum_{k=1}^{k_n} \int_{\{|x|\leq\varepsilon\}} [1-\cos(tx)] dP_{n,k}(x) \right| = \limsup_n \left| \sum_{k=1}^{k_n} \int_{\{|x|>\varepsilon\}} [1-\cos(tx)] dP_{n,k}(x) \right|$$

$$\leq \limsup_n \left| \sum_{k=1}^{k_n} \int_{\{|x|>\varepsilon\}} 2dP_{n,k}(x) \right| \leq 2 \limsup_n \sum_{k=1}^{k_n} \int \frac{|x|^2}{\varepsilon^2} dP_{n,k}(x) = \frac{2}{\varepsilon^2}.$$

Segue que

$$\frac{2}{\varepsilon^2} \geq \limsup_n \left| \frac{t^2}{2} - \sum_{k=1}^{k_n} \int_{\{|x|\leq\varepsilon\}} [1 - \cos(tx)] dP_{n,k}(x) \right|$$

$$\geq \limsup_n \left(\frac{t^2}{2} - \sum_{k=1}^{k_n} \int_{\{|x|\leq\varepsilon\}} \frac{t^2x^2}{2} dP_{n,k}(x) \right).$$

Então,

$$\frac{4}{\varepsilon^2 t^2} \geq \limsup_n \Big(1 - \sum_{k=1}^{k_n} \int_{\{|x|\leq\varepsilon\}} x^2 dP_{n,k}(x) \Big) \geq 0.$$

Faça $t \to \infty$, para obter $\sum_{k=1}^{k_n} \int_{\{|x|\leq\varepsilon\}} x^2 dP_{n,k}(x) \to 1$. Conclui-se que a soma

$$\sum_{k=1}^{k_n} \int_{\{|x|>\varepsilon\}} x^2 dP_{n,k}(x) \to 0,$$

pois $\sum_{k=1}^{k_n} \int x^2 dP_{n,k}(x) = 1$. $\square$

O seguinte teorema foi provado independentemente por Berry (1941) e Esseen (1942). Veja Feller (1966) para uma prova.

Teorema 8.3. (Berry-Esseen) *Sejam $\{X_n, n \geq 1\}$ v.a's i.i.d, de média zero e variância σ^2 e suponha $E(|X_1|^3) < \infty$. Seja F_n a f.d de $(X_1 + \ldots + X_n)/(\sigma\sqrt{n})$ e Φ a f.d de uma normal padrão. Então,*

$$\sup_{x\in\mathbb{R}} |F_n(x) - \Phi(x)| \leq \frac{33}{4} \frac{E(|X_1|^3)}{\sigma^3} \frac{1}{\sqrt{n}}.$$

O teorema é caso particular de um resultado mais geral. Seja $\Delta_n(x) = |F_n(x) - \Phi(x)|$. Sob as condições do teorema (suponha $\sigma = 1$), existe uma constante absoluta $C_0(\delta)$, para $\delta \in (0, 1]$, tal que

$$\sup_x \Delta_n(x) \leq C_0(\delta) L_n^{2+\delta}, \quad \text{onde } L_n^{2+\delta} = \frac{E(|X_1|^{2+\delta})}{n^{\delta/2}}.$$

Observe que o teorema anterior é um caso particular para $\delta = 1$. Vários trabalhos subsequentes foram provados no sentido de tornar mais preciso o limite superior do resultado. Veja Korolev e Shevtsova (2010) para uma resenha histórica.

8.2 Distribuições infinitamente divisíveis

Vamos considerar os seguintes exemplos:

(a) Sejam $\{X_n, n \geq 1\}$ v.a's i.i.d, média zero e variância 1. Então,

$$\frac{X_1 + \ldots + X_n}{\sqrt{n}} \xrightarrow{\mathcal{D}} N(0,1), \quad \text{f.c. } e^{-t^2/2}.$$

(b) Sejam $\{X_n, n \geq 1\}$ v.a's i.i.d, média μ. Então,

$$\frac{X_1 + \ldots + X_n}{n} \xrightarrow{\mathcal{D}} \mu, \quad \text{f.c. } e^{it\mu}.$$

(c) Suponha que X_n tenha distribuição binomial, com parâmteros n e $p = \lambda/n$, com $\lambda > 0$. Então, sabemos que $X_n \xrightarrow{\mathcal{D}} P(\lambda)$, ou seja, uma Poisson com parâmetro λ, e f.c $e^{\lambda(e^{it}-1)}$. Se, para cada n, considerarmos $X_{n,1}, \ldots, X_{n,n}$ independentes, cada uma Bernoulli, com $p = \lambda/n$, então $X_n \sim X_{n,1} + \ldots + X_{n,n}$ e teremos $X_{n,1} + \ldots + X_{n,n} \xrightarrow{\mathcal{D}} P(\lambda)$.

Esses exemplos são instâncias da seguinte situação. Temos um arranjo triangular

$$\begin{array}{lll} X_{1,1} & & \\ X_{2,1}, & X_{2,2} & \\ X_{3,1}, & X_{3,2}, & X_{3,3} \\ \cdots & & \end{array}$$

onde, para cada n, $X_{n,1}, \ldots, X_{n,n}$ são i.i.d. Seja $S_n = \sum_{i=1}^{n} X_{n,i}$. Em cada um dos exemplos acima, S_n converge para alguma v.a. Quais outras variáveis aparecem em situações como essas?

Suponha que $S_n \xrightarrow{\mathcal{D}} X$. Seja φ f.c de X. Temos que $S_{2n} \xrightarrow{\mathcal{D}} X$, mas $S_{2n} = (X_{2n,1} + \ldots + X_{2n,n}) + (X_{2n,n+1} + \ldots + X_{2n,2n}) =: Y_n + Y_n^{'}$. As v.a's $\{Y_n\}$ formam uma família fechada, pois

$$\begin{aligned} P\{|Y_n| > K\}^2 &\leq 2[P\{Y_n \geq K\}^2 + P\{Y_n < -K\}^2] \\ &= 2[P\{Y_n \geq K, Y_n' \geq K\} + P\{Y_n < -K, Y_n' < -K\}] \\ &\leq 2P\{|S_{2n}| > K\} \end{aligned}$$

e como S_{2n} converge em lei, a família é fechada. Portanto, usando o Teorema de Prokhorov, existe uma subsequência $\{n_k\}$ tal que Y_{n_k} converge em lei para Y e $Y_{n_k}^{'}$ converge em lei para $Y^{'}$, e pela independência $Y_{n_k} + Y_{n_k}^{'}$ converge em lei para $Y + Y^{'}$, com Y e Y' independentes. Como $Y_{n_k} + Y_{n_k}^{'}$ converge em lei para X, temos que $X \sim Y + Y^{'}$.

Seja φ_Y a f.c de Y. Segue-se que $\varphi(t) = [\varphi_Y(t)]^2$. De modo similar, $\varphi(t) = [\varphi_Z(t)]^k$, para alguma v.a Z, e cada k.

Definição 8.1. *Uma v.a X diz-se infinitamente divisível se, para cada n, existe uma f.c φ_n, tal que $\varphi(t) = [\varphi_n(t)]^n$, sendo φ a f.c de X.*

De modo equivalente, podemos dizer que X é infinitamente divisível se, para cada n, existem v.a's $X_{n,1}, \ldots, X_{n,n}$, que são i.i.d e tais que X tem a mesma distribuição que $X_{n,1} + \ldots + X_{n,n}$.

Exemplo 8.2. São infinitamente divisíveis as distribuições:
(1) Normal
(2) Cauchy
(3) Poisson
(4) exponencial
(5) Gama

Teorema 8.4. *X é infinitamente divisível se, e somente se, X for o limite em distribuição de uma soma $S_n = X_{n,1} + \ldots + X_{n,n}$ de v.a's i.i.d.*

Prova: ($\Leftarrow$) feito acima.
($\Rightarrow$) Óbvio, pois se X for infinitamente divisível, então, para cada n, $X \sim X_{n,1} + \ldots + X_{n,n}$. $\square$

Teorema 8.5. *A classe das distribuições infinitamente divisíveis é fracamente sequencialmente fechada (isto é, se $P_n \Rightarrow P$, e se P_n for infinitamente divisível, então P também o será).*

Prova: Para cada n, sejam $X_{n,1}, \ldots, X_{n,n}$ infinitamente divisíveis, tais que $\sum_{i=1}^n X_{n,i}$ tenha P_n como sua distribuição. Como $P_n \Rightarrow P$, $\sum_{i=1}^n X_{n,i} \xrightarrow{\mathcal{D}} X$, onde X tem distribuição P. Logo, P é infinitamente divisível, pelo teorema anterior. $\square$

Teorema 8.6. *Seja φ a f.c de uma distribuição infinitamente divisível. Então, $\varphi(t) \neq 0$, para todo $t \in \mathbb{R}$.*

Prova: É suficiente mostrar que $|\varphi(t)|^2 \neq 0$, para todo t. Observe que $|\varphi(t)|^2$ é a f.c de uma distribuição infinitamente divisível (i.d). De fato, seja X i.d com f.c φ e X' independente de X e com a mesma distribuição que X. Então, $X - X'$ é i.d e sua f.c é $|\varphi(t)|^2$.

Seja $g(t) = |\varphi(t)|^2$ e seja $h_n(t)$ a n-ésima raiz real de $g(t)$: $h_n(t) = [g(t)]^{1/n}$. Então, $h_n(t)$ é uma f.c e $\lim_{n\to\infty} h_n(t)$ existe e é igual a zero, se e somente se $g(t) = 0$ e igual a 1, caso contrário. Também, como g é uma f.c, existe $\varepsilon > 0$ tal que $g(t) > 0$, para todo $|t| < \varepsilon$ (pois g é contínua na origem). Segue que $\lim_{n\to\infty} h_n(t) = 1$, para $|t| < \varepsilon$. Se $h(t)$ é o limite, $h(t)$ é contínua no zero e portanto é uma f.c. Logo, h é contínua, donde $h(t) = 1$, para todo t. Logo, $g(t)$ não pode ser zero. $\square$

Definição 8.2. *Seja X uma v.a com f.c φ. Sejam $X_1, X_2, \ldots$ v.a's i.i.d, $X_i \sim X$, e seja $X_0 = 0$. Seja N uma v.a independente de X_i, para todo i, tendo distribuição de Poisson, $P(\lambda)$. Defina $Y = \sum_{i=0}^N X_i$. Dizemos que Y tem distribuição de Poisson composta.*

A f.c de Y é dada por $\psi(t) = e^{\lambda[\varphi(t)-1]}$. Além disso, Y é infinitamente divisível. Veja o Problema 2.

Para o resultado a seguir, necessitamos de alguns fatos sobre funções complexas. Seja S um espaço topológico e $f : S \to \mathbb{C}$, contínua. Dizemos que f tem logaritmo contínuo se existir uma função contínua $g : S \to \mathbb{C}$ tal que $f = e^g$. A função g é única sobre cada componente conectado de S, a menos de uma constante da forma $2\pi i m$, $m \in \mathbb{Z}$. O resultado vale para $S = \mathbb{R}$ e f contínua, não nula.

Teorema 8.7. *X é infinitamente divisível se, e somente se, X é o limite em distribuição de uma sequência de distribuições de Poisson compostas.*

Prova: ($\Leftarrow$) Segue do Problema 2 e Teorema 8.5.

($\Rightarrow$) A f.c de X, $\varphi(t)$, não se anula nunca, pois X é infinitamente divisível. Logo φ admite um logaritmo contínuo g, ou seja, $\varphi(t) = e^{g(t)}$. Como $\varphi(0) = 1$ e como g é única a menos de uma constante, $2\pi i m$, podemos escolher uma g única com $g(0) = 0$.

Também sabemos que $\varphi(t) = [\varphi_n(t)]^n$, para cada n, onde φ_n é uma f.c; $\varphi_n(t)$ nunca se anula, logo pelo mesmo argumento, existe um único logaritmo contínuo g_n, tal que $\varphi_n(t) = e^{g_n(t)}$ e $g_n(0) = 0$.

Note que e^{g_n} é uma n-ésima raiz de φ, logo pela unicidade do logaritmo contínuo, obtemos $g = ng_n + 2\pi i m$ para algum m inteiro. Como $g(0) = g_n(0) = 0$, segue-se que $m = 0$. Portanto,

$$\lim_{n\to\infty} e^{n(\varphi_n - 1)} = \lim_{n\to\infty} e^{n[e^{g/n} - 1]} = e^g,$$

usando $e^z - 1 \sim z$ quando $z \to 0$. Mas $e^{n(\varphi_n - 1)}$ é a f.c de uma distribuição de Poisson composta. $\square$

Queremos encontrar a forma geral da f.c de uma distribuição infinitamente divisível. Sabemos que, se X for infinitamente divisível e φ é a sua f.c, $[\varphi_n]^n = \varphi$, então $e^{n(\varphi_n - 1)} \to e^g = \varphi$, ou seja, $n(\varphi_n - 1) \to \log \varphi$. Portanto, podemos escrever

$$\lim_{n\to\infty} \int (e^{itx} - 1)\, n dF_n(x) = \log \varphi(t),$$

onde F_n é a f.d correspondente a φ_n.

Como $P\{|X_{n,k}| > \varepsilon\} \to 0$, as medidas $ndF_n(x)$ colocam mais e mais massa no zero, quando $n \to \infty$. Para estudar o limite acima vamos considerar a seguinte decomposição

$$\lim_{n\to\infty} \int \left(e^{itx} - 1 - \frac{itx}{1+x^2}\right) \frac{1+x^2}{x^2} dG_n(x) + it \int \frac{x}{x^2+1} ndF_n(x),$$

chamando $\frac{x^2}{1+x^2} ndF_n(x) = dG_n(x)$, e notando que o termo entre parêntesis na primeira integral é da ordem de t^2x^2 perto de zero e a segunda integral é aproximadamente igual a $it \int x^{-1} dG_n(x)$.

O resultado a seguir dá a representação da f.c de uma distribuição infinitamente divisível.

Teorema 8.8. (Lévy-Khintchine)
(a) X é infinitamente divisível, com f.c φ se, e somente se $\varphi = e^{\psi}$, onde

$$\psi(t) = it\gamma + \int_{-\infty}^{\infty} \Big(e^{itx} - 1 - \frac{itx}{1+x^2}\Big)\frac{1+x^2}{x^2} dG(x), \tag{8.13}$$

onde γ é uma constante real e G é uma função crescente e de variação limitada.

(b) A f.c φ determina univocamente γ e G, isto é, a representação de Lévy-Khintchine é única.

Observação: O integrando em (8.13) é definido como $-t^2/2$ em $x = 0$. Se G coloca massa σ^2 em zero, podemos escrever

$$\psi(t) = it\gamma - \frac{\sigma^2 t^2}{2} + \int_{-\infty}^{\infty} \Big(e^{itx} - 1 - \frac{itx}{1+x^2}\Big)\frac{1+x^2}{x^2} d\hat{G}(x), \tag{8.14}$$

onde $\hat{G}$ é uma medida sem massa no zero. Outra maneira de escrever é

$$\psi(t) = it\gamma - \frac{\sigma^2 t^2}{2} + \int_{-\infty}^{\infty} \Big(e^{itx} - 1 - \frac{itx}{1+x^2}\Big)\nu(dx), \tag{8.15}$$

onde ν é chamada *medida de Lévy*.

Se X satisfaz (8.13), escreveremos $X \sim (\gamma, G)$.

Prova do Teorema: (a) ($\Rightarrow$) Seja φ a f.c de X, então $\varphi(t) = [\varphi_n(t)]^n$. Também, $\varphi(t) = e^g$, $\varphi_n(t) = e^{g_n}$ e $n[\varphi_n(t) - 1] \to g(t)$.

Defina

$$H_n(t) = \int_{-\infty}^{t} \frac{x^2}{1+x^2} n dF_n(x),$$

na qual F_n é a f.d correspondente a φ_n.

Antes de prosseguir, vamos considerar os seguintes fatos.

[1] Temos que

$$\limsup_{n\to\infty} nP_n\{[-a,a]^c\} \le a \int_{-2/a}^{2/a} |g(t)| dt,$$

sendo P_n a distribuição de probabilidade de F_n.

De fato, usando a prova do Lema 7.4,

$$nP_n\{[-a,a]^c\} \le a \int_{-2/a}^{2/a} n|\varphi_n(t) - 1| dt.$$

Mas $n[|\varphi_n(t) - 1|] \to g(t)$, logo pelo TCD,

$$\int_{-2/a}^{2/a} n[|\varphi_n(t)-1|]dt \to \int_{-2/a}^{2/a} |g(t)|dt.$$

[2] Existe uma constante A tal que

$$\int_{-1}^{1} x^2 n dF_n(x) \leq A, \quad \forall\, n.$$

De fato, temos que para todo n,

$$n[1-\mathcal{R}(\varphi_n(1))] \geq \int_{-1}^{1}[1-\cos x] n dF_n(x) \geq \frac{1}{\pi}\int_{-1}^{1} x^2 n dF_n(x),$$

sendo que na última integral usamos que $1-\cos x \geq x^2/\pi$ para $|x| \leq 1$. Mas $n[1-\mathcal{R}(\varphi_n(1))] \to -\mathcal{R}(g(1)) \geq 0$, logo existe uma constante A tal que $A \geq \int_{-1}^{1} x^2 n dF_n(x)$, para todo n.

[3] Seja $H_n(\infty) = \int x^2/(1+x^2) n dF_n(x)$. Por [1] e [2], $\{H_n(\infty), n \geq 1\}$ é limitada. Defina $G_n(t) = H_n(t)/H_n(\infty)$, de modo que G_n é uma medida de probabilidade. Além disto, se $\liminf_n H_n(\infty) > 0$, por [1], temos que $\{G_n, n \geq 1\}$ é fechada. Logo, pelo Teorema de Prokhorov, existe uma subsequência $\{n_k\}$ e uma distribuição de probabilidade G tal que $G_{n_k} \to G$ em pontos de continuidade de G. Podemos também escolher n_k tal que $H_{n_k}(\infty) \to \liminf_n H_n(\infty) = L \in (0,\infty)$.

Portanto, temos que

$$\log\varphi(t) = g(t) = \lim_{n_k\to\infty} n[\varphi_n(t)-1] = \lim_{n_k\to\infty}\int [e^{itx}-1] n_k dF_{n_k}(x) =$$

$$= \lim_{n_k\to\infty}\left(H_{n_k}(\infty)\left\{\int\left[e^{itx}-1-\frac{itx}{1+x^2}\right]\frac{1+x^2}{x^2} dG_{n_k}(x)\right\} + it\gamma_{n_k}\right).$$

Como $G_{n_k} \to G$ e o integrando é contínuo e limitado, esse tende para uma integral com dG_{n_k} substituída por dG. Como $H_{n_k}(\infty) \to L$, uma parte do limite em questão resulta $L\{\int[e^{itx}-1-\frac{itx}{1+x^2}]\frac{1+x^2}{x^2}dG(x)\}$, logo $\gamma_{n_k} \to \gamma$, para algum $\gamma \in \mathbb{R}$. No caso em que $\liminf_n H_n(\infty) = 0$, consideramos uma subsequência $\{n_k\}$ tal que $\lim_k H_{n_k}(\infty) = 0$. Neste caso o primeiro termo no limite acima vai para zero, o que força a convergência da sequência γ_{n_k} para um valor $\gamma \in \mathbb{R}$. Obtemos assim $\log\varphi(t) = it\gamma$.

($\Leftarrow$) Suponha que X tenha f.c. $\varphi = e^{\psi}$, com (8.13) válida. Mostraremos que X é i.d. Escreva ψ como um limite de somas da forma

$$\psi_n = \sum_k \left[e^{ita_k}-1-\frac{ita_k}{1+a_k^2}\right]\frac{1+a_k^2}{a_k^2}[G(a_k)-G(a_{k-1})].$$

Esse é o logaritmo da f.c de uma soma de v.a's independentes, com distribuições de Poisson compostas. O limite de e^{ψ_n} é e^{ψ}, que é contínua no zero, logo e^{ψ} é uma f.c. Segue que X é i.d., pois é o limite em distribuição de v.a's i.d's (uma soma de v.a's independentes e com distribuições de Poisson compostas é i.d.).

(b) (Unicidade) Seja $\varphi = e^g$, onde g tem a forma (8.13). Queremos provar que g determina γ e G univocamente. Defina

$$h(t) = \int_{t-1}^{t+1} g(x)dx - 2g(t).$$

Então, g determina h univocamente. Defina, agora,

$$H(t) = 2\int_{-\infty}^{t} \Big(1 - \frac{\sin x}{x}\Big)\frac{1+x^2}{x^2} dG(x).$$

Então, $h(t) = \int e^{itx} dH(x)$, logo h determina H univocamente (pois h é a transformada de Fourier de H). Mas o integrando em $H(t)$ é positivo, logo H determina G univocamente. Segue que g determina G e portanto γ. $\square$

Exemplo 8.3. (a) Se $X \sim N(0, \sigma^2)$, então G coloca massa pontual σ^2 no zero, e $\gamma = 0$. Se $E(X) = \mu$, então $\gamma = \mu$.

(b) Se $X \sim P(\lambda)$, então G tem massa pontual de tamanho $\lambda/2$ em 1 e $\gamma = \lambda/2$.

Teorema 8.9. (da continuidade) *Seja X_n infinitamente divisível com parâmetros (γ_n, G_n) e X com parâmetros (γ, G). Então, $X_n \xrightarrow{\mathcal{D}} X$ se, e somente se, $\gamma_n \to \gamma$, $G_n \to G$ nos pontos de continuidade de G e para todo $\varepsilon > 0$, existe $a > 0$ tal que $\sup_n G_n([-a, a]^c) < \varepsilon$.*

Prova: ($\Leftarrow$) Imediata
($\Rightarrow$) Se $X_n \xrightarrow{\mathcal{D}} X$, então $\varphi_n(t) \to \varphi(t)$, para todo t; como φ_n, φ não são nunca nulas, temos que $\log \varphi_n(t) \to \log \varphi(t)$. Ou seja $it\gamma_n + \int [\ \cdots] dG_n$ converge para $it\gamma + \int [\ \cdots] dG$. Argumentando como na prova do teorema anterior, mostra-se que a sequência $\{G_n(\infty), n \geq 1\}$ é limitada e existem uma subsequência n_k e uma medida $\hat{G}$ tal que $(G_{n_k}(x))/(G_{n_k}(\infty)) \to \hat{G}(x)$ e $G_{n_k}(\infty) \to L$. Segue-se que

$$G_{n_k} \int [\ \cdots] \frac{dG_{n_k}}{G_{n_k}} \to L \int [\ \cdots] d\hat{G}, \quad k \to \infty,$$

pela definição de convergência fraca, pois o integrando é limitado e contínuo. Como, acrescentando-se $it\gamma_{n_k}$ ao primeiro termo da relação anterior e $it\gamma$ ao segundo, continuamos a ter convergência, por unicidade devemos ter $G = L\hat{G}$, de modo que $G_{n_k}(x) \to G(x)$ nos pontos de continuidade de G. De fato, $G_n(x) \to G(x)$ e, portanto, $\gamma_n \to \gamma$, pois $e^{it\gamma_n} \to e^{it\gamma}$, para todo t. $\square$

8.3 Distribuições estáveis

Sabemos que, se $X_1, X_2, \ldots$ são v.a's i.i.d, média μ e variância σ^2, então

$$\frac{X_1 + \ldots + X_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{\mathcal{D}} N(0,1).$$

Esse é um teorema limite da seguinte forma: se $\{X_i, i \geq 1\}$ são i.i.d, então $\frac{X_1+\ldots+X_n}{A_n} - B_n \xrightarrow{\mathcal{D}} X$.

Gostaríamos de decobrir todas as leis limites que surgem dessa maneira.

Definição 8.3. *Seja X uma v.a e suponha que, para cada n, existam constantes a_n, b_n, tais que $a_n X + b_n \sim X_1 + \ldots + X_n$, onde $X_1, X_2, \ldots$ são v.a's i.i.d, $X_i \sim X$. Então, dizemos que X é uma v.a tendo uma distribuição estável.*

Como exemplos, temos as distribuições normal, Cauchy e de Lévy.

Para provar o resultado seguinte, precisamos do seguinte lema (convergência de tipos). Veja Billingsley (1966). A prova pode ser feita usando f.c's (veja Loéve, 1978). Uma prova simples aplicando o Teorema de Skorohod é dada por Fazli e Behboodian (1995).

Lema 8.3. *Suponha que $Y_n \xrightarrow{\mathcal{D}} Y$ e $a_n Y_n + b_n \xrightarrow{\mathcal{D}} \hat{Y}$. Suponha que Y e $\hat{Y}$ sejam não degeneradas e $a_n > 0$. Então, $a_n \to a > 0$, $b_n \to b$ e $\hat{Y} \sim aY + b$, isto é, Y e $\hat{Y}$ são do mesmo tipo.*

Prova: Veja Billingsley (1995).

Teorema 8.10. *(a) Sejam $X_1, X_2, \ldots$ i.i.d e sejam $A_n > 0$, B_n constantes. Se*

$$\frac{X_1 + \ldots + X_n}{A_n} - B_n \xrightarrow{\mathcal{D}} X,$$

onde X não é degenerada, então X é estável.

(b) Se X for estável, então X pode ser representada como um limite em distribuição de somas como em (a).

Prova: (a) Seja $Y_n = (X_1 + \ldots + X_n)/A_n - B_n$, então Y_n converge em distribuição para X, o que ocorre também com a sequência Y_{nk}, para k inteiro positivo. Defina:

$$\begin{aligned} S_n^{(1)} &= X_1 + \ldots + X_n, \\ S_n^{(2)} &= X_{n+1} + \ldots + X_{2n}, \\ \ldots & \quad \ldots \\ S_n^{(k)} &= X_{nk-n} + \ldots + X_{nk} \end{aligned}$$

Considere

$$\frac{S_n^{(1)}}{A_n} - B_n + \frac{S_n^{(2)}}{A_n} - B_n + \ldots + \frac{S_n^{(k)}}{A_n} - B_n =: \frac{A_{nk} Y_{nk}}{A_n} + C_{n,k}.$$

O lado esquerdo converge em distribuição para $X^{(1)} + \ldots + X^{(k)}$, com $\{X_j^{(j)}\}$ i.i.d, com a mesma distribuição que X. Também, Y_{nk} converge em distribuição para X, e portanto $A_{nk}Y_{nk}/A_n + C_{n,k}$ converge em distribuição para $X^{(1)} + \ldots + X^{(k)}$. Pelo lema, $A_{nk}/A_n \to a_k$, $C_{n,k} \to b_k$ e $X^{(1)} + \ldots + X^{(k)} \sim a_k X + b_k$.

(b) Se X é estável, tome $X_1, \ldots, X_n$ i.i.d, $X_i \sim X$, de modo que $a_n X + b_n \sim X_1 + \ldots + X_n$. Então, $X \sim \frac{X_1+\ldots+X_n}{a_n} - \frac{b_n}{a_n}$ (use f.c's). A parte (b) segue. $\square$

Corolário 8.1. *Os a_k's da prova do item (a) do Teorema 8.10 satisfazem $a_{mk} = a_m \cdot a_k$ para todo $k, m \geq 1$.*

Prova: Basta notar que

$$\frac{A_{nmk}}{A_n} = \frac{A_{nmk}}{A_{nm}} \frac{A_{nm}}{A_n},$$

e tomar $n \to \infty$. $\square$

Teorema 8.11. *Seja X estável. Então, $a_n = n^{1/\alpha}$, onde $0 < \alpha \leq 2$.*

Prova: Veja a prova do Teorema 8.12 abaixo.

O número α é chamado o *índice de estabilidade* ou o *expoente* e também dizemos que X é α-estável.

Teorema 8.12. *Seja X α-estável. Então,*

(a) Ou X é normal, ou

(b) Para algum α, $0 < \alpha < 2$, a f.c de X é da forma $\varphi(t) = e^{\psi(t)}$, onde

$$\psi(t) = it\gamma + m_1 \int_0^\infty \left(e^{itx} - 1 - \frac{itx}{1+x^2}\right) \frac{1}{x^{1+\alpha}} dx + m_2 \int_{-\infty}^0 \left(e^{itx} - 1 - \frac{itx}{1+x^2}\right) \frac{1}{|x|^{1+\alpha}} dx \tag{8.16}$$

e γ, $m_1 \geq 0$, $m_2 \geq 0$ são constantes reais.

Prova: Sabemos que, se X é estável, X é i.d, de modo que sua f.c é da forma $e^{\psi(t)}$, com

$$\psi(t) = it\gamma - \frac{\sigma^2 t^2}{2} + \int_{-\infty}^{\infty} \left(e^{itx} - 1 - \frac{itx}{1+x^2}\right) \nu(dx).$$

[1] Caso 1: $\sigma^2 > 0$.

Como X é estável, $a_kX+b_k \sim X^{(1)}+\ldots+X^{(k)}$. Logo, tomando a f.c de ambos os lados, temos $k\psi(t)=\psi(a_kt)+ib_kt$. Seja $L(t,x)=[e^{itx}-1-(itx)/(1+x^2)](1+x^2)/x^2$.

Então, se $x\neq 0$, $L(t,x)/t^2\to 0$, quando $t\to\infty$, logo $\psi(t)=i\gamma t-\sigma^2t^2/2+\int L(t,x)d\hat{G}(x)$ e $\psi(t)/t^2\to -\sigma^2/2$. Por um lado, $k\psi(t)/t^2\to -k\sigma^2/2$, para $t\to\infty$, e por outro $(\psi(a_kt)+itb_k)/t^2\to -a_k^2\sigma^2/2$, do que decorre $a_k=\sqrt{k}$, pois $\sigma^2>0$. Temos, então, $\psi(t)=\psi(\sqrt{k}t)/k+(itb_k)/k$. Para $k\to\infty$, o lado esquerdo converge, logo o lado direito também converge e

$$\frac{\psi(\sqrt{k}t)}{k}=\frac{\psi(\sqrt{k}t)}{kt^2}t^2\to -t^2\sigma^2/2,$$

pois $\psi(t)/t^2\to -\sigma^2/2$. Logo, $ib_kt/k\to i\gamma t$ e $\psi(t)=i\gamma t-t^2\sigma^2/2$, portanto X é normal.

[2] Caso 2: $\sigma^2=0$

Lembremos que $k\psi(t)=ib_kt+\psi(a_kt)$, logo

$$k\psi(t)=ki\gamma t+\int\left[e^{itx}-1-\frac{itx}{1+x^2}\right]k\nu(dx)$$

e

$$ib_kt+\varphi(a_kt)=ib_kt+\int\left[e^{ia_ktx}-1-\frac{ia_ktx}{1+x^2}\right]\nu(dx)$$

e somando e subtraindo $\frac{ia_ktx}{1+(a_kx)^2}$ ao integrando, obtemos que

$$\begin{aligned}ib_kt+\varphi(a_kt)&=ib_kt+\int\left[e^{itx}-1-\frac{itx}{1+x^2}\right]\nu_1(dx)+it\gamma_k\\&=it(b_k+\gamma_k)+\int\left[e^{itx}-1-\frac{itx}{1+x^2}\right]\nu_1(dx).\end{aligned}$$

Pelo teorema da unicidade, $k\nu(dx)=\nu_1(dx)=\nu(dx/a_k)$ e também $b_k+\gamma_k=k\gamma$.

Sejam $\nu^+(x)=\nu[x,+\infty)$, se $x>0$ e $\nu^-(x)=\nu(-\infty,x]$, se $x<0$. Então, $k\nu^+(x)=\nu(x/a_k)$, uma fórmula similar para ν^-. Suponha que $a_k=k^\lambda,\ \lambda>0$. Seja $x=(k/n)^\lambda$ nessa fórmula; obtemos

$$k\nu^+\left((k/n)^\lambda\right)=\nu^+\left(\frac{1}{n^\lambda}\right). \tag{8.17}$$

Escolhendo $k=n$, obtemos

$$n\nu^+(1)=\nu^+\left(\frac{1}{n^\lambda}\right). \tag{8.18}$$

Comparando (8.17) e (8.18) obtemos $\nu^+\left((k/n)^\lambda\right)=\nu^+(1)(k/n)^{-1}$.

Logo, para x num conjunto denso,

$$\nu^+(x) = \nu^+(1)x^{-1/\lambda}. \tag{8.19}$$

Como γ é decrescente, (8.19) vale para todo x. Isso prova o resultado, desde que mostremos que $1/\lambda = \alpha$ satisfaz $0 < \alpha < 2$.

Sabemos que $\int_{-1}^{1} x^2\nu(dx) < \infty$, logo $\int_{-1}^{1} x^2 x^{-1/\lambda-1}dx < \infty$, do que segue $2 - 1/\lambda - 1 > -1$ e daqui $1/\lambda < 2$ e $\lambda > 0$.

Resta provar que $a_k = k^\lambda$, $\lambda > 0$ Para $x = 1$, $k\nu^+(1) = \nu^+(1/a_k)$. Quando k cresce, o lado esquerdo dessa igualdade cresce, logo $1/a_k$ decresce e portanto a_k deve crescer para $+\infty$.

Lembremos também que $a_{mn} = a_m a_n$. Fixando k, seja n um inteiro tal que $k^j \leq n \leq k^{j+1}$. Como a_j é crescente, $a_{k^j} \leq a_n \leq a_{k^{j+1}}$, ou seja, $(a_k)^j \leq a_n \leq (a_k)^{j+1}$. Tomando logaritmos, $j \log a_k \leq \log a_n \leq (j+1)\log a_k$, de onde segue

$$\frac{\log a_k}{j \log k} \leq \frac{\log a_n}{j \log k} \leq \frac{j+1}{j}\frac{\log a_k}{\log k}.$$

Para $j \to \infty$, $\lim_{j\to\infty} \frac{\log a_n}{j \log k} = \lambda$, independentemente de k e, portanto, $\frac{\log a_k}{\log k} = \lambda$, logo $a_k = k^\lambda$. $\square$

As integrais do Teorema 8.12 podem ser calculadas explicitamente e temos o seguinte resultado (para uma prova, veja Breiman (1968)).

Teorema 8.13. *Se $0 < \alpha < 2$ e se X tem uma distribuição α-estável, então o logaritmo da f.c de X é dado por:*

$$\psi(t) = it\mu - \sigma|t|^\alpha\Big[1 + i\beta \text{sinal}(t)\tan(\frac{\pi}{2}\alpha)\Big], \quad \text{se } \alpha \neq 1. \tag{8.20}$$

Se $\alpha = 1$, então,

$$\psi(t) = it\mu - \sigma|t|\Big[1 + i\beta \text{sinal}(t)\frac{2}{\pi}\log(|t|)\Big]. \tag{8.21}$$

Em (8.20) e (8.21), μ é um parâmetro de localização real, $\sigma > 0$ é um parâmetro de escala , β é um parâmetro de simetria real, $|\beta| \leq 1$.

Usualmente, usamos a notação $X \sim S_\alpha(\sigma, \beta, \mu)$ para denotar uma v.a com distribuição estável, com parâmetros $(\alpha, \sigma, \beta, \mu)$.

Se α decresce de 2 a 0, as caudas de X tornam-se mais pesadas que a normal. Se $1 < \alpha < 2$ a média de X é μ, mas se $0 < \alpha \leq 1$, a média é infinita. Se $\beta = 0$, X é simétrica, ao passo que se $\beta > 0$ ($\beta < 0$), então X é assimétrica à direita (à esquerda).

Proposição 8.1. *Se X é α-estável, então X tem uma função densidade de probabilidade limitada e contínua.*

Prova: De fato, $|\varphi(t)| \leq e^{-|t|^\alpha}$, integrável. $\square$

Proposição 8.2. *Suponha X simétrica, α-estável. Então, a f.c de X é da forma $\varphi(t) = e^{-c|t|^\alpha}$.*

Prova: Imediata. □

Problemas

1. Para cada n, suponha que $X_{n,1}, \ldots, X_{n,n}$ sejam v.a's i.i.d e suponha que $X_{n,1} + \ldots + X_{n,n} \xrightarrow{\mathcal{D}} X$. Então, $\lim_{n\to\infty} \max_{1\le k\le n} P\{|X_{n,k}| > \varepsilon\} = 0$.
2. Mostre que a f.c de distribuição de Poisson composta Y é dada por $\psi(t) = e^{\lambda[\varphi(t)-1]}$. Mostre que Y é infinitamente divisível.
3. Prove a Proposição 8.2.
4. Suponha que X tenha distribuição gama, $\Gamma(\alpha)$, com densidade $f(x) = [\Gamma(\alpha)]^{-1}x^{\alpha-1}e^{-x}$, para $x > 0$.

   (i) Prove que X é infinitamente divisível.

   (ii) Encontre explicitamente a densidade da v.a correspondente a ϕ_n, a raiz n-ésima de ϕ: $\phi(t) = [\varphi_n(t)]^n$. Aqui, φ é a f.c de X.

   (iii) Encontre G na representação canônica de φ, por meio da obtenção da mesma, e calculando o limite dado nessa representação.
5. Se a medida de Lévy ν for concentrada sobre um intervalo finito $[-a, a]$, então X tem momentos de qualquer ordem.
6. Prove que qualquer v.a X i.d pode ser escrita como $X \sim c_1X_1 + c_2X_2 + c_3X_3$, onde as X_i são independentes, i.d, c_i são constantes, possivelmente nulas em alguns casos, e: (a) X_1 é normal; (b) X_2 é Poisson composta; (c) X_3 tem momentos de todas as ordens. [Sugestão: escreva a integral definindo $\psi(t)$ como a soma de duas integrais, sobre $|x| < 1$ e $|x| > 1$.]
7. Suponha que X seja i.d. Prove:

   (a) X é simétrica se, e somente se, ν for simétrica.

   (b) Se $X \ge 0$, então ν é concentrada em $(0, +\infty)$ e ψ tem a forma
   $$\psi(t) = it\gamma_1 + \int_0^\infty [e^{itx} - 1]\nu(dx),$$
   onde γ_1 é uma constante.

   (c) Se $\int_0^\infty |x|\nu(dx) < \infty$, e se ν for concentrada em $(0, +\infty)$, então $X \ge 0$. (Note que (b) e (c) mostram que, não é verdade, em geral, que se ν for concentrada em $(0, \infty)$, então $X \ge 0$.)
8. (a) Mostre que se $X \sim N(0, 1)$, então X não é Poisson composta, mas dê uma sequência explícita de distribuições de Poisson compostas convergindo para X.

   (b) Mostre que no Teorema de Berry-Esseen, a taxa de convergência $1/\sqrt{n}$ é a melhor possível (Considere o caso Bernoulli).

9. Considere $X_{n,1},\ldots,X_{n,n}$ v.a's i.i.d, cada uma com distribuição uniforme em $[-n,n]$ e seja $Y_n=\sum_{i=1}^n \text{sinal}(X_{n,i})/X_{n,i}^2$. Prove que Y_n converge em distribuição para uma v.a estável com $\alpha=1/2$ (Calcule a f.c.)

10. Sejam $X_1,X_2,\ldots$ independentes, com:

$$\begin{aligned}
P(X_j=j^2) &= P(X_j=-j^2)=\frac{1}{12j^2},\\
P(X_j=j) &= P(X_j=-j)=\frac{1}{12}\\
P(X_j=0) &= 1-\frac{2}{12}-\frac{2}{12j^2}.
\end{aligned}$$

(a) Prove que a condição de Lindeberg não está satisfeita, mas existe uma constante absoluta C tal que

$$\frac{X_1+\ldots+X_n}{Cn^{3/2}}\xrightarrow{\mathcal{D}} N(0,1).$$

[Sugestão: use truncamento.]

(b) Explique porque (a) não contradiz o Teorema de Feller.

11. Sejam $X_{n,1},\ldots,X_{n,n}$ i.i.d e suponha que $X_{n,1}+\ldots+X_{n,n}\xrightarrow{\mathcal{D}} X$. Seja P_n a distribuição de $X_{n,1}$.

   (a) X é normal (ou degenerada) se, e somente se, $nP_n([-a,a]^c)\to 0$, para todo $a>0$.

   (b) Use (a) para provar o TLC ordinário.

   (c) Prove que, para quaisquer v.a's $\{X_n, n\geq 1\}$ que sejam i.i.d, teremos que $(X_1+\ldots+X_n)/\sqrt{n}$ converge em distribuição para uma v.a $N(0,1)$ se, e somente se $\max_{1\leq k\leq n}|X_k|/\sqrt{n}$ converge para zero, em probabilidade.

   (d) Prove que $X_{n,1}+\ldots+X_{n,n}$ converge para uma distribuição de Poisson (ou uma v.a degenerada) se, e somente se $\int_{-\infty}^t x^2/(1+x^2)ndP_n(x)\to 0$, para $t<1$ e $\int_t^\infty x^2/(1+x^2)ndP_n(x)\to 0$, para $t>1$.

Capítulo 9

O Princípio da Invariância

Na teoria de somas de variáveis aleatórias independentes, há resultados que podem ser provados de acordo com o seguinte esquema: primeiramente, prova-se que a distribuição limite não depende das distribuições das variáveis aleatórias, desde que certa condições sejam válidas. A seguir, a distribuição limite é calculada para uma escolha especial de variáveis aleatórias. A essa propriedade foi dado o nome de *princípio da invariância*.

Esse princípio foi introduzido por Kolmogorov (1931). Em 1933, Kolmogorov provou uma versão mais forte que foi usada para obter a distribuição limite da diferença entre uma f.d empírica e a f.d teórica correspondente. Para mais detalhes sobre os trabalhos subsequentes, veja Kruglov (1998).

Neste capítulo apresentaremos o princípio da invariância de Donsker e estudaremos com um pouco de mais detalhes o movimento browniano ou processo de Wiener. A independência dos somandos no Teorema de Lindeberg-Feller garante também a convergência fraca de todas as distribuições finito-dimensionais de um processo estocástico contínuo q.c para aquelas de um processo Gaussiano com incrementos independentes, ou seja, o movimento browniano. Além disso, essas distribuições convergem fracamente para a medida de Wiener sobre $C([0,1])$, fato esse também conhecido como *TLC funcional*, uma ideia originada em trabalhos de Erdos e Kac (1946) e Donsker (1951), depois desenvolvidas por Billingsley, Prokhorov, Skorohod e outros.

O movimento browniano tem aplicações relevantes em finanças, como na fórmula de Black-Scholes (veja Capítulo 12), para apreçamento de opções e em equações diferenciais estocásticas, particularmente em problemas relacionados a difusões, que descrevem o comportamento da volatilidade de séries financeiras.

9.1 Introdução

Nesta seção desenvolveremos noções de convergência fraca no espaço $C([0,1])$. Algumas das definições e propriedades a seguir já foram vistas no Capítulo 6.

Definição 9.1. *$C([0,1])$ é o espaço de todas as funções contínuas definidas em $[0,1]$ com valores reais. Definimos sobre esse espaço a métrica:*

$$d(x,y) = \sup_{0\le t\le 1} |x(t)-y(t)|, \quad x,y \in C([0,1]). \tag{9.1}$$

Com esta métrica, o espaço $C([0,1])$ é um espaço métrico completo e separável. Observamos também que esta métrica é induzida pela norma infinita em $C([0,1])$.

Definição 9.2. *Para $x \in C([0,1])$, definimos o módulo de continuidade*

$$w_x(\delta) = \sup_{|t-s|<\delta} |x(t)-x(s)|, \quad 0<\delta<1. \tag{9.2}$$

Alguns fatos sobre o módulo de continuidade:

[1] $w_x(\delta)$ é uma função contínua de x, para δ fixo.

De fato,

$$\begin{aligned} w_x(\delta) &= \sup_{|t-s|<\delta} |x(t)-x(s)| = \sup_{|t-s|<\delta} |x(t)-y(t)+y(t)-y(s)+y(s)-x(s)| \\ &\le \sup_{|t-s|<\delta} |x(t)-y(t)| + \sup_{|t-s|<\delta} |y(t)-y(s)| + \sup_{|t-s|<\delta} |y(s)-x(s)| \\ &\le 2d(x,y) + w_y(\delta), \end{aligned}$$

logo $w_x(\delta) - w_y(\delta) \le 2d(x,y)$, de modo que $|w_x(\delta)-w_y(\delta)| \le 2d(x,y)$, por simetria.

[2] Para x fixo, $w_x(\delta) \to 0$, quando $\delta \to 0$.

De fato, toda função contínua sobre $[0,1]$ é uniformemente contínua.

Definição 9.3. *Para $t_1 < t_2 < \cdots < t_n$ defina $\pi_{t_1,\ldots,t_n} : C([0,1]) \to \mathbb{R}^n$ por*

$$\pi_{t_1,\ldots,t_n}(x) = (x(t_1),\ldots,x(t_n)). \tag{9.3}$$

Dizemos que $\pi_{t_1,\ldots,t_n}$ é uma projeção.

De maneira genérica, denotaremos projeções por π.

Definição 9.4. *Se h é alguma função de $C([0,1])$ em $(S,\mathcal{S})$, mensurável, defina Ph^{-1} por $Ph^{-1}(A) = P\{h^{-1}(A)\}$, onde A é um subconjunto de $\mathcal{S}$ e P é uma medida* *sobre* $\mathcal{C}$*, os conjuntos de Borel de* $C([0,1])$*. Segue-se que* Ph^{-1} *é uma medida sobre* $(S,\mathcal{S})$ *chamada* medida imagem *de* P *por* h*.*

Definição 9.5. *Se* P *é uma medida de probabilidade sobre* $(C([0,1]),\mathcal{C})$*, então* $\{P\pi^{-1} : \pi$ *é uma projeção*$\}$ *é a distribuição finito-dimensional de* P*.*

Teorema 9.1. *Seja* $\mathcal{F}_\pi$ *a menor* σ*-álgebra tornando cada* π *mensurável. Então:*

(a) $\mathcal{C} = \mathcal{F}_\pi$*;*

(b) Se P, Q *são duas medidas de probabilidade sobre* $C([0,1])$*,* $\mathcal{C}$ *e se* $P\pi^{-1} = Q\pi^{-1}$*, para todas as projeções* π*, então* $P = Q$*.*

(c) Se P *é uma probabilidade sobre* $C([0,1])$*, então* P *é fechada, ou seja, para todo* $\varepsilon > 0$*, existe um compacto* K *tal que* $P(K) \geq 1-\varepsilon$*.*

Prova: (a) Como cada π_t é contínua, e $\mathcal{C}$ contém os conjuntos abertos, segue-se que $\mathcal{F}_\pi \subset \mathcal{C}$. Para provar que $\mathcal{C} \subset \mathcal{F}_\pi$, basta provar que $\mathcal{F}_\pi$ contém conjuntos abertos. Como $C([0,1])$ é separável, todo conjunto aberto é uma reunião enumerável de bolas fechadas. Logo, é suficiente provar que $\mathcal{F}_\pi$ contém bolas fechadas. Tome x_0, $\varepsilon > 0$ e seja $B_\varepsilon(x_0)$ uma bola fechada com centro em x_0, de raio ε. Temos que

$$B_\varepsilon(x_0) = \{x : d(x,x_0) \leq \varepsilon\} = \left\{x : \sup_{0\leq t\leq 1} |x(t) - x_0(t)| \leq \varepsilon\right\}$$

$$= \left\{x : \sup_{r\in[0,1]\cap\mathbb{Q}} |x(r) - x_0(r)| \leq \varepsilon\right\} = \bigcap_{r\in[0,1]\cap\mathbb{Q}} \{x : |x(r) - x_0(r)| \leq \varepsilon\}.$$

Mas, $\{x : |x(r) - x_0(r)| \leq \varepsilon\} = \{x : -\varepsilon + x_0(r) \leq x(r) \leq \varepsilon + x_0(r)\} = \pi_r^{-1}\{[-\varepsilon + x_0(r), \varepsilon + x_0(r)]\}$, que pertence a $\mathcal{F}_\pi$, pela sua definição.

(b) Suponha $Q\pi^{-1} = P\pi^{-1}$, para toda π. Seja $\mathcal{F}_{t_1,\ldots,t_n} = \{\pi^{-1}_{t_1,\ldots,t_n}(A) : A \in \mathcal{B}(\mathbb{R}^n)\}$. Seja $\hat{\mathcal{F}} = \cup_{t_1,\ldots,t_n}\mathcal{F}_{t_1,\ldots,t_n}$. Então, $\hat{\mathcal{F}}$ é uma álgebra.

De $Q\pi^{-1} = P\pi^{-1}$ temos que $Q(B) = P(B)$, para todo $B \in \hat{\mathcal{F}}$, e como $\hat{\mathcal{F}} \subset \mathcal{F}_\pi$ e gera $\mathcal{F}_\pi$, temos que $Q(B) = P(B)$, para todo $B \in \mathcal{F}_\pi$, do que decorre que $Q(B) = P(B)$, para todo $B \in \mathcal{C}$.

(c) Seja $(S,\mathcal{S})$ um espaço métrico completo e separável. Mostramos que se P é uma probabilidade sobre $(S,\mathcal{S})$, então P é fechada. Para todo $n \geq 1$, sejam $B^1_{1/n}, B^2_{1/n}, \ldots$ bolas abertas de raios $1/n$ que cobrem S. Seja $\varepsilon > 0$. Tome um inteiro k_n tal que $P(\cup_{j=1}^{k_n} B^j_{1/n}) \geq 1-\varepsilon/2^{n+1}$. Defina $K = \overline{\cap_n \cup_j^{k_n} B^j_{1/n}}$. Como $\cap_n\cup_j^{k_n} B^j_{1/n}$ é totalmente limitado, o fecho desse conjunto é compacto e $P(K) \geq 1-\varepsilon$. $\square$

Sabemos que:

(a) Se P_n, P são medidas de probabilidade sobre $C([0,1])$, então $P_n\pi^{-1} \Rightarrow P\pi^{-1}$, desde que $P_n \Rightarrow P$;

(b) Se P_n, P são medidas de probabilidade sobre $C([0,1])$, e se $P_n\pi^{-1} \Rightarrow P\pi^{-1}$ para todas as projeções π, não é necessariamente verdade que $P_n \Rightarrow P$.

Mas, o resultado seguinte é válido.

Teorema 9.2. *Sejam P_n, P medidas de probabilidade sobre $C([0,1])$, e $P_n\pi^{-1} \Rightarrow P\pi^{-1}$ para todas as projeções π. Suponha, também, que a família $\{P_n, n \geq 1\}$ seja fechada. Então, $P_n \Rightarrow P$.*

Prova: Como $\{P_n\}$ é fechada, existe uma subsequência n_k e uma probabilidade Q tal que $P_{n_k} \Rightarrow Q$, pelo teorema de Prokhorov. A probabilidade Q poderia depender de n_k. Contudo, não depende. Porque a convergência para Q implica que $P_{n_k}\pi^{-1} \Rightarrow Q\pi^{-1}$, para toda π, e por hipótese $P_n\pi^{-1} \Rightarrow P\pi^{-1}$, do que segue $P\pi^{-1} = Q\pi^{-1}$, para toda π. Logo, $Q = P$, isto é, Q é independente da sequência n_k. Logo, $P_n \Rightarrow P$, pelo Lema 6.2. $\square$

Lema 9.1. (Arzela-Ascoli) *$K \in \mathcal{C}$ tem um fecho compacto se, e somente se, $\sup_{x\in K} |x(0)| \leq M < \infty$ e $\lim_{\delta\to 0} \sup_{x\in K} w_x(\delta) = 0$ (equicontinuidade).*

Teorema 9.3. *A sequência de medidas de probabilidade $\{P_n, n \geq 1\}$ é fechada se, e somente se, o seguinte vale:*

(a) Para todo $\Delta > 0$, existe λ tal que $P_n\{x : |x(0)| > \lambda\} \leq \Delta$, para todo n;

(b) Para todo $\varepsilon > 0$, $\Delta > 0$, existe δ, $0 < \delta < 1$ e um inteiro n_0, tal que $P_n\{x : w_x(\delta) > \varepsilon\} \leq \Delta$, para todo $n \geq n_0$.

Prova: $(\Rightarrow)$ Sejam $\Delta > 0$ e K um conjunto compacto, tal que $P_n(K) \geq 1 - \Delta$, para todo n. Como K é compacto, pelo Lema 9.1, $\sup_{x\in K} |x(0)| \leq M$, para algum M. Então,

$$K \subset \{x : |x(0)| \leq M\} \tag{9.4}$$

Também, $\lim_{\delta\to 0} \sup_{x\in K} w_x(\delta) = 0$, logo

$$K \subset \{x : w_x(\delta) \leq \varepsilon\}, \tag{9.5}$$

para δ suficientemente pequeno. Logo, por (9.4)

$$P_n\{x : |x(0) > M\} \leq P_n(K^c\} \leq \Delta, \quad \text{para todo } n$$

e por (9.5),

$$P_n\{x : w_x(\delta) > \varepsilon\} \leq P_n(K^c) \leq \Delta, \quad \text{para todo } n.$$

($\Leftarrow$) Podemos supor na condição (b) que $n_0 = 1$. Pela condição (a), se $\Delta > 0$ é escolhido, podemos encontrar M tal que $P_n(A) \geq 1 - \Delta/2$, onde $A = \{x : |x(0)| \leq M\}$. Se Δ e k inteiros são dados, podemos encontrar δ_k tal que $P_n(A_k) \geq 1-\Delta/2^{k+1}$, com $A_k = \{x : w_x(\delta_k) \leq 1/2^k\}$, usando (b). Defina $K = A \cap (\cap_{k\geq 1} A_k)$. Então, K tem fecho compacto, pelo Lema 9.1 e $P_n(K) \geq 1 - \Delta$, de modo que a família $\{P_n\}$ é fechada. $\square$

Teorema 9.4. *Suponha que as seguintes afirmações sejam válidas:*

(a) Para todo $\Delta > 0$, existe λ tal que $P_n\{x : |x(0)| > \lambda\} \leq \Delta$, para todo n;

(b) Para todo $\varepsilon > 0$, $\Delta > 0$, existe δ, $0 < \delta < 1$ e um inteiro n_0, tal que

$$\frac{1}{\delta} \sup_{t\in[0,1-\delta]} P_n\Big\{x : \sup_{t\leq s\leq t+\delta} |x(s) - x(t)| > \varepsilon\Big\} \leq \Delta,$$

para todo $n \geq n_0$.

Então, $\{P_n, n \geq 1\}$ é fechada.

Prova: Vamos mostrar que (b) implica (b) do Teorema 9.3. Fixemos t, que pertence a algum intervalo da forma $[i\delta, (i+1)\delta]$. Se $s \geq t$, então $s \in [i\delta, (i+2)\delta]$, logo

$$\begin{aligned}\Big\{x : \sup_{t\leq s\leq t+\delta} |x(s) - x(t)| > \varepsilon\Big\} \;&\subset\; \Big\{x : \sup_{t\leq s\leq t+\delta} |x(s) - x(i\delta)| + |x(t) - x(i\delta)| > \varepsilon\Big\} \\ &\subset \Big\{x : 2 \sup_{i\delta\leq s\leq (i+2)\delta} |x(s) - x(i\delta)| > \varepsilon\Big\}. \qquad (9.6)\end{aligned}$$

Então,

$$\begin{aligned}P_n\{x : w_x(\delta) > \varepsilon\} &= P_n\Big\{x : \sup_{|t-s|<\delta} |x(s) - x(t)| > \varepsilon\Big\} \\ &\leq \sum_{i=0}^{\lfloor 1/\delta\rfloor} P_n\Big\{x : \sup_{|t-s|<\delta,\; i\delta\leq t\leq (i+1)\delta} |x(s) - x(t)| > \varepsilon\Big\} \\ &\leq \sum_{i=0}^{\lfloor 1/\delta\rfloor} P_n\Big\{x : \sup_{i\delta\leq s\leq (i+2)\delta} |x(s) - x(i\delta)| > \varepsilon/2\Big\} \leq \frac{1}{\delta}(2\delta)\Delta = 2\Delta\end{aligned}$$

onde na segunda desigualdade usamos (9.6) e na última a hipótese (b). $\square$

Seja $(\Omega, \mathcal{F}, P)$ um espaço de probabilidade e $X : \Omega \to C([0,1])$. Logo, para cada $\omega \in \Omega$, $X(\omega)$ é uma função contínua. Seja $X_t(\omega)$ o valor dessa função em cada ponto t. Suponha que X seja mensurável.

Proposição 9.1. *Se $X = \{X_t, 0 \leq t \leq 1\}$, então X é um processo estocástico com trajetórias contínuas.*

Prova: Claramente, X tem trajetórias contínuas. Resta provar que cada X_t é uma v.a. Mas $X_t = \pi_t \circ X$, logo se $A \in \mathcal{B}(\mathbb{R})$, então $X_t^{-1}(A) = X^{-1} \circ \pi_t^{-1}(A)$. Mas $B = \pi_t^{-1}(A) \in \mathcal{C}$, pois π_t é contínua e $X^{-1}(B) \in \mathcal{F}$, pois X é mensurável, como uma aplicação de Ω em $C([0,1])$. $\square$

A distribuição de X é a medida de probabilidade sobre $C([0,1])$, P_X, definida por $P_X(A) = P\{\omega : X(\omega) \in A\}$ para todo $A \in \mathcal{C}$.

Dada qualquer medida de probabilidade P' definida sobre $C([0,1])$, existe um processo estocástico tendo trajetórias contínuas e P' como sua distribuição. De fato, tome $\Omega = C([0,1])$, $\mathcal{F} = \mathcal{C}$ e $P = P'$. Defina X por $X(\omega) = \omega$ (note que pontos de Ω são identificados com pontos da trajetória).

Reciprocamente, dado qualquer p.e $X = \{X_t, 0 \leq t \leq 1\}$, com trajetórias contínuas, existe uma probabilidade P sobre $C([0,1])$, que é a distribuição de X.

Os dois teoremas precedentes podem ser reescritos em termos de processos estocásticos.

Definição 9.6. *A medida de Wiener W é uma medida de probabilidade sobre $C([0,1])$ tal que, se $X = \{X_t, 0 \leq t \leq 1\}$ for um processo estocástico com trajetórias contínuas tendo W como sua distribuição, então:*

(1) $X_0 = 0$, *q.c.*

(2) Para cada t, $X_t \sim N(0,t)$.

(3) Se $0 \leq t_1 < t_2 < \cdots < t_n \leq 1$, *então* $X_{t_1}, X_{t_2} - X_{t_1}, \ldots, X_{t_n} - X_{t_{n-1}}$ *são independentes.*

Consequências:

[1] Se $X = \{X_t, 0 \leq t \leq 1\}$ é como descrito na definição anterior, então se $s < t$, teremos que $X_t - X_s \sim N(0, t-s)$.

De fato, $X_t = X_s + X_t - X_s$ e como X_s e $X_t - X_s$ são independentes e se $\varphi(u)$ é a f.c de $X_t - X_s$, teremos $e^{-tu^2/2} = \varphi(u) e^{-su^2/2}$, do que decorre que $\varphi(u) = e^{-\frac{u^2}{2}(t-s)}$.

2] Suponha que X satisfaça a definição anterior. Sejam $t_1 < t_2 < \ldots < t_n$ e o vetor aleatório $(X_{t_1}, \ldots, X_{t_n})$, com f.c $\Gamma(s_1, \ldots, s_n)$. Então,

$$\Gamma(s_1, \ldots, s_n) = E\left[e^{is_1 X_{t_1} + \ldots + is_n X_{t_n}}\right]$$

$$= E \exp\left\{ i \left[s_n (X_{t_n} - X_{t_{n-1}}) + (s_n + s_{n-1})(X_{t_{n-1}} - X_{t_{n-2}}) + \ldots + (s_1 + \ldots + s_n) X_{t_1} \right] \right\}$$

$$= \exp\left[-\frac{s_n^2}{2}(t_n - t_{n-1}) - \frac{(s_n + s_{n-1})^2}{2}(t_{n-1} - t_{n-2}) - \ldots - \frac{(s_1 + \ldots + s_n)^2}{2} t_1\right].$$

A última igualdade segue da normalidade de X.

Definição 9.7. *O processo $X = \{X_t, 0 \leq t \leq 1\}$ com trajetórias contínuas, tendo como distribuição a medida W da definição anterior, é chamado Movimento Browniano (MB)* ou *processo de Wiener.*

Teorema 9.4. (Donsker) *Sejam $Y_1, Y_2, \ldots$ v.a's i.i.d, com média zero e variância σ^2. Seja $S_n = Y_1 + \ldots + Y_n$. Defina uma família de processos estocásticos $X_n = \{X_n(t), 0 \leq t \leq 1\}$ como segue:*

$$X_n(t) = \frac{S_{\lfloor nt \rfloor}}{\sigma\sqrt{n}} + \frac{nt - \lfloor nt \rfloor}{\sigma\sqrt{n}} Y_{\lfloor nt \rfloor + 1}, \tag{9.7}$$

onde $\lfloor a \rfloor$ representa o maior inteiro menor ou igual a a. Então, a sequência $\{X_n, n \geq 1\}$ converge fracamente em $C([0,1])$ para um MB.

Prova: A prova tem duas partes:
(a) Para $t_1 < t_2 < \cdots < t_k$, mostramos que

$$(X_n(t_1), \ldots, X_n(t_k)) \xrightarrow{\mathcal{D}} \mu_{t_1,\ldots,t_k},$$

onde $\mu_{t_1,\ldots,t_k}$ é a medida correspondente à f.c $\Gamma(s_1, \ldots, s_k)$.
(b) Depois mostramos que a família $\{X_n, n \geq 1\}$ é fechada (*tight*).

Segue-se que X_n converge em distribuição para X, onde X satisfaz as propriedades (1)-(3) da definição 9.6. De fato, seja n' qualquer subsequência de inteiros e P_n a distribuição de X_n. Por (b), existe uma outra subsequência n'' e uma probabilidade Q sobre $C([0,1])$ tal que $P_{n''} \Rightarrow Q$ (por Prokhorov). A medida Q não depende da subsequência, pois por (a), cada distribuição limite Q tem a mesma distribuição finito-dimensional. Como isso vale para cada subsequência n', segue que $P_n \Rightarrow Q$ e $Q = W$.

Prova de (a). Sejam $t_1, \ldots, t_k$ dados; devemos mostrar que $E(\exp i\{s_1 X_n(t_1) + \ldots + s_k X_n(t_k)\})$ converge para $\Gamma(s_1, \ldots, s_k)$. Mas

$$E(\exp i\{s_1 X_n(t_1) + \ldots + s_k X_n(t_k)\}) =$$

$$E(\exp i\{s_k[X_n(t_k) - X_n(t_{k-1})] + [s_k + s_{k-1}][X_n(t_{k-1}) - X_n(t_{k-2})] + \ldots + [s_1 + \ldots + s_k] X_n(t_1)\}).$$

Note que, para cada t,

$$\left| X_n(t) - \frac{S_{\lfloor nt \rfloor}}{\sigma\sqrt{n}} \right| \leq \frac{nt - \lfloor nt \rfloor}{\sigma\sqrt{n}} |Y_{\lfloor nt \rfloor + 1}|.$$

Quando $n \to \infty$, o lado direito da desigualdade tende a zero em probabilidade, pela desigualdade de Chebyshev. Portanto, é suficiente provar que

$$E(\exp i\{s_k(S_{\lfloor nt_k\rfloor}-S_{\lfloor nt_{k-1}\rfloor})/\sigma\sqrt{n}+\ldots+(s_1+\ldots+s_k)S_{\lfloor nt_1\rfloor}/\sigma\sqrt{n}\}) \to \Gamma(s_1,\ldots,s_k).$$

Mas $(S_{\lfloor nt_k\rfloor}-S_{\lfloor nt_{k-1}\rfloor})/\sigma\sqrt{n}$ converge em distribuição para uma $N(0,t_k-t_{k-1})$, pelo TLC e essas parcelas são independentes, logo o limite é $\exp\{-s_k^2(t_k-t_{k-1})/2-\ldots-(s_1+\ldots+s_k)^2t_1/2\}=\Gamma(s_1,\ldots,s_k)$.

Prova de (b). Temos que provar que a família $X_n=\{X_n(t), 0\le t\le 1\}$ é fechada. É suficiente provar que:

(i) para todo $\Delta>0$ e todo $\varepsilon>0$, existe δ, $0<\delta<1$ e um inteiro n_0, tal que

$$\frac{1}{\delta}P\Big\{\omega : \sup_{t\le s\le t+\delta}|X_n(s)-X_n(t)|>\varepsilon\Big\}\le\Delta,\ \ n>n_0,\ \ \forall t;$$

(ii) Para todo $\Delta>0$, existe $\lambda>0$, tal que $P\{\omega : |X_n(0)|>\lambda\}\le\Delta,\quad \forall n$.

A condição (ii) é trivial, pois $X_n(0)=0$, logo basta provar (i).

Fixemos t e tome inteiros j,k tais que $k/n\le t\le (k+1)/n$ e $(j-1)/n\le t+\delta\le j/n$. Então, $j/n-k/n\le\delta+2/n$.

A desigualdade de Lévy (Teorema 2.17), com variáveis i.i.d, média zero e variância comum σ^2, pode ser re-escrita como

$$P\left\{\sup_{1\le k\le n}\frac{|S_k|}{\sigma\sqrt{n}}\ge\lambda\right\}\le 2P\left\{\frac{|S_n|}{\sigma\sqrt{n}}\ge\lambda-1\right\}, \tag{9.8}$$

usando $|m(X)-E(X)|\le\sqrt{Var(X)}$. Agora fixe $\varepsilon>0$ e $\delta<\varepsilon^2/16$. Para n suficientemente grande, temos que

$$\frac{1}{\delta}P\Big\{\omega : \sup_{t\le s\le t+\delta}|X_n(s)-X_n(t)|\ge\varepsilon\Big\}$$

$$\le\frac{1}{\delta}P\Big\{\omega : \sup_{0\le i\le\lfloor n\delta+2\rfloor}\frac{|S_i|}{\sigma\sqrt{\lfloor n\delta+2\rfloor}}\ge\frac{\varepsilon}{2}\frac{1}{\sqrt{\delta}}\Big\}$$

$$\le\frac{2}{\delta}P\left\{\frac{|S_{\lfloor n\delta+2\rfloor}|}{\sigma\sqrt{\lfloor n\delta+2\rfloor}}\ge\frac{\varepsilon}{2}\frac{1}{\sqrt{\delta}}-1\right\}\le\frac{2}{\delta}P\left\{\frac{|S_{\lfloor n\delta+2\rfloor}|}{\sigma\sqrt{\lfloor n\delta+2\rfloor}}\ge\frac{\varepsilon}{4}\frac{1}{\sqrt{\delta}}\right\}.$$

Pelo TLC, obtemos se $Z\sim N(0,1)$,

$$\lim_{n\to\infty} \frac{2}{\delta} P\left\{ \frac{|S_{\lfloor n\delta+2 \rfloor}|}{\sigma\sqrt{\lfloor n\delta+2 \rfloor}} > \frac{\varepsilon}{4}\frac{1}{\sqrt{\delta}} \right\} = \frac{2}{\delta} P\Big\{ |Z| > \frac{\varepsilon}{4}\frac{1}{\sqrt{\delta}} \Big\} \leq \frac{2}{\delta}\frac{E(|Z|^3)4^3}{\varepsilon^3}\delta^{3/2},$$

pois $P(|Z| > a) \leq E(|Z|^3)/a^3$. Dado $\varepsilon > 0$, dado $\Delta > 0$, podemos tomar δ tão pequeno de modo que este último termo seja menor do que Δ. $\square$

Corolário 9.1. *Seja h uma função mensurável, $h : C([0,1]) \xrightarrow{\mathcal{D}} (S, \mathcal{S})$, um espaço métrico. Suponha que D_h, conjunto das desontinuidades de h, tenha medida de Wiener zero. Então, $h(X_n) \xrightarrow{\mathcal{D}} h(X)$, onde X é um MB.*

9.2 Movimento browniano

Definimos, na seção anterior, o MB $X = \{X(t)\}$, para $t \in [0,1]$. Queremos estender esse processo para o conjunto paramétrico $[0,\infty)$.

Sejam $X^1, X^2, \ldots$ MBs independentes, com espaço paramétrico $[0,1]$. Defina o processo $X(t)$ como:

$$\begin{aligned} X(t) &= X^1(t), \quad 0 \leq t \leq 1, \\ &= X^1(1) + X^2(t-1), \quad 1 \leq t \leq 2, \\ &\ldots \\ &= X^1(1) + X^2(1) + \ldots + X^n(1) + X^{n+1}(t-n), \quad n \leq t \leq n+1. \end{aligned}$$

Então, $X = \{X(t), 0 \leq t < \infty\}$ tem a propriedade desejada. X é chamado de MB sobre $[0,\infty)$ ou simplesmente de MB. Nessa seção, trataremos de tal processo. Usaremos as notações $X(t)$ ou X_t.

Teorema 9.5. *(a) Se $\mathcal{F}_t = \mathcal{F}\{X_s, s \leq t\}$, então X é um martingale relativamente a $\{\mathcal{F}_t, t \geq 0\}$.*
(b) $E(X_sX_t) = \min\{s,t\}$.

Prova: (a) Para $s < t$, temos que $E(X_t|\mathcal{F}_s) = E(X_t - X_s|\mathcal{F}_s) + X_s = E(X_t - X_s) + X_s = 0 + X_s = X_s$, usando o fato de X ter incrementos independentes e média zero.
(b) Se $s < t$, $E(X_sX_t) = E[X_s(X_t - X_s)] + E(X_s^2) = E(X_t - X_s)E(X_s) + s = s = \min\{s,t\}$. $\square$

O seguinte teorema pode ser provado usando resultados anteriores.

Teorema 9.6. *Seja T um tempo de parada finito. Defina $Y_t = X_{T+t} - X_T$. Então, $\{Y_t, t \geq 0\}$ é um MB, independente de $\mathcal{F}_T$ (propriedade Forte de Markov).*

Teorema 9.7. (Princípio da reflexão). *Seja T um tempo de parada finito. Defina $\{U_t, t \geq 0\}$ por:*

$$\begin{aligned} U_t &= X_t, \ \ se\ \ t < T; \\ &= 2X_T - X_t, \ \ se\ \ t \geq T. \end{aligned}$$

Então, $\{U_t, t \geq 0\}$ é um MB.

Prova: É suficiente provar as seguintes propriedades. (a) U_t tem trajetórias contínuas (por construção); (b) as distribuições finito-dimensionais de U_t são as mesmas que aquelas de X_t.

De fato, (a) é imediata. Para (b), consideramos em primeiro lugar o caso $n = 1$. Seja T um tempo de parada com um número enumerável de valores. Provaremos (b) para esse caso primeiramente. Temos que

$$E\{e^{isU_t}\} = E\{e^{isU_t} I_{\{T \leq t\}}\} + E\{e^{isU_t} I_{\{T > t\}}\}$$

$$= E\{e^{is(2X_T - X_t)} I_{\{T \leq t\}}\} + E\{e^{isX_t} I_{\{T > t\}}\}.$$

Agora,

$$E\{e^{is(2X_T - X_t)} I_{\{T \leq t\}}\} = \sum_{i \geq 1, a_i \leq t} E\Big\{e^{is(2X_{a_i} - X_t)} I_{\{T = a_i\}}\Big\},$$

onde $a_1, a_2, \ldots$ são os valores de T. Por sua vez, o último termo é igual a

$$\sum_{i \geq 1, a_i \leq t} E\Big\{e^{is[X_{a_i} - (X_t - X_{a_i})]} I_{\{T = a_i\}}\Big\} = \sum_{i \geq 1, a_i \leq t} E\Big\{e^{isX_{a_i}} I_{\{T = a_i\}}\Big\} E\Big\{e^{-is(X_t - X_{a_i})}\Big\},$$

por independência. Como X_t é simétrico (normal), $X_{a_i} - X_t$ tem a mesma distribuição que $X_t - X_{a_i}$, portanto da última igualdade obtemos

$$\sum_{i \geq 1, a_i \leq t} E\Big\{e^{isX_t} I_{\{T = a_i\}}\Big\} = E\Big\{e^{isX_t} I_{\{T \leq t\}}\Big\},$$

logo

$$E\{e^{isU_t}\} = E\{e^{isX_t} I_{\{T \leq t\}}\} + E\{e^{isX_t} I_{\{T > t\}}\} = E\{e^{isX_t}\}.$$

Caso de T arbitrário. Construa uma sequência T_n de tempos de parada, com um número enumerável de valores, tais que $T_n \downarrow T$. Por exemplo, defina $T_n = k/2^n$, se $(k-1)/2^n \leq T \leq k/2^n$.

Seja U_t^n o processo obtido usando T_n no lugar de T. Então, $U_t^n \to U_t$, quando $n \to \infty$, pois U_t tem trajetórias contínuas. Pelo TCD, $E\{e^{isU_t^n}\} \to E\{e^{isU_t}\}$, do que segue $E\{e^{isU_t}\} = E\{e^{isX_t}\}$ e portanto $U_t \sim N(0,t)$, para cada t.

Agora, observe que se $t_1 < t_2$, temos que $E\{e^{is_1U_{t_1}+is_2U_{t_2}}\} = E\{e^{is_1X_{t_1}+is_2X_{t_2}}\}$, veja o Problema 5. De modo similar, a mesma relação vale para $t_1,\ldots,t_n$, de modo que $(X_{t_1},\ldots,X_{t_n})$ tem a mesma distribuição que $(U_{t_1},\ldots,U_{t_n})$, para quaisquer $t_1,\ldots,t_n$, isto é, U_t e X_t têm as mesmas distribuições finito-dimensionais. $\square$

Teorema 9.8. *Seja $\{X_t, t \geq 0\}$ um MB e λ, R reais positivos. Então,*

$$P\Big\{\max_{0\leq t\leq R} X_t \geq \lambda\Big\} = 2P\{X_R \geq \lambda\}.$$

Prova: Seja T definido por $T = \inf\{t \leq R : X_t = \lambda\}$ e $T = R+1$, se esse conjunto for vazio. Então, T é um tempo de parada para $\{X_t\}$ e $P\{\max_{0\leq t\leq R} X_t > \lambda\} = P\{T \leq R\}$.

Defina $\{U_t\}$ por $U_t = X_t$, se $t < T$ e $U_t = 2X_T - X_t$, se $t \geq T$. Seja $T_1 = \inf\{t : U_t = \lambda\}$ e igual a $R+1$, se o conjunto for vazio. Então,

$$P\{T \leq R, X_R < \lambda\} = P\{T_1 \leq R, U_R > \lambda\} = P\{T \leq R, X_R > \lambda\},$$

pois $U \sim X$. Agora,

$$P\{\max_{0\leq t\leq R} X_t > \lambda\} = P\{T \leq R\} =$$

$$P\{T \leq R, X_R \geq \lambda\} + P\{T \leq R, X_R < \lambda\} =$$

$$P\{X_R \geq \lambda\} + P\{T \leq R, X_R > \lambda\} = P\{X_R \geq \lambda\} + P\{X_R > \lambda\},$$

pois $\{T \leq R\} \supset \{X_R \geq \lambda\}$. Como X é normal, obtemos o resultado. $\square$

Aplicações

[1] Seja $X = \{X_t, t \geq 0\}$ um MB. Então, quase todas as trajetórias de X são não limitadas.

De fato pelo teorema anterior, $P(\max_{0\leq t\leq R} X_t \geq \lambda) = 2P(\frac{X_R}{\sqrt{R}} \geq \frac{\lambda}{\sqrt{R}}) = 2P(Z \geq \lambda/\sqrt{R})$, onde $Z \sim N(0,1)$. Para $R \to \infty$, o último termo tende a $2(1/2) = 1$.

[2] Seja $T_a = \inf\{t \geq 0 : X_t = a\}$, sendo X um MB e $a > 0$. Pela Aplicação [1], $T_a < \infty$ q.c. Então,

$$P(T_a \leq t) = P(\max_{0\leq s\leq t} X_s \geq a) = 2P(X_t \geq a) = \frac{2}{\sqrt{2\pi t}}\int_a^\infty e^{-x^2/2t}dx.$$

A densidade é $f(t) = \frac{a}{\sqrt{2\pi}} t^{-3/2} e^{-a^2/2t}$, para $t > 0$. Veja o Problema 3.

Teorema 9.9. (Lei do Logaritmo Iterado) *Seja* $X = \{X_t, t \geq 0\}$ *um MB. Então,*

$$P\Big\{\limsup_{t\to\infty} \frac{X_t}{\sqrt{2t\log\log t}} = 1\Big\} = 1,$$

e

$$P\Big\{\liminf_{t\to\infty} \frac{X_t}{\sqrt{2t\log\log t}} = -1\Big\} = 1.$$

Prova: Basta provar a parte do limite superior.

(a) Provamos, primeiramente, que

$$\limsup_{t\to\infty} \frac{X_t}{\sqrt{2t\log\log t}} \leq 1, \text{ q.c.}$$

Seja $c > 1$. É suficiente provar que $\limsup_{t\to\infty} X_t/c\sqrt{2t\log\log t} \leq 1$, q.c. Tome $t_n = \alpha^n$, para $\alpha > 1$ a ser escolhido depois. Seja $M_n = \max_{0\leq t\leq t_n} X_t$. Como $P(M_n \geq \lambda) = 2P(X_{t_n} \geq \lambda)$, obtemos

$$P\{M_n \geq x\sqrt{t_n}\} = 2P\{X_{t_n} \geq x\sqrt{t_n}\} = \frac{2}{\sqrt{2\pi}}\int_x^\infty e^{-y^2/2}dy$$

$$\leq \frac{2}{\sqrt{2\pi}}\int_x^\infty \frac{y}{x}e^{-y^2/2}dy \leq \frac{2}{\sqrt{2\pi}}\frac{1}{x}e^{-x^2/2}.$$

Escolha $x = x_n = (c/\sqrt{t_n})\sqrt{2t_{n-1}\log\log t_{n-1}}$. Então,

$$P\{M_n \geq x_n\sqrt{t_n}\} \leq \frac{2}{\sqrt{2\pi}}\frac{1}{x_n}e^{-x_n^2/2}$$

$$\leq \frac{\sqrt{\alpha}}{\sqrt{\pi\log\log t_{n-1}}}(\log t_{n-1})^{-c^2/\alpha} \leq \frac{\sqrt{\alpha}}{\sqrt{\pi\log(n-1)}}(\log\alpha)^{-c^2/\alpha}(n-1)^{-c^2/\alpha}.$$

Como $c > 1$, escolha α tal que $1 < \alpha < c^2$, de modo que

$$\sum_n P\Big\{M_n > c\sqrt{2t_{n-1}\log\log t_{n-1}}\Big\} \leq K\sum_n (n-1)^{-c^2/\alpha} < \infty.$$

Logo, por Borel-Cantelli, se $t_{n-1} < t \leq t_n$,

$$X(t) \leq M_n \leq c\sqrt{2t_{n-1}\log\log t_{n-1}} \leq c\sqrt{2t\log\log t},$$

sendo que a segunda desigualdade vale para todos exceto um número finito de n. Ou seja,

$$P\Big\{\lim_{t\to\infty}\sup \frac{X_t}{\sqrt{2t\log\log t}} \leq c\Big\} = 1, \quad \forall c > 1.$$

(b) Basta provar que, se $c < 1$, então

$$P\Big\{\lim_{t\to\infty} \sup \frac{X_t}{\sqrt{2t\log\log t}} \geq c\Big\} = 1.$$

Coloquemos $x_n = c'\sqrt{2\log\log t_n}$, com $c' < 1$, arbitrário, e seja $Y_n := X_{t_n} - X_{t_{n-1}}$, $n \geq 1$, que são independentes. Agora, para n suficientemente grande temos

$$P\{Y_n > x_n\sqrt{t_n - t_{n-1}}\} = \frac{1}{\sqrt{2\pi}}\int_{x_n}^{\infty} e^{-y^2/2}dy$$

$$\geq \frac{1}{\sqrt{2\pi}}\frac{1}{x_n}e^{-x_n^2/2} \geq [c'\sqrt{4\pi\log n}]^{-1}(n\log\alpha)^{-(c')^2},$$

logo

$$\sum_n P\{Y_n > x_n\sqrt{t_n - t_{n-1}}\} \geq C\sum_n \frac{n^{-(c')^2}}{\sqrt{\log n}} = \infty,$$

do que seque, por Borel-Cantelli, que $\{Y_n > c'\sqrt{2\log\log t_n}\sqrt{t_n - t_{n-1}}\}$ ocorre i.v, ou ainda $\{Y_n > c'\sqrt{(\alpha-1)/\alpha}\sqrt{2t_n\log\log t_n}\}$ ocorre i.v. Portanto,

$$X_{t_n} \geq Y_n - |X_{t_{n-1}}| \geq \left(c'\sqrt{\frac{\alpha-1}{\alpha}} - \frac{2}{\sqrt{\alpha}}\right)\sqrt{2t_n\log\log t_n}$$

ocorre i.v, pois por (a),

$$P\Big\{X_{t_{n-1}} \leq 2\sqrt{2t_{n-1}\log\log t_{n-1}} \text{ para todos exceto um número finito de } n\Big\} = 1.$$

Ou seja, como $t_{n-1} = \alpha^{n-1} = t_n/\alpha$,

$$P\Big\{X_{t_{n-1}} \leq \frac{2}{\sqrt{\alpha}}\sqrt{2t_n\log\log t_n}, \text{para todos exceto un número finito de } n\Big\} = 1.$$

Escolha $c < c'$. Para α suficientemente grande, temos também que

$$c < c'\sqrt{(\alpha-1)/\alpha} - 2/\sqrt{\alpha},$$

logo

$$P\Big\{X_{t_n} \geq c\sqrt{2t_n\log\log t_n} \text{ i.v}\Big\} = 1.$$

Como $c' < 1$, arbitrário, c pode ser escolhido menor do que 1 arbitrariamente, perto de 1. $\square$

Nota: Na prova foi usado o fato que $\int_x^\infty e^{-u^2/2}du \geq (1/x)e^{-x^2/2}$, para x suficientemente grande. Para provar esta desigualdade, basta integrar por partes.

Teorema 9.10. *Suponha X MB e seja $Y_t = t \cdot X(1/t)$. Então, $Y = \{Y_t, t \geq 0\}$ é também MB.*

Prova: Primeiramente, Y tem trajetórias contínuas, exceto possivelmente em $t = 0$. De fato, Y é contínuo em $t = 0$:

$$\lim_{t\to 0} Y_t = \lim_{t\to 0} tX(1/t) = \lim_{t\to 0} \frac{X(1/t)}{\sqrt{(2/t)\log\log t^{-1}}}\sqrt{(2/t)\log\log t^{-1}} \cdot t.$$

O primeiro termo afetado pelo limite é limitado, pela lei do logaritmo iterado, e o segundo termo do produto tende a zero, para $t \to 0$. Também, as distribuições finito-dimensionais de Y são normais. Seja $s < t$. Então, $E(X_sX_t) = s$ and $E(Y_sY_t) = stE[X(1/s)X(1/t)]st(1/t) = s$, logo X e Y têm a mesma função de covariância, logo têm as mesmas distribuições finito-dimensionais. $\square$

Corolário 9.2. (Lei Local do Logaritmo Iterado) *Suponha t_0 fixado. Então,*

$$\limsup_{h\to 0} \frac{X(t_0+h) - X(t_0)}{\sqrt{2h\log\log h^{-1}}} = 1, \ q.c$$

e

$$\liminf_{h\to 0} \frac{X(t_0+h) - X(t_0)}{\sqrt{2h\log\log h^{-1}}} = -1, \ q.c.$$

Prova: É suficiente provar o caso $t_0 = 0$. Temos que

$$P\left\{\limsup_{h\to 0} \frac{X(h)}{\sqrt{2h\log\log h^{-1}}} = 1\right\} = P\left\{\limsup_{h\to 0} \frac{hX(h^{-1})}{\sqrt{2h\log\log h^{-1}}} = 1\right\} =$$

$$P\left\{\limsup_{h\to 0} \frac{X(h^{-1})}{\sqrt{2h^{-1}\log\log h^{-1}}} = 1\right\} = 1,$$

pela lei ordinária do logaritmo iterado. $\square$

Aplicações

[1] $\limsup_{t\to 0} \frac{X_t}{t} = +\infty$ e $\liminf_{t\to 0} \frac{X_t}{t} = -\infty$.

Basta escrever $X(t)/t$ como

$$\frac{X(t)}{t} = \frac{X(t)}{\sqrt{2t\log\log t^{-1}}}\frac{\sqrt{2t\log\log t^{-1}}}{t}.$$

O primeiro termo do produto é limitado, pelo Corolário 9.2 e o segundo termo do produto tende para $+\infty$, quando $t \to 0$.

[2] Quando $t \to 0$, X_t cruza o eixo das abscissas infinitas vezes.

Segue de [1], ou $X(t)/\sqrt{2t \log \log t^{-1}}$ deve mudar de sinal infinitas vezes, mas como as trajetórias são contínuas, deve cruzar o eixo x infinitas vezes.

Teorema 9.11. *O MB tem quase todas as trajetórias de variação não limitada.*

Necessitamos do seguinte lema.

Lema 9.2. *Sejam $\frac{1}{2^n}, \frac{2}{2^n}, \ldots, \frac{2^n}{2^n}$ pontos de $[0,1]$ e seja $S_n = \sum_{k=1}^{2^n} |X(k/2^n) - X((k-1)/2^n)|^2$. Então, $S_n \to 1$ q.c, quando $n \to \infty$.*

Prova: De fato, temos que

$$S_n - 1 = \sum_{k=1}^{2^n} \{|X(k2^{-n}) - X((k-1)2^{-n})|^2 - 2^{-n}\},$$

logo

$$P\left(|S_n - 1| > \varepsilon\right) \leq \frac{E(S_n-1)^2}{\varepsilon^2} = \frac{\text{Var}(S_n)}{\varepsilon^2} =$$

$$2^n \frac{\text{Var}(X(2^{-n})^2)}{\varepsilon^2} = \frac{\text{Var}(Z^2)}{2^n \varepsilon^2},$$

onde $Z \sim N(0,1)$. Segue-se que

$$\sum_{n \geq 1} P\{|S_n - 1| > \varepsilon\} \leq \frac{C}{\varepsilon^2} \sum_{n \geq 1} 2^{-n} < \infty.$$

Por Borel-Cantelli, $S_n \to 1$ q.c. $\square$

Prova do Teorema: Mostramos que quase todas as trajetórias sobre $[0,1]$ são de variação não limitada. Suponha que existe um conjunto $\tilde{\Omega}$ tal que $P(\tilde{\Omega}) > 0$, e tal que para todo $\omega \in \tilde{\Omega}$ temos

$$\sum_{k=1}^{2^n} |X(k2^{-n}, \omega) - X((k-1)2^{-n}, \omega)| \leq M(\omega) < \infty, \quad \text{para todo } n.$$

Então,

$$\begin{aligned}\sum_{k=1}^{2^n} |X(k2^{-n},\omega) - X((k-1)2^{-n},\omega)|^2 &\leq \max_{1\leq k\leq 2^n} |X(k2^{-n},\omega) - X((k-1)2^{-n},\omega)| \\ &\quad \times \sum_{k=1}^{2^n} |X(k2^{-n},\omega) - X((k-1)2^{-n},\omega)| \\ &\leq M(\omega) \max_{1\leq k\leq 2^n} |X(k2^{-n},\omega) - X((k-1)2^{-n},\omega)|.\end{aligned}$$

Logo,

$$\frac{\sum_{k=1}^{2^n} |X(k2^{-n},\omega) - X((k-1)2^{-n},\omega)|^2}{\max_{1\leq k\leq 2^n} |X(k2^{-n},\omega) - X((k-1)2^{-n},\omega)|} \leq M(\omega).$$

Mas para $n \to \infty$, o numerador tende a 1 q.c e o denominador tende a zero, pois o MB tem trajetórias contínuas. Isso é uma contradição. $\square$

Teorema 9.12. *Seja X um MB sobre $[0,1]$. Então, $\lambda\{t \in [0,1] : X_t(\omega) = 0\} = 0$, para quase todo ω, sendo λ a medida de Lesbegue no intervalo $[0,1]$.*

Prova: Seja $I_{\{0\}}$ o indicador do $\{0\}$. Então, $0 = \int_0^1 [\int_\Omega I_{\{0\}}(X_t) dP] dt$, pois a integral interna reduz-se a $P\{\omega : X_t(\omega) = 0\} = 0$, dado que X_t é normal, para cada t. Agora, $X(t,\omega)$ é mensurável, como uma aplicação de $[0,1] \times \Omega \to \mathbb{R}$ (veja o Capítulo 5). Logo, $I_{\{0\}}(X(t,\omega))$ é mensurável em (t,ω) como uma aplicação entre os mesmo conjuntos anteriores. Segue-se, portanto, por Fubini, que $0 = \int_\Omega [\int_0^1 I_{\{0\}}(X_t) dt] dP$. Logo, para quase todo ω, $\int_0^1 I_{\{0\}}(X(t,\omega)) dt = 0$, ou ainda, $\lambda\{t \in [0,1] : X(t,\omega) = 0\} = 0$, para quase todo ω. $\square$

9.3 Aplicações do Teorema de Donsker

Nesta seção estudamos algumas aplicações do Teorema de Donsker, a saber, a estatística de Kolmogorov-Smirnov, a lei do arco seno e uma lei do logaritmo iterado derivada do Teorema de Skorohod.

[1] **Estatística de Kolmogorov-Smirnov**

Sejam $Y_1, Y_2, \ldots$ v.a's i.i.d, com f.d comum F, suposta contínua e seja $F_n(x)$ a f.d. empírica, baseada em uma amostra de F de tamanho n. O Teorema de Glivenko-Cantelli nos diz que

$$\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| \to 0, \quad n \to \infty.$$

Seja

$$D_n = \sqrt{n} \sup_{x \in \mathbb{R}} |F_n(x) - F(x)|. \tag{9.9}$$

Um intervalo $I = [a, b]$ é chamado *intervalo de constância* para F se $P\{Y_1 \in [a, b]\} = 0$ e nenhum intervalo contendo este intervalo tem essa propriedade.

Seja B a reunião de todos os intervalos de constância para F. Então, $D_n = \sqrt{n} \sup_{x \in B^c} |F_n(x) - F(x)|$, q.c. Seja $U_k = F(Y_k)$. Então U_k é uniforme em $[0, 1]$.

Vamos usar os seguintes lemas.

Lema 9.3. *A distribuição de D_n não depende de F.*

Prova: Temos que

$$D_n = \sqrt{n} \sup_{x \in B^c} |F_n(x) - F(x)| = \sqrt{n} \sup_{x \in B^c} |G_n(F(x)) - F(x)|,$$

onde G_n é a função de distribuição empírica de $U_1, \ldots, U_n$. Agora observe que o último termo é igual a $\sqrt{n} \sup_{0<x<1} |G_n(x) - x|$, porque quando x varia em B^c, $F(x)$ varia em $(0, 1)$ e isso envolve somente a distribuição uniforme. $\square$

Lema 9.4. *Sejam $\mathcal{E}_1, \mathcal{E}_2, \ldots$ v.a's i.i.d exponencias com média 1 e $R_k = \mathcal{E}_1 + \ldots + \mathcal{E}_k$. Sejam $U_{(1,n)}, \ldots, U_{(n,n)}$ as estatísticas de ordem de n v.a's com distribuição uniforme no intervalo $[0, 1]$, $U_1, \ldots, U_n$. Então:*

(a) A distribuição conjunta de $(U_{(1,n)}, \ldots, U_{(n,n)})$ tem densidade $f(y_1, \ldots, y_n) = n!$, se $0 \le y_1 < y_2 < \cdots < y_n \le 1$ e 0 senão;

(b) $(R_1/R_{n+1}, \ldots, R_n/R_{n+1}) \sim (U_{(1,n)}, \ldots, U_{(n,n)})$.

Prova. Veja o Problema 8.

Teorema 9.13. $D_n \xrightarrow{\mathcal{D}} Y$, *onde* $Y = \sup_{0 \le t \le 1} |W(t) - tW(1)|$, *sendo* $W(t)$ *um MB.*

O processo $B(t) = W(t) - tW(1)$ é chamado *ponte browniana*. Vemos que $B(0) = B(1) = 0$. A distribuição de Y é dada por

$$P(Y \le x) = 1 + 2 \sum_{k=1}^{\infty} (-1)^k e^{-2k^2 x^2}, \;\; x > 0.$$

Prova: Temos que $Y_k \le x$ se, e somente se, $F(Y_k) \le F(x)$, se $x \in B^c$. Para provar essa afirmação, a parte $\Rightarrow$ é óbvia, pois F é crescente. Para a parte $\Leftarrow$, suponha que F seja estritamente crescente à direita de x; se $Y_k \ge x$, então, $F(Y_k) > F(x)$, uma contradição. Suponha, agora, que F não seja estritamente crescente à direita de x. Como $x \in B^c$ e supondo $Y_k > x$, então $F(Y_k) \ge F(x)$ e portanto $F(Y_k) = F(x)$. Logo, Y_k está no intervalo de constância $[x, x_0]$ e portanto $P(x \le Y_k \le x_0) = 0$.

Pelo Lema 9.3, para provar o teorema, basta provar o caso em que $Y_1, Y_2, \ldots$ são uniformes em $[0, 1]$. Note que $G_n(x)$ é constante sobre os intervalos $I_k = [U_{(k,n)}, U_{(k+1,n)}]$ e cresce em saltos de $1/n$. Logo,

$$D_n = \sqrt{n} \sup_{k \leq n} \left\{ \sup_{x \in I_k} |G_n(x) - x| \right\} =$$

$$\sqrt{n} \sup_{k \leq n} \left\{ \sup_{x \in I_k} |G_n(U_{(k,n)}) - x| \right\} =$$

$$\sqrt{n} \sup_{k \leq n} \left\{ \sup_{x \in I_k} \left| \frac{k}{n} - x \right| \right\}.$$

Temos que, ou o último termo é igual a $|k/n - U_{(k,n)}|$ ou é igual a $|k/n - U_{(k+1,n)}| = |(k+1)/n - U_{(k+1,n)}| + M/n$, com $|M| \leq 1$.

Portanto, para provar que D_n converge em distribuição para Y, é suficiente provar que $\sqrt{n} \sup_{k \leq n} |k/n - U_{(k,n)}|$ converge em distribuição para $\sup_{0 \leq t \leq 1} |B(t)|$.

Usando o Lema 9.4, devemos mostrar que $\sqrt{n} \sup_{k \leq n} |k/n - R_k/R_{n+1}|$ converge em distribuição para o que se deseja. Mas essa quantidade é igual a

$$\frac{n}{R_{n+1}} \sup_{k \leq n} \left| \frac{R_k - k}{\sqrt{n}} - \frac{k}{n} \frac{R_{n+1} - n}{\sqrt{n}} \right|.$$

Note que $n/R_{n+1} \to 1$, pois $R_n = \sum_{i=1}^n \mathcal{E}_i$ e $R_n/n \to E(\mathcal{E}_1) = 1$. Como $\mathcal{E}_{n+1}/\sqrt{n} \to 0$ em probabilidade, é suficiente provar que $\sup_{k \leq n} |(R_k - k)/\sqrt{n} - (k/n)(R_{n+1} - n)/\sqrt{n}|$ converge em distribuição. Seja $S_k = R_k - k$ e seja

$$X^{(n)}(t) = \frac{S_{[nt]}}{\sqrt{n}} + \frac{nt - [nt]}{\sqrt{n}} \left[\mathcal{E}_{[nt]+1} - 1 \right].$$

Pelo teorema de Donsker, $X^{(n)}(t)$ converge em distribuição para $W(t)$. Seja $h : C([0, 1]) \to \mathbb{R}$ definida por $h(x) = \sup_{0 \leq t \leq 1} |x(t) - tx(1)|$. Segue-se que h é contínua, logo $h(X^{(n)}(t))$ converge em distribuição para $h(W(t))$. Por outro lado,

$$\sup_{k \leq n} \left| \frac{R_k - k}{\sqrt{n}} - \frac{k}{n} \frac{R_{n+1} - n}{\sqrt{n}} \right| = \sup_{0 \leq t \leq 1} |X^{(n)}(t) - tX^{(n)}(1)|$$

pois, note que o supremo é determinado pelos vértices da linha poligonal. $\square$

[2] **Lei do Arco seno**

Seja $x \in C([0, 1])$ e defina uma função $h : C([0, 1]) \to \mathbb{R}$ como segue, sendo λ a medida de Lebesgue:

$$h(x) = \lambda\{t \in [0, 1] : x(t) > 0\}.$$

Então, temos:

(a) h é uma função mensurável de $C([0,1])$ em $\mathbb{R}$.

(b) Seja D_h o conjunto dos pontos de descontinuidade de h. Então,

$$D_h = \{x \in C([0,1]) : \lambda\{t \in [0,1] : x(t) = 0\} > 0\}.$$

(c) $W(D_h) = 0$.

(d) Sejam $Y_i, i \geq 1$ v.a's i.i.d, média zero e variância σ^2. Defina $X^n(t)$ como usualmente foi feito antes. Então, $h(X^n) \to h(W)$. Seja $S_n = Y_1 + \ldots + Y_n$. Denotemos por $\#S_n$ o número de índices $k \leq n$ tais que $S_k > 0$. Seja $R_n = \#S_n/n$ a proporção das vezes que $S_k > 0$. Então, $h(X^n) - R_n$ converge para 0 em probabilidade quando $n \to \infty$.

(e) Logo,

$$P(R_n \leq x) \to P\{\omega : \lambda\{t \in [0,1] : W(t) > 0\} \leq x\} = P(h(W) \leq x) \tag{9.10}$$

quando $n \to \infty$. Teríamos, então, a distribuição limite de R_n, desde que pudéssemos calcular (9.10).

(f) Para calcular (9.10), consideremos um caso especial,ou seja, tomemos os Y_i tais que $P(Y_1 = 1) = P(Y_1 = -1) = 1/2$. Para esse caso, sabemos que (veja Feller, 1968) $P(R_n \leq x) \to \frac{2}{\pi} \arcsin \sqrt{x}$, logo

$$P[h(W) \leq x] = \frac{2}{\pi} \arcsin \sqrt{x}.$$

[3] **Lei do Logaritmo Iterado: Teorema de Skorokhod**

A seguinte versão do teorema de Skorokhod não será provada. Veja Billingsley (1999).

Teorema 9.14. (Skorokhod). *Sejam $X_i, i \geq 1$ v.a's i.i.d, média zero e variância σ^2 e seja $S_n = \sum_{i=1}^n X_i$. Então, existe um espaço de probabilidade sobre o qual podemos definir:*

(i) v.a's i.i.d, não negativas $T_i, i \geq 1$, com $E(T_1) = \sigma^2$;

(ii) um MB $W = \{W(t)), t \geq 0\}$ tal que

$$(S_1, S_2, \ldots) \sim (W(T_1), W(T_1 + T_2), \ldots).$$

As v.a's $T_1, T_1 + T_2, \ldots$ podem ser vistas como tempos de parada para W.

Teorema 9.15. (Lei do Logaritmo Iterado) *Sejam $X_i, i \geq 1$, i.i.d, média zero, variância 1 e $S_n = \sum_{i=1}^n X_i$. Então,*

$$P\Big\{\lim_{n\to\infty}\sup\frac{S_n}{\sqrt{2n\log\log n}}=1\Big\}=1,\quad q.c$$

e

$$P\Big\{\lim_{n\to\infty}\inf\frac{S_n}{\sqrt{2n\log\log n}}=-1\Big\}=1\ \ q.c.$$

Prova: Sejam W e $T_1, T_2, \ldots$ como no Teorema de Skorokhod. Para provar o teorema é suficiente provar que:

1) $\limsup_{n\to\infty}\frac{W(T_1+\ldots+T_n)}{\sqrt{2n\log\log n}}=1$, q.c. Para provar isso, devemos provar

2) $\frac{W(T_1+\ldots+T_{\lfloor t\rfloor})-W(t)}{\sqrt{2t\log\log t}}\to 0$, q.c.

Pela lei do logaritmo iterado para MB, temos que $\limsup_{t\to\infty}\frac{W(t)}{\sqrt{2t\log\log t}}=1$, e se 2) vale, então $\limsup\frac{W(T_1+\ldots+T_{\lfloor t\rfloor})}{\sqrt{2t\log\log t}}=1$, q.c. Portanto, o resultado segue pois $\frac{\sqrt{2t\log\log t}}{\sqrt{2\lfloor t\rfloor\log\log\lfloor t\rfloor}}\to 1$, $t\to\infty$. Provemos 2).

Como $T_1, T_2, \ldots$ são i.i.d e $E(T_1)=1$, a LFGN nos dá $(T_1+\ldots+T_{\lfloor t\rfloor})/t\to 1$, q.c, quando $t\to\infty$. Seja $\varepsilon>0$. Para cada ω, existe um número $\tau(\omega)$ tal que $t>\tau$ implica

$$\frac{1}{1+\varepsilon}\le\frac{T_1+\ldots+T_{[t]}}{t}\le 1+\varepsilon,$$

ou

$$\frac{t}{1+\varepsilon}\le T_1+\ldots+T_{[t]}\le(1+\varepsilon)t,$$

portanto para tal t temos

$$|W(T_1+\ldots+T_{\lfloor t\rfloor})-W(t)|\le\sup_{t(1+\varepsilon)^{-1}\le s\le t(1+\varepsilon)}|W(s)-W(t)|.$$

Seja $t_k=(1+\varepsilon)^k$. Então, se $t_k\le t\le t_{k+1}$, então

$$\begin{aligned}|W(T_1+\ldots+T_{\lfloor t\rfloor})-W(t)|&\le\sup_{t_{k-1}\le s\le t_{k+2}}|W(s)-W(t)|\\&\le 2\sup_{t_{k-1}\le s\le t_{k+2}}|W(s)-W(t_{k-1})|\\&=:2M_k.\end{aligned}$$

Para mostrar que $\frac{|W(T_1+\ldots+T_{\lfloor t\rfloor})-W(t)|}{D_t} \to 0$, onde $D_t = \sqrt{2t\log\log t}$, é suficiente provar que, para $\Delta > 0$ arbitrário, $\limsup_k M_k/D_{t_k} \leq \Delta$, para $D_{t_k} \leq D_t$. Para tanto, usamos Borel-Cantelli. Note que

$$P\{M_k > x\} = P\Big\{\sup_{t_{k-1}\leq s\leq t_{k+2}} |W(s)-W(t_{k-1})| > x\Big\} = 2P\Big\{|W(t_{k+2})-W(t_{k-1})| > x\Big\}.$$

Note, também, que

$$t_{k+2} - t_{k-1} = (1+\varepsilon)^k[(1+\varepsilon)^2 - (1+\varepsilon)^{-1}] = t_k\delta$$

com $\delta := (1+\varepsilon)^2 - (1+\varepsilon)^{-1} = 2\varepsilon + \varepsilon^2$. Vamos considerar

$$P\left\{\frac{M_k}{D_{t_k}} > \sqrt{2\delta}\right\} = 4P\left\{\frac{W(t_{k+2}) - W(t_{k-1})}{\sqrt{t_{k+2} - t_{k-1}}} > 2\sqrt{\log\log t_k}\right\}$$
$$= \frac{4}{\sqrt{2\pi}}\int_{2\sqrt{\log\log t_k}}^{\infty} e^{-t^2/2}dt \leq \frac{4}{\sqrt{2\pi}}e^{-(2\sqrt{\log\log t_k})^2/2} = \frac{4}{\sqrt{2\pi}}\frac{1}{k^2[\log(1+\varepsilon)]^2}$$

para k suficientemente grande.

Segue que

$$\sum_k P\Big\{\frac{M_k}{D_{t_k}} > \sqrt{2\delta}\Big\} < \infty,$$

ou seja $M_k/D_{t_k} \leq \sqrt{2\delta}$ para todo k suficientemente grande. Como δ pode ser feito arbitrariamente pequeno, tomando-se ε arbitrariamente pequeno, obtemos o resultado. $\square$

Problemas

1. Prove que uma projeção π_t é uma função contínua definida sobre $C([0,1])$.
2. Prove o Teorema 9.6.
3. Prove que T_a da Aplicação [2] seguindo o Teorema 9.8 tem uma distribuição estável, com índice $\alpha = 1/2$. Esta distribuição é chamada *lei de Lévy*.
4. Prove (9.8).
5. Na prova do Teorema 9.7, prove a afirmação: se $t_1 < t_2$, temos que $E\{e^{is_1U_{t_1}+is_2U_{t_2}}\} = E\{e^{is_1X_{t_1}+is_2X_{t_2}}\}$.
6. Prove o afirmado na Nota depois da prova do Teorema 9.9.
7. Usando o Teorema 9.4 (Donsker) e Corolário 9.1, prove que

$$P\Big\{\max_{k\leq n}\frac{S_k}{\sigma\sqrt{n}} \leq x\Big\} \to \frac{2}{\sqrt{2\pi}}\int_0^x e^{-t^2/2}dt,$$

onde $S_n = \sum_{i=1}^n Y_i$, e $Y_1, Y_2, \ldots$ são v.a's i.i.d, com média zero e variância σ^2.

8. (Teorema de Donsker no caso Lindeberg) Sejam $Y_{n,1}, Y_{n,2}, \ldots, Y_{n,k_n}$ v.a's independentes, de média zero, variância $\sigma^2_{n,i}$ e $S_{n,i} = \sum_{j=1}^{i} Y_{n,j}$, $s^2_{n,i} = \sum_{j=1}^{i} \sigma^2_{n,j}$. Seja X_n o processo estocástico com trajetórias contínuas, definido por $X_n = \{X_n(t), 0 \leq t \leq 1\}$,

$$X_n(t) = \frac{S_{n,i}}{s_{n,i}}, \quad \text{se } t = \frac{s^2_{n,i}}{s^2_{n,k_n}},$$

e linear entre os valores. Suponha que os Y's satisfaçam as condições do Teorema de Lindeberg. Prove o Teorema de Donsker nesse caso.

9. Seja W a medida de Wiener. Encontre:

(a) $W\{x \in C([0,1]) : \sup_{0 \leq t \leq 1/2} x(t) \leq 1\}$;

(b) $W\{x \in C([0,1]) : 0 \leq x(t) \leq 1, \text{ para todo } t\}$.

10. (a) Seja T um tempo de parada, com $E(T) < \infty$. Se X for um MB, prove que $E(X_T) = 0$ (Use o TAO).

(b) Usando (a), prove que se $b > 0$, então o tempo esperado para que o MB atinja b é infinito (mesmo que o MB atingirá b com probabilidade 1).

Capítulo 10

Cadeias de Markov

Neste capítulo estudaremos os conceitos e propriedades principais sobre cadeias de Markov discretas. As referências que serão usadas são Chung (1967), Freedman (1983) e Norris (1997).

10.1 A propriedade de Markov

Definição 10.1. *Seja $X = \{X_n, n \geq 0\}$ um p.e com parâmetro discreto e $\mathcal{F}_n = \mathcal{F}\{X_1, \ldots, X_n\}$. Dizemos que X é um processo de Markov se, para todo conjunto de Borel B e todo n, temos*

$$P\{X_{n+1} \in B | \mathcal{F}_n\} = P\{X_{n+1} \in B | X_n\}. \tag{10.1}$$

A equação (10.1) é chamada *Propriedade de Markov.*

Note que X é um p.e de Markov se, e somente se,

$$E\{f(X_{n+1})|\mathcal{F}_n\} = E\{f(X_{n+1})|X_n\},$$

para toda função f de Borel limitada.

Exemplo 10.1. São exemplos triviais de processos de Markov:

(1) Uma sequência $\{X_n, n \geq 1\}$ de v.a's independentes.

(2) Uma sequência $\{X_n, n \geq 1\}$ de v.a's tais que $X_n = Y_1 + \ldots + Y_n$, sendo $Y_i, i \geq 1$ independentes.

As duas proposições a seguir dão critérios para se saber se X é um p.e de Markov.

Proposição 10.1. *Seja $\mathcal{F}_n^{'} = \mathcal{F}\{X_{n+1}, X_{n+2}, \ldots\}$. X é um processo de Markov se, e somente se, para todo $M \in \mathcal{F}_n^{'}$, tivermos $P(M|\mathcal{F}_n) = P(M|X_n)$.*

Prova: ($\Leftarrow$) trivial

($\Rightarrow$) Sabemos que, para toda função de Borel f, limitada,

$$E\{f(X_{n+1})|\mathcal{F}\} = E\{f(X_{n+1})|X_n\}^{\cdot} \tag{10.2}$$

Considere $f : \mathbb{R}^2 \to \mathbb{R}$, Borel, limitada. Vamos provar a relação correspondente a (10.2), para $f(X_{n+1}, X_{n+2})$. Para isso, considere primeiramente $f(x, y) = g(x)h(y)$, sendo f e g funções borelianas limitadas. Temos que

$$E\{f(X_{n+1}, X_{n+2})|\mathcal{F}_n\} = E\{g(X_{n+1})h(X_{n+2})|\mathcal{F}_n\} = E\{g(X_{n+1})E[h(X_{n+2})|\mathcal{F}_{n+1}]|\mathcal{F}_n\},$$

pois $\mathcal{F}_{n+1} \supset \mathcal{F}_n$. Por (10.2), o último termo da igualdade acima é igual a

$$E\{g(X_{n+1})E[(h(X_{n+2})|\mathcal{F}_{n+1}]|X_n\} = E\{g(X_{n+1})E[(h(X_{n+2})|X_{n+1}]|X_n\} =$$

$$= E\{g(X_{n+1})E[h(X_{n+2})|X_{n+1}, X_n]|X_n\} = E\{E[g(X_{n+1})h(X_{n+2})|X_{n+1}, X_n]|X_n\} =$$
$$E\{g(X_{n+1})h(X_{n+2})|X_n\},$$

pois $\mathcal{F}\{X_n, X_{n+1}\} \supset \mathcal{F}\{X_n\}$. Portanto,

$$E\{f(X_{n+1}, X_{n+2})|\mathcal{F}_n\} = E\{f(X_{n+1}, X_{n+2})|X_n\}, \tag{10.3}$$

para f da forma acima. Por um argumento de classe monotônica (10.3) vale para qualquer f boreliana, limitada de $\mathbb{R}^2$ em $\mathbb{R}$. De modo similar, podemos provar que

$$E\{f(X_{n+1}, X_{n+2}, \ldots, X_{n+k})|\mathcal{F}_n\} = E\{f(X_{n+1}, X_{n+2}, \ldots, X_{n+k})|X_n\},$$

para toda função de Borel limitada $f : \mathbb{R}^k \to \mathbb{R}$.

Em particular, se $A \in \mathcal{F}\{X_{n+1}, \ldots, X_{n+k}\}$, então $P\{A|\mathcal{F}_n\} = P\{A|X_n\}$, tomando $f = I_A$. Como $\mathcal{F}_n^{'} = \mathcal{F}\{X_{n+1}, X_{n+2}, \ldots\} = \vee_{k=1}^{\infty}\mathcal{F}\{X_{n+1}, \ldots, X_{n+k}\}$, temos o resultado por um argumento de classe monotônica. $\square$

Proposição 10.2. *X é um processo de Markov se, e somente se, sempre que $M \in \mathcal{F}_n^{'}$ e $N \in \mathcal{F}_n$, tivermos $P\{M \cap N\}|X_n\} = P\{M|X_n\}P\{N|X_n\}$.*

Prova: Será suficiente provar que a proposição é equivalente à Proposição 10.1.

(a) Suponha que

$$P(M|\mathcal{F}_n) = P(M|X_n), \quad M \in \mathcal{F}_n^{'}. \tag{10.4}$$

Mostremos que $P(MN|X_n) = P(M|X_n)P(N|X_n)$, para $N \in \mathcal{F}_n$. Temos que

$$P(M|X_n)P(N|X_n) = E(I_M|X_n)E(I_N|X_n) =$$

$$E(I_N E(I_M|X_n)|X_n) = E(I_N E(I_M|\mathcal{F}_n)|X_n),$$

pois $E(I_M|X_n)$ é $\mathcal{F}\{X_n\}$-mensurável e por (10.4). O último termo é igual a

$$E(E(I_N I_M|\mathcal{F}_n)|X_n) = E(I_N I_M|X_n) = P(M \cap N|X_n),$$

pois I_N é $\mathcal{F}_n$-mensurável.

(b) Suponha, agora, que

$$P(M \cap N|X_n) = P(M|X_n)P(N|X_n). \tag{10.5}$$

Mostremos que (10.4) vale. Tome qualquer conjunto $N \in \mathcal{F}_n$ e considere

$$\int_N P(M|X_n)dP = E(I_N E(I_M|X_n)) = E(E(I_N|X_n)E(I_M|X_n)) =$$

$$= E(P(N|X_n)P(M|X_n)) = E(P(M \cap N|X_n)) = P(M \cap N),$$

usando (10.5). Considere, agora,

$$\int_N P(M|\mathcal{F}_n)dP = \int_N E(I_M|\mathcal{F}_n)dP = \int_N I_M dP = P(M \cap N)$$

pela definição de esperança condicional, dado que $N \in \mathcal{F}_n$. Portanto, para qualquer conjunto $N \in \mathcal{F}_n$, temos

$$\int_N P(M|X_n)dP = \int_N P(M|\mathcal{F}_n)dP,$$

o que implica que $P(M|\mathcal{F}_n) = P(M|X_n)$. $\square$

Esta proposição nos diz que, dado o presente, o passado e o futuro são independentes.

Corolário 10.1 *Se $X_0, X_1, \ldots, X_n$ é um processo de Markov, também o será X_n, $X_{n-1}, \ldots, X_0$.*

10.2 Cadeias de Markov

Definição 10.2. *Seja $X = \{X_n, n \geq 1\}$ um processo de Markov. Seja I o conjunto de todos os possíveis valores de X_n, o chamado espaço de estados de X. Suponha que I seja enumerável. Nesse caso, chamamos X de Cadeia de Markov (CM).*

No caso de uma CM, a propriedade de Markov fica

$$P\{X_{n+1} = i_{n+1}|X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n\} = P\{X_{n+1} = i_{n+1}|X_n = i_n\},$$

ou, alternativamente para todo $k \geq 1$ e $0 \leq n_0 < n_1 \cdots < n_k$,

$$P\{M|X_{n_0} = i_0, \ldots, X_{n_k} = i_k\} = P\{M|X_{n_k} = i_k\}, \quad \text{se } M \in \mathcal{F}'_{n_k}.$$

Por exemplo,

$$P\{X_8 = a, X_5 = b|X_2 = c, X_3 = d\} = P\{X_8 = a, X_5 = b|X_3 = d\}.$$

Não é verdade, em geral, que se $M \in \mathcal{F}'_n$, então

$$P\{M|X_1 \in A_1, \ldots, X_n \in A_n\} = P\{M|X_n \in A_n\},$$

contudo é verdade que, se $N \in \mathcal{F}_n$, então

$$P\{M|N, X_n = i\} = P\{M|X_n = i\}.$$

Definição 10.3. *Seja $X = \{X_n, n \geq 0\}$ uma CM. Seja $p_k := P\{X_0 = k\}$. Então, $\{p_k, k \in I\}$ é chamada a distribuição inicial de X. Se $P\{X_0 = k\} = 1$, para algum $k \in I$, dizemos que X começa em k.*

Definição 10.4. *A probabilidade de transição do estado i para o estado j no tempo n é dada por $P\{X_{n+1} = j|X_n = i\}$. A seguir, sempre consideraremos cadeias de Markov homogêneas, isto é, cadeias para as quais essa probabilidade não depende de n e a denotaremos por p_{ij}. A matriz $[p_{ij}]_{(i,j)\in I^2}$, é chamada matriz de transição.*

Uma matriz $[a_{ij}]_{(i,j)\in I^2}$ é uma *matriz estocástica* se $a_{ij} \geq 0$ e $\sum_{j\in I} a_{ij} = 1$. Uma *matriz subestocástica* satisfaz $a_{ij} \geq 0$ e $\sum_{j\in I} a_{ij} \leq 1$. Se $C = AB$, e A e B são estocásticas, C também é.

Suponha que queiramos calcular $P\{X_0 = i_0, \ldots, X_n = i_n\}$. Temos

$$\begin{aligned} P\{X_0 = i_0, \ldots, X_n = i_n\} &= P\{X_n = i_n|X_0 = i_0, \ldots, X_{n-1} = i_{n-1}\} \\ &\quad \times P\{X_0 = i_0, \ldots, X_{n-1} = i_{n-1}\} \\ &= P\{X_n = i_n|X_{n-1} = i_{n-1}\}P\{X_{n-1} = i_{n-1}|X_0 = i_0, \ldots, X_{n-2} = i_{n-2}\} \\ &\quad \times P\{X_0 = i_0, \ldots, X_{n-2} = i_{n-2}\}, \end{aligned}$$

e prosseguindo, obtemos no final

$$P\{X_0 = i_0, \ldots, X_n = i_n\} = p_{i_0} p_{i_0 i_1} p_{i_1 i_2} \cdots p_{i_{n-1} i_n}.$$

Teorema 10.1. *Seja I um conjunto enumerável e $\{\hat{p}_i, i \in I\}$ uma distribuição de probabilidade sobre I. Seja $[\hat{p}_{ij}]_{(i,j)\in I^2}$, uma matriz estocástica. Então, existe uma CM $X = \{X_n, n \geq 0\}$ tendo $\{\hat{p}_i\}$ como distribuição inicial e $[\hat{p}_{ij}]_{(i,j)\in I^2}$ como matriz de transição.*

Prova: Seja $\Omega = I^\infty$, ou seja Ω é o conjunto de todas as sequências $\omega = (i_0, i_1, \ldots)$, com $i_k \in I$. Tome $\mathcal{F}$ como a σ-álgebra produto sobre Ω.

Um cilindro baseado nas coordenadas $0, 1, \ldots, k$ é um conjunto da forma $\{\omega \in \Omega : \omega_0 = i_0, \omega_1 = i_1, \ldots, \omega_k = i_k\}$. Seja $\mathcal{F}_k$ a σ-álgebra em Ω gerada por tais cilindros. Defina P_k sobre $(\Omega, \mathcal{F}_k)$ por: se $A_k = \{\omega \in \Omega : \omega_0 = i_0, \omega_1 = i_1, \ldots, \omega_k = i_k\}$, então

$$P_k(A_k) = \hat{p}_{i_0}\hat{p}_{i_0 i_1} \cdots \hat{p}_{i_{k-1} i_k}.$$

Então, as P_k são consistentes, isto é, P_{k+1} restrita a $(\Omega, \mathcal{F}_k)$ é P_k. Logo, pelo teorema de Kolmogorov, existe uma probabilidade P sobre $(\Omega, \mathcal{F})$ tal que P restrita a $(\Omega, \mathcal{F}_k)$ é P_k.

Assim, nosso espaço de probabilidades básico é composto por: $\Omega = I^\infty$, P a probabilidade gerada por Kolmogorov de acordo com o exposto acima, $\mathcal{F}$ é a σ-álgebra produto.

Defina $X = \{X_n, n \geq 0\}$ como segue: se $\omega = (\omega_0, \omega_1, \ldots)$, então $X_n(\omega) = \omega_n$. Segue que X é Markov, com as distribuições corretas. De fato:

(a) É imediato que X tem a distribuição inicial correta.

(b) X tem as probabilidades de transição corretas:

$$P\{X_{n+1} = j | X_n = i\} = \frac{P\{X_n = i, X_{n+1} = j\}}{P\{X_n = i\}} =$$

$$\frac{\sum_{i_0, \ldots, i_{n-1}} P\{X_{n+1} = j, X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0\}}{\sum_{i_0, \ldots, i_{n-1}} P\{X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0\}} =$$

$$\frac{\sum \hat{p}_{i_0}\hat{p}_{i_0 i_1} \cdots \hat{p}_{i_{n-1} i}\hat{p}_{ij}}{\sum \hat{p}_{i_0}\hat{p}_{i_0 i_1} \cdots \hat{p}_{i_{n-1} i}} = \hat{p}_{ij}.$$

(c) X é Markov:

$$P\{X_{n+1} = j | X_0 = i_0, \ldots, X_{n-1} = i_{n-1}, X_n = i\} =$$

$$\frac{P\{X_0 = i_0, \ldots, X_{n-1} = i_{n-1}, X_n = i, X_{n+1} = j\}}{P\{X_0 = i_0, \ldots, X_n = i\}} = \frac{\hat{p}_{i_0}\hat{p}_{i_0 i_1} \ldots \hat{p}_{i_{n-1} i}\hat{p}_{ij}}{\hat{p}_{i_0}\hat{p}_{i_0 i_1} \ldots \hat{p}_{i_{n-1} i}}$$

$$= \hat{p}_{ij} = P\{X_{n+1} = j | X_n = i\}. \quad \square$$

Definição 10.5. *As probabilidades de transição em n passos do estado i para o estado j são definidas por:*

$$
\begin{aligned}
p_{ij}^{(0)} &= 1, \quad se \;\; i=j, \\
&= 0, \quad se \;\; i \neq j, \\
p_{ij}^{(1)} &= p_{ij}, \\
p_{ij}^{(n+1)} &= \sum_{k \in I} p_{ik}^{(n)} p_{kj}.
\end{aligned} \tag{10.6}
$$

Por exemplo,

$$P\{X_{n+2}=j|X_n=i\} = \sum_{s\in I} P\{X_{n+2}=j, X_{n+1}=s|X_n=i\}$$

$$= \sum_{s\in I} P\{X_{n+2}=j|X_{n+1}=s, X_n=i\} P\{X_{n+1}=s|X_n=i\}$$

$$= \sum_{s\in I} p_{is}p_{sj} = p_{ij}^{(2)},$$

onde usamos a propriedade de Markov na primeira probabilidade da soma, desprezando $X_k = i$. Usando (10.6) repetidamente, encontramos

$$p_{ij}^{(m+n)} = \sum_{k\in I} p_{ik}^{(n)} p_{kj}^{(m)}.$$

Esta é chamada *equação de Chapman-Kolmogorov.*

Sejam $\mathbb{P} := [p_{ij}]$ e $\mathbb{P}^{(n)} := \left[p_{ij}^{(n)}\right]$ para todo $n \geq 0$. Então, de (10.6) obtemos que $\mathbb{P}^{(n)} = \mathbb{P}^n$ para todo n e finalmente,

$$\mathbb{P}^{(n+m)} = \mathbb{P}^{n+m} = \mathbb{P}^n\mathbb{P}^m = \mathbb{P}^{(n)}\mathbb{P}^{(m)}.$$

Exemplo 10.2. Vamos considerar alguns exemplos de CMs.

(a) Uma CM com probabilidades de transição p_{ij} diz-se *espacialmente homogênea* se p_{ij} é uma função de $j-i$ somente. Por exemplo, considere $\{Y_i, i \geq 1\}$, uma sequência de v.a's discretas, i.i.d, e suponha que Y_0 seja independente de $\{Y_k, k \geq 1\}$. Defina $X = Y_0 + Y_1 + \ldots + Y_n$. Então, X tem incrementos independentes e é espacialmente homogênea.

Uma recíproca parcial desse resultado é a seguinte. Seja X uma CM com espaço de estados I e suponha que p_{ij} seja uma função de $j-i$ e que I seja um grupo aditivo. Então, X tem incrementos independentes. Ou seja, se $Y_0 = X_0$, $Y_1 = X_1 - X_0, \ldots, Y_n = X_n - X_{n-1}$, então $Y_0, \ldots, Y_n$ são independentes e $\{Y_k, k \geq 1\}$ são i.i.d.

De fato, observe que $\sum_{k\in I} p_{i,i+k} = \sum_{j\in I} p_{i,j} = 1$. Seja $q_k = p_{i,i+k}$, de modo que $\sum_k q_k = 1$. Então,

$$P\{Y_0 = i_0, Y_1 = i_1, \dots, Y_n = i_n\} =$$

$$P\{X_0 = i_0, X_1 = i_1 + i_0, X_2 = i_0 + i_1 + i_2, \dots, X_n = i_0 + \dots + i_n\} =$$

$$= p_{i_0} \cdot p_{i_0, i_0+i_1} \cdot \dots \cdot p_{i_0+\dots+i_{n-1}, i_0+\dots+i_n} = p_{i_0} \cdot q_{i_1} \cdot q_{i_2} \cdots q_{i_n}.$$

Somando, obtemos que para $j \geq 1$, $P\{Y_j = i_j\} = q_{i_j}$ e finalmente

$$P\{Y_0 = i_0, \dots, Y_n = i_n\} = P\{Y_0 = i_0\} P\{Y_1 = i_1\} \dots P\{Y_n = i_n\}.$$

Como um outro exemplo, se I é um conjunto de inteiros positivos e se p_{ij} depende de $j - i$ somente, então X tem incrementos independentes.

(b) *Passeio aleatório começando em i_0.* Nesse caso, I é o conjunto dos inteiros, a distribuição inicial é dada por $P(X_0 = i_0) = 1$ e a matriz de transição é definida por:

$$\begin{aligned} p_{i,i+1} &= p, \\ p_{i,i-1} &= q = 1 - p, \\ p_{ij} &= 0, \text{ caso contrário.} \end{aligned}$$

Pelo exemplo (a), essa CM tem incrementos independentes. De fato, se $Y_1, Y_2, \dots$ são i.i.d, cada uma com distribuição $P(Y_1 = 1) = p,\ P(Y_1 = -1) = 1 - p$, então $X_n = i_0 + Y_1 + \dots + Y_n$.

(c) Considere, agora, o *passeio aleatório com absorção no zero, começando em i_0.* Aqui, I é o conjunto dos inteiros não negativos, $p_{i,i+1} = p,\ i > 0,\ p_{i,i-1} = 1 - p,\ i > 0$, e $p_{0,0} = 1$.

(d) *Passeio aleatório com reflexão no zero.* Aqui, $p_{i,i+1} = p,\ i > 0,\ p_{i,i-1} = q,\ i > 0,\ p_{0,0} = q,\ p_{0,1} = p$.

10.3 Propriedade forte de Markov

Seja $X = \{X_n, n \geq 0\}$ uma CM sobre $(\Omega, \mathcal{F}, P)$ e $\mathcal{F}_n = \mathcal{F}\{X_0, \dots, X_n\}$. Seja T um tempo de parada para $\{\mathcal{F}_n\}$, possivelmente infinito. Lembremos que $\{T = n\} \in \mathcal{F}_n$ e defina $\mathcal{F}_T$ como a classe dos conjuntos $A \in \mathcal{F}_\Delta$ tais que $A \cap \{T = n\} \in \mathcal{F}_n$, sendo que Δ e $\mathcal{F}_\Delta$ são tais que $\Delta = \{\omega : T(\omega) < \infty\}$ $(\Delta \in \mathcal{F})$ e $\mathcal{F}_\Delta$ é a restrição de $\mathcal{F}$ a Δ, ou seja, $\mathcal{F}_\Delta = \{A \cap \Delta : A \in \mathcal{F}\}$.

Defina P_Δ sobre $(\Omega, \mathcal{F}_\Delta)$ como segue, supondo $P(\Delta) > 0$. Se $A \in \mathcal{F}_\Delta$, então $P_\Delta(A) = \frac{P(A)}{P(\Delta)}$. Então, $(\Delta, \mathcal{F}_\Delta, P_\Delta)$ é um espaço de probabilidade.

Defina $Y = \{Y_n, n \geq 0\}$ por $Y_n(\omega) = X_{T+n}(\omega)$, se $T < \infty$. Ou seja, $Y_n(\omega) = X_{k+n}(\omega)$, se $T(\omega) = k$. Então, Y é um processo estocástico sobre $(\Delta, \mathcal{F}_\Delta, P_\Delta)$, chamado o *processo pós-T*.

Teorema 10.2 (Propriedade forte de Markov) *Seja $X = \{X_n, n \geq 0\}$ uma CM, T um tempo de parada e Y o processo pós-T. Então, Y é uma CM sobre $(\Delta, \mathcal{F}_\Delta, P_\Delta)$ tendo distribuição inicial*

$$P_\Delta\{Y_0 = k\} = \frac{1}{P(\Delta)} \sum_{n \geq 0} P\{X_n = k, T = n\}. \tag{10.7}$$

Além disso, se $A \in \mathcal{F}_T$, teremos

$$P_\Delta\{A, Y_0 = i_0, \ldots, Y_n = i_n\} = P_\Delta\{A, Y_0 = i_0\} P\{X_1 = i_1, \ldots, X_n = i_n | X_0 = i_0\}. \tag{10.8}$$

De modo equivalente,

$$P_\Delta\{Y_1 = i_1, \ldots, Y_n = i_n | A, Y_0 = i_0\} = P\{X_1 = i_1, \ldots, X_n = i_n | X_0 = i_0\}.$$

Em particular, Y tem as mesmas probabilidades de transição que X.

Prova: Primeiramente, se (10.8) vale, então Y é um processo de Markov sobre $(\Delta, \mathcal{F}_\Delta, P_\Delta)$ (é óbvio que Y tem distribuição inicial dada por (10.7)). De fato, se tomarmos $A = \Delta$ em (10.8),

$$P_\Delta\{Y_n = i_n | Y_0 = i_0, \ldots, Y_{n-1} = i_{n-1}\} = \frac{P_\Delta\{Y_0 = i_0, \ldots, Y_n = i_n\}}{P_\Delta\{Y_0 = i_0, \ldots, Y_{n-1} = i_{n-1}\}}$$

$$= \frac{P\{X_1 = i_1, \ldots, X_n = i_n | X_0 = i_0\}}{P\{X_1 = i_1, \ldots, X_{n-1} = i_{n-1} | X_0 = i_0\}}$$

$$= \frac{P\{X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n\}}{P\{X_0 = i_0, X_1 = i_1, \ldots, X_{n-1} = i_{n-1}\}}$$

$$= P\{X_n = i_n | X_{n-1} = i_{n-1}, \ldots, X_0 = i_0\} = P\{X_n = i_n | X_{n-1} = i_{n-1}\}$$

$$= P\{X_1 = i_n | X_0 = i_{n-1}\} = P_\Delta\{Y_1 = i_n | Y_0 = i_{n-1}\}.$$

Resta verificar (10.8). Tome $A \in \mathcal{F}_T$. Temos

$$P_\Delta\{A, Y_0 = i_0, \ldots, Y_n = i_n\} = \sum_{k \geq 0} P_\Delta\{A, Y_0 = i_0, \ldots, Y_n = i_n, T = k\}$$

$$= \frac{1}{P(\Delta)} \sum_{k \geq 0} P\{A, T = k, X_k = i_0, \ldots, X_{k+n} = i_n\}$$

$$= \frac{1}{P(\Delta)} \sum_{k \geq 0} P\{X_{k+1} = i_1, \ldots, X_{k+n} = i_n | X_k = i_0, A, T = k\} P\{A, T = k, X_k = i_0\},$$

sendo que a segunda igualdade decorre da definição de Y_n. Como $A \in \mathcal{F}_T$, temos que $A \cap \{T = k\} \in \mathcal{F}_k$, logo a última soma é igual a

$$\sum_{k \geq 0} P\{X_{k+1} = i_1, \ldots, X_{k+n} = i_n | X_k = i_0\} P\{A, T = k, X_k = i_0\}$$

$$= P\{X_1 = i_1, \ldots, X_n = i_n | X_0 = i_0\} \sum_{k \geq 0} P\{A, T = k, X_k = i_0\}$$

$$= P\{X_1 = i_1, \ldots, X_n = i_n | X_0 = i_0\} P(\Delta) P_\Delta\{A, Y_0 = i_0\}. \quad \square$$

Algumas extensões

[1] Suponha que T seja um tempo de parada tal que $P_\Delta\{Y_0 = i_0\} = 1$. Seja $A \in \mathcal{F}_T$ e $B \in \mathcal{F}\{Y_0, Y_1, Y_2, \ldots\}$. Então, $P_\Delta(A \cap B) = P_\Delta(A) P_\Delta(B)$.
De fato, seja $A \in \mathcal{F}_T$. Então,

$$P_\Delta\{A, Y_0 = i_0, \ldots, Y_n = i_n\} = P_\Delta\{A, Y_0 = i_0\} P\{X_1 = i_1, \ldots, X_n = i_n | X_0 = i_0\},$$

por (10.8). Temos que $P_\Delta\{A, Y_0 = i_0\} = P_\Delta(A)$, usando $P_\Delta(Y_0 = i_0) = 1$, logo a última probabilidade é igual a

$$P_\Delta(A) P_\Delta\{Y_1 = i_1, \ldots, Y_n = i_n | Y_0 = i_0\} = P_\Delta(A) \frac{P_\Delta\{Y_0 = i_0, \ldots, Y_n = i_n\}}{P_\Delta\{Y_0 = i_0\}}$$

$$= P_\Delta(A) P_\Delta\{Y_0 = i_0, \ldots, Y_n = i_n\},$$

pois a distribuição condicional de Y é igual à dsitribuição condicional de X. Segue que $P_\Delta(A \cap B) = P_\Delta(A) P_\Delta(B)$, onde $B = \{Y_0 = i_0, \ldots, Y_n = i_n\}$, portanto essa igualdade vale para todo $B \in \mathcal{F}\{Y_i, i \geq 0\}$.

[2] Suponha que T e Y satisfaçam às hipóteses de [1]. Suponha, ainda, que $P(T < \infty) = 1$. Então, $P_\Delta = P$ (pois $P(\Delta) = 1$). Logo, se A e B são como em [1], $P(A \cap B) = P(A) P(B)$.

Isso significa, claro, que A e B são independentes, isto é, passado e futuro são independentes nesse caso.

[3] Sejam $T_1 \leq T_2 \leq \cdots \leq T_N$ tempos de parada para X. Seja $\Delta_n = \{T_n < \infty\}$. Então, $\Delta_1 \supset \Delta_2 \supset \cdots \supset \Delta_N$.

Suponha que $P_{\Delta_i}\{X_{T_i} = k_i\} = 1$, para cada i e que $T_i < T_{i+1}$ sobre Δ_i. Sejam $A_1, A_2, \ldots, A_{N+1}$ conjuntos tais que:

$$
\begin{aligned}
A_1 &\in \mathcal{F}_{T_1}, \\
A_2 &\in \mathcal{F}\{X_{T_1}, X_{T_1+1}, \ldots\} \cap \mathcal{F}_{T_2}, \\
A_3 &\in \mathcal{F}\{X_{T_2}, X_{T_2+1}, \ldots\} \cap \mathcal{F}_{T_3},
\end{aligned}
$$

e assim por diante, até

$$
A_{N+1} \in \mathcal{F}\{X_{T_N}, X_{T_N+1}, \ldots\}.
$$

Então,

$$
P\Big(\bigcap_{i=1}^{N+1} A_i\Big) = P(A_{N+1}) \prod_{i=1}^{N} P_{\Delta_i}(A_i).
$$

De fato, temos que

$$
P_{\Delta_1}\Big(\bigcap_{i=1}^{N+1} A_i\Big) = P_{\Delta_1}\Big(A_1 \cap \Big(\bigcap_{i=2}^{N+1} A_i\Big)\Big) = P_{\Delta_1}(A_1) P_{\Delta_1}\Big(\bigcap_{i=2}^{N+1} A_i\Big)
$$

$$
= P_{\Delta_1}(A_1) P_{\Delta_2}(A_2) \frac{P(\Delta_3)}{P(\Delta_1)} P_{\Delta_3}(A_3) P_{\Delta_3}\Big(\bigcap_{i=4}^{N+1} A_i\Big) = \ldots
$$

$$
= P_{\Delta_1}(A_1) P_{\Delta_2}(A_2) \cdots P_{\Delta_{N-1}}(A_{N-1}) \frac{P(\Delta_N)}{P(\Delta_1)} P_{\Delta_N}(A_N) P_{\Delta_N}(A_{N+1}).
$$

[4] Suponha que os T_i's são como em [3], mas também que $T_i < \infty$ q.c., para $1 \leq i \leq N$. Se os A_i são como em [3], temos

$$
P\Big(\bigcap_{i=1}^{N+1} A_i\Big) = \prod_{i=1}^{N+1} P(A_i).
$$

Exemplo 10.3. Suponha que $X = \{X_n, n \geq 0\}$ seja um passeio aleatório usual, começando no zero. Seja $T_1 = \inf\{n > 0 : X_n = 0\}$, T_2 o próximo tempo depois de T_1 tal que $X_n = 0$, e assim por diante. $P\{T_1 < \infty\} = 1$ e portanto $P\{T_i < \infty\} = 1$. Também, $X_{T_i} = 0$, para todo i. Sejam: M_1 o máximo de X no intervalo $[0, T_1)$, M_2 o máximo de X em $[T_1, T_2)$ etc. Então, pelo visto acima, $M_1, M_2, \ldots$ são independentes (e identicamente distribuídas).

10.4 Classificação de estados

Definição 10.6. *Seja I o espaço dos estados da CM X e $\mathbb{P} = [p_{ij}]$ a sua matriz de transição. Dizemos que i atinge j ($i \to j$) se $p_{ij}^{(n)} > 0$, para algum $n \geq 0$. Dizemos que i e j comunicam ($i \leftrightarrow j$) se $i \to j$ e $j \to i$. $\mathbb{P}$ (ou a CM X) é dita iredutível se todos os estados de I comumicam entre si.*

Defina a classe C_i como o conjunto de todos os estados j que se comunicam com i; C_i é chamada de classe comunicante.

Um estado i é essencial se ele se comunica com todo estado j que é atingido por ele. Caso contrário, o estado é não essencial.

Exemplo 10.4. (a) No passeio aleatório simples, todo estado é essencial.

(b) No passeio aleatório com o zero sendo estado absorvente, o estado 0 é essencial, todos os demais são não essenciais.

Fatos sobre Classes Comunicantes

[1] Para todo $i \in I$, $i \leftrightarrow i$.

[2] Se $i \to j$ e $j \to k$, então $i \to k$.

Prova: use Chapman-Kolmogorov (veja o Problema 4).

[3] Se $i \leftrightarrow j$, então $C_i = C_j$.

Prova: Suponha $k \in C_i$. Por hipótese, $i \leftrightarrow j$ e $i \to k$ (pois $k \in C_i$). Portanto, $j \to i \to k$, logo $j \to k$, por [2]. Também, $k \to i \to j$, logo $k \to j$, de onde $C_i \subset C_j$. De modo similar, $C_j \subset C_i$.

[4] Suponha que C_i seja uma classe comunicante. Se $j \in C_i$ é essencial, então todos os estados de C_i são essenciais (veja o Problema 5).

Definição 10.7. *Seja j um estado para o qual $\{n \geq 1 : p_{jj}^{(n)} > 0\} \neq \emptyset$. O período de j é o máximo divisor comum de $\{n \geq 1 : p_{jj}^{(n)} > 0\}$. Caso o período de j é igual a 1, o estado j é dito aperiódico.*

[5] Seja C_i uma classe comunicante tal que $|C_i| \geq 2$. Então, todos os estados de C_i têm o mesmo período.

Prova: Em primeiro lugar, observe que como $|C_i| \geq 2$, para todo $j \in C_i$ o conjunto $\{n \geq 1 : p_{jj}^{(n)} > 0\} \neq \emptyset$. Agora, sejam d_i e d_j os períodos de i e j, respectivamente. Seja $n_0 \geq 1$ tal que $p_{ii}^{(n_0)} > 0$. Como $i \leftrightarrow j$, existem m e n tais que $p_{ij}^{(m)} > 0$ e $p_{ji}^{(n)} > 0$. Portanto, $p_{jj}^{(m+n+n_0)} \geq p_{ji}^{(n)} p_{ii}^{(n_0)} p_{ij}^{(m)} > 0$, de modo que d_j divide $n + n_0 + m$. Mas também, $p_{ii}^{(2n_0)} \geq p_{ii}^{(n_0)} p_{ii}^{(n_0)} > 0$, logo pelo mesmo argumento,

$p_{jj}^{(n+2n_0+m)} \geq p_{ji}^{(n)} p_{ii}^{(2n_0)} p_{ij}^{(m)} > 0$, de modo que d_j divide $n + 2n_0 + m$. Segue que d_j divide $n + 2n_0 + m - (n + n_0 + m)$, ou seja, d_j divide n_0. Conclui-se que d_j divide tudo que d_i divide. Similarmente, d_i divide tudo que d_j divide, logo $d_i = d_j$. $\square$

Exemplo 10.5. (a) Passeio aleatório simples: há uma classe comunicante, a saber, a classe de todos os inteiros. O período é dois.

(b) Seja X uma CM com $p_{i,i+2} = 1/2,\ p_{i,i-2} = 1/2$. Aqui, o espaço dos estados é o conjuntos dos inteiros. Então, há duas classes comunicantes: C_0, a classe dos inteiros pares e C_1, a classe dos inteiros ímpares, ambas com período 2.

Definição 10.8. *Seja i um estado com período $d \geq 1$. Para todo $r \geq 0$, seja*

$$C_i^r := \{j \in C_i : p_{ij}^{(nd+r)} > 0,\ \textit{para algum}\ n \geq 0\}.$$

Essas classes são chamadas subclasses movendo-se ciclicamente. Observe que $C_i^{r+d} = C_i^r$ para todo r.

[6] As classes C_i^r, $r \in \{0, 1, \ldots, d-1\}$, formam uma partição de C_i.

Prova: Em primeiro lugar, observe que como C_i é uma classe comunicante, $C_i^0 \cup \cdots \cup C_i^{d-1} = C_i$. Agora, seja $j \in C_i^r \cap C_i^{r'}$ com $0 \leq r < r' < d$. Temos que existem $n \geq 0$ e $m \geq 0$ tais que $p_{ij}^{(nd+r)} > 0$ e $p_{ij}^{(md+r')} > 0$. Por outro lado, como $j \in C_i$ existe $n' \geq 1$ tal que $p_{ji}^{(n')} > 0$. Deduzimos que $p_{ii}^{(nd+n'+r)} > 0$ e $p_{ii}^{(md+n'+r')} > 0$. Logo d divide $md + n' + r' - (nd + n' + r) = (m-n)d + r' - r$ e portanto d divide $r' - r$ o que é impossível. Portanto, $C_i^r \cap C_i^{r'} = \emptyset$. $\square$

[7] $p_{jk}^{(n)} > 0$ somente se $j \in C_i^r$ e $k \in C_i^{r+n}$ para algum r.

Prova: Suponha que $p_{jk}^{(n)} > 0$ com $j \in C_i^r$ para algum $r \in \{0, \ldots, d-1\}$. Por definição de C_i^r existe m tal que $p_{ij}^{(md+r)} > 0$. Deduzimos que $p_{ik}^{(md+r+n)} > 0$, logo $k \in C_i^{r+n}$. $\square$

[8] Se $j, k \in C_i^r$, então $p_{jk}^{(nd)} > 0$ para todo n suficientemente grande.

Prova: É uma consequência do seguinte resultado clássico da arithmética: Seja $S \subset \mathbb{N}$ estável por adição com maior divisor comum M. Então para todo n grande o suficiente $nM \in S$. Aplicamos este resultado ao conjunto $\{n \geq 1 : p_{ii}^{(n)} > 0\}$. Agora, escolhe m_1 e m_2 tais que $p_{ji}^{(m_1)} > 0$ e $p_{ik}^{(m_2)} > 0$. Obtemos

$$p_{jk}^{(m_1+nd+m_2)} \geq p_{ji}^{(m_1)} p_{ii}^{(nd)} p_{ik}^{(m_2)} > 0$$

para n grande o suficiente. Como, por [7], $m_1 + m_2$ é um multiplo de d, provamos o resultado. $\square$

Exemplo 10.6. Considere o passeio aleatório ordinário: C_0^0 contém todos os inteiros pares, C_0^1 todos os inteiros ímpares, C_0^2 todos os inteiros pares etc. Depois C_1^0 contém todos os inteiros ímpares, C_1^1 todos os inteiros pares, etc.

10.5 Recorrência

Chamemos de $f_{ij}^{(n)}$ a probabilidade de alcançar o estado j pela primeira vez em $n \geq 1$ passos, dado que o processo começou no estado i, ou seja

$$f_{ij}^{(n)} = P\{X_1 \neq j, X_2 \neq j, \ldots, X_{n-1} \neq j, X_n = j \mid X_0 = i\}.$$

Seja, também, $f_{ij}^* = \sum_{n \geq 1} f_{ij}^{(n)}$ a probabilidade de que X_n atinja j, começando em i. Denote por $T_j = \inf\{n > 0 : X_n = j\}$. Então,

$$f_{ij}^{(n)} = P\{T_j = n | X_0 = i\}, \quad f_{ij}^* = P\{T_j < \infty | X_0 = i\}.$$

Denotemos por U_{ij} o número esperado de visitas ao estado j, começando em i (se $i = j$, contamos o estado inicial).

Lembrando que $p_{ij}^{(0)} = 1$, se $j = i$ e $p_{ij}^{(0)} = 0$, se $j \neq i$, então $U_{ij} = \sum_{n \geq 0} p_{ij}^{(n)}$. De fato, se chamarmos $u_j = \sum_{n \geq 0} I_{\{j\}}(X_n)$ o número de vezes que $X_n = j$, teremos

$$U_{ij} = E(u_j | X_0 = i) = \sum_{n \geq 0} E(I_{\{j\}}(X_n) | X_0 = i)$$

$$= \sum_{n \geq 0} P\{X_n = j | X_0 = i\} = \sum_{n \geq 0} p_{ij}^{(n)}.$$

Definição 10.9. *Um estado i é chamado recorrente se $P\{X_n = i,\ i.v\, | X_0 = i\} = 1$ e é chamado transitório se $P\{X_n = i,\ i.v\, | X_0 = i\} = 0$.*

Mostraremos que essas são as duas únicas possibilidades.

Teorema 10.3. *(a) Suponha $P\{T_i < \infty | X_0 = i\} = 1$. Então, o estado i é recorrente. (b) Nesse caso, $\sum_{n \geq 0} p_{ii}^{(n)} = \infty$.*

Prova: (a) Sejam:

$$T_i^{(1)} = \inf\{n > 0 : X_n = i\},$$

$$T_i^{(2)} = \inf\{n > T_i^{(1)} : X_n = i\},$$

e assim por diante. Então, pela propriedade forte de Markov, os T_i são i.i.d, finitos. Logo,

$$P\{X_n = i, \text{ i.v } |X_0 = i\} = P\{T_i^{(1)} < \infty, T_i^{(2)} < \infty, \ldots |X_0 = i\} =$$

$$\lim_{n\to\infty} P\{T^{(1)} < \infty, \ldots, T_i^{(n)} < \infty\} = \lim_{n\to\infty} P^n\{T_i < \infty | X_0 = i\} = 1,$$

usando a hipótese (a).

(b) Temos que a soma $\sum p_{ii}^{(n)}$ nos dá o número esperado de vezes que visitamos o estado i, começando em i, logo $\sum p_{ii}^{(n)} = E\{\sum I_{\{i\}}(X_n)|X_0 = i\} = \infty$, dado que a soma afetada pelo valor esperado é $+\infty$ q.c. □

Teorema 10.4. *(a) Suponha que $P\{T_i < \infty | X_0 = i\} < 1$. Então, o estado i é transitório.*
(b) Nesse caso, $\sum_{n\geq 0} p_{ii}^{(n)} = [1 - f_{ii}^*]^{-1} < \infty$.

Prova: (a) Defina $T_i^{(1)}, T_i^{(2)}, \ldots$ como na prova do teorema anterior. Note que $T_i^{(1)} = +\infty$ é possível. Seja $\Delta_n = \{T_1 < \infty, \ldots, T_{n-1} < \infty\}$, $n \geq 2$. Temos que

$$\begin{aligned}
P\{X_n = i, \text{ i.v } |X_0 = i\} &= P\{T_i^{(1)} < \infty, T_i^{(2)} < \infty, \ldots |X_0 = i\} \\
&= \lim_{n\to\infty} P\{T_i^{(1)} < \infty, \ldots, T_i^{(n)} < \infty | X_0 = i\} \\
&= \lim_{n\to\infty} P_{\Delta_{n-1}}(T_i^{(n)} < \infty) \cdots P_{\Delta_2}(T_i^{(2)} < \infty) P(T_i^{(1)} < \infty) \\
&= \lim_{n\to\infty} (P\{T_i^{(1)} < \infty | X_0 = i\})^n = 0,
\end{aligned}$$

usando a propriedade forte de Markov.

(b) Temos

$$\begin{aligned}
\sum_{n\geq 0} p_{ii}^{(n)} &= E_i(\text{número de visitas ao estado } i) \\
&= \sum kP\{\text{houve exatamente } k \text{ visitas } |X_0 = i\} \\
&= \sum kP\{T_i^{(j)} < \infty, \ j \leq k-1, \ T_i^{(k)} = +\infty | X_0 = i\} \\
&= \sum kP\{T_i^{(j)} = \infty | X_0 = i\}(P\{T_i^{(1)} < \infty | X_0 = i\})^{k-1} \\
&= [1 - f_{ii}^*] \sum_{k\geq 1} k (f_{ii}^*)^{k-1} \\
&= 1/(1 - f_{ii}^*).
\end{aligned}$$

□

Note que i é recorrente se, e somente se, $\sum p_{ii}^{(n)} = \infty$, que por sua vez é equivalente a $P\{T_i < \infty | X_0 = i\} = 1$. Também, i é transitório se, e somente se $\sum_{n\geq 0} p_{ii}^{(n)} < \infty$, que é equivalente a $P\{T_i < \infty | X_0 = i\} < 1$.

Teorema 10.5. *Suponha que C seja uma classe comunicante e que $j \in C$ seja recorrente. Então, todos os estados de C são recorrentes.*

Prova: Suponha que j seja recorrente, então j é essencial. Segue que todo estado em C é essencial. Portanto, se i é algum estado de C, então $p_{ij}^{(n)} > 0$, $p_{ji}^{(m)} > 0$, $p_{jj}^{(k)} > 0$, de onde $p_{ii}^{(n+k+m)} \geq p_{ij}^{(n)} p_{jj}^{(k)} p_{ji}^{(m)} > 0$. Considere

$$\sum_{n\geq 0} p_{ii}^{(n)} \geq \sum_{k\geq 0} p_{ii}^{(n+k+m)} = p_{ij}^{(n)} p_{ji}^{(m)} \sum_{k\geq 0} p_{jj}^{(k)} = +\infty,$$

pois j é recorrente. Logo i é recorrente. $\square$

Teorema 10.6. *Suponha que X seja uma CM, com espaço de estados I finito. Então:*
(a) o estado i é recorrente se, e somente se, i for essencial;
(b) existe pelo menos um estado essencial.

Note que é sempre verdade que, se i for recorrente, então i é essencial. A parte (b) do teorema não vale se I for infinito. Veja o Problema 6.

Prova: (a) Pela observação anterior, é suficiente provar a parte ($\Leftarrow$). Suponha que i seja essencial. Como I é finito, existe um estado j tal que $P\{X_n = j, \text{ i.v } | X_0 = i\} > 0$. Seja $T_j = \inf\{n > 0 : X_n = j\}$. Então, $0 < P\{X_n = j, \text{ i.v } | X_0 = i\} = P\{T_j < \infty, X_{T_j} = j, X_{T_j+n} = j, \text{ i.v } | X_0 = i\} = P\{X_{T_j+n} = j, \text{ i.v } | X_{T_j} = j, T_j < \infty, X_0 = i\} P\{T_j < \infty | X_0 = i\} = P\{X_n = j, \text{ i.v } | X_0 = j\} P\{T_j < \infty | X_0 = i\}$, pela propriedade forte de Markov. Segue-se que $P\{X_n = j, \text{ i.v } | X_0 = j\} > 0$, logo $P\{X_n = j, \text{ i.v } | X_0 = j\} = 1$. Portanto, j é recorrente. Como i se comunica com j deduzimos que i é recorrente.

(b) Imediata. $\square$

Seja C uma classe comunicante. Então, nós mostramos que ou todos os estados são recorrentes, ou todos são transitórios. Se todos os estados forem recorrentes, dizemos que C é uma *classe recorrente* senão C é uma *classe transitória*.

10.6 Recorrência positiva

Começamos com os conceitos de estado recorrente positivo e de classe recorrente positiva. Nesta seção usamos a notação $P_i\{\cdot\} = P\{\ \cdot \mid X_0 = i\}$ para $i \in I$ e E_i é a esperança correspondente.

Definição 10.10. *Um estado recorrente j é recorrente positivo se $E_j(T_j) < \infty$ e é recorrente nulo se $E_j(T_j) = \infty$. Uma classe recorrente é chamada recorrente positiva (resp. nula) se todos os seus estados são recorrentes positivos (resp. nulos).*

Teorema 10.7. *Seja i recorrente positivo e $j \leftrightarrow i$, então $E_i(T_j) < \infty$.*

Prova: Sejam $T_i^{(1)} = \inf\{n > 0 : X_n = i\}$, $T_i^{(2)} = \inf\{n > T_i^{(1)} : X_n = i\}$ e $T_i^{(n)}$ definido similarmente para todo $n \geq 3$. Defina $U_1, U_2, \ldots$ como segue: $U_k = 1$ se X_n está no estado j para algum n tal que $T_i^{(1)} + \ldots + T_i^{(k)} < n \leq T_i^{(1)} + \ldots + T_i^{(k+1)}$ e $U_k = 0$ caso contrário. Pela propriedade forte de Markov, $U_1, U_2, \ldots$ são i.i.d. Seja $S = \inf\{n : U_n = 1\}$; temos que $E_i(S) < \infty$, pois

$$P_i\{S = k\} = P_i\{U_1 = 0, \ldots, U_{k-1} = 0, U_k = 1\} = P_i\{U_1 = 0\}^{k-1} P_i\{U_1 = 1\}$$

e $P_i\{U_1 = 1\} > 0$. Como $T_j \leq T_i^{(1)} + T_i^{(2)} + \ldots + T_i^{(S)}$, temos que $E_i(T_j) \leq E_i(T_i^{(1)} + \ldots + T_i^{(S)}) = E_i(T_i)E_i(S) < \infty$, pela identidade de Wald. $\square$

Teorema 10.8. *Seja C uma classe recorrente. Então ou todos os seus estados são recorrentes positivos ou todos são recorrentes nulos.*

Prova: Suponha que i seja recorrente positivo e $j \leftrightarrow i$ com $i \neq j$. Seja $p := P_i\{T_j < T_i\}$. Como i e j comunicam temos que $p > 0$. Logo, pela propriedade forte de Markov notamos que

$$E_i(T_i) \geq E_i(T_i I_{\{T_j < T_i\}}) = pE_j(T_i)$$

e portanto deduzimos que $E_j(T_i) < \infty$. Agora observe que pela propriedade forte de Markov

$$E_j(T_j) \leq E_j(T_i) + E_i(T_j).$$

Usando o Teorema 10.7, deduzimos que $E_j(T_j) < \infty$. $\square$

Exemplo 10.7. (a) Os estados de um passeio aleatório padrão, com $p_{i,i+1} = p_{i,i-1} = 1/2$ são todos recorrentes nulos.

(b) Seja X uma CM com conjunto de estados I finito. Se o estado j for recorrente, então j é recorrente positivo. De fato, seja C a classe recorrente contendo j. Temos que $\sum_{i \in C} p_{ji}^{(n)} = 1$ para todo $n \geq 0$. Supomos que C tem período d. Para $n \to \infty$, pelo Corolário 10.2 a seguir, como essa soma é uma soma finita, existe um estado i e $0 \leq r < d$, tais que $\lim p_{ji}^{(nd+r)} = d/E_i(T_i) > 0$, logo $E_i(T_i) < \infty$, ou seja, i é recorrente positivo. Mas como i e j estão na mesma classe, j é recorrente positivo.

10.7 Medidas estacionárias

Lembremos que um processo $X = \{X_t, t \in T\}$ é estritamente estacionário se, para todo h e todo conjunto $t_1 < t_2 < \cdots < t_n$, temos que $(X_{t_1}, X_{t_2}, \ldots, X_{t_n})$ tem a mesma distribuição que $(X_{t_1+h}, X_{t_2+h}, \ldots, X_{t_n+h})$ (sempre que $t_i \in T$, para todo i e $t_i + h \in T$, para todo i).

Como consequência temos que, se X for uma CM homogênea, com distribuição inicial α e distribuição de X_n $\{p_k^{(n)}, k \in I\}$, então X é estritamente estacionária se, e somente se, $\alpha_k = p_k^{(n)}$, para todo k e todo n.

Definição 10.11. *Seja $\mathbb{P} = [p_{ij}]$ a matriz de transição para a CM $X = \{X_n, n \geq 0\}$, com espaço de estados I. Uma medida estacionária μ sobre I é uma medida não trivial satisfazendo*

$$\mu_j = \sum_{i \in I} \mu_i p_{ij}, \quad \text{para todo } j \in I.$$

Em termos de matriz $\mathbb{P}$ as equações acima pode ser reescritas $\mu\mathbb{P} = \mu$. Outros nomes são *medida regular* ou *medida invariante*. Por medida não trivial entendemos uma medida diferente de $\mu \equiv 0$ e $\mu \equiv \infty$.

Teorema 10.9. *(a) Suponha que μ seja uma medida de probabilidade estacionária sobre I. Se X for uma CM com distribuição inicial μ, então X é estritamente estacionária.*

(b) Suponha que X seja uma CM estritamente estacionária. Então, a distribuição inicial de X é uma medida de probabilidade estacionária.

Prova: (a) $\mu_j = \sum_i \mu_i p_{ij}$, pois μ é estacionária. Se X é qualquer CM, $p_k^{(n)} = \sum_i \alpha_i p_{ik}^{(n)} = \sum_i \mu_i p_{ik}^{(n)}$, pois $\alpha_i = \mu_i$ por hipótese. Agora,

$$\mu_j = \sum_i \mu_i p_{ij} = \sum_i (\sum_k \mu_k p_{ki}) p_{ij} = \sum_k \mu_k \sum_i p_{ki} p_{ij}$$

$$= \sum_k \mu_k p_{kj}^{(2)} = \cdots = \sum_k \mu_k p_{kj}^{(n)},$$

logo $p_j^{(n)} = \mu_j = \alpha_j$, ou seja $X_n \sim X_0$.

(b) Para toda CM temos $p_k^{(1)} = \sum_j \alpha_j p_{jk}$. Como X é estritamente estacionária, $\alpha_k = p_k^{(1)}$, para todo k. Logo $\alpha_k = \sum_j \alpha_j p_{jk}$, de modo que α é uma medida estacionária. $\square$

Teorema 10.10. *Seja X uma CM irredutível então os itens a seguir são equivalentes:*
(a) X é recorrente positiva;
(b) $\mathbb{P}$ tem uma probabilidade invariante π.

Em ambos os casos, a probabilidade invariante π é única e

$$\pi_i = \frac{1}{E_i(T_i)}, \quad \textit{para todo } i \in I.$$

Prova: Provaremos primeiro que $(b) \Rightarrow (a)$. Para isto, vamos começar descartando o caso transitório. Suponha que X seja transitória. Neste caso pelo Teorema 10.4, temos que para todo $i \in I$, $\lim_{n\to\infty} p_{ii}^{(n)} = 0$. Por outro lado, como π é invariante temos

$$\sum_{i\in I} \pi_i p_{ii}^{(n)} = \pi_i, \quad \text{para todo } n.$$

Pelo TCD obtemos que $\pi_i = 0$, para todo $i \in I$. Como $\sum_{i\in I} \pi_i = 1$, temos uma contradição. Portanto, X não pode ser transitória.

Vamos considerar agora que X é recorrente. Fixe um elemento $i \in I$ e considere a medida

$$\nu_i(j) = E_i\left[\sum_{n=0}^{T_i-1} 1_{\{X_n=j\}}\right], \quad \forall j \in I.$$

Observe que $\nu_i(i) = 1$ e pela irredutibilidade de X, $\nu_i(j) > 0$ para todo $j \in I$. Podemos ver também que ν_i é invariante para $\mathbb{P}$, de fato temos

$$\begin{aligned}
\nu_i(j) &= E_i\left[\sum_{n=1}^{T_i} 1_{\{X_n=j\}}\right] \\
&= \sum_{k\in I} E_i\left[\sum_{n=1}^{T_i} 1_{\{X_{n-1}=k, X_n=j\}}\right] \\
&= \sum_{k\in I}\sum_{n=1}^{\infty} E_i\left[1_{\{n\le T_i, X_{n-1}=k\}} 1_{\{X_n=j\}}\right] \\
&= \sum_{k\in I}\sum_{n=1}^{\infty} E_i\left[1_{\{n\le T_i, X_{n-1}=k\}}\right] p_{kj} \\
&= \sum_{k\in I} E_i\left[\sum_{n=1}^{T_i} 1_{\{X_{n-1}=k\}}\right] p_{kj} \\
&= \sum_{k\in I} \nu_i(k) p_{kj}.
\end{aligned}$$

Agora suponha que μ é uma medida invariante para $\mathbb{P}$ (lembramos que por definição μ é não trivial). Mostraremos a seguir que μ é proporcional à medida ν_i. Podemos provar por indução que para todo $p \ge 0$,

$$\mu(j) \ge \mu(i) E_i\left[\sum_{n=0}^{p\wedge(T_i-1)} 1_{\{X_n=j\}}\right]. \tag{10.9}$$

Basta considerar o caso $i \neq j$. Para $p = 0$, a desigualdade (10.9) é trivial. Agora vamos supor que (10.9) vale para p e vamos mostrar que ainda vale para $p+1$. De fato, temos que

$$\begin{aligned}
\mu(j) &= \sum_{k \in I} \mu(k) p_{kj} \\
&\geq \mu(i) \sum_{k \in I} E_i \left[\sum_{n=0}^{p \wedge (T_i - 1)} 1_{\{X_n = k\}} \right] p_{kj} \\
&= \mu(i) \sum_{k \in I} \sum_{n=0}^{p} E_i \left[1_{\{X_n = k,\, n \leq T_i - 1\}} \right] p_{kj} \\
&= \mu(i) \sum_{k \in I} \sum_{n=0}^{p} E_i \left[1_{\{X_n = k,\, n \leq T_i - 1\}} 1_{\{X_{n+1} = j\}} \right] \\
&= \mu(i) E_i \left[\sum_{n=0}^{p \wedge (T_i - 1)} 1_{\{X_{n+1} = j\}} \right] \\
&= \mu(i) E_i \left[\sum_{n=0}^{(p+1) \wedge (T_i - 1)} 1_{\{X_n = j\}} \right].
\end{aligned}$$

Tomando $p \to +\infty$ em (10.9) obtemos

$$\mu(j) \geq \mu(i) E_i \left[\sum_{n=0}^{T_i - 1} 1_{\{X_n = j\}} \right] = \mu(i) \nu_i(j).$$

A medida ν_i é invariante e temos que $\mu(j) \geq \mu(i)\nu_i(j)$ para todo $j \in I$. Assim para todo $n \geq 1$,

$$\mu(i) = \sum_{k \in I} \mu(k) p_{ki}^{(n)} \geq \sum_{k \in I} \mu(i) \nu_i(k) p_{ki}^{(n)} = \mu(i)\nu_i(i) = \mu(i).$$

Deduzimos que a última desigualdade é na verdade uma igualdade, o que significa que $\mu(k) = \mu(i)\nu_i(k)$ para todo k tal que $p_{ki}^{(n)} > 0$. A irredutibilidade de X garante que, para todo $k \in I$, podemos encontrar um inteiro n tal que $p_{ki}^{(n)} > 0$. Concluimos que $\mu = \mu(i)\nu_i$, isto é, μ é proporcional a ν_i (como μ é não trivial, necessariamente $\mu(i) \in (0, \infty)$).

Finalmente, aplicando o resultado acima a π, obtemos que $1 = \pi(I) = \pi(i)\nu_i(I) = \pi(i)E_i(T_i)$. Deduzimos que $\pi(i) = 1/E_i(T_i)$ e portanto $E_j(T_j) < \infty$ para algum j. Assim, por irredutibilidade, X é recorrente positiva.

Falta provar que $(a) \Rightarrow (b)$. Supomos agora que X seja recorrente positiva e definimos a medida μ tal que

$$\mu(j) = \frac{1}{E_i(T_i)} \nu_i(j) \ \text{ para todo } j \in I.$$

Neste caso, μ é uma probabilidade invariante para $\mathbb{P}$. $\square$

10.8 Limite de $\mathbb{P}^n$

Nesta seção, estudaremos o comportamento limite de $p_{ij}^{(n)}$ quando $n \to \infty$. Como o seguinte exemplo mostra, o limite nem sempre existe. Considere a cadeia de Markov com dois estados e matriz de transição

$$\mathbb{P} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Então $\mathbb{P}^2 = I$, portanto $\mathbb{P}^{2n} = I$ e $\mathbb{P}^{2n+1} = \mathbb{P}$ para todo n. Logo, $p_{ij}^{(n)}$ diverge para todo i, j.

Como veremos, o comportamento da cadeia acima é atrelado à noção de periodicidade. A seguir apresentamos o principal resultado desta seção.

Teorema 10.11. *Seja $\mathbb{P} = [p_{ij}]$ irredutível e aperiódica com probabilidade estacionária π. Suponha que $\{X_n, n \geq 0\}$ seja uma cadeia de Markov com lei inicial arbitrária μ e matriz de transição $\mathbb{P}$. Então,*

$$\lim_{n\to\infty} P(X_n = j) \to \pi_j, \quad \text{para todo } j.$$

Em particular,

$$\lim_{n\to\infty} p_{ij}^{(n)} \to \pi_j, \quad \text{para todo } i, j.$$

Prova. Seja $\{Y_n, n \geq 0\}$ uma CM com lei inicial π, matriz de transição $\mathbb{P}$ e independente de $\{X_n, n \geq 0\}$. Fixe um estado de referência $a \in I$ e defina

$$T = \inf\{n \geq 1 : X_n = Y_n = a\}.$$

1) Mostramos que $P\{T < \infty\} = 1$. O processo $W_n = (X_n, Y_n)$ é uma CM em $I \times I$ com probabilidades de transição

$$\widetilde{p}_{(i,k)(j,l)} = p_{ij} p_{kl}$$

e distribuição inicial

$$\tilde{\mu}_{(i,k)} = \mu_i \pi_k.$$

Como $\mathbb{P}$ é aperiódica, para todos os estados i, j, k, l temos que

$$\widetilde{p}^{(n)}_{(i,k)(j,l)} = p_{ij}^{(n)} p_{kl}^{(n)} > 0$$

para todo n suficientemente grande. Deduzimos que $\widetilde{\mathbb{P}} := [\tilde{p}_{(i,k)(j,l)}]$ é irredutível. Além disso, $\widetilde{\mathbb{P}}$ tem uma distribuição invariante dada por

$$\widetilde{\pi}_{(i,k)} = \pi_i \pi_k,$$

assim, pelo Teorema 10.10, $\widetilde{\mathbb{P}}$ é recorrente positiva. Mas T é o primeiro tempo de passagem de W_n para (a,a) então $P\{T < \infty\} = 1$.

2) Defina

$$Z_n = \begin{cases} X_n, & \text{se } n < T, \\ Y_n, & \text{se } n \geq T. \end{cases}$$

Mostramos a seguir que $\{Z_n, n \geq 0\}$ é uma CM com lei inicial μ e matriz de transição $\mathbb{P}$. A propriedade forte de Markov aplica-se a $\{W_n, n \geq 0\}$ no tempo T, então $(X_{T+n}, Y_{T+n})_{n\geq 0}$ é ume CM com lei inicial $\delta(a,a)$, matriz de transição $\widetilde{\mathbb{P}}$ e independente de

$$(X_0, Y_0), (X_1, Y_1), \ldots, (X_T, Y_T).$$

Por simetria, podemos substituir o processo $(X_{T+n}, Y_{T+n})_{n\geq 0}$ por $(Y_{T+n}, X_{T+n})_{n\geq 0}$ que também é uma CM com lei inicial $\delta(a,a)$, matriz de transição $\widetilde{\mathbb{P}}$ e permanece independente de $(X_0, Y_0), \ldots, (X_T, Y_T)$. Agora seja

$$Z'_n = \begin{cases} Y_n, & \text{se } n < T, \\ X_n, & \text{se } n \geq T. \end{cases}$$

Obtemos que $W'_n = (Z_n, Z'_n)$ é uma CM com mesma lei que W_n. Assim, deduzimos que $\{Z_n, n \geq 0\}$ é uma CM com lei inicial μ e matriz de transição $\mathbb{P}$.

3) Temos

$$P\{Z_n = j\} = P\{X_n = j, n < T\} + P\{Y_n = j, n \geq T\},$$

assim

$$\begin{aligned} |P\{X_n = j\} - \pi_j| &= |P\{Z_n = j\} - P\{Y_n = j\}| \\ &= |P\{X_n = j, n < T\} - P\{Y_n = j, n < T\}| \\ &\leq P\{T > n\} \end{aligned}$$

e $P\{T > n\} \to 0$ quando $n \to \infty$. □

Vamos ver agora o que dá errado na prova acima quando $\mathbb{P}$ não é aperiódica. Considere a cadeia de dois estados do início desta seção que tem $(1/2, 1/2)$ como sua única distribuição invariante. Começamos $\{X_n, n \geq 0\}$ de 0 e $\{Y_n, n \geq 0\}$ com igual probabilidades de 0 ou 1. Se $Y_0 = 1$, devido à periodicidade, $\{X_n\}$ e $\{Y_n\}$ nunca se encontrarão, e a prova falha. Consideramos agora os casos que foram excluídos no último teorema.

Teorema 10.12. *Seja $\mathbb{P}$ irredutível de período $d \geq 1$ e sejam $C^0, C^1, \ldots, C^{d-1}$ as classes cíclicas. Seja $\{X_n, n \geq 0\}$ uma CM de com lei inicial μ e matriz de transição $\mathbb{P}$. Suponha que $\sum_{i \in C^0} \mu_i = 1$. Então para $r \in \{0, 1, \ldots, d-1\}$ e $j \in C^r$ temos*

$$\lim_{n\to\infty} P\{X_{nd+r} = j\} = \frac{d}{E_j(T_j)}.$$

Em particular, para $i \in C^0$ e $j \in C^r$, temos

$$\lim_{n\to\infty} p_{ij}^{(nd+r)} = \frac{d}{E_j(T_j)}.$$

Prova. A prova será feita em três passos.

1) Reduzimos o problema ao caso aperiódico. Seja $\nu = \mu\mathbb{P}^r$, então temos

$$\sum_{i\in C^r} \nu_i = 1$$

pelo fato [7] da Seção 10.4. Seja $Y_n = X_{nd+r}$, então $\{Y_n, n \geq 0\}$ é uma CM com lei inicial ν e matriz de transição $\mathbb{P}^d$. Pelo fato [8] da Seção 10.4, $\mathbb{P}^d$ é irredutível e aperiódica em C^r. Para $j \in C^r$ o tempo de retorno esperado de $\{Y_n\}$ a j é $E_j(T_j)/d$. Portanto se o teorema vale no caso aperiódico, então

$$P\{X_{nd+r} = j\} = P\{Y_n = j\} \to \frac{d}{E_j(T_j)} \quad \text{quando } n \to \infty,$$

e logo o teorema vale no caso periódico.

2) Assume que $\mathbb{P}$ é aperiódica. Se $\mathbb{P}$ é recorrente positiva então $1/E_j(T_j) = \pi_j$, onde π é a unica probabilidade invariante, portanto o resultado segue do Teorema 10.11. Senão $E_j(T_j) = \infty$ e temos que mostrar que

$$P\{X_n = j\} \to 0 \quad \text{quando } n \to \infty.$$

Se $\mathbb{P}$ é transitória o enunciado acima é imediato. Desta forma, só temos que considerar o caso recorrente nulo.

3) Suponha que $\mathbb{P}$ seja aperiódica e recorrente nula. Então

$$\sum_{k=0}^{\infty} P_j\left\{T_j > k\right\} = E_j\left(T_j\right) = \infty.$$

Dado $\varepsilon > 0$ escolhe K tal que

$$\sum_{k=0}^{K-1} P_j\left\{T_j > k\right\} > \frac{2}{\varepsilon}.$$

Então para $n \geq K-1$, temos

$$\begin{aligned}
1 &\geq \sum_{k=n-K+1}^{n} P\left\{X_k = j \text{ e } X_m \neq j \text{ para } m = k+1, \ldots, n\right\} \\
&= \sum_{k=n-K+1}^{n} P\left\{X_k = j\right\} P_j\left\{T_j > n-k\right\} \\
&= \sum_{k=0}^{K-1} P\left\{X_{n-k} = j\right\} P_j\left\{T_j > k\right\}
\end{aligned}$$

portanto devemos ter $P\{X_{n-k} = j\} \le \varepsilon/2$ para algum $k \in \{0, 1, \ldots, K-1\}$.

Agora, seja $\{Y_n, n \ge 0\}$ uma CM com lei inicial λ e matriz de transição $\mathbb{P}$, onde λ será escolhido depois. Seja $W_n = (X_n, Y_n)$. Usando um argumento similar a 1) na prova do Teorema 10.11, a aperiodicidade de $\{X_n\}$ garante a irredutibilidade de $\{W_n, n \ge 0\}$. Se $\{W_n\}$ é transitória, tomando $\lambda = \mu$, obtemos

$$P\{X_n = j\}^2 = P\{W_n = (j, j)\} \to 0,$$

quando $n \to \infty$. Suponha agora que $\{W_n\}$ seja recorrente. Então, usando a notação da prova do Teorema 10.11, temos $P\{T < \infty\} = 1$ e um argumento similar ao argumento 3) da prova mostra que

$$|P\{X_n = j\} - P\{Y_n = j\}| \to 0, \quad \text{quando } n \to \infty.$$

Tome $\lambda = \mu P^k$ para $k \in \{1, \ldots, K-1\}$, de tal maneira que $P\{Y_n = j\} = P\{X_{n+k} = j\}$. Podemos encontrar N tal que para $n \ge N$ e $k \in \{1, \ldots, K-1\}$,

$$|P\{X_n = j\} - P\{X_{n+k} = j\}| \le \frac{\varepsilon}{2}.$$

Mas para todo n, podemos encontrar $k \in \{0, 1, \ldots, K-1\}$ tal que $P\{X_{n+k} = j\} \le \varepsilon/2$. Logo, para $n \ge N$

$$P\{X_n = j\} \le \varepsilon.$$

Como $\varepsilon > 0$ é arbitrário, mostramos que $P\{X_n = j\} \to 0$ quando $n \to \infty$. □

Corolário 10.2. *Seja X uma CM e $j \in I$.*
Se j é transitório ou recorrente nulo, então para todo $i \in I$

$$\lim_{n\to\infty} p_{ij}^{(n)} = 0.$$

Se j é recorrente positivo com período $d \ge 1$, então para todo $i \in I$ e todo $r \in \{0, \ldots, d-1\}$,

$$\lim_{n\to\infty} p_{ij}^{(nd+r)} = f_{ij}^*(r) \frac{d}{E_j(T_j)},$$

onde $f_{ij}^*(r) := \sum_{n=1}^{\infty} f_{ij}^{(nd+r)}$.

Prova. Considere o caso j transitório. Se $i = j$ o resultado segue do Teorema 10.4, parte (b). Se $i \ne j$, pela propriedade forte de Markov temos que

$$U_{ij} \le P_i\{T_j < \infty\} U_{jj} < \infty.$$

Mas $U_{ij} = \sum_{n \ge 0} p_{ij}^{(n)}$, logo $p_{ij}^{(n)} \to 0$, quando $n \to \infty$.

Agora consideramos o caso j recorrente. Observe que o caso $i = j$ é uma consequência do Teorema 10.12. O caso $i \ne j$ se trata com a relação elementar

$$p_{ij}^{(n)} = \sum_{m=1}^{n} f_{ij}^{(m)} p_{jj}^{(n-m)}$$

usando o TCD. □

10.9 σ-álgebras caudais

Seja $X = \{X_n, n \geq 0\}$ um processo estocástico, com espaço de estados I e seja I^∞, munido com a σ-álgebra produto.

Definição 10.12. *A σ-álgebra caudal de X é a σ-álgebra $\mathcal{F}_\infty$ definida por $\mathcal{F}_\infty = \cap_{n\geq 0}\mathcal{F}_n^{'}$, onde $\mathcal{F}_n^{'} = \mathcal{F}\{X_n, X_{n+1}, \ldots\}$.*

Defina $\xi = \{\xi_n, n \geq 0\}$ por meio de $\xi_n(\omega) = \omega_n$, com $\omega = (\omega_0, \omega_1, \ldots) \in I^\infty$. A σ-álgebra caudal de ξ é a σ-álgebra sobre I^∞ gerada da mesma maneira, e denotada por $\mathcal{H}$.

A *σ-álgebra invariante* $\mathcal{I}$ é a σ-álgebra formada pelos conjuntos mensuráveis $B \in I^\infty$ tal que se $\omega = (\omega_0, \omega_1, \ldots)$, então $\omega \in B$ se e somente se $(\omega_1, \omega_2, \ldots) \in B$. A *$\sigma$-álgebra invariante de* X é a σ-álgebra sobre $(\Omega, \mathcal{F}, P)$ consistindo de conjuntos da forma $X^{-1}(B), B \in \mathcal{I}$.

Definição 10.13. *Defina uma aplicação $T : I^\infty \to I^\infty$ como segue: se $\omega = (\omega_0, \omega_1, \ldots)$, então $T(\omega) = (\omega_1, \omega_2, \ldots)$. Dizemos que T é uma translação. Defina $T^{n+1} = T(T^n)$, com $T^n(\omega) = (\omega_n, \omega_{n+1}, \ldots)$.*

Segue-se que $B \in \mathcal{I}$ se, e somente se, $T^{-1}B = B$, ou seja, conjuntos de $\mathcal{I}$ são invariantes com respeito a translações.

Definição 10.14. *A σ-álgebra permutável $\mathcal{G}$ é a σ-álgebra sobre I^∞ definida como segue: $B \in \mathcal{G}$ significa $(\omega_0, \omega_1, \ldots) \in B$ se, e somente se $(\omega_{\pi_0}, \omega_{\pi_1}, \ldots) \in B$, onde π é uma permutação finita de $\{0, 1, 2, \ldots\}$. A σ-álgebra permutável de X é aquela que contém $X^{-1}(B), B \in \mathcal{G}$.*

Observações: (a) T é mensurável, quando I^∞ é munido da σ-álgebra produto.

(b) $\mathcal{I} \subset \mathcal{H} \subset \mathcal{G}$.

(c) Sejam $I = \mathbb{Z}$ e $B = \{\omega : \omega_i = 1, \text{ i.v}\}$. Então $B \in \mathcal{I}$.

(d) Sejam $I = \mathbb{Z}$ e $B = \{\omega : \omega_{2i} = 0, \text{ i.v}\}$. Então, $B \in \mathcal{H}$, mas $B \notin \mathcal{I}$.

(e) Sejam $I = \{-1, 1\}$ e $B = \{\omega : \omega_0 + \ldots + \omega_n = 0, \text{ i.v}\}$. Então, $B \in \mathcal{G}$, mas $B \notin \mathcal{H}$.

Veja o Problema 8.

Teorema 10.13. *Seja X uma CM e i um estado recorrente. Se A for um conjunto de $\mathcal{H}$, então $P_i(A) = 0$ ou $P_i(A) = 1$.*

Prova: Seja B o subconjunto de I^∞ consistindo de todos os pontos ω tais que $\omega_0 = i$ e $\omega_n = i$ para infinitas coordenadas. Então, $P_i(B) = 1$, pois i é recorrente. Portanto, $P_i(A) = P_i(A \cap B)$, de modo que temos que calcular $P_i(A \cap B)$. Sejam

$$T_1 = \inf\{n > 0 : X_n = i\},$$

$$T_2 = \inf\{n > T_1 : X_n = i\},$$

e assim por diante. Sejam

$$Z_1 = (X_0, X_1, \ldots, X_{T_1-1}),$$

$$Z_2 = (X_{T_1}, X_{T_1+1}, \ldots, X_{T_2-1}),$$

etc. Pela propriedade forte de Markov, as Z_i são i.i.d.

Sejam $\beta_0, \beta_1, \ldots$ sequências de comprimentos finitos, cada uma começando com i (todas as coordenadas são pontos de I), contendo somente um i. Existe uma correspondência entre esses objetos e pontos de B. Se $\omega \in B$, seja $\beta\omega$ representando ω vista como uma sequência de sequências (finitas) como descrito acima. Então,

$$\begin{aligned} P_i(A \cap B) &= P_i\{(X_0, X_1, \ldots, X_n, \ldots) \in A \cap B\} \\ &= P_i\{(Z_1, Z_2, \ldots) \in \beta(A \cap B)\}. \end{aligned}$$

Seja π uma permutação finita dos inteiros não negativos, isto é, π permuta somente um número finito de inteiros. Sabemos que

$$(Z_1, Z_2, \ldots) \in \beta(A \cap B) \;\Leftrightarrow\; (X_0, X_1, \ldots) \in A \cap B. \tag{10.10}$$

Também,

$$\begin{aligned} (Z_{\pi_1}, Z_{\pi_2}, \ldots) \in &\beta(A \cap B) \;\Leftrightarrow \\ &\text{alguma permutação finita de } (X_0, X_1, \ldots) \text{ pertence a } A \cap B. \end{aligned} \tag{10.11}$$

Mas (10.10) é equivalente a (10.11), pois $A \cap B \in \mathcal{H}$. Portanto, $(Z_1, Z_2, \ldots) \in \beta(A \cap B)$ se, e somente se, $(Z_{\pi_1}, Z_{\pi_2}, \ldots) \in \beta(A \cap B)$. Pela lei 0-1 de Hewitt-Savage aplicada a $(Z_1, Z_2, \ldots)$ temos que $P_i\{(Z_1, Z_2, \ldots) \in \beta(A \cap B)\} = 0$ ou 1, ou seja, $P_i(A \cap B) = 0$ ou 1. $\square$

Corolário 10.3. *Se $A \in \mathcal{G}$, então $P_i(A) = 0$ ou $P_i(A) = 1$.*

Prova: Mesma prova, pois a hipótese de que $A \in \mathcal{H}$ entra somente em (10.10) e (10.11). $\square$

Teorema 10.14. *Seja X uma CM com todos os seus estados recorrentes. Sejam $\{I_m, m \in M\}$ a partição de I pelas subclasses movendo-se ciclicamente. Então temos que*

(a) Todo conjunto $A \in \mathcal{H}$ difere por um conjunto de probabilidade nula de uma reunião de conjuntos da forma $\{\omega : \omega_0 \in I_m\}$, $m \in M$.

(b) Reciprocamente, todo conjunto $\{\omega : \omega_0 \in I_m\}$ difere de um conjunto de $\mathcal{H}$ por um conjunto de probabilidade nula.

Prova: (a) Primeiramente, se i e j estão na mesma classe I_m, então $P_i(A) = P_j(A)$, se $A \in \mathcal{H}$. De fato, se $i, j \in I_m$, existem k e n tais que $p_{ik}^{(n)} > 0$ e $p_{jk}^{(n)} > 0$.

Suponha $P_i(A) > 0$ (ou seja, $P_i(A) = 1$). Mostremos que $P_j(A) > 0$. Temos que

$$1 = P_i(A) = p_{ik}^{(n)} P_i(A|X_n = k) + [1 - p_{ik}^{(n)}] P_i(A|X_n \neq k).$$

Como consequência da propriedade de Markov, temos $P_j(A|X_n = k) = P_i(A|X_n = k) > 0$. Como

$$P_j(A) = p_{jk}^{(n)} P_j(A|X_n = k) + [1 - p_{jk}^{(n)}] P_j(A|X_n \neq k),$$

vemos que $P_j(A) > 0$.

Para terminar a prova, sejam $M_0 = \{m \in M : i \in I_m \Rightarrow P_i(A) = 0\}$, $M_1 = \{m \in M : i \in I_m \Rightarrow P_i(A) = 1\}$. Seja de P_α a lei da cadeia X (sobre I^∞), onde α é a distribuição inicial. Então, $P_\alpha(A, \xi_0 \in I_m) = 0$, se $m \in M_0$ e $P_\alpha(A, \xi_0 \in I_m) = P_\alpha(\xi_0 \in I_m)$, se $m \in M_1$. Segue-se que, a um conjunto de probabilidade nula, A é a reunião de conjuntos da forma $\{\omega : \xi_0(\omega) \in I_m\}$, com $m \in M_1$.

(b) Recíproca: suponha que se $i \in I_m$ então i tem período d. O conjunto $\{\omega : \xi_0(\omega) \in I_m\}$ difere por um conjunto de probabilidade nula do conjunto

$$\{\omega : \xi_{nd}(\omega) \in I_m, \text{ para uma infinidade de valores de } n\}. \quad \square$$

Exemplo 10.8 Seja X um passeio aleatório, com $p_{i,i+1} = p,\ p_{i,i-1} = q,\ p+q = 1$ e sejam I_0, I_1 as classes movendo-se ciclicamente. Suponha que a distribuiçmao inicial seja $P(X_0 = 0) = 1/3,\ P(X_0 = 1) = 2/3$. Pelo Teorema 10.14, todo conjunto na σ-álgebra caudal tem probabilidade $0, 1/3, 2/3$ ou 1. Cada evento caudal difere por um conjunto de medida nula de um desses eventos: $\emptyset,\ \{\omega : \omega_0 = 1\},\ \{\omega : \omega_0 = 0\},\ \Omega$.

Teorema 10.15. *Seja X uma CM com todos os seus estados recorrentes. Seja $\{I_c, c \in \mathcal{C}\}$ o conjunto das classes recorrentes de X. Então temos que*

(a) Todo conjunto $A \in \mathcal{I}$ difere de um conjunto de probabilidade nula, de uma reunião de conjuntos da forma $\{\omega : \omega_0 \in I_c\}$, $c \in \mathcal{C}$.

(b) Todo conjunto $\{\omega : \omega_0 \in I_c\}$ difere de um conjunto de probabilidade nula de um conjunto de $\mathcal{I}$.

Prova: (a) Seja $A \in \mathcal{I}$, então $A \in \mathcal{H}$. Logo, $A = \cup_{m \in M_1} \{\omega : \omega_0 \in I_m\}$, q.c, sendo M_1, definido na prova anterior. Afirmamos que $A = \cup_{c \in N} \{\omega : \omega_0 \in I_c\}$, onde $N = \{c \in \mathcal{C} : I_c \supset I_m, \text{ para algum } m \in M_1\}$. Claramente $A \subset \cup_{c \in N} \{\omega : \omega_0 \in I_c\}$, q.c. Basta provar a inclusão em sentido contrário. Suponha que $\omega \in \{\omega : \omega_0 \in$

$I_m\} \cap A$, para algum $m \in M_1$, logo $\omega \in A$. Mas, se $\omega = (\omega_0, \omega_1, \ldots) \in A$, então $(\omega_1, \omega_2, \ldots) \in A$. Portanto, como $\omega_0 \in I_m$, $\omega_1 \in I_{m+1}$ etc. Portanto, A deve conter q.c a classe recorrente contendo I_m.

(b) Recíproca: O conjunto $\{\omega : \omega_0 \in I_c\}$ difere por um conjunto de probabilidade nula do conjunto

$$\{\omega : \xi_n(\omega) \in I_c, \text{ para uma infinidade de valores de } n\}. \quad \square$$

Teorema 10.16. (Blackwell) *Seja X uma CM com espaço de estados I e seja $A \in \mathcal{I}$. Então, existe um conjunto $\hat{A} \subset I$ tal que A difere de um conjunto de probabilidade nula de cada um dos conjuntos:* $\{\omega : \omega_n \in \hat{A}, \text{ i.v}\}$, $\{\omega : \omega_n \in \hat{A}$ *para todo n, com exceção de um número finito deles*$\}$. *Aqui não supomos que todos os estados de X sejam recorrentes.*

Prova: Temos que

$$P\{(X_0, X_1, \ldots) \in A | X_0 = i_0, \ldots, X_n = i_n\} =$$

$$P\{(X_{n+1}, X_{n+2}, \ldots) \in A | X_0 = i_0, \ldots, X_n = i_n\},$$

pois $A \in \mathcal{I}$. Pela propriedade de Markov, o último termo é igual a

$$P\{(X_{n+1}, X_{n+2}, \ldots) \in A | X_n = i_n\} = P\{(X_1, X_2, \ldots) \in A | X_0 = i_n\},$$

usando o fato que a CM X é homogênea. Portanto, existe uma função boreliana h tal que $P\{(X_0, X_1, \ldots) \in A | X_0, \ldots, X_n\} = h(X_n)$.

Agora, $P\{(X_0, X_1, \ldots) \in A | X_0, \ldots, X_n\} = E(I_{X^{-1}(A)} | X_0, \ldots, X_n)$ é um martingale, que converge q.c para $E(I_{X^{-1}(A)} | X_0, X_1, \ldots)$, quando $n \to \infty$. Mas $A \in X^{-1}(\mathcal{I}) \subset \mathcal{F}\{X_0, X_1, \ldots\}$, portanto

$$P\{X^{-1}(A) | X_0, \ldots, X_n\} \to I_{X^{-1}(A)}, \text{ q.c.}$$

Segue que $h(X_n) \to I_{X^{-1}(A)}$, q.c. Como podemos re-escrever o precedente em termos de ξ_n, temos que

$$h(\xi_n) \to I_A, \; P_X\text{-q.c.} \tag{10.12}$$

Seja a qualquer número tal que $0 < a < 1$ e defina $\hat{A} = \{i \in I : h(i) > a\}$ e seja N o conjunto nulo envolvido em (10.12). Suponha que $\omega \in A - N$. Então, $h[\xi_n(\omega)] \to I_A(\omega) = 1$, donde $\omega \in \{\omega : \xi_n(\omega) \in \hat{A}, \text{ para todos os } n \text{ exceto um número finito deles}\}$. Suponha que $\omega \notin A$, $\omega \notin N$, então $h[\xi_n(\omega)] \to I_A(\omega) = 0$, logo $\omega \in \{\omega : \xi_n(\omega) \in \hat{A}, \text{i.v}\}^c$. $\square$

Problemas

1. Prove o Corolário 10.1.
2. Prove a afirmação contida no Exemplo 10.3.
3. O passeio aleatório do Exemplo 10.4 (b) tem incrementos independentes?
4. Prove o fato [2]
5. Prove o fato [4].
6. Mostre, por meio de um exemplo, que se I for infinito, a parte (b) do Teorema 10.6 não vale.

   [Sugestão: considere a CM com I o conjunto dos inteiros positivos e $p_{i,i+1} = 1$.]
7. Prove a afirmação do Exemplo 10.8.
8. Prove as afirmações (a)-(e) das Observações feitas após a Definição 10.14.
9. Suponha que $\{X_n, n \geq 0\}$ seja uma CM. Prove que $\lim_{n\to\infty} E\{E(f(X_0)|X_n)|X_0\}$ existe, onde f é uma função boreliana limitada.
10. Suponha que $\{X_n, n \geq 0\}$ seja uma CM e f uma função boreliana. Mostre, por meio de um exemplo, que $\{f(X_n), n \geq 0\}$ não é, necessariamente, um processo de Markov. Prove que $f(X_n)$ é um processo de Markov se f for 1-1.
11. Seja X uma CM com $I \subset \mathbb{R}$. Suponha que $X_0 = 0$, $\{X_{n+1} - X_n, n \geq 0\}$ sejam i.i.d. (a) Prove que ou todos os estados são recorrentes ou nenhum o é.

    (b) Se todos os estados forem recorrentes, então I é um grupo aditivo e todos os pontos de I são da forma $\{nd, n \in \mathbb{Z}\}$, com $d > 0$ e $\mathbb{Z}$ é o conjunto dos inteiros.
12. Seja X um processo estocástico, com espaço de estados I enumerável e suponha que exista uma função φ tal que, para cada n,

    $$P\{X_{n+1} = j|X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0\} = \varphi(i, j).$$

    Prove que X é uma CM homogênea.
13. Sabemos que se $X_0, X_1, \ldots$ é um processo de Markov, então $X_n, X_{n-1}, \ldots, X_1$ tem a propriedade de Markov. Verifique isso diretamente para o caso de uma CM e encontre as probabilidades de transição. Suponha I enumerável e que a CM é homogênea.
14. Suponha que P_m denote a distribuição de Poisson com parâmetro m. Escolha um inteiro n_1 de acordo com a distribuição P_1; depois escolha um segundo inteiro n_2 de acordo com a distribuição P_{n_1}, e assim por diante. Prove que esse processo de Markov atinge o estado 0 (e permanece lá).
15. Seja $[p_{ij}]$ uma matriz de transição e I o espaço dos estados. Uma medida μ sobre I é:

    (i) invariante à diretita (ID) se $\sum_j p_{ij}\mu(j) = \mu(i)$;

    (ii) super-invariante à direita (SID) se $\sum_j p_{ij}\mu(j) \leq \mu(j)$;

    (ii) invariante à esquerda (IE) se $\sum_i \mu(i)p_{ij} = \mu(j)$;

(iv) super-invariante à esquerda (SIE) se $\sum_i \mu(i)p_{ij} \leq \mu(j)$.

(a) Suponha que todos os estados sejam transitórios. Prove que existem sempre (muitas) medidas SID não-constantes.

[Sugestão: fixe j_0, tente $\mu(i) = \sum_{n=0}^{\infty} p_{i,j_0}^{(n)}$. Lembre-se que $\sum p_{ij}^{(n)} < \infty$.]

(b) Suponha que todos os estados comunicam-se e são recorrentes. Prove que todas as medidas SID não negativas são constantes.

[Sugestão: Seja μ uma medida SID, não negativa; seja X_n a CM começando em i, Mostre que $\mu(X_n)$ é um super-martingale não negativo. Conclua usando seu conhecimento da álgebra caudal.]

(c) Suponha que que todos os estados sejam comunicantes e que exista uma medida IE finita. Prove que a a CM é recorrente positiva.

16. Suponha que todos os estados sejam comunicantes e recorrentes positivos. Suponha que a distribuição inicial seja a distribuição estacionária. Prove que a cadeia reversa tem transições estacionárias e é recorrente.

Capítulo 11

Teoria Ergódica

Neste capítulo trataremos dos aspectos principais da teoria ergódica, como transformações invariantes (ou que preservam a medida), recorrência e os teoremas ergódicos pontual e médio. As referências principais que serão usadas são Billingsley (1978), Halmos (2006) e Garsia (1970).

11.1 Transformações invariantes

Definição 11.1. *Seja* $(\Omega, \mathcal{F}, P)$ *um espaço de medida (*P *não precisa ser uma medida de probabilidade, ou mesmo finita em várias situações). Seja* $T : \Omega \to \Omega$ *uma função mensurável sobre* Ω*, isto é,* $T^{-1}(B) \in \mathcal{F}$ *se* $B \in \mathcal{F}$*. Dizemos que* T *preserva a medida* P *se* $P(T^{-1}(A)) = P(A)$*, para todo* $A \in \mathcal{F}$*.*

Exemplo 11.1. (i) Seja $\Omega = \{a_1, a_2, \ldots, a_n\}$, $\mathcal{F}$ a classe de todos os subconjuntos de Ω. Defina T por $Ta_k = a_{k+1}$, se $k < n$ e $Ta_n = a_1$ (permutação cíclica). Então, T preserva P se, e somente se, $P(\{a_i\})$ não depende de i.

De modo geral, se T for qualquer permutação de Ω, T pode ser expandida como um produto de ciclos disjuntos $C_1, \ldots, C_k$. Nesse caso, T preserva P se, e somente se, dentro de cada ciclo, P associa pesos iguais a cada ponto.

(ii) Suponha $\Omega = \mathbb{R}$, $\mathcal{F} = \mathcal{B}$ e P a medida de Lebesgue na reta. Defina T por $Tx = x + a$, sendo a um número real fixo. Então, T preserva P.

(iii) Seja Ω o círculo unitário no plano complexo, $\mathcal{F}$ a σ-álgebra de Borel sobre Ω e P a medida de Lebesgue sobre Ω dividida por 2π. Defina T por $Te^{i\theta} = e^{i(\theta+\alpha)}$, α fixo. Então, T é uma rotação e preserva P.

(iv) Seja $\Omega = [0, 1)$, P a medida de Lebesgue sobre Ω, e $\mathcal{F} = \mathcal{B}([0, 1])$. Suponha $T\omega = 2\omega$ (mod 1)(ou seja, $T\omega = 2\omega$ se $\omega < 1/2$ e $T\omega = 2\omega - 1$, se $1/2 \leq \omega < 1$). Então, T preserva P. T é chamada *transformação diádica.* Em outras palavras, se $\omega \in \Omega$ e tem a expansão diádica $\omega = 0, \omega_1\omega_2 \cdots$, então $T\omega = 0, \omega_2\omega_3 \cdots$.

Definição 11.2. *Seja $\Omega = \mathbb{R}^\infty$, ou seja, o conjunto de todas as sequências $(\omega_0, \omega_1, \omega_2, \cdots)$ de números reais, e $\mathcal{F}$ a σ-álgebra produto sobre Ω, ou seja, $\mathcal{B}^\infty$. Seja P uma probabilidade sobre $(\Omega, \mathcal{F})$. A medida P diz-se estacionária se, para todo $B \in \mathcal{F}$, tivermos $P\{\omega : (\omega_0, \omega_1, \cdots) \in B\} = P\{\omega : (\omega_1, \omega_2, \cdots) \in B\}$. A translação unilateral T é a aplicação $T : \Omega \to \Omega$ definida por $T(\omega_0, \omega_1, \cdots) = (\omega_1, \omega_2, \cdots)$.*

Exemplo 11.2. (a) Considere T e P como na definição 11.2. Então, T preserva P se, e somente se, P é estacionária. Veja o Problema 2.

(b) São exemplos de medidas estacionárias sobre $(\mathbb{R}^\infty, \mathcal{B}^\infty)$:

(i) Seja $X = \{X_n, n \geq 0\}$ um processo estritamente estacionário definido sobre $(\Omega, \mathcal{F}, P)$. Seja P_X a probabilidade definida sobre $(\mathbb{R}^\infty, \mathcal{B}^\infty)$ definida por: se $B \in \mathcal{B}^\infty$, então $P_X(B) = P\{\omega : (X_0(\omega), X_1(\omega), \ldots) \in B\}$. Segue-se que P_X é a distribuição de X. Então, P_X é estacionária e a translação T sobre $\mathbb{R}^\infty$ preserva P_X.

Logo, começando com qualquer processo estocástico estritamente estacionário, podemos construir uma transformação invariante.

Reciprocamente, dada qualquer transformação T que preserva a medida, podemos construir um processo estritamente estacionário como segue: considere T sobre $(\Omega, \mathcal{F}, P)$ e seja X uma v.a. sobre esse espaço. O processo $Y = \{Y_n, n \geq 0\}$ definido por $Y_n(\omega) = X(T^n(\omega))$ é estritamente estacionário. Veja Breiman (1968). T^0 é definida como a identidade.

(2) Seja $X = \{X_n, n \geq 0\}$ uma cadeia de Markov com pelo menos uma classe recorrente positiva. Seja π uma medida estacionária. Considere a distribuição π como distribuição inicial de X. Então X é estritamente estacionário.

Definição 11.3. *Seja $(\Omega, \mathcal{F}, P)$ um espaço de probabilidade, e seja T preservando P. O conjunto $A \in \mathcal{F}$ é invariante se $A = T^{-1}A$, isto é, $x \in A$ se, e somente se, $Tx \in A$. Dizemos que A é quase-invariante se A e $T^{-1}A$ diferem q.c. Uma v.a X sobre $(\Omega, \mathcal{F}, P)$ é invariante se, e somente se, $X(\omega) = X(T\omega)$ e é quase-invariante se, e somente se, $X(\omega) = X(T\omega)$ q.c.*

Teorema 11.1. *(a) Se $\mathcal{I}$ é a classe dos conjuntos invariantes e $\mathcal{I}'$ é a classe dos conjuntos quase-invariantes, então ambas são σ-álgebras.*

(b) *Qualquer conjunto quase-invariante difere de um conjunto invariante somente por um conjunto nulo.*

(c) *Uma v.a X é invariante se, e somente se X é $\mathcal{I}$-mensurável.*

Prova: (a) Imediata.

(b) Seja A quase-invariante e $A_1 = \limsup_{n\to\infty} T^{-n}(A)$. Então, A_1 é invariante e difere de A por um conjunto nulo.

(c) ($\Rightarrow$) $\{\omega : X(\omega) \leq x\} = \{\omega : X(T\omega) \leq x\} = T^{-1}\{\omega : X(\omega) \leq x\}$.

($\Leftarrow$) Suponha que $A \in \mathcal{I}$ e seja $X = I_A$. Então, $X(T\omega) = I_A(T\omega) = I_{T^{-1}(A)}(\omega) = I_A(\omega)$, donde o resultado é verdadeiro se X é uma função indicadora em $\mathcal{I}$. Segue que o resultado é verdadeiro para toda v.a $\mathcal{I}$-mensurável, por um argumento padrão. $\square$

Definição 11.4. *Uma transformação T que preserva uma medida P é ergódica se para qualquer conjunto invariante A tivermos $P(A) = 0$ ou $P(A^c) = 0$. Se P é uma medida de probabilidade, então T é ergódica se todo conjunto invariante A for tal que $P(A) = 0$ ou $P(A) = 1$.*

Exemplo 11.3. Seja X uma CM com pelo menos uma classe recorrente positiva. Seja π uma medida estacionária concentrada em uma dessas classes. A existência de tal medida é garantida pela Seção 10.7 do Capítulo 10. Seja P_π a distribuição estacionária induzida sobre I^∞ dando a X a distribuição inicial π. Seja T a translação sobre I^∞ (I é o espaço de estados de X). Então, T preserva a medida P_π e é ergódica.

Definição 11.5. *Seja T uma transformação que preserva a medida, definida sobre $(\Omega, \mathcal{F}, P)$, um espaço de probabilidade. Dizemos que T é mixing se, para todo $A, B \in \mathcal{F}$, tivermos*

$$\lim_{n\to\infty} P(A \cap T^{-n}B) = P(A)P(B).$$

Essa é uma forma de independência assintótica. Para uma motivação intuitiva dessa noção, veja Halmos (2006).

Exemplo 11.4. Seja P_π como no exemplo anterior. Suponha, ainda, que a classe recorrente positiva mencionada lá tenha somente uma subclasse movendo-se ciclicamente. Então, a translação T sobre I^∞ é mixing (veja o Corolário 11.1 a seguir). Isso segue do fato que a σ-álgebra caudal $\mathcal{H}$ aqui é trivial (veja a seção 10.9 do Capítulo 10).

Teorema 11.2. *Suponha que T seja mixing. Então, T é ergódica (a recíproca não vale).*

Prova: Seja B invariante. Devemos provar que $P(B) = 0$ ou $P(B) = 1$. Como $T^{-n}(B) = B$, para todo n, temos $P(A \cap B) = P(A \cap T^{-n}B) \to P(A)P(B)$. Como isso vale para qualquer A, tome $A = B$. $\square$

Teorema 11.3. *Seja T uma transformação que preserva a medida sobre $(\Omega, \mathcal{F}, P)$. Seja $\mathcal{F}_0$ uma álgebra gerando $\mathcal{F}$. Se a condição de mixing vale para todo $A, B \in \mathcal{F}_0$, então a condição vale para $A, B \in \mathcal{F}$.*

Prova: Sejam A e B conjuntos em $\mathcal{F}$. Tome A_k, B_k em $\mathcal{F}_0$ tais que $P(A\triangle A_k) \to 0$, $P(B\triangle B_k) \to 0$ quando $k \to \infty$ (possível pelo Problema 16 do Capítulo 1). Agora,

$$P\{(A \cap T^{-n}B)\triangle(A_k \cap T^{-n}B_k)\}$$

$$\leq P(A\triangle A_k) + P[T^{-n}(B\triangle B_k)] = P(A\triangle A_k) + P(B\triangle B_k),$$

que tende a zero, quando $k \to \infty$. Portanto, $P(A_k \cap T^{-n}B_k) \to P(A \cap T^{-n}B)$, uniformemente em n, logo

$$\lim_{n\to\infty} P(A \cap T^{-n}B) = \lim_{n\to\infty}\lim_{k\to\infty} P(A_k \cap T^{-n}B_k) =$$

$$\lim_{k\to\infty}\lim_{n\to\infty} P(A_k \cap T^{-n}B_k) = \lim_{k\to\infty} P(A_k)P(B_k) = P(A)P(B),$$

na qual a mudança dos limites é justificada pela convergência uniforme. $\square$

Exemplo 11.5. Continuação do Exemplo 11.1.

(a) Se T for uma permutação cíclica, T é ergódica se, e somente se, T tem um só ciclo. No caso em que P é uma probabilidade, T nunca é mixing.

(b) Seja $\Omega = \mathbb{R}$, $Tx = x + a$. Então T não é ergódica pois $\cup_{n=-\infty}^{\infty}(na, (n + 1/2)a)$ é um conjunto invariante não trivial.

(c) Se T for uma rotação, $Te^{i\theta} = e^{i(\theta+\alpha)}$, então T é ergódica se, e somente se α for irracional. T não é mixing nunca. Veja o Problema 11. Veja, também, Breiman (1968) e Billingsley (1978).

(d) Seja T uma transformação diádica. Então, T é ergódica e mixing. Veja Billingsley (1978) para detalhes.

(e) Seja $\Omega = \mathbb{Z}$, $\mathcal{F}$ a classe de todos os subconjuntos de Ω e P a medida de contagem ($P(A)$ dá o número de elementos de A). Seja T tal que $T\omega = \omega + 1$. Então, T é ergódica.

Definição 11.6. *Considere $\Omega = \mathbb{R}^\infty$ (ou I^∞ como em CM), $\mathcal{F}$ a σ-álgebra produto, P uma medida de probabilidade estacionária sobre $(\Omega, \mathcal{F})$ e T a translação (unilateral). Sejam $\xi_1, \xi_2, \ldots$ as funções coordenadas e $\mathcal{H}$ a σ-álgebra caudal dos ξ_i. Dizemos que T é uma translação de Kolmogorov se $\mathcal{H}$ for trivial e que T é uma translação de Markov se ξ_n forma uma CM.*

Teorema 11.4. *Toda translação de Kolmogorov é mixing.*

Prova: Seja $A \in \mathcal{H}$ e seja B um cilindro. Então,

$$|P(A \cap T^{-n}B) - P(A)P(B)| = |E(I_A I_{T^{-n}B}) - P(A)P(B)| =$$

$$\Big|\int_{T^{-n}B} I_A dP - P(A)P(B)\Big| = \Big|\int_{T^{-n}B} E(I_A|\mathcal{F}_n^{'})dP - P(A)P(B)\Big|,$$

onde $\mathcal{F}_n^{'} = \mathcal{F}\{\xi_n, \xi_{n+1}, \ldots\}$ e usamos a definição da esperança condicional ao obter a última igualdade ($T^{-n}B \in \mathcal{F}_n^{'}$).

Como T preserva a medida, o último termo é igual a $|\int_{T^{-n}B}[E(I_A|\mathcal{F}_n^{'}) - P(A)]dP| \leq \int_\Omega |E(I_A|\mathcal{F}_n^{'}) - P(A)|dP$. Mas $E(I_A|\mathcal{F}_n^{'}) - P(A)$ é um martingale, que converge para $E[I_A|\mathcal{H}] - P(A)$. Mas $\mathcal{H}$ é trivial, por hipótese, logo, de fato, o martingale em questão converge para $P(A) - P(A) = 0$. Segue pelo TCD que $P(A \cap T^{-n}B) \to P(A)P(B)$, se B for um cilindro. Pelo teorema anterior, a convergência vale para todo B. $\square$

Corolário 11.1. *A translação de Markov do Exemplo 11.4 é mixing.*

Existem translações que são mixing, mas não de Kolmogorov.

O teorema a seguir mostra que, a fim de responder a muitas questões da teoria ergódica, podemos restringir atenção, sem perda de generalidade, somente a translações.

Teorema 11.5. *Seja T_0 uma transformação que preserva a medida, sobre o espaço de probabilidade $(\Omega_0, \mathcal{F}_0, P_0)$. Seja $\Omega = \Omega_0^\infty$, o conjunto de todas as sequências $(\omega_0, \omega_1, \ldots)$, com $\omega_n \in \Omega_0$, $\mathcal{F}$ a σ-álgebra produto e P a probabilidade sobre $(\Omega, \mathcal{F})$ dada por:*

$$P(B) = P_0\{x \in \Omega_0 : (x, T_0 x, T_0^2 x, \ldots) \in B\}, \quad B \in \mathcal{F}.$$

Seja T a translação sobre $(\Omega, \mathcal{F}, P)$. Então:

(a) *T preserva P;*

(b) *T_0 é ergódica (mixing) $\Leftrightarrow$ T é ergódica (mixing);*

(c) *$P\{\omega \in \Omega : \omega_k = T_0^k \omega_0,$ para todo $k \geq 1\} = 1$.*

Prova: Veja o Problema 10.

11.2 Recorrência

Seja $(\Omega, \mathcal{F}, P)$ um espaço de probabilidade e T uma transformação mensurável, $T : \Omega \to \Omega$. Tomemos $A \in \mathcal{F}$ e $\omega \in \Omega$. Considere a sequência $\omega, T\omega, T^2\omega, \cdots$. Duas questões básicas dessa seção: essa sequência entra no conjunto A? Entra infinitas vezes?

Definição 11.7. *(a) Dizemos que T é recorrente se definindo $A^{(r)} = \{\omega \in A : T^n\omega \in A,$ para algum $n \geq 1\}$, então $P(A - A^{(r)}) = 0$, para todo $A \in \mathcal{F}$.*

(b) *Dizemos que T é infinitamente recorrente se $A^{(i)} = \{\omega \in A : T^n\omega \in A \text{ i.v}\}$, então $P(A - A^{(i)}) = 0$, para todo $A \in \mathcal{F}$.*

(c) *$B \in \mathcal{F}$ diz-se wandering se $B, T^{-1}B, T^{-2}B, \cdots$ são disjuntos.*

(d) *T é conservativa se todos os conjuntos wandering têm medida P zero.*

(e) *T é incompressível se $T^{-1}(A) \subset A$ implica $P(A - T^{-1}(A)) = 0$.*

Teorema 11.6. *As seguintes afirmações são equivalentes:*

(1) T é incompressível;

(2) T é conservativa;

(3) T é recorrente;

(4) T é infinitamente recorrente.

Prova: (1) $\Rightarrow$ (2) Suponha que A seja *wandering* e seja $B = \cup_{n=0}^{\infty} T^{-n}A$. Então, $T^{-1}(B) = \cup_{n=1}^{\infty} T^{-n}A$ é um subconjunto de B, logo por (1) $P(B - T^{-1}B) = 0$. Mas $B - T^{-1}B = A$, pois os $T^{-n}A$ são disjuntos, portanto $P(A) = 0$.
(2) $\Rightarrow$ (3) Seja $A \in \mathcal{F}$, $C = A - A^r$. Então, $T^{-n}C = \{\omega : T^n\omega \in A - A^r\} = \{\omega : T^n\omega \in A$ mas $T^k\omega \notin A, k > n\}$. Logo, os $T^{-n}C$ são disjuntos, logo C é *wandering* e portanto $P(C) = 0$.
(3) $\Rightarrow$ (1) Seja $T^{-1}A \subset A$. Então, $T^{-2}A = T^{-1}(T^{-1}A) \subset T^{-1}A \subset A$ e, portanto, $T^{-n}A \subset T^{-1}A$, para $n \geq 1$. Segue que $T^{-1}A = \cup_{n=1}^{\infty} T^{-n}A$, mas $A - T^{-1}A = A - \cup_{n=1}^{\infty} T^{-n}A = A - A^r$, e este conjunto tem probabilidade zero.
(4) $\Rightarrow$ (3) Imediato.
(1) $\Rightarrow$ (4). Seja $A \in \mathcal{F}$ e $B = \cup_{n=0}^{\infty} T^{-n}A$. Então, $T^{-1}B \subset B$, donde $P(B - T^{-1}B) = 0$. Similarmente, $T^{-(k+1)}B \subset T^{-k}B$, implicando que $P(T^{-k}B - T^{-(k+1)}B) = 0$. Mas $T^{-k}B - T^{-(k+1)}B = \cup_{n=k} T^{-n}A - \cup_{n=k+1}^{\infty} T^{-n}A$, que é igual ao conjunto dos ω tais que $T^n\omega$ entra em A pela última vez em $n = k$. Como $A - A^i = A \cap \cup_{k=0}^{\infty}\{\omega : T^n\omega \in A$ pela últiva vez em $n = k\}$, temos que $P(A - A^i) \leq \sum_k P\{T^{-k}B - T^{-(k+1)}B\} = 0$.
$\square$

Corolário 11.2. (Poincaré) *Seja T uma transformação que preserva a medida sobre um e.p $(\Omega, \mathcal{F}, P)$. Então, T é infinitamente recorrente.*

Prova: Seja A wandering. Como $\sum_n P(A) = \sum_n P(T^{-n}A) = P(\cup T^{-n}A) \leq 1$, segue-se que $P(A) = 0$. Logo, T é conservativa, e portanto infinitamente recorrente.
$\square$

Observação: Seja X qualquer processo estocástico que seja estritamente estacionário. Seja T a translação associada. Então, T é infinitamente recorrente, pois preserva a medida. Em termos do processo X temos: se X for um p.e estritamente estacionário e se B for um conjunto de Borel, então

$$P\{X_0 \in B \text{ e } X_n \in B \text{ i.v } \} = P\{X_0 \in B\}. \tag{11.1}$$

Observe que já tínhamos notado isso como verdade para CM estritamente estacioária. Ou, se $P\{X_0 \in B\} > 0$, então (11.1) reduz-se a

$$P\{X_n \in B \text{ i.v } | X_0 \in B\} = 1,$$

para qualquer p.e estritamente estacionário.

Teorema 11.7. (Kac, 1947) *Seja T uma transformação que preserva a medida sobre um e.p $(\Omega, \mathcal{F}, P)$. Seja $A \in \mathcal{F}$ e $A_k = \{\omega \in A : T^n\omega \notin A,\ 1 \leq n \leq k-1,\ \text{e}\ T^k\omega \in A\}$. Seja $r_A(\omega) = k$ se $\omega \in A_k$ (r_A é o tempo de recorrência, e $r_A(\omega) < \infty$ q.c, pelo Teorema de Poincaré). Então:*

$$\int_A r_A(\omega) dP(\omega) = P\{\cup_{n=0}^{\infty} T^{-n} A\}. \tag{11.2}$$

Note que A_k é o conjunto no qual o primeiro retorno a A ocorre no tempo k. Note, também, que a probabilidade do lado direito de (11.2) é a probabilidade de que $T^n\omega$ esteja em A para algum $n \geq 0$. Veja abaixo exemplos para algumas interpretações do teorema.

Prova: $P\{\omega \in A : r_A(\omega) = k+1\} = P\{\omega \in A : T^n\omega \notin A, 1 \leq n \leq k, T^{k+1}\omega \in A\}$. Se $B_k = (T^{-k}A)^c, C_k = T^{-k}A$, então

$$P\{\omega \in A : r_A(\omega) = k+1\} = P\{C_0 \cap B_1 \cap \cdots \cap B_k \cap C_{k+1}\}$$

$$= P\{C_0 \cap B_1 \cap \cdots \cap B_k\} - P\{C_0 \cap B_1 \cap \cdots \cap B_{k+1}\}$$

$$= P\{B_1 \cap \cdots \cap B_k\} - P\{B_0 \cap B_1 \cap \cdots \cap B_k\}$$

$$-P\{B_1 \cap \cdots \cap B_{k+1}\} + P\{B_0 \cap B_1 \cap \cdots \cap B_{k+1}\}$$

$$= P\{B_0 \cap \cdots \cap B_{k-1}\} - 2P\{B_0 \cap \cdots \cap B_k\} + P\{B_0 \cap \cdots \cap B_{k+1}\},$$

pois T preserva P. Logo, podemos escrever

$$P\{\omega \in A : r_A(\omega) = k+1\} =: b_k - 2b_{k+1} + b_{k+2}, \quad k \geq 1,$$

usando uma notação óbvia. Definindo $b_0 = 1$, isso vale para todo $k \geq 0$. Agora,

$$\int_A r_A dP = \sum_{k \geq 0} (k+1)(b_k - 2b_{k+1} + b_{k+2})$$

$$= \lim_{n\to\infty} \sum_{k=0}^{n-1} (k+1)(b_k - 2b_{k+1} + b_{k+2})$$

$$= \lim_{n\to\infty} [b_0 - (n+1)b_n + nb_{n+1}].$$

O limite deve existir (em $\bar{\mathbb{R}}_+$), pois temos a n-ésima soma parcial de uma série não negativa, logo

$$\int_A r_A dP = 1 - \lim_{n\to\infty} [n(b_n - b_{n+1}) + b_n].$$

Mas $b_n - b_{n+1} = P(B_0 \cap \cdots \cap B_{n-1}) - P(B_0 \cap \cdots \cap B_n) = P(B_0 \cap \cdots \cap B_{n-1} \cap C_n) \geq 0$. Como $B_0 \cap \cdots \cap B_{n-1} \cap C_n$, $n \geq 1$, são disjuntos, $\sum_n (b_n - b_{n+1}) < \infty$. Como o $\lim_{n\to\infty} [n(b_n - b_{n+1}) + b_n]$ existe e $b_n \downarrow P(\cap_k B_k)$, o $\lim_{n\to\infty} n(b_n - b_{n+1})$ existe. Portanto, se $a_n = b_n - b_{n+1}$, então $a_n \geq 0$, $\sum a_n < \infty$ e $\lim_{n\to\infty} na_n$ existe. Finalmente, temos que $na_n \to 0$, pois se não, $na_n \geq \varepsilon$, para todo n suficientemente grande, e então $\sum a_n \geq \varepsilon \sum 1/n$, contradizendo o fato que $\sum a_n < \infty$. Segue que

$$\int_A r_A dP = 1 - \lim_{n\to\infty} b_n = 1 - P(\cap_{n=0}^{\infty} B_n) = P(\cup_{n=0}^{\infty} T^{-n} A). \quad \square$$

Exemplo 11.6. Suponha que T seja também ergódica e que $P(A) > 0$. Seja $E = \cup_{n=0}^{\infty} T^{-n} A$. Se $T\omega \in E$, então $T\omega$ pertence a algum $T^{-n}A$, de modo que $\omega \in T^{-(n+1)}A$, logo $\omega \in E$. Ou seja, $T^{-1}E \subset E$. Como T é incompressível, $P(E - T^{-1}E) = 0$, logo $T^{-1}E = E$ q.c. Segue-se que E é quase invariante e portanto $\int_A r_A(\omega) dP(\omega) = 1$ ($P(E) \geq P(A) > 0$, ou seja, $P(E) = 1$, pois E é invariante). De outro modo,

$$\frac{1}{P(A)} \int_A r_A(\omega) dP(\omega) = \frac{1}{P(A)}.$$

Ou seja, dado que o ponto inicial está em A, a amplitude média de tempo para retornar a A é $1/P(A)$.

Exemplo 11.7. Seja $X = \{X_n, n \geq 0\}$ um p.e estritamente estacionário e $S = \inf\{n \geq 1 : X_n \in A\}$. Pela observação acima, $P\{S < \infty | X_0 \in A\} = 1$, se $P\{X_0 \in A\} > 0$.

Suponha que X seja ergódico (isto é, a translação associada é ergódica). Então,

$$E\{S | X_0 \in A\} = \frac{1}{P\{X_0 \in A\}}.$$

Um caso especial é: se X é uma CM com distribuição inicial concentrada numa classe recorrente positiva única, então, $E_i S = 1/\pi_i$, um resultado já conhecido.

Teorema 11.8. *Seja $X = \{X_n, n \geq 0\}$ estritamente estacionário e ergódico (ou seja, a σ-álgebra invariante é trivial). Sejam $T_1, T_2, \ldots$ tempos de retornos sucessivos ao conjunto A. Então, o processo $\{T_{k+1} - T_k, k \geq 0\}$ (com $T_0 = 0$) é estritamente estacionário e ergódico, sob $P\{\cdot|X_0 \in A\}$.*

Prova: Omitida. Veja Breiman (1968).

Caso especial: Tome X como uma CM com um estado i recorrente. Sejam $\{T_k, k \geq 1\}$ tempos de sucessivos retornos ao estado i, começando em i. Então, $T_1, T_2 - T_1, T_3 - T_2, \ldots$ são i.i.d, e portanto formam um processo estritamente estacionário e ergódico.

11.3 Teoremas ergódicos

Nesta seção trataremos do teorema ergódico médio e do teorema ergódico pontual e a forma de Hopf de ambos. Depois veremos algumas recíprocas desses teoremas.

Seja $(\Omega, \mathcal{F}, P)$ um e.p e, para $p \geq 1$, considere o espaço $L_p := L_p(\Omega, \mathcal{F}, P)$ (veja o Apêndice A.3). A seguir apresentamos dois exemplos clássicos de operadores lineares sobre L_p.

Exemplo 11.8. (a) Seja $\mathcal{F}_0 \subset \mathcal{F}$ e $Tf = E(f|\mathcal{F}_0)$.

(b) Seja S uma transformação que preserva a medida sobre $(\Omega, \mathcal{F}, P)$ e defina T por $Tf(\omega) = f(S\omega)$. É fácil ver que T é linear. Mostre que $\|f\|_p = \|Tf\|_p$.

Os dois operadores do Exemplo 11.8 são contrações positivas. Note que, se T é uma contração, $\|T^n f\|_p \leq \|f\|_p$ para todo $n \geq 1$.

Vamos, agora, enunciar os dois teoremas principais dessa seção.

Teorema 11.9. (Teorema Ergódico Médio - TEM) *Seja T um operador linear sobre L_p, $p \geq 1$, sendo $(\Omega, \mathcal{F}, P)$ um e.p. Suponha que T seja uma contração positiva tal que T1=1 e seja $f \in L_p$. Então,*

$$R_n(f) := \frac{f + Tf + \ldots + T^n f}{n+1}$$

converge em norma L_p para um limite, denotado por $\hat{P}f$.

Teorema 11.10. (Teorema Ergódico Pontual - TEP) *Seja T como no Teorema 11.9. Então,*

$$R_n(f) \stackrel{q.c}{\to} \hat{P}f.$$

Antes de provar os teoremas, vamos considerar alguns exemplos.

Exemplo 11.9. (1) Seja $X = \{X_n, n \geq 0\}$ um processo estritamente estacionário. Seja S a translação sobre $(\mathbb{R}^\infty, \mathcal{B}^\infty, P_X)$, sendo P_X a distribuição de X. Então, S preserva P_X, de modo que S é uma transformação que preserva a medida sobre esse espaço. Defina um operador T por $Tf = f(S\omega)$. Então, T é um operador linear positivo (isometria), $T1 = 1$.

Seja ξ a projeção de $\omega = (\omega_0, \omega_1, \ldots)$ sobre a primeira coordenada. Observe que $X_0(\omega), X_1(\omega), \ldots) \sim (\xi(\omega), \xi(S\omega), \ldots)$. Logo, para provar que $(X_0 + X_1 + \ldots + X_n)/(n+1)$ converge, é suficiente provar que $(\xi(\omega) + \xi(S\omega) + \ldots + \xi(S^{(n)}(\omega))/(n+1)$ converge. Mas o último é igual a $(\xi + T\xi + \ldots + T^n\xi)/(n+1)$.

Suponha, agora, que X_0 seja integrável. Então ξ é integrável, logo pelos teoremas ergódicos, $(\xi + T\xi + \ldots + T^n\xi)/(n+1)$ converge q.c e em L_1.

Ou seja, se $X = \{X_n, n \geq 0\}$ é um processo estritamente estacionário, e se X_0 for integrável, então $(X_0 + \ldots + X_n)/(n+1)$ converge q.c e em L_1.

Como caso especial, se $X_0, X_1, \ldots$ são v.a's i.i.d, integráveis, então $(X_0 + \ldots + X_n)/(n+1)$ converge q.c e em L_1. Temos, pois, uma outra prova da LFGN. Provaremos, também, que o limite acima é $E(X_0|\mathcal{I})$, onde $\mathcal{I}$ é a σ-álgebra invariante para X.

(2) Seja I um conjunto enumerável e $\mathbb{P} = [p_{ij}]_{i,j}$ uma matriz de transição. Tome $\Omega = I, \mathcal{F}$ como a σ-álgebra de todos os subconjuntos de Ω e P qualquer probabilidade sobre I que coloca massa positiva em cada ponto de I. Defina um operador T por meio de

$$Tf(i) = \sum_{j \in I} p_{ij} f(j).$$

Esse operador satisfaz todas as hipóteses do teorema ergódico. Escolhamos f como segue: $f = I_{\{j_0\}}$, ou seja, o indicador do estado j_0. Notemos que

$$Tf(i) = \sum_j p_{ij} f(j) = p_{ij_0},$$

$$T^2 f(i) = T(Tf(i)) = \sum_j p_{ij}(Tf(j)) = \sum_j p_{ij} p_{ij_0} = p_{ij_0}^{(2)},$$

e de modo similar, $T^n f(i) = p_{ij_0}^{(n)}$. Logo, pelo teorema ergódico, $(f + Tf + \ldots + T^n f)/(n+1)$ converge. Ou seja, isso implica que $(p_{ij_0} + p_{ij_0}^{(2)} + \ldots + p_{ij_0}^{(n)})/(n+1)$ converge.

Antes de provar o TEM, vamos prová-lo para o caso de T em L_2.

Teorema 11.11. (TEM para o caso L_2) Seja T uma contração linear positiva em L_2.
(a) Se $f \in L_2$, então $R_n(f)$ converge em norma L_2 para um limite, denotado $\hat{P}f$.

(b) O operador $\hat{P}$ definido em (a) é linear, positivo e uma contração, com $T\hat{P} = \hat{P}$ e $\hat{P}^2 = \hat{P}$.

Prova: (a) Mostraremos que $R_n(f)$ é uma sequência de Cauchy em L_2. Defina

$$\mu_N = \inf_{\sum r_i = 1, r_i \geq 0} \|r_0 f + \ldots + r_N T^N f\|_2,$$

e defina $\mu = \inf_N \mu_N$. Vamos provar, primeiramnete, que $\|R_n(f)\|_2 \to \mu$.

Tome $g = r_0 f + r_1 Tf + \ldots + r_N T^N f$, tal que $\|g\|_2 \leq \mu + \varepsilon$. Temos que

$$R_n(g) = \frac{g + Tg + \ldots + T^n g}{n+1}$$

$$= \frac{r_0(f + Tf + \ldots + T^n f) + \ldots + r_N(T^N f + \ldots + T^{N+n} f)}{n+1}.$$

Portanto,

$$\|R_n(g) - R_n(f)\|_2 = \left\| R_n(g) - \frac{f + Tf + \ldots + T^n f}{n+1} \right\|_2$$

$$= \frac{1}{n+1} \Big\| \sum_{i=0}^{N} r_i \Big(\sum_{j=i}^{i+n} T^j - \sum_{j=0}^{n} T^j \Big) f \Big\|_2 \leq \frac{2N\|f\|_2}{n+1},$$

pois $\|T^k f\|_2 \leq \|f\|_2$.

Segue que

$$\mu \leq \|R_n(f)\|_2 \leq \|R_n(f) - R_n(g)\|_2 + \|R_n(g)\|_2$$

$$\leq \frac{2N\|f\|_2}{n+1} + \|R_n(g)\|_2 \leq \frac{2N\|f\|_2}{n+1} + \mu + \varepsilon.$$

Como N é fixo para ε escolhido, faça $n \to \infty$ para obter

$$\mu \leq \lim_{n \to \infty} \|R_n(f)\|_2 \leq \mu + \varepsilon,$$

Isso prova a afirmação feita acima.

Para provar a parte (a) do teorema, observe que pela igualdade do paralelograma

$$\|R_n(f) - R_m(f)\|_2^2 + \|R_n(f) + R_m(f)\|_2^2 = 2\|R_n(f)\|_2^2 + 2\|R_m(f)\|_2^2,$$

e portanto,

$$\|R_n(f) - R_m(f)\|_2^2 \leq 2\left(\|R_n(f)\|_2^2 - \mu^2\right) + 2\left(\|R_m(f)\|_2^2 - \mu^2\right), \tag{11.3}$$

pois, por definição de μ

$$\left\|\frac{R_n(f)+R_m(f)}{2}\right\|_2 \geq \mu.$$

Usando (11.3) e o fato que $\lim_{n\to\infty}\|R_n(f)\|_2 = \mu$, provamos que a sequência $\{R_n(f), n \geq 1\}$ é de Cauchy e portanto converge para algum elemento de L_2.

(b) $\hat{P}$ é uma contração linear positiva, pois R_n é, para cada n. Provaremos somente que $T\hat{P} = \hat{P}$, ou seja, para cada $f \in L_2$, mostramos que $\hat{P}f = T(\hat{P}f)$. Agora, $R_n(f) \to \hat{P}f$, em L_2 e $TR_n(f) = (Tf + T^2 f + \ldots + T^{n+1}f)/(n+1)$ converge para $\hat{P}f$, pois $f/n \to 0$. Contudo, $TR_n(f) \to T(\hat{P}f)$, pois

$$\|T(R_n f) - T(\hat{P}f)\|_2 = \|T(R_n f - \hat{P}f)\|_2 \leq \|R_n f - \hat{P}f\|_2,$$

dado que T é uma contração, e o último termo tende a zero, pela parte (a). Segue que $TR_n(f) \to \hat{P}f$ e $TR_n(f) \to T(\hat{P}f)$, logo $T\hat{P} = \hat{P}$. $\square$

Teorema 11.12. (TEM para o caso L_1) *Seja T uma contração linear positiva em L_1 satisfazendo*

$$\text{se } |f| \leq C, \ \ \text{então } \ |Tf| \leq C. \tag{11.4}$$

Então, $R_n(f)$ converge em norma L_1 para $\hat{P}f$, sempre que $f \in L_1$. $\hat{P}$ é a extensão a L_1 do operador obtido no Teorema 11.11.

Para provarmos esse teorema, precisamos do seguinte

Lema 11.1. Suponha que T seja uma contração em L_1, satisfazendo (11.4). Então, T é uma contração em L_2.

Prova: Seja $g \in L_2$. Mostremos, inicialmente, que se c for uma constante, $(Tg - c)_+ \leq T(g-c)_+$. Defina g_c como segue:

$$g_c = \begin{cases} g, & |g| \leq c, \\ c, & g > c, \\ -c, & g < -c. \end{cases}$$

Seja $h_c = g - g_c$. Note que $h_c \leq (g-c)_+$. Temos, então,

$$Tg = Tg_c + Th_c \leq c + T(g-c)_+,$$

usando (11.4). Logo $Tg - c \leq T(g-c)_+$, de modo que $(Tg-c)_+ \leq T(g-c)_+$.

Usando esse resultado, temos que

$$E((Tg-c)_+) \leq E(T(g-c)_+) \leq E((g-c)_+),$$

pois T é uma contração. Integrando a desigualdade em c, temos que

$$\int_0^\infty E(Tg-c)_+dc \leq \int_0^\infty E(g-c)_+dc.$$

Pelo teorema de Fubini, o lado direito fica

$$\int_\Omega \int_0^\infty (g-c)_+dcdP = \int_\Omega \int_{\{0\leq c\leq g\}} (g-c)dcdP = \frac{1}{2}\int_\Omega g^2 dP.$$

Por sua vez, novamente usando Fubini, o lado esquerdo fica

$$\int_\Omega \int_0^\infty (Tg-c)_+dcdP = \int_\Omega \int_{\{0\leq c\leq Tg\}} (Tg-c)dcdP = \frac{1}{2}\int_\Omega (Tg)^2 dP.$$

Portanto,

$$\int_\Omega (Tg)^2 dP \leq \int_\Omega g^2 dP,$$

ou seja, $\|Tg\|_2 \leq \|g\|_2$, $g \in L_2$, logo T é uma contração em L_2. □

Prova do Teorema 11.12: Seja $\varepsilon > 0$ e tome $f_\varepsilon \in L_2$ tal que $\|f - f_\varepsilon\|_1 \leq \varepsilon$, o que é possível pois L_2 é denso em L_1. Então,

$$\|R_n(f)-R_m(f)\|_1 \leq \|R_n(f)-R_n(f_\varepsilon)\|_1+\|R_n(f_\varepsilon)-R_m(f_\varepsilon)\|_1+\|R_m(f_\varepsilon)-R_m(f)\|_1$$

$$= \|R_n(f-f_\varepsilon)\|_1 + \|R_n(f_\varepsilon)-R_m(f_\varepsilon)\|_1 + \|R_m(f-f_\varepsilon)\|_1$$
$$\leq 2\|f-f_\varepsilon\|_1 + \|R_n(f_\varepsilon)-R_m(f_\varepsilon)\|_1 \leq 2\varepsilon,$$

fazendo $n, m \to \infty$ e usando o Teorema 11.11. (a). Observe que R_n, visto como um operador, é uma contração em L_1, pois T também o é. Como ε é arbitrário, $\{R_n(f)\}$ é uma sequência de Cauchy em L_1, logo converge em norma L_1 a um limite, $\hat{P}f$, digamos. Como $\hat{P}$ tem as mesmas propriedades dadas no Teorema 11.11. (b), temos em particular que $T\hat{P} = \hat{P}$ e $\hat{P}$ é a única extensão a L_1 do operador obtido no Teorema 11.11. □

Corolário 11.1 *Seja $(\Omega, \mathcal{F}, P)$ um e.p e S uma transformação que preserva a medida. Seja T o operador usual associado a S: se $f \in L_1$, $Tf(\omega) = f(S\omega)$. Então,*

$$\frac{f + f(S\omega) + \ldots + f(S^n\omega))}{n+1} \xrightarrow{L_1} E(f|\mathcal{I}),$$

onde $\mathcal{I}$ é a σ-álgebra invariante para S.

Prova: Sabemos que

$$\frac{f + f(S\omega) + \ldots + f(S^n\omega)}{n+1} = \frac{f + Tf + \ldots + T^n f}{n+1} \to \hat{P}f,$$

a convergência sendo em L_1, logo para todo conjunto $A \in \mathcal{F}$, teremos

$$\int_A \frac{f + Tf + \ldots + T^n f}{n+1} \to \int_A \hat{P} f.$$

Seja $A \in \mathcal{I}$. Então,

$$\int_A T^n f = \int_A f(S^n \omega) = \int_{S^{-n}A} f = \int_A f.$$

Logo temos

$$\int_A f = \int_A \hat{P} f. \tag{11.5}$$

Também, sabemos que $T(\hat{P}f) = \hat{P}(f)$, o que significa $\hat{P}f(S\omega) = \hat{P}f(\omega)$, ou seja, $\hat{P}f$ é uma função invariante. Por um resultado anterior, $\hat{P}f$ é $\mathcal{I}$-mensurável. Usando este fato e (11.5), temos que $\hat{P}f = E(f|\mathcal{I})$. $\square$

Corolário 11.2. Suponha que $\{X_n, n \geq 0\}$ seja um processo estritamente estacionário. Se X_0 é integrável, então

$$\frac{X_0 + \ldots + X_n}{n+1} \xrightarrow{L_1} E(X_0|\mathcal{I}),$$

onde $\mathcal{I}$ é a σ-álgebra invariante de $\{X_n, n \geq 0\}$.

Prova: Use o corolário anterior com a translação S apropriada. $\square$

Teorema 11.13. (Teorema ergódico maximal de Hopf) *Seja T uma contração linear positiva em L_1. Então, se $f \in L_1$, $\int_{E_n} f dP \geq 0$, onde $E_n = \{\max_{0 \leq k \leq n} R_k(f) > 0\}$.*

Prova: Seguimos Garsia (1970) para a prova do teorema. Primeiramente, notemos que se $f_1, \ldots, f_n \in L_1$, então

$$\max_{1 \leq k \leq n} T f_k \leq T\Big(\max_{1 \leq k \leq n} f_k\Big).$$

De fato, para todo $1 \leq k \leq n$, $f_k \leq \max_{1 \leq k \leq n} f_k$, logo $Tf_k \leq T(\max_{1 \leq k \leq n} f_k)$, pois T é positiva. Logo, $\max_{1 \leq k \leq n}(Tf_k) \leq T(\max_{1 \leq k \leq n} f_k)$. Observe também que

$$f + \max_{1 \leq k \leq n} (Tf + \ldots + T^{k+1} f)_+ \geq \max_{0 \leq k \leq n} (f + \ldots + T^k f)_+,$$

sobre o conjunto E_n. Usando esses fatos, obtemos

$$\int_{E_n} f dP \geq \int_{E_n} \max_{0 \leq k \leq n} (f + \ldots + T^k f)_+ dP - \int_{E_n} \max_{1 \leq k \leq n} (Tf + \ldots + T^{k+1} f)_+ dP$$

$$\geq \int_{E_n} \max_{0\leq k\leq n}(f+\ldots+T^k f)_+ dP - \int_{E_n} T(\max_{0\leq k\leq n}(f+\ldots+T^k f)_+)dP.$$

Seja $\varphi = \max_{0\leq k\leq n}(f+Tf+\ldots+T^k f)_+$. Temos, então,

$$\int_{E_n}(\varphi - T\varphi)dP = \int_{\{\varphi>0\}}(\varphi - T\varphi)dP \geq \int_{\Omega}(\varphi - T\varphi)dP.$$

De fato,

$$\int_{\Omega}(\varphi - T\varphi)dP = \int_{\{\varphi>0\}}(\varphi - T\varphi)dP + \int_{\{\varphi=0\}}(\varphi - T\varphi)dP$$

$$= \int_{\{\varphi>0\}}(\varphi - T\varphi)dP - \int_{\{\varphi=0\}} T\varphi dP.$$

Como $\varphi \geq 0$, $T\varphi \geq 0$ (T é positiva), o último termo da igualdade acima é negativo. Logo,

$$\int_{E_n} f dP \geq \int_{\Omega}(\varphi - T\varphi)dP \geq 0,$$

pois T é uma contração, ou seja $\int_\Omega T\varphi \leq \int_\Omega \varphi$. $\square$

Corolário 11.3. *Seja T uma contração linear positiva e suponha que $T1 = 1$. Se $R_n(f) = (f+Tf+\ldots+T^n f)/(n+1)$, teremos que*

$$\lambda P\Big\{\max_{0\leq k\leq n} R_k(f) > \lambda\Big\} \leq \int_{\max_{0\leq k\leq n} R_k(f)>\lambda} f dP.$$

Prova: Observe que

$$\Big\{\omega : \max_{0\leq k\leq n}(f+Tf+\ldots+T^k f) > 0\Big\} = \Big\{\omega : \max_{0\leq k\leq n} R_k(f) > 0\Big\}.$$

Também note que, como $T1 = 1$, teremos

$$R_n(f-\lambda) = R_n(f) - \lambda, \tag{11.6}$$

se λ for uma constante. Logo, chamando

$$F_n = \Big\{\max_{0\leq k\leq n}\Big((f-\lambda)+\ldots+T^k(f-\lambda)\Big) > 0\Big\},$$

teremos pelo teorema anterior

$$0 \leq \int_{F_n}(f-\lambda)dP = \int_{\{\max_{0\leq k\leq n} R_k(f-\lambda)>0\}}(f-\lambda)dP$$

$$= \int_{\{\max_{0\leq k\leq n} R_k(f)>\lambda\}} (f-\lambda)dP,$$

por (11.6). Ou seja, temos

$$0 \leq \int_{\{\max_{0\leq k\leq n} R_k(f)>\lambda\}} fdP - \lambda P\Big\{\max_{0\leq k\leq n} R_k(f) > \lambda\Big\}. \quad \square$$

Observação: Seja $\{X_n, n \geq 1\}$ um submartingale. Sabemos que

$$\lambda P\Big\{\max_{1\leq k\leq n} X_k > \lambda\Big\} \leq \int_{\max_{1\leq k\leq n} X_k>\lambda} X_n dP.$$

Essa é uma desigualdade do mesmo tipo daquela do corolário.

Teorema 11.14. *Sejam $X \geq 0$, $Y \geq 0$ variáveis aleatórias. Suponha que $\lambda P\{Y > \lambda\} \leq \int_{\{Y>\lambda\}} XdP$. Se $p > 1$, então,*

$$E(Y^p) \leq \left(\frac{p}{p-1}\right)^p E(X^p).$$

Prova: Basta considerar o caso $X \in L_p$. Para começar, supomos também que $Y \in L^p$. Então,

$$\lambda P\{Y > \lambda\} \leq \int_{\{Y>\lambda\}} XdP,$$

multiplicando por λ^{p-2},

$$\lambda^{p-1} P\{Y > \lambda\} \leq \lambda^{p-2} \int_{\{Y>\lambda\}} XdP,$$

e integrando em λ obtemos

$$\int_0^\infty \lambda^{p-1} P\{Y > \lambda\} d\lambda \leq \int_0^\infty \lambda^{p-2}\Big(\int_{\{Y>\lambda\}} XdP\Big)d\lambda.$$

Usando Fubini, o lado direito fica

$$\int X\Big(\int_{\{0\leq\lambda<Y\}} \lambda^{p-2}d\lambda\Big)dP \leq \frac{1}{p-1}\int XY^{p-1}dP$$
$$\leq \frac{1}{p-1}\left(\int X^p dP\right)^{1/p}\left(\int Y^p dP\right)^{(p-1)/p},$$

por Hölder. Por outro lado, o lado esquerdo é igual a

$$\int\Big(\int_{\{0\leq\lambda<Y\}} \lambda^{p-1}d\lambda\Big)dP = \frac{1}{p}\int Y^p dP.$$

Logo,

$$\frac{1}{p}\int Y^p dP \le \frac{1}{p-1}\left(\int X^p dP\right)^{1/p}\left(\int Y^p dP\right)^{(p-1)/p},$$

do que segue

$$\int Y^p dP \le \left(\frac{p}{p-1}\right)^p \int X^p dP.$$

Se $Y \notin L_p$, substitua Y por Y_r, ond $Y_r = Y$, se $Y \le r$ e $Y_r = r$, se $Y > r$. Logo, como Y_r satisfaz as mesmas desigualdades do que Y, obtemos

$$E(Y_r^p) \le \left(\frac{p}{p-1}\right)^p E(X^p).$$

Basta fazer $r \to \infty$. $\square$

Aplicação: Seja $\{X_n, n \ge 1\}$ um martingale, que seja limitado em L_p, $p > 1$, isto é, $\sup_n E(|X_n|^p) < \infty$. Sabemos que vale a desigualdade da observação feita acima, logo pelo teorema, se $p > 1$,

$$E\Big(\sup_{1\le k\le n} |X_k|^p\Big) \le \left(\frac{p}{p-1}\right)^p E(|X_n|^p).$$

Para $n \to \infty$, obtemos

$$E\Big(\sup_{n\ge 1} |X_n|^p\Big) \le \left(\frac{p}{p-1}\right)^p \sup_{n\ge 1} E(|X_n|^p) < \infty.$$

Segue que o martingale $\{X_n\}$ converge em norma L_p (e também q.c), pois $|X_n|^p \le \sup_n |X_n|^p$, que é integrável, logo $\{|X_n|^p\}$ é uniformemente integrável.

Corolário 11.4. *Seja T uma contração linear positiva em L_1, com $T1 = 1$ e seja $R^*(f) = \sup_{n\ge 0} |R_n(f)|$. Então, se $p > 1$,*

$$E\left(R^*(f)^p\right) \le E\left(R^*(|f|)^p\right) \le \left(\frac{p}{p-1}\right)^p E(|f|^p).$$

Prova: Sabemos que para $\lambda > 0$,

$$\lambda P\{\max_{0\le k\le n} R_k(|f|) > \lambda\} \le \int_{\max_{0\le k\le n} R_k(|f|)>\lambda} |f| dP,$$

logo pelo teorema anterior

$$E\Big(\max_{0\le k\le n} R_k(|f|)^p\Big) \le \left(\frac{p}{p-1}\right)^p E(|f|^p).$$

Logo, para $n \to \infty$ e pelo TCM,

$$E(R^*(|f|)^p) \leq \left(\frac{p}{p-1}\right)^p E(|f|^p). \quad \square$$

Teorema 11.15. (TEP - forma de Hopf) *Seja T uma contração linear positiva em L_1, com $T1 = 1$. Então, $R_n(f) \to \hat{P}f$ q.c, onde $\hat{P}f$ é definida pelo TEM.*

Prova: Suponha, inicialmente, que $f \in L_2$ e sejam $g(f) := \limsup_{n\to\infty} R_n(f),\ h(f) := \liminf_{n\to\infty} R_n(f)$. Então, usando $T\hat{P} = \hat{P}$, é fácil verificar que para todo $k \geq 0$

$$g\left(\frac{f + Tf + \ldots + T^k f}{k+1} - \hat{P}f\right) = g(f) - \hat{P}f,$$

$$h\left(\frac{f + Tf + \ldots + T^k f}{k+1} - \hat{P}f\right) = h(f) - \hat{P}f.$$

Considere

$$E\left(|g(f) - \hat{P}f|^2\right) = E\left(\left|g\left(\frac{f + Tf + \ldots + T^k f}{k+1} - \hat{P}f\right)\right|^2\right)$$

$$\leq E\left(\left|R^*\left(\frac{f + Tf + \ldots + T^k f}{k+1} - \hat{P}f\right)\right|^2\right) \leq 4E\left(\left|\frac{f + Tf + \ldots + T^k f}{k+1} - \hat{P}f\right|^2\right),$$

pelo Corolário 11.4, com $p = 2$. Isso é verdade para todo k. Para $k \to \infty$, o lado direito da última desigualdade converge para zero, pelo TEM, logo $g(f) = \hat{P}f$, q.c, isto é, $\limsup_n R_n(f) = \hat{P}f$ q.c. Uma prova similar resulta em $\liminf_n R_n(f) = \hat{P}f$ q.c. Logo, $R_n(f)$ converge para $\hat{P}f$ q.c.

Para o caso geral, suponha $f \in L_1$. Seja $\varepsilon > 0$ e tome $f_\varepsilon \in L_2$ tal que $\|f - f_\varepsilon\|_1 \leq \varepsilon^2$. Considere

$$\frac{f + Tf + \ldots + T^n f}{n+1} - \hat{P}f = \frac{(f - f_\varepsilon) + \ldots + T^n(f - f_\varepsilon)}{n+1} - \hat{P}(f - f_\varepsilon)$$

$$-\hat{P}f_\varepsilon + \frac{f_\varepsilon + \ldots + T^n f_\varepsilon}{n+1}.$$

Então,

$$\limsup_n \left|\frac{f + Tf + \ldots + T^n f}{n+1} - \hat{P}f\right| =$$

$$\limsup_n \left|\frac{(f - f_\varepsilon) + \ldots + T^n(f - f_\varepsilon)}{n+1} - \hat{P}(f - f_\varepsilon)\right|$$

$$\leq R^*(f - f_\varepsilon) + \hat{P}(|f - f_\varepsilon|),$$

pelo caso L_2. Segue que

$$P\left\{\limsup_n \left|\frac{f + Tf + \ldots + T^n f}{n+1} - \hat{P}f\right| > \varepsilon\right\} \leq P\{R^*(f - f_\varepsilon) + \hat{P}(|f - f_\varepsilon|) > \varepsilon\}$$

$$\leq P\{R^*(f - f_\varepsilon) > \varepsilon/2\} + P\{\hat{P}(|f - f_\varepsilon|) > \varepsilon/2\}$$

$$\leq \frac{2}{\varepsilon}\int |f - f_\varepsilon| dP + \frac{2}{\varepsilon} E\hat{P}(|f - f_\varepsilon|),$$

usando o Corolário 11.3 e a desigualdade de Markov. Finalmente, o último termo é

$$\leq \frac{2}{\varepsilon}\|f - f_\varepsilon\|_1 + \frac{2}{\varepsilon}\|f - f_\varepsilon\|_1 \leq \frac{2}{\varepsilon}\varepsilon^2 + \frac{2}{\varepsilon}\varepsilon^2 \leq 4\varepsilon,$$

Portanto,

$$\limsup_n \left|\frac{f + Tf + \ldots + T^n f}{n+1} - \hat{P}f\right| = 0, \text{ q.c.} \quad \square$$

Corolário 11.5. (Versão final do TEM) *Seja $f \in L_p$, $p \geq 1$ e T uma contração linear positiva em L_1, com $T1 = 1$. Então, $R_n(f) \to \hat{P}f$ em norma L_p.*

Prova: Sabemos que o corolário vale para $p = 1$. Se $f \in L_p$, $p > 1$, então $f \in L_1$, pois $(\Omega, \mathcal{F}, P)$ é um e.p. Pelo TEP, $R_n(f) \to \hat{P}f$ q.c. Contudo, pelo Corolário 11.4, temos que $E(\sup_n |R_n(f)|^p) < \infty$, logo a família $\{|R_n(f)|^p, n \geq 1\}$ é uniformemente integrável. Portanto, $R_n(f) \to \hat{P}f$ em L_p. $\square$

11.4 Recíprocas dos teoremas ergódicos

Sabemos que, se T for uma transformação que preserva a medida sobre um e.p $(\Omega, \mathcal{F}, P)$ e se T for ergódica, então se $f \in L_1$, teremos

$$\frac{f + Tf + \ldots + T^n f}{n+1} \xrightarrow{\text{q.c}} \text{constante.} \tag{11.7}$$

Teorema 11.16. (Recíproca) *Suponha que T seja uma transformação preservando a medida e, para toda $f \in L_1$, tenhamos (11.7). Então, T é ergódica.*

Prova: Seja $A \in \mathcal{I}$ e $f = I_A$. Sabemos que

$$\frac{f + Tf + \ldots + T^n f}{n+1} \xrightarrow{\text{q.c}} E(f|\mathcal{I}).$$

Nesse caso, $E(f|\mathcal{I}) = I_A$. Por hipótese, todos os limites são constantes q.c, portanto I_A é constante q.c, logo $P(A) = 0$ ou $P(A) = 1$, ou seja S é ergódica. $\square$

Teorema 11.17. *Seja $(\Omega, \mathcal{F}, P)$ um e.p e T uma transformação que preserva a medida e ergódica. Suponha que $f \geq 0$ e $\lim_{n\to\infty}(f + Tf + \ldots + T^n f)/(n+1)$ exista e seja finito q.c. Então f é integrável.*

Prova: Se $\lim_{n\to\infty}(f + Tf + \ldots + T^n f)/(n+1)$ converge q.c para $\hat{f}$, então

$$\frac{Tf + T^2 f + \ldots + T^{n+1} f}{n+1} \to T\hat{f} = \hat{f},$$

de modo que o limite $\hat{f}$ é invariante. Como T é ergódica, $\hat{f}$ é constante, digamos $\hat{f} = d$. Defina $f_r = f$, se $f \leq r$ e $f_r = r$, se $f > r$. Então, $f \in L_1$ e portanto

$$\frac{f_r + Tf_r + \ldots + T^n f_r}{n+1} \to E(f_r|\mathcal{I}) = E(f_r),$$

pois T é ergódica. Além disso, como $f_r \leq f$, $Tf_r \leq Tf$, etc, de modo que

$$\lim_n \frac{f_r + Tf_r + \ldots + T^n f_r}{n+1} \leq \lim_n \frac{f + Tf + \ldots + T^n f}{n+1} = d,$$

ou seja, $E(f_r) \leq d < \infty$. Mas $f_r \uparrow f$, logo $E(f) = \lim_{r\uparrow\infty} E(f_r) \leq d$, pelo TCM, ou seja $E(f) < \infty$. $\square$

Observação: Lembremos a LFGN: se $X_i, i \geq 1$ são i.id, se $E(X_1)$ existe, então $(X_1 + \ldots + X_n)/n$ converge q.c. Também provamos que, se $X_i, i \geq 1$ são i.i.d e se $(X_1 + \ldots + X_n)/n$ converge para um limite finito, então $E(X_1)$ existe.

Problemas

1. Prove que a aplicação T da definição 11.2 é mensurável.

2. Prove a afirmação do Exemplo 11.2 (a).

3. Seja Ω o conjunto de todas as sequências da forma $(\cdots, \omega_{-1}, \omega_0, \omega_1, \cdots)$, com ω_i real, e $\mathcal{F}$ a menor σ-álgebra contendo todos os conjuntos da forma $\{\omega : (\omega_k, \omega_{k+1}, \ldots, \omega_{k+n-1}) \in B_n\}$, onde B_n é um conjunto de Borel do espaço Euclidiano n-dimensional, $k = 0, \pm 1, \pm 2, \ldots$. Uma translação bilateral T é definida por $T(\cdots, \omega_{-1}, \omega_0, \omega_1, \cdots) = (\cdots, \omega_0, \omega_1, \cdots)$, ou seja, se $\xi_1, \xi_2, \ldots$ são as funções coordenadas, temos $\xi_k(T\omega) = \xi_{k+1}(\omega)$. Mostre que T preserva a medida se, e somente se P é estacionária. A definição de estacionária aqui é análoga à dada no texto, exceto que agora temos sequência bilateral.

4. Prove a afirmação do Exemplo 11.5 (a).

5. Encontre uma translação T que seja ergódica mas não de Kolmogorov (de fato, T pode ser escolhida como uma translação de Markov).

6. Mostre que, se $Tx = x+1$, sobre a reta real, então qualquer intervalo de comprimento menor do que um é um conjunto *wandering* não trivial.

7. Para o Exemplo 11.8 (b), prove que T é linear, que T transforma funções de L_p em funções de L_p e portanto $\|f\|_p = \|Tf\|_p$.

8. Sejam $(\Omega, \mathcal{F}, P_1)$ e $(\Omega, \mathcal{F}, P_2)$ dois e.p com $P_1 \neq P_2$ e S uma transformação ergódica sobre $(\Omega, \mathcal{F}, P_1)$ e $(\Omega, \mathcal{F}, P_2)$. Mostre que $P_1 \perp P_2$.

9. Prove o Teorema 11.5.

10. Mostre que as transformações a seguir são mensuráveis e preservam a medida. Depois, decida se são ergódicas ou mixing.

    (a) Ω é o círculo unitário no plano complexo, $\mathcal{F}$ é a σ-álgebra de Borel no círculo e P é dada pela medida de Lebesgue no círculo dividida por 2π. Seja $Te^{i\theta} = e^{i(\theta+\alpha)}$, α irracional. O que acontece se α for racional?

    (b) $\Omega = [0,1]$, P é a medida de Lesbesgue, $\mathcal{F}$ a σ-álgebra de Borel e $T\omega = 2\omega$ (mod1). O que acontece se $T\omega = k\omega$ (mod1), k inteiro, $k > 2$?

11. Provamos que, se S é ergódica, e se $f \geq 0$, $\sum_{k=1}^{n} f(S^k\omega)/n$ converge q.c para um limite finito, então $f \in L_1$. Mostre que isso pode não ser verdade se: (a) $f \geq 0$ não valer; ou (b) se S não for ergódica.

12. Sejam $X = \{X_n, n \geq 0\}$ e $Y = \{Y_n, n \geq 0\}$ dois processos estacionários, ergódicos. Lance uma moeda independentemente de X e Y. Se ocorrer cara, observe o processo X e se ocorrer coroa, observe o processo Y.

    (a) O processo resultante é estritamente estacionário?

    (b) O processo resultante é ergódico?

13. Seja $X = \{X_n, n \geq 0\}$ um processo tal que $(X_1, \ldots, X_n)$ seja normal, para cada n, $E(X_i) = 0$, $\text{Cov}(X_i, X_j) = R(i,j)$:

    (a) Prove que X é estacionário se e somente se $R(i,j)$ depende somente de $|i-j|$;

    (b) Suponha que $R(i,j) = r(|i-j|)$. Prove que $\lim_n r(n) = 0$ implica que X seja ergódico.

Capítulo 12

Introdução ao Cálculo Estocástico

O objetivo deste capítulo é apresentar uma construção da integral estocástica de Itô,

$$I(f)(\omega) = \int_0^T f(\omega, t) dW(\omega, t),$$

na qual W é o Movimento Browniano (MB) e f é uma função satisfazendo certos critérios. Como as trajetórias de W não tem variação finita em intervalos compactos, não é possivel usar a integral de Riemann-Stieltjes para integrar com respeito ao MB. Em vez disto vamos usar técnicas de extensão de isometrias em espaços de Hilbert para definir a integral estocástica. Veremos em seguida a fórmua de Itô, o equivalente do teorema fundamental do cálculo para o cálculo estocástico. Finalmente, provaremos a fórmula de Girsanov e daremos alguns exemplos de aplicação deste resultado. Indicamos como referências para este capítulo os livros de Comets e Meyre (2015) e Le Gall (2016).

12.1 Integral estocástica

Nesta seção, vamos definir a integral estocástica em várias etapas, começando com a definição da integral para a classe das funções em escada até chegar na classe de funções H_2^{loc} (veja a Definição 12.5).

Iniciamos com o cálculo da variação quadrática do MB W nos intervalos da forma $[0, T]$. Esta propriedade será fundamental na construção da integral estocástica.

Proposição 12.1 (Variação quadrática do MB). *Seja $T \geq 0$, $t_0 = 0 < t_1 < \cdots < t_n = T$ uma partição do intervalo $[0, T]$ e $\delta = \max_i\{t_i - t_{i-1}\}$. Temos que*

$$\lim_{\delta \to 0} \Big\| \sum_{i=1}^n (W(t_i) - W(t_{i-1}))^2 - T \Big\|_2 = 0.$$

Prova: Lembramos para começar que se $X \sim N(0, \sigma^2)$, $E(X^2) = \sigma^2$, $Var(X^2) = 2\sigma^4$. Aplicando estes resultados aos termos da soma

$$V := \sum_{i=1}^{n} (W(t_i) - W(t_{i-1}))^2,$$

obtemos

$$E(V) = T, \ \ Var(V) = 2\sum_{i=1}^{n} (t_i - t_{i-1})^2 \leq 2T\delta.$$

Deduzimos que $\|V - T\|_2 = Var(V) \to 0$, quando $\delta \to 0$. $\square$

Seja $(\Omega, \mathcal{F}, P)$ um e.p e $\{\mathcal{F}_t, t \geq 0\}$ uma sequência não decrescente de σ-álgebras tais que $\mathcal{F}_t \subset \mathcal{F}$ para todo $t \geq 0$. A sequência $\{\mathcal{F}_t, t \geq 0\}$ é também chamada de *filtração* sobre $(\Omega, \mathcal{F}, P)$.

Definição 12.1. *Um MB W definido sobre $(\Omega, \mathcal{F}, P)$ é um $\{\mathcal{F}_t\}$-MB se ele é adaptado a $\{\mathcal{F}_t, t \geq 0\}$ e se $W(t) - W(s)$ é independente de $\mathcal{F}_s$, para todo $s \geq 0$ e $t \geq s$.*

Neste capítulo sempre iremos supor que a filtração $\{\mathcal{F}_t, t \geq 0\}$ é completa, i.e, $\mathcal{F}_0$ contém todos os conjuntos nulos de $\mathcal{F}$.

Definição 12.2. *Uma função $\phi : \mathbb{R}_+ \times \Omega \to \mathbb{R}$ é progressivamente mensurável se para todo $t \in \mathbb{R}_+$, a aplicação $(\omega, t) \mapsto f(\omega, t)$ de $[0, t] \times \Omega \to \mathbb{R}$ é $\mathcal{B}[0, t] \otimes \mathcal{F}_t$-mensurável.*

Proposição 12.2. *Uma função ϕ de $\mathbb{R}_+ \times \Omega \to \mathbb{R}$ adaptada e contínua é progressivamente mensurável.*

Prova: Para todo $n \geq 1$, definimos

$$\phi_n(s, \omega) = \phi(kT/n, \omega), \ \ \text{para} \ \ \frac{kT}{n} < s \leq \frac{(k+1)T}{n}.$$

É fácil ver que ϕ_n é progressivamente mensurável, e, por continuidade, ϕ_n converge para ϕ em todo ponto $(s, \omega) \in [0, T] \times \Omega$ quando $n \to \infty$. Isto mostra que ϕ é progressiamente mensurável. $\square$

Definição 12.3. *Denotamos por $H_2(\mathbb{R}_+)$ (resp. $H_2([0, T])$, $T > 0$), o espaço das funções ϕ progressivamente mensuráveis tais que*

$$E\Big(\int_{\mathbb{R}_+} \phi^2(t, \omega) dt\Big) < \infty \quad (\textit{resp. } E\Big(\int_{[0,T]} \phi^2(t, \omega) dt\Big) < \infty).$$

Também introduzimos $H_2 = \bigcap_{T>0} H_2([0, T])$.

Identificando duas funções iguais fora de um conjunto de $P \otimes \lambda_+$-medida nula, obtemos que $H_2(\mathbb{R}_+)$ (resp. $H_2([0,T])$) é um espaços de Hilbert com norma

$$\|\phi\|_{H_2(\mathbb{R}_+)} := E\Big(\int_{\mathbb{R}_+} \phi^2(t,\omega)dt\Big)$$

$$(\text{resp. } \|\phi\|_{H_2([0,T])} := E\Big(\textstyle\int_{[0,T]} \phi^2(t,\omega)dt\Big)).$$

Integral de funções em escada

Uma *função em escada* é uma função real da forma

$$\phi(t,\omega) = \sum_{k=0}^{n-1} X_i(\omega)\mathbf{1}_{(t_i,t_{i+1}]}(t)$$

com $0 = t_0 < t_1 < t_2 < \cdots < t_n$ e $X_i \in L_2(\Omega, \mathcal{F}_{t_i}, P)$. Uma função deste tipo é progressivamente mensurável. Observamos que as funções em escada pertencem a $H_2(\mathbb{R}_+)$ pois

$$E\Big(\int_{\mathbb{R}_+} \phi(t)^2 dt\Big) = \sum_{i=0}^{n-1} E(X_i^2)(t_{i+1} - t_i) < \infty.$$

Para estas funções, definimos a integral de Itô com respeito a um $\{\mathcal{F}_t\}$-MB W por

$$\int_{\mathbb{R}_+} \phi\, dW = \sum_{i=0}^{n-1} X_i(\omega)(W(t_{i+1},\omega) - W(t_i,\omega)).$$

A aplicação $\phi \mapsto \int_{\mathbb{R}_+} \phi\, dW$ é linear sobre o espaço vetorial $\mathcal{E}$ das funções em escada e tem valores em $L_2(\Omega, \mathcal{F}, P)$ pois

$$\begin{aligned} E\Big[\Big(\int_{\mathbb{R}_+} \phi\, dW\Big)^2\Big] &= \sum_{i=0}^{n-1} E(X_i^2 E([W(t_{i+1}) - W(t_i)]^2 \mid \mathcal{F}_{t_i})) \\ &\quad + 2\sum_{0\le i<j<n} E(X_i X_j [W(t_{i+1}) - W(t_i)] E([W(t_{j+1}) - W(t_j)] \mid \mathcal{F}_{t_j})) \\ &= E\Big(\int_{\mathbb{R}_+} \phi(t)^2 dt\Big) = \|\phi\|^2_{H_2(\mathbb{R}_+)}. \end{aligned}$$

Integral de funções em $H_2(\mathbb{R}_+)$

O espaço vetorial $\mathcal{E}$ das funções em escadas é denso em $H_2(\mathbb{R}_+)$ (veja Problema 1). Pelo teorema de extensão de isometrias (veja o Apêndice A.3), $\phi \in \mathcal{E} \mapsto \int_{\mathbb{R}_+} \phi\, dW \in L_2(\Omega, \mathcal{F}, P)$, pode ser estendida de maneira única à aderência de $\mathcal{E}$, ou seja $H_2(\mathbb{R}_+)$.

Definição 12.4. *A variável aleatória* $\int_{\mathbb{R}_+} \phi\, dW \in L_2(\Omega, \mathcal{F}, P)$ *é chamada integral estocástica de* ϕ.

Como consequência imediata da propriedade de isometria temos o

Teorema 12.1. *Para* ϕ *e* $\psi \in H_2(\mathbb{R}_+)$ *temos que*

$$E\Big(\int_{\mathbb{R}_+} \phi\, dW\Big) = 0, \quad E\Big[(\int_{\mathbb{R}_+} \phi\, dW)^2\Big] = E\Big(\int_{\mathbb{R}_+} \phi^2 dt\Big),$$
$$E\Big[(\int_{\mathbb{R}_+} \phi\, dW)(\int_{\mathbb{R}_+} \psi\, dW)\Big] = E\Big(\int_{\mathbb{R}_+} \phi\psi\, dt\Big).$$

Integral de funções em H_2

Vamos, agora, estender a integral estocástica ao espaço H_2. Quando $\phi \in H_2$, definimos para todo $t \geq 0$,

$$\int_0^t \phi\, dW = \int_{\mathbb{R}_+} \mathbf{1}_{(0,t]}\phi\, dW.$$

A integral acima está bem definida para todo $t \geq 0$, pois podemos observar que $\mathbf{1}_{(0,t]}\phi$ é progressivamente mensurável (como produto de funções progressivamente mensuráveis) e pertence a $H_2(\mathbb{R}_+)$. Definimos $\int_s^t \phi\, dW = \int_{\mathbb{R}_+} \mathbf{1}_{(s,t]}\phi\, dW$ e deduzimos imediatamente a relação de Chasles

$$\int_0^t \phi\, dW = \int_0^s \phi\, dW + \int_s^t \phi\, dW$$

por linearidade. A próxima proposição mostra que a integral estocástica é uma função contínua do seu limite superior.

Proposição 12.3. *A aplicação de* $\mathbb{R}_+ \to L_2$, $t \mapsto \int_0^t \phi\, dW$ *é contínua em todo ponto* $t_0 \geq 0$ *e ela possui uma versão contínua sobre* $\mathbb{R}_+$.

Prova: A continuidade em L_2 é uma consequência imediata da propriedade de isometria. Vamos mostrar que existe uma versão contínua desta aplicação. Podemos usar a sequência $(\Pi_n)_n$ do Problema 1. Neste caso , $M_n(t) := \int_0^t P_n\phi\, dW$ é contínua para quase todo ω. Como a filtração $\{\mathcal{F}_t, t \geq 0\}$ contém os conjuntos nulos de $\mathcal{F}$, podemos modificar M_n para que todas as suas trajetórias sejam contínuas para todo n. A desigualdade de Doob aplicada à martingale contínua $M_m - M_n$ (veja a Proposição 12.4 abaixo) com $T < \infty$ dá, para todo $\varepsilon > 0$,

$$P\Big(\max_{t\in[0,T]} |M_n(t) - M_m(t)| > \varepsilon\Big) \leq \varepsilon^{-2}\|P_n\phi - P_m\phi\|^2_{H_2([0,T])}.$$

Assim, em probabilidade, a sequência $(M_n)_n$ é de Cauchy no espaço das funções contínuas sobre $[0,T]$ com a norma da convergência uniforme. Podemos portanto encontrar uma subsequência $(M_{n_k})_k$ tal que, quase certamente, esta subsequência converge uniformemente sobre os compactos de $\mathbb{R}_+$. O limite M_∞ pode ser tomado contínuo como limite (quase certamente) uniforme de funções contínuas. Como $M_n(t) \to \int_0^t \phi\, dW$ em L_2, temos que $M_\infty(t) = \int_0^t \phi\, dW$ q.c., para todo $t \geq 0$, e portanto M_∞ é uma versão contínua da integral estocástica. $\square$

Proposição 12.4. *Para $\phi \in H_2$, o processo estocástico $M(t) = \int_0^t \phi\, dW$ é um martingale de quadrado integrável tal que*

$$M^2(t) - \int_0^t \phi^2 ds$$

é um martingale.

Observação: o processo $\{\int_0^t \phi^2 ds, t \geq 0\}$, é tradicionalmente chamado de "colchete"do martingale M e é denotado por $\langle M \rangle$.

Prova: Em primeiro lugar, vamos mostrar que M é um martingale com respeito à filtração $\{\mathcal{F}_t, t \geq 0\}$. Se $\phi \in \mathcal{E}$, obtemos para $s = t_0 < t_1 < \cdots < t_n = t$,

$$\begin{aligned} E\Big(\int_s^t \phi\, dW \mid \mathcal{F}_s\Big) &= \sum_{i=0}^{n-1} E(X_i[W(t_{i+1}) - W(t_i)] \mid \mathcal{F}_s) \\ &= \sum_{i=0}^{n-1} E(X_i E([W(t_{i+1}) - W(t_i)] \mid \mathcal{F}_{t_i}) \mid \mathcal{F}_s) \\ &= 0. \end{aligned}$$

O resultado geral para $\phi \in H_2$ é obtido por continuidade da esperança condicional em L_2 e a densidade de $\mathcal{E}$ em $H_2([0,T])$ para todo $T > 0$. A seguir, mostramos que $M(t)^2 - \int_0^t \phi(s)^2 ds$ é um martingale. Como $E(M(t)^2 - M(s)^2 \mid \mathcal{F}_s) = E((M(t) - M(s))^2 \mid \mathcal{F}_s)$, é suficiente verificar que

$$E\Big((\int_s^t \phi\, dW)^2 \mid \mathcal{F}_s\Big) = E\Big(\int_s^t \phi(u)^2 du \mid \mathcal{F}_s\Big). \tag{12.1}$$

Mas esta última igualdade é facilmente obtida por um argumento similar a aquele usado para provar que M é um martingale (veja Problema 2). $\square$

Localização

Para certas aplicações o espaço H_2 não é suficiente, assim é necessário estender a integral estocástica a um espaço maior.

Definição 12.5. *Denotamos por H_2^{loc} o conjunto das funções progressivamente mensuráveis $\phi : \mathbb{R}_+ \times \Omega \to \mathbb{R}$ tais que, para todo $T > 0$,*

$$P\Big(\int_0^T \phi(t,\omega)^2 dt < \infty\Big) = 1.$$

Observamos que $H_2 \subset H_2^{\text{loc}}$. Se ϕ é progressivamente mensurável, o tempo aleatório

$$\tau_n = \inf\Big\{t \geq 0 : \int_0^t \phi(s)^2 ds \geq n\Big\} \in [0,\infty]$$

é um tempo de parada da filtração $\{\mathcal{F}_t, t \geq 0\}$. Além disto se $\phi \in H_2^{\text{loc}}$, a sequência τ_n diverge q.c. para ∞ e $\mathbf{1}_{[0,\tau_n]}\phi \in H_2$ para todo n. Isto sugere a seguinte extensão da integral estocástica ao espaço H_2^{loc}.

Definição 12.6. *A integral estocástica para $\phi \in H_2^{\text{loc}}$ é definida para todo $t \geq 0$ pelo limite quase certo*

$$\int_0^t \phi \, dW = \lim_{n\to\infty} \int_0^t \mathbf{1}_{[0,\tau_n]}\phi \, dW.$$

A integral estocástica em H_2^{loc} não é, em geral, um martingale mas somente uma versão mais fraca, chamada martingale local (veja o Problema 3).

Definição 12.7 (Martingale local)**.** *Um processo estocástico $X = \{X(t), t \geq 0\}$, adaptado a uma filtração $\{\mathcal{F}_t, t \geq 0\}$, é um martingale local, se existe uma sequência de tempos de parada τ_n, tal que $\lim_{n\to\infty} \tau_n = \infty$, q.c., e tal que $\{X(t \wedge \tau_n), t \geq 0\}$ seja um martingale para todo n.*

Para concluir, podemos observar que a integral estocástica em H_2^{loc} permanece, quase certamente, com trajetórias contínuas. Com respeito à propriedade de isometria, ela é substituida pela desigualdade trivial

$$E\Big((\int_0^t \phi \, dW)^2\Big) \leq E\Big(\int_0^t \phi^2 ds\Big).$$

12.2 Fórmula de Itô

A fórmula de Itô é, para o cálculo estocástico, o análogo do Teorema Fundamental do Cálculo (TFC). Em primeiro lugar veremos que o TFC é falso no contexto do cáclulo estocástico. Começamos considerando $f, g \in C^1$, o espaço das funções continuamente diferenciáveis. Temos que o TFC implica que

$$g(f(t)) = g(f(0)) + \int_0^t g'(f(s))f'(s)ds = g(f(0)) + \int_0^t g'(f(s))df(s). \qquad (12.2)$$

Vamos verificar se esta fórmula ainda vale no contexto do cálculo estocástico. Para isto, tomamos $g(x) = x^2$. Com $0 = t_0 < \cdots < t_n = t$, temos que

$$\begin{aligned} W^2(t) &= \sum_{i=1}^{n} W^2(t_i) - W^2(t_{i-1}) \\ &= 2\sum_{i=1}^{n} W(t_{i-1})(W(t_i) - W(t_{i-1})) + \sum_{i=1}^{n}(W(t_i) - W(t_{i-1}))^2. \end{aligned}$$

Quando $n \to \infty$, o primeiro termo converge para $2\int_0^t W dW$ e o segundo para a variação quadrática t do movimento browniano. Portanto obtemos,

$$W(t)^2 = 2\int_0^t W dW + t.$$

Isto mostra que a fórmula (12.2) não se aplica neste caso.

Começamos para enunciar a fórmula de Itô no caso mais simples a seguir. Seja C^2 o espaço das funções duas vezes continuamente diferenciáveis e C_b^2 o espaço das funções $f \in C^2$ tais que f, f' e f'' são limitadas.

Proposição 12.5. *Para $f \in C_b^2$, temos q.c,*

$$f(W(t)) = f(W(0)) + \int_0^t f'(W)dW + \frac{1}{2}\int_0^t f''(W)ds, \tag{12.3}$$

para todo $t \geq 0$.

Prova: Aplicando a fórmula de Taylor de ordem dois, obtemos que

$$\begin{aligned} f(W(t)) &= f(W(0)) + \sum_{i=1}^{n}[f(W(t_i)) - f(W(t_{i-1}))] \\ &= f(W(0)) + \sum_{i=1}^{n} f'(W(t_{i-1}))[W(t_i) - W(t_{i-1})] \\ &\quad + \frac{1}{2}\sum_{i=1}^{n} f''(W(\theta_i))[W(t_i) - W(t_{i-1})]^2, \end{aligned}$$

onde $\theta_i = \theta_i(\omega) \in (t_{i-1}, t_i)$. O segundo termo do membro da direita converge, quando $n \to \infty$, para a integral estocástica $\int_0^t f'(W)dW$. Vamos mostar que o terceiro termo converge para $\frac{1}{2}\int_0^t f''(W)ds$. Para isto, denotemos por

$$A_n = \sum_{i=1}^{n} f''(W(\theta_i))[W(t_i) - W(t_{i-1})]^2,$$

$$B_n = \sum_{i=1}^{n} f''(W(t_{i-1}))[W(t_i) - W(t_{i-1})]^2$$

e

$$C_n = \sum_{i=1}^{n} f''(W(t_{i-1}))[t_i - t_{i-1}].$$

Quando $n \to \infty$, $C_n \to \int_0^t f''(W)ds$ q.c. e em L_1. Vamos mostrar que A_n converge em L_1 para o mesmo limite. Pela desigualdade triangular temos,

$$E(|A_n - C_n|) \leq E(|A_n - B_n|) + E(|B_n - C_n|).$$

Pela desigualdade de Cauchy-Schwarz, obtemos

$$\begin{aligned} E(|A_n - B_n|) &\leq E\Big(\sup_i |f''(W(t_{i-1})) - f''(W(\theta_i))| \times \sum_i [W(t_i) - W(t_{i-1})]^2\Big) \\ &\leq \Big[E(\sup_i |f''(W(t_{i-1})) - f''(W(\theta_i))|^2) \\ &\qquad\qquad \times E\Big(\Big(\sum_i [W(t_i) - W(t_{i-1})]^2\Big)^2\Big)\Big]^{1/2} \end{aligned}$$

que tende a 0, quando $n \to \infty$, usando o TCD e a Proposição 12.1. Por outro lado,

$$\begin{aligned} E(|B_n - C_n|^2) &= E\left[\Big|\sum_{i=1}^{n} f''(W(t_{i-1}))([W(t_i) - W(t_{i-1})]^2 - (t_i - t_{i-1}))\Big|^2\right] \\ &\leq \sum_{i=1}^{n} E\Big[|f''(W(t_{i-1}))([W(t_i) - W(t_{i-1})]^2 - (t_i - t_{i-1}))|^2\Big] \\ &\leq \sup(f'')^2 \times \sum_{i=1}^{n} E\Big[|[W(t_i) - W(t_{i-1})]^2 - (t_i - t_{i-1})|^2\Big] \\ &= \sup(f'')^2 \times 2\sum_{i=1}^{n} (t_i - t_{i-1})^2, \end{aligned}$$

que tende a 0, o que implica que $E(|B_n - C_n|) \to 0$, quando $n \to \infty$. Assim, para todo $t \geq 0$, a igualdade (12.3) ocorre em L_1 e portanto q.c. Usando a continuidade das trajetorias, podemos inverter as expressões "para todo $t \geq 0$" e "q.c." na frase anterior. □

Exemplo 12.1. Considere a integral estocástica

$$\int_0^t \text{sen}(W)\, dW.$$

Usando a Proposição 12.5, podemos obter uma expressão para a integral acima que não involve nenhuma integral estocástica. Para isto, basta aplicar a proposição com $f(x) = -\cos x$. Usando que $W(0) = 0$, obtemos que

$$\int_0^t \operatorname{sen}(W)\, dW = 1 - \cos(W(t)) - \frac{1}{2}\int_0^t \cos(W(s))\, ds.$$

Notação diferencial: Tradicionalmente (12.3) é escrita sob a "forma diferencial" mais compacta a seguir:

$$df(W) = f'(W)dW + \frac{1}{2}f''(W)dt,$$

embora apenas a forma integral (12.3) tem um sentido matemático rigoroso.

Todo processo estocástico da forma

$$X(t) = X(0) + \int_0^t \phi\, dW + \int_0^t \psi\, ds,$$

com $\phi, \psi \in H_2$ e $X(0) \in L_2(\Omega, \mathcal{F}_0, P)$ é chamado de *processo de Itô* e denotaremos por

$$dX = \phi dW + \psi ds$$

a sua forma diferencial. Os mesmos argumentos que na prova da proposição anterior podem ser usados para provar que os processos de Itô satisfazem: para $f \in C_b^2$,

$$f(X(t)) = f(X(0)) + \int_0^t f'(X)\phi\, dW + \int_0^t f'(X)\psi\, ds + \frac{1}{2}\int_0^t f''(X)\phi^2 ds,$$

ou seja, em notação diferencial,

$$df(X) = f'(X)dX + \frac{1}{2}f''(X)d\langle X\rangle,$$

com

$$\langle X\rangle(t) := \int_0^t \phi^2 ds.$$

Exemplo 12.2. Considere o processo dado por $X(t) = \exp\{W(t)\}$, para todo $t \geq 0$. Usando a Proposição 12.5 com $f(x) = e^x$ obtemos que

$$dX = XdW + \frac{X}{2}dt.$$

Assim, X é um processo de Itô pois $X \in H_2$.

Para terminar este paragrafo, enunciamos uma generalização da fórmula de Itô anterior para os espaços H_2^{loc}, H_1^{loc} e C^2 (veja Karatzas and Schreve (1988)). Introduzimos em primeiro lugar o espaço H_1^{loc} das funções $f : \mathbb{R}_+ \times \Omega \to \mathbb{R}$ progressivamente mensuráveis tais que $\int_0^T |f(t,\omega)|dt$ seja finito q.c. para todo $T > 0$.

Teorema 12.2. *Se $X(t) = X(0) + \int_0^t \phi \, dW + \int_0^t \psi \, ds$ é um processo de Itô com $\phi \in H_2^{\text{loc}}$, $\psi \in H_1^{\text{loc}}$ e $f \in C^2$, então, q.c., para todo $t \geq 0$,*

$$f(X(t)) = f(X(0)) + \int_0^t f'(X)\phi \, dW + \int_0^t f'(X)\psi \, ds + \frac{1}{2}\int_0^t f''(X)\phi^2 ds.$$

12.3 Transformação de Girsanov

Seja $(\Omega, \mathcal{F}, P)$ um e.p com uma filtração completa $\{\mathcal{F}_t, t \geq 0\}$, em que está definido um $\{\mathcal{F}_t\}$-MB W.

A densidade gaussiana $g(x) = (2\pi)^{-1/2} \exp -\{x^2/2\}$ possui a propriedade $g(x - a) = g(x)e^{ax - a^2/2}$, para todo $a \in \mathbb{R}$. Assim, se V é uma variável aleatória com lei normal padrão, temos para toda f mensurável e limitada

$$E(f(V+a)) = E\Big(f(V)\exp\{aV - a^2/2\}\Big).$$

Isto significa que a lei de $V + a$ é a mesma que a lei de V sob a nova probabilidade $dQ(\omega) = \exp\{aV(\omega) - a^2/2\}dP(\omega)$. Esta fórmula foi generalizada para o MB por Cameron e Martin em 1944, e de novo estendida em 1960 por Girsanov.

Sejam $\phi \in H_2^{\text{loc}}$ e para todo $t \geq 0$,

$$V_\phi(t) := \exp\Big\{\int_0^t \phi \, dW - \frac{1}{2}\int_0^t \phi^2 ds\Big\}.$$

Teorema 12.3 (Fórmula de Girsanov)**.** *Suponha que*

$$E(V_\phi(t)) = 1, \ \textit{para todo}\, t \geq 0.$$

Então existe uma única probabilidade Q sobre $\mathcal{F}_\infty := \bigvee_{t\geq 0} \mathcal{F}_t$, definida por

$$Q(A) = E(V_\phi(t)\mathbf{1}_A), \ A \in \mathcal{F}_t, \ t \geq 0.$$

Além disto, o processo estocástico $\overline{W}$ definido por

$$\overline{W}(t) = W(t) - \int_0^t \phi \, ds$$

é um $\{\mathcal{F}_t\}$-MB sob Q.

No resto desta seção daremos uma prova da fórmula de Girsanov. Começamos com a prova do seguinte

Lema 12.1. *Seja $\phi \in H_2^{\mathrm{loc}}$ com valores complexos, se existe uma constante finita C tal que a condição*

$$\int_0^t |\phi(s)|^2 ds \leq C, \ q.c,$$

então $V_\phi(t) \in L_2$ e $E(V_\phi(t)) = 1$.

Prova: Começamos com a prova do resultado no caso ϕ real. A provar será feito em duas etapas. Na primeira etapa, consideramos que ϕ é limitada no intervalo $[0,t]$, ou seja existe K tal que $\sup_{s\in[0,t]} |\phi(s)| \leq K$. Aplicando a fórmula de Itô obtemos

$$V_\phi(t) = 1 + \int_0^t \phi V_\phi \, dW. \tag{12.4}$$

Vamos mostrar que $\phi V_\phi \in H_2([0,t])$ para obter que a integral estocástica $V_\phi(t)$ acima é um martingal e que portanto $E(V_\phi(t)) = 1$. Para isto, usando a desigualdade $(a+b)^2 \leq 2a^2 + 2b^2$, temos que

$$V_\phi(u)^2 \leq 2\Big(1 + \Big[\int_0^u \phi V_\phi \, dW\Big]^2\Big),$$

para todo $u \leq t$. Usando o Problema 5, obtemos que $V_\phi(s) \in L_2$ e é limitado em L_2 para $s \in [0,t]$. Deduzimos que $\phi V_\phi \in H_2([0,t])$ e, finalmente, $E(V_\phi(t)) = 1$. Na segunda etapa, consideramos ϕ real sem a hipótese de limitação. Neste caso, usamos o truncamento $\phi_n := \phi \mathbf{1}_{[-n,n]}$ para $n \geq 1$. Obtemos

$$\begin{aligned} E(V_{\phi_n}(t)^2) &\leq \Big(E\Big[\exp\Big\{4\int_0^t \phi_n \, dW - 8\int_0^t \phi_n^2 ds\Big\}\Big] \\ &\qquad \times E\Big[\exp\Big\{6\int_0^t \phi_n^2 ds\Big\}\Big]\Big)^{1/2} \\ &\leq 1 \times \exp\{3C\}, \end{aligned}$$

usando a desigualdade de Cauchy-Schwarz e a primeira etapa. Além disto,

$$V_{\phi_n}(t) \to V_\phi(t),$$

em probabilidade pois $\int_0^t \phi_n^2 du \to \int_0^t \phi^2 du$ pelo TCM e $\int_0^t \phi_n \, dW$ converge em L_2 para $\int_0^t \phi \, dW$. Usando o Problema 6 item (ii), obtemos que

$$E(V_\phi(t)) = \lim_{n\to\infty} E(V_{\phi_n}(t)) = 1.$$

Consideramos, para terminar, um integrando com valores complexos $\phi = \rho + i\theta$.

Temos que

$$\begin{aligned}
E(|V_\phi(t)|^2) &= E\Big[\Big|\exp\Big\{\int_0^t (\rho+i\theta)dW - \frac{1}{2}\int_0^t (\rho+i\theta)^2 ds\Big\}\Big|\Big] \\
&\leq E\Big[V_\rho(t)^2 \times \exp\{\frac{1}{2}\int_0^t \theta^2 ds\}\Big] \\
&\leq E(V_\rho(t)^2) \times \exp\{C\},
\end{aligned}$$

usando a versão real do lema provada acima. Da mesma forma, temos que

$$\begin{aligned}
\int_0^t E(|V_\phi|^2|\phi|^2)ds &\leq \exp\{C\}\int_0^t E(V_\rho^2|\phi|^2)ds \\
&\leq \exp\{C\}\int_0^t E(E(V_\rho(t)^2 \mid \mathcal{F}_s)|\phi(s)|^2)ds \\
&= \exp\{C\}\int_0^t E(V_\rho(t)^2|\phi(s)|^2)ds \\
&= \exp\{C\}E\Big(V_\rho(t)^2\int_0^t |\phi(s)|^2 ds\Big) \\
&\leq C\exp\{C\}E(V_\rho(t)^2) < \infty.
\end{aligned}$$

Deduzimos que o integrando da equação (12.4) pertence a $H_2([0,t])$ e portanto que $E(V_\phi(t)) = 1$. $\square$

Prova do Teorema 12.3:

Em primeiro lugar, a fórmula $Q_t(A) = E(V_\phi(t)\mathbf{1}_A)$ para $A \in \mathcal{F}_t$ e $t \geq 0$, define uma família de probabilidades consistente (veja o Problema 7). Pelo teorema de extensão de Kolmogorov, existe uma única probabilidade Q sobre $\mathcal{F}_\infty$ cuja restrição à $\mathcal{F}_t$ seja Q_t. Vamos mostrar que $\overline{W}$ é um movimento browniano sob Q, ou seja para todo $0 \leq t_1 < \cdots < t_p = t$

$$E\Big(V_\phi(t)\exp\Big\{i\sum_{j=1}^p u_j\overline{W}(t_j)\Big\}\Big) = \exp\Big\{-\frac{1}{2}\sum_{j,k=1}^p u_ju_k(t_j\wedge t_k)\Big\}, \tag{12.5}$$

para todo $u_1,\ldots,u_p$.

Inicialmente, provaremos (12.5) sob a condição $\int_0^t \phi^2 ds \leq C$, q.c. A função $\psi(s,\omega) = \phi(s,\omega) + i\sum_{j=1}^p u_j\mathbf{1}_{[0,t_j]}(s)$ com valores complexos é tal que $\int_0^t |\psi|^2 ds \leq C'$ para uma constante finita C'. Usando o Lema 12.1, obtemos que $E(V_\psi(t)) = 1$. Assim, expandindo ψ^2 temos

$$E\Big(V_\phi(t)\exp\Big\{i\sum_{j=1}^p u_j\overline{W}(t_j)\Big\}\Big)\exp\Big\{\frac{1}{2}\sum_{j,k=1}^p u_ju_k(t_j\wedge t_k)\Big\} = 1 \tag{12.6}$$

o que prova (12.5).

No caso geral, usamos um argumento de localização com os tempos de parada $\tau_n = \inf\{t \geq 0 : \int_0^t \phi^2 ds \geq n\}$. As funções $\phi_n(s,\omega) = \phi(s,\omega)\mathbf{1}_{[0,\tau_n]}(s)$ verificam:

- $\int_0^t \phi_n^2 ds \leq n$ e pelo Lema 12.1, $E(V_{\phi_n}(t)) = 1$;
- q.c., $\int_0^t \phi_n \, dW \to \int_0^t \phi \, dW$ e $\int_0^t \phi_n^2 ds \to \int_0^t \phi^2 ds$.

Assim, pelo Problema 6, item (i), obtemos que

$$E_{\phi_n}(t) \to V_\phi(t) \ \text{ em } \ L_1. \tag{12.7}$$

Como (12.6) vale para ϕ_n, temos para todo $n \geq 1$,

$$E\Big(V_{\phi_n}(t) \exp\Big\{ i \sum_{j=1}^p u_j \Big(\overline{W}(t_j) - \int_0^{t_j} \phi_n ds\Big)\Big\}\Big) = \exp\Big\{ -\frac{1}{2} \sum_{j,k=1}^p u_j u_k (t_j \wedge t_k)\Big\}. \tag{12.8}$$

Agora, usando a decomposição

$$\begin{aligned}
&V_{\phi_n}(t) \exp\Big\{ i \sum_{j=1}^p u_j \Big(W(t_j) - \int_0^{t_j} \phi_n ds\Big)\Big\} - V_\phi(t) \exp\Big\{ i \sum_{j=1}^p u_j \overline{W}(t_j)\Big\} \\
&\quad = (V_{\phi_n}(t) - V_\phi(t)) \exp\Big\{ i \sum_{j=1}^p u_j \Big(W(t_j) - \int_0^{t_j} \phi_n ds\Big)\Big\} \\
&\quad + V_\phi(t)\Big(\exp\Big\{ i \sum_{j=1}^p u_j \Big(\overline{W}(t_j) - \int_0^{t_j} \phi_n ds\Big)\Big\} - \exp\Big\{ i \sum_{j=1}^p u_j \overline{W}(t_j)\Big\}\Big)
\end{aligned}$$

concluimos por (12.7) e pelo TCD que

$$V_{\phi_n}(t) \exp\Big\{ i \sum_{j=1}^p u_j \Big(W(t_j) - \int_0^{t_j} \phi_n ds\Big)\Big\} \to V_\phi(t) \exp\Big\{ i \sum_{j=1}^p u_j \overline{W}(t_j)\Big\}$$

em L_1, o que implica a convergência das esperanças e portanto, por (12.8) o resultado desejado. $\square$

A seguir, daremos algumas aplicações clássicas da fórmula de Girsanov.

Continuidade absoluta de funcionais aditivos do MB

Seja $f : [0,T] \to \mathbb{R}$ uma função de quadrado integrável e $d(t) := \int_0^t f(s)ds$, $t \leq T$. Consideramos a lei ν de $W + d$ no intervalo $[0,T]$, isto é, a probabilidade no espaço $C([0,T])$ definida por,

$$\nu(A) = P(W + d \in A), \quad A \text{ boreliano de } C([0,T]).$$

Denotaremos a seguir por μ a lei do MB W sobre $C([0,T])$. A proposição a seguir mostra que ν e μ são equivalentes.

Proposição 12.6. *A lei ν é absolutamente contínua com respeito a μ e a sua derivada de Radon-Nikodym é dada por*

$$\frac{d\nu}{d\mu}(x) = \exp\Big\{ \int_0^T f(t)dx(t) - \frac{1}{2}\int_0^T f(t)^2 dt \Big\}.$$

Observação: A primeira integral aparecendo na exponencial é simplesmente a integral estocástica com respeito ao MB sobre $C([0,T])$.

Prova: Seja Q a probabilidade definida por $dQ = Z(T)dP$, com

$$Z(T) = \exp\Big\{ \int_0^T f \, dW - \frac{1}{2}\int_0^T f^2 dt \Big\}.$$

Pela fórmula de Girsanov, sob Q, $W - d$ é um MB, isto implique que $Q(W \in \cdot) = P(W + d \in \cdot)$. Assim, obtemos para todo boreliano A de $C([0,T])$,

$$\begin{aligned}
\nu(A) &= P(W + d \in A) \\
&= Q(W \in A) \\
&= E(\mathbf{1}_A(W)Z(T)) \\
&= \int \mathbf{1}_A(x) \exp\Big\{ \int_0^T f(t)dx(t) - \frac{1}{2}\int_0^T f(t)^2 dt \Big\} d\mu(x). \quad \square
\end{aligned}$$

Exemplo 12.3. Sejam $a \in \mathbb{R}$ e $X(t) = W(t) + at$ para $t \in \mathbb{R}_+$. X é um MB com drift linear. Aplicando a Proposição 12.6 com $f = a$, obtemos que, para todo $T \in \mathbb{R}_+$, a lei de X no intervalo $[0,T]$ é absolutamente contínua com respeito à lei de W no mesmo intervalo com densidade dada por

$$\exp\Big\{ \mu W(T) - \frac{\mu^2}{2} T \Big\}.$$

Fórmula de Black-Scholes

A fórmula de Black-Scholes é uma famosa fórmula em finanças que permite calcular o preço teórico de uma opção europeia. Consideramos, a seguir, que temos um mercado com dois tipos de ativos financeiros. Um ativo sem risco com taxa de juros r. Matematicamente, temos que o preço ao longo do tempo deste ativo sem risco S satisfaz a equação diferencial a seguir

$$dS = rSdt, \tag{12.9}$$

ou seja

$$S(t) = S(0)\exp\{rt\}.$$

Além disto, temos um ativo com risco R, com tendência (*drift*) μ e volatilidade σ. Isto significa que R é um processo de Itô que pode ser representado em notação diferencial por

$$dR = R(\mu dt + \sigma dW).$$

Consideramos aqui a filtração canônica (completa) $\{\mathcal{F}_t, t \geq 0\}$ do MB W. Usando a fórmula de Itô, podemos verificar que

$$R(t) = R(0)\exp\Big\{\sigma W(t) + \Big(\mu - \frac{\sigma^2}{2}\Big)t\Big\}. \tag{12.10}$$

A seguir consideramos um produto finaceiro comum chamado *opção de compra europeia*. Esta opção permite comprar o ativo com risco em algum momento posterior T a um preço K. Portanto, o detentor da opção aposta que, no momento T, o ativo valerá mais do que K. Nesse caso, o detentor pode comprar o ativo pelo preço K e vende-lo imediatamente para obter um lucro $R(T) - K$. Se o ativo está com um preço inferior a K, o detentor da opção tem a possibilidade de não exercer a sua opção de compra. Assim, temos que o preço $\Psi(T)$ da opção no tempo T é dado por $\Psi(T) = (R_T - K)_+$.

A questão central aqui é determinar o preço de venda da opção para cada $t < T$. No desenvolvimento a seguir, vamos supor a *ausência de arbitragem*, ou seja, em nenhum momento é possivel comprar o ativo com risco e vendo-lo em seguida a um preço mais alto. Assim, supomos que existe um preço justo para a opção europeia.

Para responder à questão central, consideramos uma carteira da forma

$$dV = \alpha dR + \beta dS,$$

na qual α e β são dois processos estocásticos em $H_2([0,T])$, representando, respectivamente, a quantitidade de ações e títulos da carteira. Este tipo de carteira é chamado de *auto-financiadora*. Isto significa que somente é possível comprar ações e títulos com o dinheiro disponível na carteira. Para calular o preço da opção, supomos que exista uma carteira deste tipo que replica o preço da opção, ou seja, tal que $V(T) = \Psi(T)$.

Lema 12.2. *O preço $\Psi(t)$ da opção europeia no tempo t é igual a $V(t)$.*

Prova: Como $V(T) = \Psi(T)$ por hipótese, investindo $V(t)$ no tempo t na carteira podemos obter o mesmo lucro da opção. Assim, se vendessemos a opção a um valor $x > V(t)$ no tempo t, ninguém a compraria pois seria possível investir a (menor) quantidade $V(t)$ na carteira e obter o mesmo lucro da opção no tempo T. Por outro lado, se $x < V(t)$ no tempo t, todo mundo compraria a opção, pois seria possível obter o mesmo lucro $\Psi(T)$ com um menor investimento. □

Com o lema acima, a questão central se reduz agora ao cálculo de $V(t)$ para $t < T$. Pela fórmula de Itô, temos

$$\begin{aligned} d(e^{-rt}V(t)) &= -re^{-rt}V(t)dt + e^{-rt}dV \\ &= e^{-rt}(-rV(t)dt + \alpha(t)dR + \beta(t)dS) \\ &= e^{-rt}(-r(\alpha(t)R(t) + \beta(t)S(t))dt + \alpha(t)R(t)(\mu dt + \sigma dW) + \beta(t)dS \\ &= \alpha(t)S(t)e^{-rt}(-rdt + \mu dt + \sigma dW) + e^{-rt}\beta(t)(-rP(t)dt + dS). \end{aligned}$$

Por (12.9), o segundo termo é nulo e obtemos

$$d(e^{-rt}V(t)) = \sigma\alpha(t)S(t)e^{-rt}\Big(\frac{\mu - r}{\sigma}dt + dW\Big).$$

Devido à presença do termo em "dt", o processo $e^{-rt}V(t)$, não é um martingale (se $\mu \neq r$). No entanto, usando a fórmula de Girsanov, sob a nova probabilidade $dQ = M(T)dP$, sendo

$$M(T) = \exp\Big\{\frac{r-\mu}{\sigma}W(t) - \frac{(r-\mu)^2}{2\sigma^2}t\Big\},$$

obtemos que $e^{-rt}V(t)$ é um martingale com respeito à filtração $\{\mathcal{F}_t, t \geq 0\}$. Em particular, temos que

$$\begin{aligned} V(t) &= e^{rt}e^{-rt}V(t) \\ &= e^{rt}E_Q(e^{-rT}V(T) \mid \mathcal{F}_t) = e^{r(t-T)}E_Q((R(T) - K)_+ \mid \mathcal{F}_t). \end{aligned} \tag{12.11}$$

Agora, usando (12.10) obtemos

$$\begin{aligned} R(t) &= R(0)\exp\Big(\sigma W(t) + (\mu - \frac{\sigma^2}{2})t\Big) \\ &= R(0)\exp\Big(\sigma W(t) + (r - \frac{\sigma^2}{2})t\Big). \end{aligned}$$

Podemos agora usar esta última expressão em (12.11) para calcular $V(t)$ para $t < T$. Em seguida, apresentamos somente a expressão de $V(0)$ deixando ao leitor o cálculo de $V(t)$ (veja Problema 8).

Teorema 12.4 (Fórmula de Black-Scholes para $t = 0$)**.** *O preço da opção europeia no tempo $t = 0$ é dado por*

$$\Psi(0) = R(0)\Phi(d_1) - Ke^{-rT}\Phi(d_2),$$

na qual

$$d_1 = \frac{-\ln(K/R(0)) + (r + \sigma^2/2)T}{\sigma\sqrt{T}}, \qquad d_2 = d_1 - \sigma\sqrt{T},$$

e Φ é função de distribuição acumulada da lei normal padrão.

Prova: Como $\mathcal{F}_0$ é trivial, temos que

$$\begin{aligned} V(0) &= e^{-rT}E_Q[(R(T) - K)_+] \\ &= e^{-rT}(E_Q[R(T)\mathbf{1}_{\{R(T)\geq K\}}] - KE_Q[\mathbf{1}_{\{R(T)\geq K\}}]). \end{aligned} \tag{12.12}$$

Como $W(t) + \frac{\mu - r}{\sigma}t$ é um MB sob Q, $W(T)$ tem mesma lei que $\sqrt{T}N$, em que N é uma v.a. com lei normal padrão. Assim, a lei de $R(T)$ é a mesma que

$$R(0)\exp\Big(\sigma\sqrt{T}N + (r - \frac{\sigma^2}{2})T\Big) = R(0)e^{\gamma N + \delta}, \tag{12.13}$$

em que $\gamma = \sigma\sqrt{T}$ e $\delta = (r - \sigma^2/2)T$. Assim, temos que

$$\begin{aligned} E_Q[\mathbf{1}_{\{R(T)\geq K\}}] &= Q(R(T) \geq K) \\ &= P(R(0)e^{\gamma N + \delta} \geq K) \\ &= P\Big(N \geq \frac{\ln(K/R(0)) - \delta}{\gamma}\Big) \\ &= P\Big(N \leq \frac{-\ln(K/R(0)) + \delta}{\gamma}\Big) \\ &= \Phi(d_2). \end{aligned}$$

De maneira análoga obtemos

$$\begin{aligned} E_Q[R(T)\mathbf{1}_{\{R(T)\geq K\}}] &= R(0)E(e^{\gamma N + \delta}\mathbf{1}_{\{R(0)e^{\gamma N + \delta}\geq K\}}) \\ &= R(0)E(e^{-\gamma N + \delta}\mathbf{1}_{\{N \leq d_2\}}) \\ &= R(0)e^{\delta}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{d_2} e^{-\gamma x - x^2/2}dx \\ &= R(0)e^{\delta + \gamma^2/2}\Phi(d_2 + \gamma) \\ &= R(0)e^{rT}\Phi(d_1). \end{aligned}$$

Colocando as duas últimas expressões em (12.12), obtemos o resultado desejado. □

Problemas

1. Para todo $n \geq 1$, considere o operador linear Π_n sobre $H^2(\mathbb{R}_+)$ definido para P-q.t. $\omega \in \Omega$ e todo $t \geq 0$ por

$$\Pi_n f(t,\omega) = n \sum_{i=1}^{n^2} \Big(\int_{(i-1)/n}^{i/n} f(u,\omega)du \Big) \mathbf{1}_{(i/n,\,(i+1)/n]}(t),$$

   Mostrar que $\Pi_n f \in \mathcal{E}$ e que $\Pi_n f \to f$ em $H^2(\mathbb{R}_+)$. Concluir que $\mathcal{E}$ é denso em $H^2(\mathbb{R}_+)$ e em $H^2([0,T])$ par todo $T > 0$.

2. Provar (12.1).

3. Dar um exemplo de um martingale local que não é um martingale.

4. Seja W um $\{\mathcal{F}_t\}$-MB. Usando a fórmula de Itô, verificar se os processos a seguir são martingales.
   (i) $X(t) = W(t)^2$;
   (ii) $X(t) = t^2 W(t) - 2\int_0^t sW(s)ds$.

5. (Lema de Gronwall) Seja f uma função não negativa localmente integrável definida sobre $\mathbb{R}_+$ tal que
$$f(t) \leq a + b\int_0^t f(s)ds, \quad t \geq 0, \tag{12.14}$$
   com a, b constantes não negativas. Mostrar que
$$f(t) \leq a\exp\{bt\}.$$

6. Seja $(X_n)_{n\geq 1}$ uma sequência de variáveis aleatórias e X uma variável aleatória definidas no mesmo espaço de probabilidade. Mostrar as seguintes implicações
   (i) Se $X_n \geq 0$, $E(X_n) \to E(X) < \infty$ e $X_n \to X$ q.c. então $X_n \to X$ em L^1.
   (ii) Se $\sup_n E(X_n^2) < \infty$ e $X_n \to X$ em probabilidade então $X_n \to X$ em L^1.

7. Mostrar que a família $\{Q_t, t \geq 0\}$ é consistente.

8. (Fórmula de Black-Scholes) Seja $\lambda = (r-\mu)/\sigma$ e $Y(t) = W(t) - \lambda t$. Usando a decomposição $Y(T) = (Y(T) - Y(t)) + Y(t)$, calcular $V(t)$ para todo $t \in [0,T]$.

Apêndice
Alguns conceitos matemáticos

A.1 Medida

Seja Ω um conjunto não vazio e $\{A_n, n \geq 1\}$ uma sequência de subconjuntos de Ω. Dizemos que esta sequência é *crescente* se $A_n \subset A_{n+1}$, para todo $n \geq 1$ e que ela é *decrescente* se $A_{n+1} \subset A_n$, para todo $n \geq 1$. Dizemos que $A_n \uparrow A$ se a sequência é crescente e $A = \cup_n A_n$. Analogamente, $A_n \downarrow A$ se a sequência é decrescente e $A = \cap_n A_n$. Definimos o $\limsup_n A_n$ como o conjunto de todos os elementos de Ω que pertencem a um número infinito de conjuntos A_n, e o $\liminf_n A_n$ como o conjunto dos elementos de Ω que pertencem a todos os conjuntos A_n com exceção de um número finito deles. Obviamente, $\liminf_n A_n \subset \limsup_n A_n$ e vale a seguinte proposição.

Proposição A.1.1 Temos as seguintes propriedades

(i) $\limsup_n A_n = \bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} A_n$;

(ii) $\liminf_n A_n = \bigcup_{m=1}^{\infty} \bigcap_{n=m}^{\infty} A_n$;

(iii) $(\limsup_n A_n^c)^c = \liminf_n A_n$.

A seguir, denotemos por $\varnothing$ o conjunto vazio.

Definição A.1.1. Uma classe $\mathcal{F}$ de subconjuntos de Ω é uma *álgebra* se:

(i) Ω pertence a $\mathcal{F}$;

(ii) se $A \in \mathcal{F}$, então $A^c \in \mathcal{F}$;

(iii) se $A, B \in \mathcal{F}$, então $A \cup B \in \mathcal{F}$.

Como consequências temos:

(a) $\varnothing \in \mathcal{F}$;

(b) uma álgebra é fechada sob reuniões finitas;

(c) uma álgebra é fechada sob intersecções finitas;

(d) uma álgebra é fechada sob todas as operações finitas com conjuntos.

Exemplo A.1.1. São exemplos de álgebras:

(i) Dado qualquer Ω dado, seja $\mathcal{F} = 2^{\Omega}$ o conjunto de todos os subconjuntos de Ω.

(ii) Seja $\Omega = \mathbb{R}$ e $\mathcal{F}$ a classe contendo todas as reuniões finitas disjuntas de intervalos da forma $(a, b]$, $(-\infty, a]$ ou $(b, +\infty)$, $a, b \in \mathbb{R}$.

(iii) Seja $\Omega = \mathbb{R}^p$, $p > 1$, e $\mathcal{F}$ a classe contendo todas as reuniões finitas disjuntas dos retângulos $\prod_{i=1}^{p} I_i$, onde para todo i, I_i é um intervalo da forma descrita em (ii).

Definição A.1.2. Uma classe $\mathcal{F}$ de subconjuntos de Ω é uma *σ-álgebra* se:

(i) Ω pertence a $\mathcal{F}$;

(ii) se $A \in \mathcal{F}$, então $A^c \in \mathcal{F}$;

(iii) se $A_n \in \mathcal{F},\ n \geq 1$, então $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$.

Como consequências da definição acima temos:

(a) $\varnothing \in \mathcal{F}$;

(b) (iii) e (a) implicam que se $A_n \in \mathcal{F},\ n = 1, \ldots, p$, então $\cup_{n=1}^{p} A_n \in \mathcal{F}$;

(c) uma σ-álgebra é uma álgebra fechada sob reuniões enumeráveis;

(d) a reunião em (iii) pode ser substituída por intersecção;

(e) uma σ-álgebra é fechada sob todas as operações enumeráveis de conjuntos;

(f) uma σ-álgebra é uma álgebra.

Exemplo A.1.2. São exemplos de σ-álgebras:

(i) Para qualquer Ω dado, $\mathcal{F}_{\varnothing} = \{\varnothing, \Omega\}$ é uma σ-álgebra; é a menor σ-álgebra de subconjuntos de Ω, chamada *σ-álgebra trivial*;

(ii) Para qualquer Ω dado, o conjunto 2^{Ω} é uma σ-álgebra; é a maior σ-álgebra de subconjuntos de Ω. Para qualquer σ-álgebra $\mathcal{F}$ sobre Ω, teremos $\mathcal{F}_{\varnothing} \subset \mathcal{F} \subset 2^{\Omega}$;

(iii) A menor σ-álgebra contendo uma classe $\mathcal{A}$ de subconjuntos de Ω é chamada a *σ-álgebra gerada por* $\mathcal{A}$, e a denotamos por $\mathcal{F} = \mathcal{F}[\mathcal{A}]$;

(iv) Seja $\Omega = \mathbb{R}^p$, $p \geq 1$, e consideremos a σ-álgebra gerada pelos retângulos p-dimensionais da forma $\prod_{i=1}^{p}(a_i, b_i]$, $a_i, b_i \in \mathbb{R}$. Esta σ-álgebra é chamada *σ-álgebra de Borel* e é denotada por $\mathcal{B}^p$. Esta também é a σ-álgebra gerada pela classe dos conjuntos abertos de $\mathbb{R}^p$; $\mathcal{B}^1$, ou simplesmente $\mathcal{B}$, é a σ-álgebra de Borel de $\mathbb{R}$.

A intersecção de uma família arbitrária de σ-álgebras $\{\mathcal{F}_\alpha, \alpha \in \Gamma\}$, onde Γ é um conjunto de índices, é uma σ-álgebra denotada por $\bigcap_{\alpha\in\Gamma} \mathcal{F}_\alpha$ ou $\bigwedge_{\alpha\in\Gamma} \mathcal{F}_\alpha$. Contudo, $\bigcup_{\alpha\in\Gamma} \mathcal{F}_\alpha$ não precisa ser uma σ-álgebra em geral. Pelo item (iii) do Exemplo A.1.2, existe uma σ-álgebra minimal contendo todas elas, denotada $\bigvee_{\alpha\in\Gamma} \mathcal{F}_\alpha$.

A seguir, apresentamos outros sistemas de conjuntos importantes.

Sistema-π e sistema de Dynkin. Dizemos que $\mathcal{F}$ é um *sistema-π* se $A, B \in \mathcal{F}$, então $A \cap B \in \mathcal{F}$. Dizemos que $\mathcal{F}$ é uma *sistema de Dynkin* se:

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

(ii) $A, B \in \mathcal{F}$, $B \subset A$, então $A - B = A \cap B^c \in \mathcal{F}$;

(iii) Se $A_n \in \mathcal{F}$ e $A_n \subset A_{n+1}$ para todo $n \geq 1$, então $\cup_{n=1}^{\infty} A_n \in \mathcal{F}$.

Um sistema de Dynkin é também chamado sistema-λ. Temos as seguintes propriedades:

(a) Dada qualquer coleção de conjuntos, existe um sistema de Dynkin minimal contendo esses conjuntos, a saber, a intersecção de todos os sistemas de Dynkin contendo a coleção. Esse sistema de Dynkin minimal é dito ser *gerado* pela coleção de conjuntos.

(b) Toda σ-álgebra é um sistema de Dynkin;

(c) Um sistema de Dynkin é uma σ-álgebra se, e somente se, é um sistema-π.

Classes monotônicas. Uma coleção de conjuntos $\mathcal{C}$ é uma *classe monotônica* se:

(i) $E_n \uparrow E, \;\; E_n \in \mathcal{C} \;\Rightarrow\; E \in \mathcal{C}$;

(ii) $E_n \downarrow E, \;\; E_n \in \mathcal{C} \;\Rightarrow\; E \in \mathcal{C}$.

Alguns fatos sobre classes monotônicas são:

(a) Dada qualquer coleção de conjuntos, existe uma classe monotônica minimal contendo esses conjuntos, a saber, a intersecção de todas as classes monotônicas contendo a coleção. Essa classe monotônica minimal é dita ser *gerada* pela coleção de conjuntos.

(b) Uma álgebra é uma σ-álgebra se e somente for uma classe monotônica.

(c) Seja $\mathcal{F}_0$ uma álgebra e $\mathcal{F}$ a σ-álgebra gerada por $\mathcal{F}_0$. Seja $\mathcal{C}$ a classe monotônica gerada por $\mathcal{F}_0$. Então $\mathcal{C} = \mathcal{F}$.

Para mais detalhes, veja Chung (1974).

Para introduzir o conceito de medida consideremos um exemplo. Seja Ω arbitrário e $\mathcal{F} = 2^{\Omega}$. Seja n uma função definida em $\mathcal{F}$ com valores em $[0, \infty]$, ou seja, uma função de conjunto, tal que $n(A)$ seja o número de elementos de $A \in \mathcal{F}$, se A for finito e $n(A) = \infty$ se A for infinito. Então n tem as propriedades:

(i) $n(\varnothing) = 0$;

(ii) Se $A, B \in \mathcal{F},\ A \cap B = \varnothing$, então $n(A \cup B) = n(A) + n(B)$;

(iii) Se $A_i \in \mathcal{F},\ i \geq 1$, disjuntos, então $n(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} n(A_i)$.

Observe que a propriedade (iii) implica a propridade (ii). A função n é uma medida sobre $\mathcal{F}$. Precisamente, temos a seguinte definição no caso em que $\mathcal{F}$ é somente uma álgebra.

Definição A.1.3. Seja $\mathcal{F}$ uma álgebra de subconjuntos de Ω. Uma função de conjuntos $\mu : \mathcal{F} \to [0, \infty]$ é uma *medida* sobre $\mathcal{F}$ se:

(i) $\mu(\varnothing) = 0$;

(ii) Sejam $A_i \in \mathcal{F}$, $i \geq 1$, disjuntos e a reunião deles pertence a $\mathcal{F}$, então $\mu(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i)$.

Uma medida μ é *finita* se $\mu(\Omega) < \infty$. No caso em que $\mu(\Omega) = 1$, μ é chamadada de *medida de probabilidade* ou simplesmente de *probabilidade*. Uma medida μ é σ-*finita* se existir uma sequência de conjuntos $\{A_n, n \geq 1\}$ de $\mathcal{F}$ com $\Omega = \bigcup_n A_n$ e $\mu(A_n) < \infty$, para todo $n \geq 1$. Sempre podemos tomar a sequência $\{A_n, n \geq 1\}$ como constituída de conjuntos disjuntos. Uma propriedade é verdadeira em μ-quase todo ponto (μ-q.t.p) se existe um conjunto mensurável A com $\mu(A^c) = 0$ tal que a propriedade vala para todo $x \in A$.

Da definição de uma medida seguem as seguintes propriedades elementares:

(a) Se $A, B \in \mathcal{F},\ A \subset B$, então $\mu(A) \leq \mu(B)$;

(b) Se $A, B \in \mathcal{F},\ A \subset B,\ \mu(A) < \infty$, então $\mu(B - A) = \mu(B) - \mu(A)$.

Exemplo A.1.3. Seja $\Omega = \mathbb{R}$ e $\mathcal{F}$ a classe de todas as reuniões finitas e disjuntas de intervalos do tipo $(a, b]$, $(-\infty, a]$ ou $(b, +\infty)$, $a, b \in \mathbb{R}$. Defina $\lambda : \mathcal{F} \to [0, \infty]$ tal que:

(i) $\lambda((a,b]) = b - a,\ a \le b$;

(ii) $\lambda(I) = \infty$, se I for um intervalo não limitado;

(iii) Seja $A \in \mathcal{F}$; então, $A = \bigcup_{r=1}^{n} I_r$, onde os I_r são disjuntos e do tipo acima descrito. Coloque $\lambda(A) = \sum_{r=1}^{n} \lambda(I_r)$.

Então, λ é uma medida sobre $\mathcal{F}$.

Exemplo A.1.4. De maneira geral, dada uma função F uma função crescente, contínua à direita tal que $\lim_{x\to\infty} F(x) = 1,\ \lim_{x\to-\infty} F(x) = 0$, defina uma função de conjunto μ_F sobre a álgebra $\mathcal{F}$ da seguinte maneira:

$$
\begin{aligned}
\mu_F((a,b]) &= F(b) - F(a),\ a \le b,\\
\mu_F((-\infty,a]) &= F(a),\\
\mu_F((b,+\infty)) &= 1 - F(b),\\
\mu_F(\mathbb{R}) &= 1.
\end{aligned}
$$

Se $A \in \mathcal{F}$, então $A = \cup_{r=1}^{m} I_r$, e defina $\mu_F(A) = \sum_{r=1}^{m} \mu_F(I_r)$. Temos que μ_F é uma medida de probabilidade sobre $\mathcal{F}$.

Consideremos, agora, uma medida sobre uma σ-álgebra.

Definição A.1.4. Por um *espaço mensurável* entendemos o par $(\Omega, \mathcal{F})$, consistindo de um conjunto Ω e de uma σ-álgebra $\mathcal{F}$ de subconjuntos de Ω. Um subconjunto A de Ω é chamado *mensurável (com respeito a* $\mathcal{F}$) se $A \in \mathcal{F}$.

Definição A.1.5. Por uma *medida* μ sobre um espaço mensurável $(\Omega, \mathcal{F})$ entendemos uma função de conjunto com valores em $[0,\infty]$ definida sobre todos os conjuntos de $\mathcal{F}$ e satisfazendo:

(i) $\mu(\varnothing) = 0$;

(ii) $\mu(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} \mu(E_i)$, para toda sequência $\{E_i, i \ge 1\}$ de conjuntos mensuráveis e disjuntos.

Um *espaço de medida* $(\Omega, \mathcal{F}, \mu)$ é um espaço mensurável $(\Omega, \mathcal{F})$ munido de uma medida μ definida sobre $\mathcal{F}$.

Exemplo A.1.5. A medida λ do Exemplo A.1.3 pode ser estendida à σ-álgebra $\mathcal{B}$. Esta extensão é chamada de medida de Lebesgue sobre $\mathbb{R}$. A prova deste fato não é simples e pode ser encontrada em Billingsley (1995). Este é um exemplo de medida σ-finita (basta tomar $A_n = (n, n+1], n = 0, \pm 1, \ldots$). A restrição da medida de Lebesgue aos borelianos do intervalo $[-1,1]$ é um exemplo de medida finita. De maneira análoga, as medidas de probabilidade do Exemplo A.1.4. podem ser estendidas à $\mathcal{B}$. Enfim, seja Ω um conjunto não enumerável e $\mathcal{F} = 2^{\Omega}$. Seja μ a medida definida por $\mu(\{x\}) = 1$, para todo $x \in \Omega$. Então, μ não é σ-finita.

A.2 Integral com respeito a uma medida

A noção de integral estende as noções de comprimento, área e volume e os trabalhos mais importantes remontam a Borel, 1898 e Lebesgue, 1902. A extensão a σ-álgebras sobre espaços abstratos foi feita por Fréchet, em 1915. A integral é definida, sucessivamente, para uma função simples, positiva e arbitrária.

Dado um espaço mensurável $(\Omega, \mathcal{F})$, uma *função simples* $f : \Omega \to \mathbb{R}$ é uma combinação linear finita de indicadores de subconjuntos mensuráveis $A_1, \ldots, A_k$ de Ω, isto é,

$$f = \sum_{i=1}^{k} c_i I_{A_i}, \quad c_i \in \mathbb{R}.$$

Evidentemente, f admite mais de uma representação como aquela acima, mas podemos associar a f uma *representação canônica*,

$$f = \sum_{j=1}^{p} a_j I_{B_j}, \tag{A.2.1}$$

tal que:

(i) os B_j são não vazios e dois a dois disjuntos, $B_j = f^{-1}(a_j)$;

(ii) $\Omega = \cup_{j=1}^{p} B_j$ (e portanto algum dos a_j pode ser nulo).

Indicamos por $S_+(\Omega)$ o conjunto das funções simples $f : \Omega \to \mathbb{R}_+$. Seja $f \in S_+(\Omega)$ com representação canônica (A.2.1). Definimos a *integral de* f com respeito à medida μ como sendo

$$\int f d\mu = \int f(\omega) d\mu(\omega) = \sum_{j=1}^{p} a_j \mu(B_j)$$

onde usamos a convenção $0 \times \infty = 0$.

Sejam $f, g \in S_+(\Omega)$ então:

(a) se $f \le g$, segue-se $\int f d\mu \le \int g d\mu$;

(b) se $a, b \ge 0$, temos $\int (af + bg) d\mu = a \int f d\mu + b \int g d\mu$.

Seja agora $g : \Omega \to [0, \infty]$ uma função mensurável positiva com valores possivelmente infinitos. Então a *integral de* g com respeito á μ é definida por

$$\int g d\mu = \int g(\omega) d\mu(\omega) = \sup \Big\{ \int f d\mu : \; f \in S_+(\Omega), f \le g \Big\}.$$

Os seguintes fatos são válidos:

(a) Se $g \geq 0$ e $h \geq 0$ e $g = h$, em μ-q.t.p então $\int g d\mu = \int h d\mu$;

(b) se $g \geq 0$, então $g = 0$ em μ-q.t.p implica que $\int g d\mu = 0$;

(c) se $g \geq 0$ e $\int g d\mu < \infty$, então $g < \infty$ em μ-quase todo ponto;

(d) se $g \geq 0$, $h \geq 0$ e $g \leq h$, então $\int g d\mu \leq \int h d\mu$.

Para toda $g \geq 0$ e A conjunto mensurável, definimos

$$\int_A g d\mu = \int I_A g d\mu.$$

Os dois teoremas a seguir são fundamentais. Para provas, veja Barttle (2001).

Teorema A.2.1. [Lema de Fatou] *Seja* $f_k : \Omega \to [0, \infty]$, $k \geq 1$, *uma sequência de funções mensuráveis. Então,*

$$\int \liminf_{k\to\infty} f_k \, d\mu \leq \liminf_{k\to\infty} \int f_k \, d\mu \leq \infty.$$

Teorema A.2.2. [da Convergência Monótona] *Seja* $f_k : \Omega \to [0, \infty]$, $k \geq 1$, *uma sequência crescente de funções mensuráveis. Então,*

$$\int \lim_{k\to\infty} f_k \, d\mu = \lim_{k\to\infty} \int f_k \, d\mu \leq \infty.$$

Consequências importantes desses teoremas são:

(e) $g \geq 0$, $h \geq 0 \;\Rightarrow\; \int (g + h) d\mu = \int g d\mu + \int h d\mu$;

(f) Dada uma sequência de funções positivas $\{f_k, k \geq 1\}$, temos que

$$\int \Big(\sum_{k=1}^{\infty} f_k\Big) d\mu = \sum_{k=1}^{\infty} \int f_k \, d\mu.$$

(g) Seja $g \geq 0$ e sejam $\{A_k, k \geq 1\}$ uma sequência de subconjuntos mensuráveis de Ω, dois a dois disjuntos e cuja reunião é Ω. Então,

$$\int g \, d\mu = \sum_{k=1}^{\infty} \int_{A_k} g \, d\mu.$$

Consideremos, agora, uma função mensurável $g : \Omega \to \mathbb{R}$. Como essa função pode ser escrita como a diferença entre sua parte positiva, f_+, e sua parte negativa, f_-, ou seja, $f = f_+ - f_-$, definamos

$$\int f d\mu = \int f_+ d\mu - \int f_- d\mu,$$

desde que tenhamos $\int f_+ d\mu < \infty$ ou $\int f_- d\mu < \infty$. No caso em que as duas integrais são finitas, dizemos que f é integrável com respeito à μ.

Os seguintes fatos podem ser facilmente provados.

(a) Se g, h são funções integráveis, então $\int (ag + bh) d\mu = a \int g d\mu + b \int h d\mu$, com a e b reais;

(b) se h é uma função mensurável tal que $|h| \leq g$ em μ-q.t.p e g é integrável, então h é integrável;

(c) se g, h são funções integráveis e $g \leq h$ em μ-q.t.p, então $\int g d\mu \leq \int h d\mu$;

(d) se g é integrável, então $|g|$ é integrável e $|\int g d\mu| \leq \int |g| d\mu$. A recíproca também vale.

O teorema seguinte é importante em muitas aplicações.

Teorema A.2.3. [da Convergência Dominada] *Sejam f e $\{f_k, k \geq 1\}$ funções mensuráveis, tal que $f_k \to f$, em μ-q.t.p. Suponha que $|f_k| \leq g$ em μ-q.t.p com g integrável. Então*

$$\lim_{k\to\infty} \int f_k d\mu = \int f d\mu.$$

Terminamos esta seção observando que quando $(\Omega, \mathcal{F}, \mu) = (\mathbb{R}, \mathcal{B}, \lambda)$ onde λ é a medidade de Lebesgue, é comum escrever (por analogia com a integral de Riemann) $\int f dx$ em vez de $\int f d\lambda$.

A.3 Análise funcional

A seguir todos os espaços vetoriais considerados serão sobre o corpo $\mathbb{R}$.

Um *espaço de Banach* é um espaço vetorial normado completo, isto é

Definição A.3.1. Um espaço vetorial normado $(X, \|\cdot\|)$ é chamado de espaço de Banach se toda sequência de Cauchy em X converge para um elemento de X, ou seja: se $\{x_n\}_{n\geq 1} \subset X$ e

$$\forall \varepsilon > 0, \exists N \geq 1 \text{ tal que } \|x_n - x_m\| < \varepsilon, \quad \forall n, m \geq N,$$

então existe $x \in X$ tal que $\lim_{n\to\infty} \|x_n - x\| = 0$.

Um *espaço de Hilbert* é um espaço de Banach cuja norma é induzida por um produto interno.

Definição A.3.2. Um espaço vetorial H é um espaço de Hilbert se ele é equipado de um produto interno

$$\langle \cdot, \cdot \rangle : H \times H \to \mathbb{R},$$

isto é:

1. $\langle \alpha x + \beta y, z \rangle = \alpha \langle x, z \rangle + \beta \langle y, z \rangle$,
2. $\langle x, y \rangle = \langle y, x \rangle$,
3. $\langle x, x \rangle \geq 0$, com igualdade somente se $x = 0$,

tal que H é completo considerando a norma induzida

$$\|x\| := \sqrt{\langle x, x \rangle}.$$

Definição A.3.3. Sejam X e Y espaços vetoriais. Uma aplicação $T : X \to Y$ é um *operador linear* (ou transformação linear) se satisfaz:

$$T(\alpha x + \beta y) = \alpha T(x) + \beta T(y), \quad \text{para todos } x, y \in X \text{ e } \alpha, \beta \in \mathbb{R}.$$

Se X e Y são espaços normados, o operador linear T é contínuo se existe uma constante $C > 0$ tal que

$$\|T(x)\|_Y \leq C \|x\|_X, \quad \forall x \in X.$$

Um operador linear T é chamado de *isometria* se $\|T(x)\|_Y = \|x\|_X$ para todo $x \in X$. T é chamado de *contração* se $\|T(x)\|_Y \leq \|x\|_X$ para todo $x \in X$.

Proposição A.3.1. Seja H um espaço de Hilbert e S um subespaço fechado de H. Existe um operador linear $\Pi : H \to H$ tal que

$$\|x - \Pi x\| = \min_{y \in S} \|x - y\|.$$

Π é chamada de projeção ortogonal sobre S. Em particular Π é uma contração.

Teorema A.3.1. (Extensão de Isometrias em Espaços de Hilbert) *Sejam H e H' espaços de Hilbert e $V \subseteq H$ um subespaço vetorial denso em H. Seja uma isometria $T : V \to H'$. Então, existe um único operador linear $\widetilde{T} : H \to H'$ tal que:*

1. *$\widetilde{T}$ é uma isometria de H em H'.*
2. *$\widetilde{T}$ estende T, isto é,*

$$\widetilde{T}(x) = T(x), \quad \forall x \in V.$$

Espaços L_p

Seja $(\Omega, \mathcal{F}, \mu)$ um espaço de medida. Definiremos os espaços $L_p(\Omega, \mathcal{F}, \mu)$ para $p \in [1, \infty)$. Para começar definimos, para $p \in [1, \infty)$, o espaço vetorial

$$\mathcal{L}_p(\Omega, \mathcal{F}, \mu) = \Big\{ f : \Omega \to \mathbb{R}, f \text{ mensurável tal que} \int |f|^p d\mu < \infty \Big\}.$$

Observe que o espaço $\mathcal{L}_1(\Omega, \mathcal{F}, \mu)$ corresponde ao espaço das funções integráveis com respeito à μ. Consideramos a relação $\sim$ em $\mathcal{L}_p(\Omega, \mathcal{F}, \mu)$ definida por: $f \sim g$ se $f = g$ em μ-quase todo ponto. É fácil mostrar que esta relação é uma relação de equivalência em $\mathcal{L}_p(\Omega, \mathcal{F}, \mu)$ e portanto podemos definir o espaço quociente

$$L_p(\Omega, \mathcal{F}, \mu) := \mathcal{L}_p(\Omega, \mathcal{F}, \mu)/ \sim .$$

Intuitivamente, no espaço $L_p(\Omega, \mathcal{F}, \mu)$ identificamos funções iguais em μ-q.t.p. Assim, é comum identificar uma função mensurável f com a sua classe de equivalência e escrever simplesmente $f \in L_p(\Omega, \mathcal{F}, \mu)$. Agora, considere a aplicação, $\| \cdot \|_p : L_p(\Omega, \mathcal{F}, \mu) \to \mathbb{R}_+$,

$$f \mapsto \Big(\int |f|^p d\mu \Big)^{1/p}.$$

Temos a seguinte

Proposição A.3.2. *Para todo $p \in [1, \infty)$, $(L_p(\Omega, \mathcal{F}, \mu), \| \cdot \|_p)$ é um espaço de Banach.*

No caso $p = 2$, a norma $\| \cdot \|_2$ é induzida pelo produto interno

$$\langle f, g \rangle := \int fg \, d\mu,$$

isto é, $\|f\|_2 = \sqrt{\langle f, f \rangle}$. Como consequência imediata da proposição anterior temos a

Proposição A.3.3. *$(L_2(\Omega, \mathcal{F}, \mu), \langle \cdot, \cdot \rangle)$ é um espaço de Hilbert.*

No caso em que μ é uma medida finita temos a seguinte

Proposição A.3.4. *Se μ é uma medida finita temos que se $p > p'$, $L_p(\Omega, \mathcal{F}, \mu) \subset L_{p'}(\Omega, \mathcal{F}, \mu)$. No caso em que μ é uma probabilidade, se $f \in L_p(\Omega, \mathcal{F}, \mu)$, $\|f\|_{p'} \leq \|f\|_p$.*

Definição A.3.4. *Seja T um operador linear de $L_p(\Omega, \mathcal{F}, \mu) \to L_q(\Omega, \mathcal{F}, \mu)$, para $p,\ q \geq 1$. Dizemos que T é um operador* positivo *se, para $f \geq 0$, tivermos $Tf \geq 0$.*

Referências

1. R. G. Bartle (2001). *A Modern Theory of Integration.* AMS.

2. A. C. Berry (1941). The accuracy of the Gaussian approximation to the sum of independent variates. *Transactions of the American Mathematical Society*, **49**, 122–136.

3. P. Billingsley (1966). Convergence of types in k-space. *Z.W. verw. Geb.*, **5**, 175–179.

4. P. Billingsley (1978). *Ergodic Theory and Information.* Krieger Publishing.

5. P. Billingsley (1995). Probability and Measure. Third Edition. New York: Wiley.

6. P. Billingsley (1999). *Convergence of Probability Measures.* Second Edition. New York: Wiley.

7. F. Black and M. Scholes (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, **81**, 635–654.

8. L. Breiman (1968). *Probability.* Addison Wesley.

9. L. Breiman (1992). *Probability.* SIAM: Classics in Applied Mathematics,7.

10. F. Comets et T. Meyre (2015). *Calcul stochastique et modèles de diffusions.* 2ème édition. Dunod.

11. H. Cramér (1937). Random variables and probability distributions. Cambridge Tracts is Mathematics, No. 36. Cambridge

12. K. L. Chung (1967). *Markov Chains with Stationary Transition Probabilities.* Second Edition. Springer.

13. K. L. Chung (1968). *A Course in Probability Theory.* New York: Harcourt, Brace & World, Inc.

14. K. L. Chung (1974). *A Course in Probability Theory.* Second Edition. Academic Press.

15. B. de Finetti (1974). *Theory of Probability*, Vol. 1. New York: Wiley.

16. M. Donsker(1951). An invariance principle for certain probability limit theorems. *Memoirs of the American Mathematical Society*, **6**.

17. J. L. Doob (1953). *Stochastic Processes.* New York: Wiley.

18. J. L. Doob (1971). What is a martingale? *American Mathematical Monthly*, **78**, 451–463.

19. P. Doukhan. (2015). *Probabilistic and Statistical Tools for Modelling Time Series.* Rio de Janeiro: IMPA.

20. P. G. Doyle and J.L. Snell (1984). *Random Walks and Electrical Networks.* Carus Mathematical Monographs, Mathematical Association of America.

21. R. Durrett (1996a). *Probability: Theory and Examples.* 2nd Edition. Duxbury.

22. R. Durrett (1996b). *Stochastic Calculus: A Practical Introduction.* CRC Press.

23. A. Dvoretzky, P. Erdös and S. Kakutani. (1950). Double points of paths of Brownian motion in n-space. *Acta Sci. Math.* (Szeged), **12**, 75–81.

24. A. Dvoretzky, P. Erdös and S. Kakutani. (1954). Multiple points of paths of Brownian motion in the plane. *Bull. Res. Council Israel*, **3**, 364-371.

25. E. B. Dynkin (1957). Inhomogeneous strong Markov processes. *Dokl. Akad. Nauk. SSSR*, **113**, 261–263.

26. P. Erdös and M. Kac (1946). On certain limit theorems in theory of probability. *Bulletin of the American Mathematical Society*, **52**, 292–302.

27. C.-G. Esseen (1942). On the Liapounoff limit of error in the theory of probability. *Ark. Mat. Astron. Fys.* **A28**, 1–19.

28. L. C. Evans. (2013). *An Introduction to Stochastic Differential Equations.* American Mathematical Society.

29. R. Fano (1961). *Transmission of Information: A Statistical Theory of Communications.* Cambridge: Massachusetts, M.I.T. Press.

30. K. Fazli and J. Behboodian (1995). Skorohod's theorem and convergence of types. *Statistics & Probability Letters*, **24**, 243–244

31. W. Feller (1968). *An Introduction to Probability Theory and Its Applications.* Volume I. Third Edition. New York: Wiley.

32. W. Feller (1966). *An Introduction to Probability Theory and Its Applications.* Volume II. New York: Wiley.

33. D. Freedman (1983). *Markov Chains*. Springer.

34. A. M. Garsia (1970). *Topics in Almost Everywhere Convergence*. Markham Publishing Co.

35. V. Glivenko (1936). Sul teorema limite della teoria delle funzioni caratteristische. *Giorn. Ist. Ital. Attvar*, 16—167.

36. B. V. Gnedenko (1989). *The Theory of Probability and The Elements of Statistics*. Chelsea Publishing Company.

37. A. Gut (2013). *Probability: A Graduate Course*. Second Edition. Springer.

38. P. R. Halmos (1976). *Measure Theory*. Springer.

39. P. R. Halmos (2006). *Lectures on Ergodic Theory*. AMS Chelsea Publishing Co.

40. E. Hewitt and L. J. Savage. (1955). Symmetric measures on Cartesian products. *Transactions of the American Mathematical Society*, **80**, 470–501.

41. H. Heyer and S. Kawakami (2005). Paul Lévy's continuity theorem: Some history and recent progress. *Bulletin Nara University Education*, **54**, 11–19.

42. G. A. Hunt (1956). Some theorems concerning Brownian motion. *Trans. Amer. Math. Soc.*, **81**, 294–319.

43. K. Itô (2006). *Essentials of Stochastic Processes*. American Mathematical Society, Mathematical Monographs, Volume 231.

44. H. Jeffreys. (1939). *Theory of Probability*. First edition. Oxford: The Clarendon Press.

45. I. Karatzas and S.E. Shreve (1988). *Brownian motion and stochastic calculus.* Graduate Texts in Mathematics, 113. New York: Springer-Verlag.

46. M. Kac (1947). On the notion of recurrence in discrete stochastic processes. *Bulletin of the American Mathematical Society*, **53**, 1002–1010.

47. O. Kallenberg (2002). *Foundations of Modern Probability*. Second Edition. Springer.

48. A. N. Kolmogorov (1931). Eine Verallgemeinerung des Laplace-Liapouno]fschen Satzes. *Izv. Akad. Nauk SSSR. Set. Fiz.-Mat.*, 959–962.

49. A. N. Kolmogorov (1933). *Grundbegriffe der Wahrscheinlichkeitsrechnung.* Berlin.

50. A. N. Kolmogorov (1956). *Foundations of The Theory of Probability.* Second English Edition. Chelsea Publishing Company.

51. A. N. Kolmogorov (1933). Sulla determinazione empirica di una legge di distribuzione. *Giorn. Inst. ital. Attuari.*, **4**, 83—91.

52. V. Yu. Korolev and I.G. Shevtsova (2010). On the upper bound for the absolute constant in the Berry-Esseen inequality. *Theory of Probability and Applications*, **54**, 638–658.

53. V. M. Kruglov (1998). To the invariance principle. *Theory of Probability and Its Applications*, **42**, 225–243.

54. J. Lamperti (1966). *Probability: A Survey of the Mathematical Theory.* New York: W. A. Benjamin, Inc.

55. J. Lamperti (1977). *Stochastic Processes: A Survey of the Mathematical Theory.* New York: Springer-Verlag.

56. P. Laplace (1795). *Essai Philosophique sur les Probabilités.* Paris: Christian.

57. J-F. Le Gall (2016). *Brownian Motion, Martingales and Stochastic Calculus.* Graduate Texts in Mathematics, 274. Springer.

58. P. Lévy (1925). *Calcul de Probabilités.* Paris: Gauthier-Villars.

59. M. Loève (1963, 1978). *Probability Theory.* Vol. I and II. Springer.

60. R. C. Merton (1973). The theory of option pricing. *Bell Journal of Economics and Management Sciences*, **4**, 141–183.

61. P. Mörters and Y. Peres (2010). *Brownian Motion.* Cambridge Series in Statistical and Probabilistic Mathematics.

62. J. Neveu (1975). *Discrete-parameter Martingales.* North-Holland.

63. J. R. Norris (1997). *Markov Chains.* Cambridge University Press.

64. K. R. Parthasaraty (2005). *Probabilty Measures on Metric Spaces.* AMS Chelsea Publishing.

65. Y. V. Prokhorov (1956). Convergence of random processes and limit theorems in probability theory. *Theory of Probability and Applications, I* (in English translation) **2**, 157–214.

66. G. F. Simmons (2003). *Introduction to Topology and Modern Analysis.* Reprint Edition. Krieger Publishing Company.

67. J. Ville (1939). *Étude Critique de la Notion de Collectif.* Paris: Gauthier-Villars.

68. R. von Mises. (1931). *Wahrscheinlichkeitsrechnung.* Leipzig and Vienna.

69. R. von Mises. (1939). *Probability, Statistics and Truth.* New York: Macmillan.

70. U.F. Wiersema (2008). *Brownian Motion Calculus.* New York: Wiley.

71. D. Williams (1991). *Probability with Martingales.* Cambridge University Press.

Soluções de Problemas Selecionados

Capítulo 1

3) Note que $f(X)$ pode ser vista como $f \circ X : \omega \to f(X(\omega))$, logo $(f \circ X)^{-1} = X^{-1} \circ f^{-1}$ e, portanto, para todo $B \in \mathcal{B}$, $(f \circ X)^{-1}(B) = (X^{-1} \circ f^{-1})(B) = X^{-1}(f^{-1}(B)) \in \mathcal{F}$, pois $f^{-1}(B) \in \mathcal{B}$.

4) (i) Se $x_1 \leq x_2$, então $(-\infty, x_1] \subseteq (-\infty, x_2]$, donde $F_X(x_1) \leq F_X(x_2)$.

   (ii) Se $x_n \to -\infty$, então $A_n = (-\infty, x_n]$ são encaixados e de intersecção vazia. Pelo Corolário 1.1, temos que $P(A_n) \to 0$. Similarmente, para $x_n \to \infty$.

   (iii) Se $x_n \downarrow x$, então $A_n = (-\infty, x_n] \downarrow (-\infty, x]$, logo $F_X(x_n) \to F_X(x)$.

10) Considere a classe $\mathcal{A} = \{A \in \mathcal{F} : \forall \varepsilon > 0, \text{existe} A_\varepsilon \text{ tal que} P(A \Delta A_\varepsilon) < \varepsilon\}$ e mostre que $\mathcal{F} \subset \mathcal{A}$. Primeiramente, note que $\mathcal{A}$ é não vazia, pois $\emptyset \in \mathcal{A}$ e $\mathcal{A}$ contém $\mathcal{F}_0$, pois é suficiente tomar $A = A_\varepsilon$. Em segundo lugar, $\mathcal{A}$ é uma classe monotônica (Prove!). Pelo teorema das CM, $\mathcal{A}$ contém $\mathcal{F}$ e portanto todo conjunto de $\mathcal{F}$ pertence a $\mathcal{A}$, e o resultado segue.

11) Devemos provar que existe uma sequência de constantes b_n tais que

$$\lim_{m \to \infty} P[\cup_{n=m}^{\infty} |X_n/b_n - 0| > \epsilon] = 0,$$

para todo $\epsilon > 0$. Mas, para cada inteiro n, podemos encontrar b_n tal que $P(|X_n| > b_n/n) \leq 1/2^n$. Logo,

$$P(\cup_{n=m}^{\infty} |\frac{X_n}{b_n}| > \frac{1}{n}) \leq \sum_{n=m}^{\infty} P(|\frac{X_n}{b_n}| > \frac{1}{n}) \leq \sum_{n=m}^{\infty} \frac{1}{2^n}$$

e quando $m \to \infty$, esta é a cauda de uma série convergente, logo tende a zero e o resultado segue com $\epsilon = 1/n$.

15) Para (a), tome $A_n = [0, 1/n]$ sobre $[0, 1]$ com medida uniforme; $P(A_n) = 1/n, \sum 1/n = \infty$ e $P(A_n \text{ i.v.}) = 0 < 1$.

    Para (b) tome $A_n = [0, 1 - 1/n]$ sobre $[0, 1]$, com medida uniforme ; $P(A_n) = 1 - 1/n$, $\sum(1 - 1/n) = \infty$ e $P(A_n \text{ i.v.}) = 1$.

16) Depois de n passos, removemos $2^n - 1$ intervalos abertos disjuntos e ficamos com 2^n intervalos fechados disjuntos. Chamemos de B_n a reunião dos removidos. Então,

$$P(B_n) = \frac{1}{3} + \frac{2}{3^2} + \frac{4}{3^3} + \ldots + \frac{2^{n-1}}{3^n} = 1 - \left(\frac{2}{3}\right)^n.$$

    Quando $n \to \infty$, $B_n \to B$, aberto, e $P(B_n) \to P(B) = 1$, logo $P(C) = 1 - P(B) = 0$.

18) (1) Prove que $\mathcal{F}$ é uma σ-álgebra sobre Ω; para isso mostre que $\Omega \in \mathcal{F}$, $\mathcal{F}$ é fechada sob complementos e é fechada sob reuniões enumeráveis.

    (2) Prove que P é uma medida de probabilidade sobre $(\Omega, \mathcal{F})$. Para isso, mostre que $P(\Omega) = 1$, $P(\emptyset) = 0$ e é enumeravelmente aditiva.

Capítulo 2

2) Cada $\omega \in [0, 1]$ tem expansão binária $\omega = 0.a_1a_2a_3\cdots$, com $a_i \in \{0, 1\}$ e $X_n(\omega) = a_n$. Temos que provar que

$$P(X_{n_1} = b_1, \ldots, X_{n_k} = b_k) = \prod_{j=1}^{k} P(X_{n_j} = b_j),$$

   para qualquer conjunto de índices $n_1, \ldots, n_k$ e valores $b_1, \ldots, b_k$. Calcule $P(X_n = b)$ para um n fixo e $b \in \{0, 1\}$. $\{X_n = b\}$ é o conjunto de números cujo n-ésimo dígito binário é b, logo a reunião de intervalos de comprimento 2^{-n}. Verifique que a probabilidade do primeiro termo acima é 2^{-k}. Mostre que o segundo termo é $\prod_{j=1}^{k}(1/2) = 1/2^k$.

6) Se $S_n/n \to Y$ *q.c*, então para quase todo ω, $S_n(\omega)/n \to Y(\omega)$. Use a Lei 0-1 de Kolmogorov; pela independência, todo evento caudal tem probabilidade zero ou um. Mostre que Y é uma variável mensuravel relativamente à σ-álgebra caudal. Isso implicará que Y é constante em q.t.p.

8) Temos que $A = \{\omega : X_n(\omega) \in B_n \text{ iv}\}$. Tome $B \in \mathcal{B}^\infty$ como segue: $B=$ conjunto de todas as sequências $(x_1, x_2, \ldots)$ tal que $x_n \in B_n$ i.v. Então,

$A = X^{-1}(B)$, onde $X = (X_1, X_2, \ldots)$. Seja σ uma permutação de $1, 2, \ldots, N$ e seja $Y = (X_{\sigma_1}, X_{\sigma_)2}, \ldots)$. Note que $X_n = X_{\sigma_n}$, para todo $n > N$. Então, $Y^{-1}(B) = \{\omega : X_{\sigma_n}(\omega) \in B_n \text{ i.v}\} = X^{-1}(B)$. De fato, seja $\omega \in X^{-1}(B)$, que vale se, e somente se, $X_n(\omega) \in B_n$, se e somente se $X_n \in B_n$ para um número infinito de n, se e somente se $X_n(\omega)$ pertence a algum dos $B_n, B_{n+1}, \ldots$,não importa quão grande seja n - a cauda da sequência B_n, se e somente $X_{\sigma_n}(\omega) \in B_n$ i.v, se e somente se $\omega \in Y^{-1}(B)$, com a escolha de $n > N$.

10) Suponha $\Lambda = \{\omega : X_n(\omega) \text{converge}\}$. Por Kolmogorov, ou $P(\Lambda) = 0$ ou $P(\Lambda) = 1$. Suponha $P(\Lambda) = 1$. Então, o valor limite deve ser uma constante c. Então, $\lim_{n\to\infty} P(|X_n - c| > \varepsilon) = 0$, para todo $\varepsilon > 0$. Como $X_1 \sim X_n$, $P(X_1 = c) = 1$, uma contradição com a hipótese que a distribuição de X não esteja concentrada num único ponto.

11) (a) Temos

$$P\left(\frac{X_n}{\log n} > 2\right) = P(X_n > \log n^2) =$$

$$1 - P(X_n \leq \log n^2) = 1 - F(\log n^2) = e^{-\log n^2} = \frac{1}{n^2}.$$

Logo,

$$\sum_{n\geq 2} P\left(\frac{X_n}{\log n}\right) = \sum_{n\geq 2} \frac{1}{n^2},$$

e por Borel-Cantelli, $P(\frac{X_n}{\log n} > 2 \text{ i.v}) = 0$.

b) Tome

$$P\left(\frac{X_n}{\log n} > 1\right) = P(X_n > \log n) = 1 - p(X_n \leq \log n)$$

$$= 1 - F(\log n) = e^{-\log n} = \frac{1}{n}.$$

Logo,

$$\sum_{n=2}^{\infty} P\left(\frac{X_n}{\log n} > 1\right) = \sum_{n\geq 2} \frac{1}{n} = \infty,$$

e lembrando que os X_i são independentes, por Borel-Cantelli o resultado segue.

16)
$$E(|X+Y|) = \int\int |x+y| dF_X(x)dF_Y(y),$$

$$= \int_{\mathbb{R}} [\int_{\mathbb{R}} |x+y| dF_X(x)] dF_Y(y)$$

a primeira igualdade pela indepenência e a segunda por Fubini.

Portanto, para pelo menos um y_0, $\int |x+y_0| dF_X(x) < \infty$. Segue que

$$\int |x| dF_X(x) \leq \int |x+y_0| dF_X(x) + |y_0| < \infty.$$

19) (a) $Var(X_1 + \ldots + X_n) = \sum_{i=1}^n \mathrm{Var}(X_i) + 2\sum\sum_{i<j} \mathrm{Cov}(X_i, X_j)$, mas como as X_i são não correlacionadas, $\mathrm{Cov}(X_i, X_j) = 0$, para todo i, j e o resultado segue.

(b) Seja $Y_n = \sum X_i/n$; pela desigualdade de Chebyshev,

$$P(|Y_n - E(Y_n)| \geq \varepsilon) \leq \frac{\mathrm{Var}(Y_n}{\varepsilon^2} =$$

$$\frac{1}{n^2\varepsilon^2} \sum_{k=1}^n \sigma_k^2.$$

Logo, pela hipótese que $\sum_{k=1}^n \sigma_k^2/n^2 \to 0$, para $n \to \infty$, obtemos que $\lim_{n\to\infty} P(|Y_n - E(Y_n)| \geq \varepsilon) \to 0$, ou seja, $Y_n - (\mu_1 + \ldots + \mu_n)/n \to 0$ em probabilidade. Também, por hipótese de (b), $E(|Y_n - E(Y_n)|^2 \to 0$, se $n \to \infty$, ou seja, $Y_n - E(Y_n) \to 0$ em L_2.

22) Para $\mathcal{F}_1$, os átomos são $A = A_1 = \{1,3,5,7\}$ and $A^c = A_2$, logo $\mathcal{F}_1 = \{\emptyset, \{1,3,5,7\}, \{2,4,6,8\}, \Omega\}$. Faça o mesmo para $\mathcal{F}_2$ e chame os átomos de B_1 e B_2. Calcule $P(A_1 \cap B_1) = 1/4 = P(A_1)P(B_1)$, etc, até $P(A_2 \cap B_2) = 1/4 = P(A_2)P(B_2)$. Verificar que $\emptyset$ and Ω satisfazem $P(\emptyset \cap B) = 0 = P(\emptyset)P(B)$ etc.

Capítulo 3

2) A σ-álgebra $\mathcal{F}$ é finita com átomos $\Lambda_1, \ldots, , \Lambda_n$. Y é uma função tomando valores c_i em cada Λ_i. Logo, Y é mensurável com respeito a $\mathcal{F}$ e, de fato, $\sigma(Y) = \mathcal{F}$ (Prove!). Por definição, $E(X|Y) = E(X|\sigma(Y)) = E(X|\mathcal{F}$.

6) Seja o e.p. $(\Omega, \mathcal{G}, P)$ e $\mathcal{F} \subset \mathcal{G}$. (a) Chamemos $Z = E(X|\mathcal{F})$; por definição, $E(X|\mathcal{F})$ é $\mathcal{F}$-mensurável e integrável. Então, como ZY é integrável e Z é $\mathcal{F}$-mensurável, teremos

$$E(ZY) = E[E(ZY|\mathrm{F})] = E[ZE(Y|\mathrm{F})] = E[E(X|\mathrm{F})E(Y|\mathrm{F})].$$

(b) De modo análogo, seja $W = E(Y|\mathrm{F})$, que é F-mensurável e integrável, de modo que

$$E(XW) = E[E(XW|\mathrm{F})] = E[WE(X|\mathrm{F})] = E[E(Y|\mathrm{F})E(X|\mathrm{F})].$$

de (a) e (b), $E(ZY) = E(XW)$, e o resultado segue.

7) Suponha, por absurdo, que $E(X|\mathcal{F})$ seja não negativa q.c. Então existe $B \in \mathcal{F}$ tal que $E(X|\mathcal{F}) < 0$ sobre B. Defina $B = \{\omega \in \Omega : E(X|\mathcal{F})(\omega) < 0\}$. Então $B \in \mathcal{F}$. Considere $\int_B E(X|\mathcal{F})dP = \int_B XdP$. Pela definição de B a primeira integral é negativa. Como $X \geq 0$ q.c., a segunda integral é não-negativa, uma contradição.

9)
$$\mathrm{Var}[E(Y|\mathrm{F})] = E[E(Y|\mathrm{F})]^2 - [E(E(Y|\mathrm{F}))]^2$$

$$\leq E[E(Y^2)|\mathrm{F})] - ([E(Y)]^2,$$

pela desigualdade de Jensen, e o lado direito da desigualdade é igual a $E(Y^2) - [E(Y)]^2 = \mathrm{Var}(Y)$.

12) Temos que $P_Y(B-X)$ é $\mathrm{F}(X)$-mensurável, logo é suficiente verificar que $P(X + Y \in B|X)$ satisfaz a definição de esperança condicional. Seja $\Lambda \in \mathrm{F}(X)$, então $\Lambda = X^{-1}(A)$, para algum $A \in \mathcal{B}^1$. Segue que

$$\int_\Lambda P_Y(B - X)dP = \int_A P_Y(B - X)P_X(dx).$$

Seja $\lambda = P_X P_Y$ e use o teorema de Fubini no lado direito:

$$\int_A P_X(dx) \int_{\{x+y \in B\}} P_Y(dy) = \int \int_{\{x \in A, x+y \in B\}} \lambda(dx, dy) =$$

$$\int_{\{X \in A, X+Y \in B\}} dP = P(X \in A, X + Y \in B)$$

e o resultado segue.

15) Tome $E(S_n|S_n, S_{n+1}, \ldots)$; como S_n é mensurável relativamente a $\mathrm{F}(S_n, S_{n+1}, \ldots)$, segue que $E(S_n|S_n, S_{n+1}, \ldots) = S_n$. Mas

$$S_n = E(S_n|S_n, S_{n+1}, \ldots) = E(X_1 + \ldots + X_n|S_n, \ldots)$$

$$= E(X_1|S_n, \ldots) + \ldots + E(X_n|S_n, \ldots) = nE(X_1|S_n, \ldots),$$

por simetria, e o resultado segue.

17) Observe que por simetria, (X, Y) tem mesma lei que $(-X, -Y)$. Portanto, temos que para toda função g boreliana e limitada,

$$E(g((X+Y)^2)X) = -E(g((X+Y)^2)X).$$

Assim, $E(g((X+Y)^2)X) = 0$ o que implica que $E(X \mid (X+Y)^2) = 0$.

19) O lado esquerdo da igualdade fica $E(X^2) - 2E[XE(X|\mathcal{F})] + E[E(X|\mathcal{F})^2]$. Como $E(X|\mathcal{F})$ é mensurável relativamente a $\mathcal{F}$, mostre que $E[XE(X|\mathcal{F})] = E[E(X|\mathcal{F})^2]$, usando a lei de esperanças iteradas. A igualdade segue.

Capítulo 4

4) Provemos para o caso de um submartingale. Nesse caso, $E(X_{n+1}) \geq E(X_n)$, para todo n. Como $E(X_n) = E(X_1)$, para todo n, então $E(X_{n+1}) = E(X_n)$, para todo n. Defina $Y_n = X_{n+1} - X_n$. Segue que $E(Y_n|\mathcal{F}_n) \geq 0$ e que $E(Y_n) \geq 0$ (Por que?). Como $E(X_{n+1}) = E(X_n)$, obtemos $E(Y_n) = 0 = E[E(Y_n|\mathcal{F}_n)]$, de onde segue $E(Y_n|\mathcal{F}_n) = 0$ e, portanto, $E(X_{n+1}|\mathcal{F}_n) = X_n$ q.c.

10) (a) $\mathcal{F}$ é gerada por reuniões de conjuntos da forma $\{n\}$ e $\Omega - \{1, 2, \ldots, n\}$, logo é suficiente verificar que $E(X|\mathcal{F}_{n-1}) = X_{n-1}$ para tais conjuntos. Claramente vale para $\Lambda = \{n\}$ e para $\Lambda = \Omega - \{1, 2, \ldots, n-1\}$ temos

$$\int_\Lambda X_n dP = \int_{\Omega - \{1,2,\ldots,n} X_n dP + \int_{\{n\}} X_n dP = nP(k > n) - P(k = n) =$$

$$\frac{n}{n+1} - (\frac{1}{n} - \frac{1}{n+1}) = \frac{n-1}{n} =$$

$$\int_{\Omega - \{1,\ldots,n-1\}} X_{n-1} dP = (n-1)P(k > n-1) = \frac{n-1}{n}.$$

Agora $\lim_{n\to\infty} X_n = -1$ e o resultado segue.

(b) Seja $\lim_a \limsup_n \int_{|X_n|>a} |X_n|dP$; tome $k > n > a$ $(a \to +\infty)$. Temos que essa expressão é igual a

$$\lim_a \limsup_n [nP(k > n)] = \lim_a \limsup_n \frac{n}{n+1} = 1$$

logo X_n não é u.i. por que o limite deveria ser zero.

(c) Mostre que $P(\sup_n |X_n| \geq \lambda) = 1/(\lambda + 1)$. Pelo teorema,

$$P(\max_{1\leq k\leq n} |X_k| \geq \lambda) \leq \frac{1}{\lambda} E(|X_n|).$$

Mas $E(|X_n|) = 2n/(n + 1)$ (prove!), portanto

$$P(\max_{1\leq k\leq n} |X_k| > \lambda) \leq \frac{1}{\lambda}\frac{2n}{n+1}.$$

Para $n \to \infty$, $P(\sup_k |X_k| \geq \lambda) \leq 2/(\lambda)$ e este limite superior é maior que $1/(\lambda + 1)$, para $\lambda > -1$.

11) (a) Considere os martingales $\{Y_1, Y_1 + Y_2\}$ e $\{Y_2, Y_2\}$. Então Y_1 e Y_2 são i.i.d, média zero (Prove!). A soma é $\{Y_1 + Y_2, Y_1 + 2Y_2\}$, que não é um martingale pois

$$E(Y_1 + 2Y_2|Y_1 + Y_2) = Y_1 + Y_2 + E(Y_2|Y_1 + Y_2) = (3/2)(Y_1 + Y_2).$$

Por exemplo, podemos tomar Y_1 e Y_2 iid, resultantes do lançamento de duas modedas honestas.

(b) Tome $X_1 = -1$, $X_2 = 0$, $\mathcal{F}_1 = \mathcal{F}_2 = \{\emptyset, \Omega\}$. Então, $E(X_2|\mathcal{F}_1) = 0 > -1 = X_1$ e $E(|X_2||\mathcal{F}_1) = 0 < 1 = |X_1|$.

13) Primeiro, observe que $|E(X|\mathcal{B})| \leq E(|X||\mathcal{B})$. Logo,

$$\int_{\{|E(X|\mathcal{B})|>\lambda\}} |E(X|\mathcal{B})|dP \leq \int_{\{|E(X|\mathcal{B})|>\lambda\}} E(|X||\mathcal{B})dP. \tag{12.15}$$

Pela definição de EC, notando que o conjunto $\{\omega : |E(X|\mathcal{B})|\lambda\}$ é $\mathcal{B}$-mensurável, temos que

$$\int_{\{|E(X|\mathcal{B})|>\lambda\}} |E(X|\mathcal{B})|dP = \int_{\{|E(X|\mathcal{B})|>\lambda\}} |X|dP \tag{12.16}$$

Por Chebyshev,

$$P\{|E(X|\mathcal{B})| > \lambda\} \leq \frac{E(|E(X|\mathcal{B})|)}{\lambda} \leq \frac{E(|X|)}{\lambda}. \tag{12.17}$$

Usando (12.19) e (12.20) temos

$$\int_{\{|E(X|\mathcal{B})|>\lambda\}} |E(X|\mathcal{B})| dP \leq \int_{\{|E(X|\mathcal{B})|>\lambda\}} |X| dP, \tag{12.18}$$

sujeitos a (12.17). Mas X é integrável, então dado $\varepsilon > 0$, existe $\delta > 0$ tal que se $P(A) < \delta$, $\int_A |X| dP < \varepsilon$. Logo, para $\lambda \to \infty$, $\lambda > \frac{E(|X|)}{\delta}$, temos que $P\{|E(X|\mathcal{B})| > \lambda\} < \delta$ e, portanto,

$$\int_{\{|E(X|\mathcal{B})|>\lambda\}} |X| dP < \varepsilon$$

e

$$\int_{\{|E(X|\mathcal{B})|>\lambda\}} |E(X|\mathcal{B})| dP < \varepsilon,$$

logo $\{E(X|\mathcal{B})\}$ é uniformemente integrável.

18) Temos que

$$E(X_T|\mathcal{F}_s) = E[E(X_\infty|\mathcal{F}_T)|\mathcal{F}_s] = E(X_\infty|\mathcal{F}_s) = X_s.$$

Logo, é suficiente provar que $E(X_\infty|\mathcal{F}_T) = X_T$

Seja $A \in \mathcal{F}_T$. É suficiente provar que $\int_{A\cap\{T<\infty\}} X_T = \int_{A\cap\{T<\infty\}} X_\infty$, porque $X_T = X_\infty$ sobre $\{T = \infty\}$.

Seja $t_k = T \wedge k$, então $A \cap \{T \leq k\} \in \mathcal{F}_{T_k}$, porque $A \cap \{T \leq k\} \cap \{T_k \leq j\} \in \mathcal{F}_j$, para todo j. Usando o TAO,

$$\int_{A\cap\{T\leq k\}} X_{T_k} = \int_{A\cap\{T\leq k\}} X_k, \quad \text{pois } T_k \leq k. \tag{12.19}$$

Também, $X_{T_k} = X_{T\wedge k} \to X_T$, q.c. quando $k \to \infty$ e $E(X_\infty|\mathcal{F}_{T_k}) = X_{T_k}$. Isso é verdade porque

$$\int_\Lambda X_{T_k} = \int_\Lambda X_k = \int_\Lambda X_{k+n} = \int_\Lambda X_\infty,$$

para todo $\Lambda \in \mathcal{F}_{T_k}$. A segunda igualdade segue pela definição de martingale e a terceira pela i.u.

Agora, a equação (12.19) pode ser escrita

$$\int_{A\cap\{T\leq k\}} X_{T_k} = \int_{A\cap\{T\leq k\}} X_{k+n} = \int_{A\cap\{T\leq k\}} X_\infty, \tag{12.20}$$

para todo n, usando propriedades de martingales, pois $A \cap \{T \leq k\} \in \mathcal{F}_{T_k}$.

Agora, pelo Problema 13, $\{X_{T_k}, k \geq 1\}$ é u.i. Para $k \to \infty$ em (12.20) obtemos

$$\int_{A\cap\{T<\infty\}} X_T = \int_{A\cap\{T<\infty\}} X_\infty.$$

Capítulo 5

7) Seja $U_{ab}^{(*)}$ o número de cruzamentos de (a, b) por $X_{s_1^{(n)}}, \ldots, X_{s_n^{(n)}}$. Pelo resultado do caso discreto, temos que

$$E(U_{ab}^{(n)}) \leq \sup_{1\leq k\leq n} \frac{E(X^+_{s_k^{(n)}}) + a}{b-a} \leq \sup_t \frac{E(X_t^*) + a}{b-a}$$

Agora, para $n \to \infty$, $U_{ab}^{(n)} \to U_{ab}$ (cresce) portanto

$$E(U_{ab}) \leq \sup_t \frac{E(X_t^+) + a}{b-a}.$$

9) O processo de Poisson é um processo de Lévy porque:

(a) tem trajetórias contínuas à direita com limites à esquerda (cadlag);

(b) tem incrementos independentes;

(c) tem incrementos estacionários;

(d) é contínuo em probabilidade.

Argumento similar para o Movimento Browniano.

12) Temos que

$$E(X_t) = E(W_t) - tE(W_t) = 0 - t \cdot 0 = 0.$$

Para a função de autocovariância, obtemos

$$\begin{aligned}
\gamma(t, s) &= \mathrm{Cov}(X_t, X_s) \\
&= E(X_t X_s) \\
&= E(W_t W_s) - sE(W_t W_1) - tE(W_s W_1) + tsE(W_1^2) \\
&= \min\{t, s\} - st - ts + ts \cdot 1 \\
&= \min\{t, s\} - ts.
\end{aligned}$$

14) Um processo de Poisson N_t é contínuo em probabilidade pois:

(a) Para cada t, N_t é finito quase certamente;

(b) N_t é cadlag;

(c) Para cada $t > 0$, $N_{t-} = N_t$ com probabilidade 1;

(d) O resultado segue de (c) pois convergência q.c implica convergência em probabilidade.

Capítulo 6

1) (i) Sejam F_n e F as f.d. de X_n e X, respectivamente, e $G_n(x)$, G as f.d. de $X_n + Y_n$ e de $X + c$, respectivamente. Seja $A_n = \{\omega : |Y_n(\omega) - c|\} \leq \varepsilon\}$, logo $P(A_n^c) \leq \varepsilon$, para $n \geq n_0(\varepsilon)$, porque $Y_n \to c$. Seja $E_n = \{\omega : X_n \leq x - Y_n\}$, para $n \geq n_0(\varepsilon)$, então,

$$P(X_n \leq x - Y_n) \leq P(X_n \leq x - Y_n, Y_n \geq c - \varepsilon) + \varepsilon \leq P(X_n \leq x - c + \varepsilon) + \varepsilon.$$

Seja $E_n^{'} = \{\omega : X_n \leq x - c - \varepsilon\}$, para $n \geq n_0(\varepsilon)$, donde segue

$$P(E_n^{'}) = P(X_n \leq x-c-\varepsilon) \leq P(X_n \leq x-c-\varepsilon, Y_n \leq c+\varepsilon)+\varepsilon \leq P(X_n \leq x-Y_n)+\varepsilon.$$

Segue-se que

$$F_n(x - c - \varepsilon) - \varepsilon \leq P(X_n + Y_n \leq x) \leq F_n(x - c + \varepsilon) + \varepsilon,$$

logo

$$F_n(x - c - \varepsilon) - \varepsilon \leq G_n(x) \leq F_n(x - c - \varepsilon) + \varepsilon.$$

Quando $n \to \infty$ e para todo $\delta > 0$, temos

$$G(x - \delta) = F(x - c - \delta) \leq \lim G_n(x) \leq F(x - c + \delta) = G(x + \delta).$$

Como δ é arbitrário, $G_n(x)$ converge para $F(x - c)$.

(iii) Considere $X_n = 0$, com probabilidade (cp) 1/2 e $X_n = 1$, cp 1/2. Também, seja $X = 0$, cp 1/2 e $X = 1$, cp 1/2. Então, $|X_n - X| = 1$, mas $F_{X_n} = F_X$, logo $X_n \to X$ em distribuição, mas X_n não converge para X em probabilidade, pois deveríamos ter $P(|X_n - X| > \varepsilon) \to 0$, para todo $\varepsilon > 0$. Para a segunda parte, seja $\varepsilon > 0$ e escolha δ talque $0 < \delta < \varepsilon$. Então,

$$P(|X_n - c| \geq \varepsilon) \leq P(X_n > c+\varepsilon) + P(X_n \leq c-\delta) = 1 - F_{X_n}(c+\varepsilon) + F_{X_n}(c-\delta).$$

Como F é contínua em $c+\varepsilon$ e $c-\delta$ e $F(c+\varepsilon) = 1$ e $F(c-\delta) = 0$, vemos que $\lim_n P(|X_n - c| \geq) \to 0$.

4) (b) Se π é fechada, para todo $\varepsilon > 0$, existe K compacto tal que $P(K) \geq 1-\varepsilon$ e $Q(K) \geq 1-\varepsilon$, para quaisquer P e Q de π. Considere $P*Q, P, Q \in \pi$. Então,

$$P*Q(K) = \int_S 1_K(x) dP*Q(x) = \int_S \int_S 1_K(x+y) dP(x) dQ(y) = \int_{K\times K} dP(x) dQ(y) =$$

$$= P \times Q(K \times K) = P(K)Q(K) \geq (1-\varepsilon)^2 \geq 1 - 2\varepsilon,$$

logo π^* é fechada.

9) $\pi^* = \{P_n \pi_k^{-1}, n \geq 1\}$ é uma família de medidas de probabilidade definidas sobre $(\mathbb{R}^k, \mathcal{B}^k)$, tal que $P_n \pi_k^{-1}(B) = P_n\{\pi_k^{-1}(B)\}$, para todo $B \in \mathcal{B}^k$.

π^* é fechada (tight) por hipótese, logo pelo teorma de Prokhorov, π^* é relativamente compacta. Então, π^* tem uma subsequência que converge fracamente para alguma medida de probabilidade. Esta subsequência pode ser escolhida por um processo diagonal (conhecido) logo temos que $P_n \pi_k^1 \Rightarrow Q_k$, para cada k e as medidas Q_k são consistentes, de tal sorte que existe uma medida de probabilidade Q sobre $\mathbb{R}^\infty$ tal que $Q\pi_k^{-1} = Q_k$, pelo teorema da extensão de Kolmogorov, portanto $P_n \pi_k^{-1} \Rightarrow Q\pi_k^{-1}$, para cada k. Isso mostra que $\pi = \{P_n, n \geq 1\}$ é relativamente compacta. Mas $\mathbb{R}^\infty$ é completo e separável, logo, de novo, pelo teorema de Prokhorov, π é fechada (tight).

12) Em geral, a classe dos conjuntos P-contínuos não é uma σ-álgebra. De fato, considere o espaço de probabilidade $(\Omega, \mathcal{F}, P)$, onde:

- $\Omega = [0,1]$,
- $\mathcal{F}$ é a σ-álgebra de Borel em $[0,1]$,
- P é a medida de Lebesgue restrita a $[0,1]$ (ou seja, a probabilidade uniforme).

Denote por $\mathcal{A}$ a classe dos subconjuntos mensuráveis de Ω que são P-contínuos, isto é:

$$\mathcal{A} = \{A \in \mathcal{F} : P(\partial A) = 0\},$$

onde ∂A é a fronteira topológica de A.
Vamos mostrar que $\mathcal{A}$ não é uma σ-álgebra. Considere os números racionais em $[0,1]$, que podem ser enumerados como:

$$\mathbb{Q} \cap [0,1] = \{q_1, q_2, q_3, \ldots\}.$$

Para cada $n \in \mathbb{N}$, defina:

$$A_n = \{q_n\}.$$

- Temos que $\partial A_n = \{q_n\}$.
- Como P é a medida de Lebesgue, temos: $P(\partial A_n) = P(\{q_n\}) = 0$.
- Logo, $A_n \in \mathcal{A}$ para todo $n \in \mathbb{N}$.

Agora, considere a união enumerável:

$$A = \bigcup_{n=1}^{\infty} A_n = \mathbb{Q} \cap [0,1].$$

Este conjunto contém todos os racionais em $[0,1]$. Vamos examinar sua fronteira.

- Como os racionais são densos em $[0,1]$, e o conjunto $\mathbb{Q} \cap [0,1]$ não contém nenhum aberto, sua fronteira é todo o intervalo:

$$\partial A = [0,1].$$

- Portanto:

$$P(\partial A) = P([0,1]) = 1.$$

- Assim, $A \notin \mathcal{A}$.

Concluimos que $\mathcal{A}$ não é uma σ-álgebra.

Capítulo 7

3) Para a normal padrão,

$$\varphi(t) = \frac{1}{\sqrt{2\pi}} \int e^{itx - x^2/2} dx = \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \int e^{-\frac{1}{2}(x-it)^2} dx.$$

Considere a integral $\int_{\mathbb{R}} e^{-z^2/2} dz$, sobre retângulos cujos vértices são $\pm k + 0i$, $\pm k - it$. A exponencial $e^{-z^2/2}$ é uma função inteira, logo

$$\int_{\mathbb{R}} e^{-z^2/2}dz = \int_{-k}^{k} e^{-x^2/2}dx + \int_{k}^{-k} e^{-(x-it)^2/2}dx + \int_{0}^{t} e^{-(k+iu)^2/2}du$$

$$+ \int_{-t}^{0} e^{-(k+iu)^2/2}du = 0.$$

As duas últimas integrasi são limitadas por $|t|e^{-k^2/2}$ e ambas tendem a zero, quando $k \to \infty$; obtemos que a primeira integral multiplicada por $\frac{1}{\sqrt{2\pi}}$ tende a 1 e $\varphi(t) = e^{-t^2/2}$.

5) Temos que $f(x) = 1/[\pi(1 + x^2)]$, for $x \in \mathbb{R}$. Então,

$$\varphi(t) = \frac{1}{\pi} \int \frac{e^{itx}}{1 + x^2} dx.$$

Considere a densidade $f_1(x) = e^{-|x|}/2$ e seja $\varphi_1(t)$ sua f.c. Não é difícil ver que

$$\varphi_1(t) = \frac{1}{1 + t^2},$$

que é absolutamente integrável sobre $\mathbb{R}$ e da inversa obtemos que

$$e^{-|x|} = \frac{1}{\pi} \int \frac{e^{-itx}}{1 + t^2} dt.$$

Mudando o sinal da exponencial dentro da integral (que não altera o resultado) e trocando os papéis de t e x obtemos

$$e^{-|t|} = \frac{1}{\pi} \int \frac{e^{itx}}{1 + x^2} dx.$$

Comparando a primeira e última integral obtemos o resultado.

8) (a) Se Φ é equicontínua no zero, então existem $\varepsilon > 0$, $\delta > 0$ tais que, para $|t - 0| < \delta$ temos $|\varphi_n(t) - \varphi_n(0)| < \varepsilon$, para todo n. Mas $\varphi_n(t)$ é uma f.c., logo $\varphi_n(0) = 1$, para todo n, logo $|1 - \varphi_n(t)| < \varepsilon$, para todo n.

Se P_n é a distribuição de probabilidades cuja f.c. é $\varphi_n(t)$, temos que

$$P_n\{|X| > 2/a\} \leq \frac{1}{a} \int_{-a}^{a} [1 - \varphi_n(t)]dt \leq \frac{1}{a} \int_{-a}^{a} \varepsilon dt = 2\varepsilon.$$

logo, $P_n\{|X| > 2/a\} \leq 2\varepsilon$, para todo n, ou seja $\{P_n, n \geq 1\}$ é fechada.

(b) Para a solução de (b), use os seguintes resultados e o teorema da continuidade para f.c.:

(i) Se $\varphi_n \to g$ em $(-T, T)$ e g é contínua no zero, então as φ_n são equicontínuas e a convergência é uniforme.

(ii) Teorema de Arzela-Ascoli: Seja φ_n uma sequência equicontínua de funções, $|\varphi_n| \leq 1$. Então existe uma subsequência $\{\varphi_{nk}\}$ convergindo a um limite contínuo f e a convergência é uniforme em todo intervalo finito.

(iii) Se $\{\varphi_n\}$ é uma sequência de funções contínuas definidas num conjunto E e se φ_n converge uniformemente para f em E, então f é contínua em E.

15) Por hipótese, $\int |\varphi(t)|^2 dt < \infty$, ou $\varphi(t) \in L_2$. Seja X com f.c. $\varphi(t)$ e considere $X - X^{'}$, com $X^{'}$ independente de X e com a mesma distribuição de X. A distribuição de $X - X^{'}$ é $G = F * F^{-}$, sendo F^{-} a f.d. de $-X^{'}$ e F a f.d. de X(ou de $X^{'}$). Usando a propriedade $G(x) = 1 - G(-x)$, temos

$$G(x) = \int F(x+y) dF(y).$$

A f.c. de $-X^{'}$ é $\varphi(-t) = \overline{\varphi(t)}$ e a f.c. de G é $\varphi(t)\overline{\varphi(t)} = |\varphi(t)|^2$.

Se $\varphi \in L_2$, segue-se que a densidade de G, digamoes $g(x) = \int f(y-x)f(y)dy$, é limitada e contínua e

$$\frac{1}{2\pi}\int |\varphi(t)|^2 e^{-itx} dx = \int f(y-x)f(y)dy = G^{'}(x).$$

Faça $x = 0$ para obter a igualdade desejada e sendo o primeiro membro finito, segue que $f \in L_2$.

Capítulo 8

1) Como os $\{X_{n,k}, k \leq n\}$ são i.i.d. para cada n, temos:

$$\mathbb{P}(|X_{n,k}| > \varepsilon) = \mathbb{P}(|X_{n,1}| > \varepsilon), \quad \text{para todo } k = 1, \ldots, n.$$

Portanto,

$$\max_{1 \leq k \leq n} \mathbb{P}(|X_{n,k}| > \varepsilon) = \mathbb{P}(|X_{n,1}| > \varepsilon),$$

e basta mostrar que $\mathbb{P}(|X_{n,1}| > \varepsilon) \to 0$.

Com $S_n := X_{n,1} + \cdots + X_{n,n}$ temos que φ_{S_n} converge para φ_X uniformemente sobre compactos de $\mathbb{R}$. Além disto sabemos pelo Teorema 8.6 que X tem lei infinitamente divisível e portanto $\varphi_X(t) \neq 0$ para todo $t \in \mathbb{R}$. Logo, para todo $t_0 > 0$ existe N tal que para todo $n \geq N$, $\varphi_{S_n}(t) \neq 0$ para todo $t \in [-t_0, t_0]$. Assim para todo $n \geq N$, existe um único logaritmo contínuo $L_n : [-t_0, t_0] \to \mathbb{C}$ tal que $L_n(0) = 0$. Analogamente, como $\varphi_{S_n} = [\varphi_{X_{n,1}}]^n$, obtemos que

$\varphi_{X_{n,1}}(t) \neq 0$ sobre $[-t_0, t_0]$ e portanto para todo $n \geq N$, existe um único logaritmo contínuo $g_n : [-t_0, t_0] \to \mathbb{C}$ tal que $g_n(0) = 0$. Pela unicidade do logaritmo, obtemos que $L_n = ng_n$. Finalmente, como $\varphi_X(t) \neq 0$, existe um único logaritmo contínuo $L : [-t_0, t_0] \to \mathbb{C}$ tal que $L(0) = 0$.

Agora, a convergência de φ_{S_n} para φ_X sobre $[-t_0, t_0]$ implica a convergência pontual de L_n para L. Assim, para todo $t \in [-t_0, t_0]$,

$$L_n(t) = ng_n(t) \to L(t), \text{ quando } n \to \infty,$$

e portanto $g_n(t) \to 0$ sobre $[-t_0, t_0]$, ou seja $\varphi_{X_{n,1}} \to 1$ sobre $[-t_0, t_0]$. Como t_0 é arbitrário, deduzimos que $\varphi_{X_{n,1}}(t) \to 1$ para todo $t \in \mathbb{R}$, ou seja $X_{n,1} \to 0$ em probabilidade.

4) (a) X é infinitamente divisível se, para cada n, existe uma f.c. φ_n tal que $\varphi(t) = [\varphi_n(t)]^n$, onde φ é a f.c. de X. Aqui, $\varphi(t) = (1 - it)^{-\alpha}$ e, portanto, $\varphi_n(t) = (1 - it)^{-\alpha/n}$ e essa é a f.c. da v.a. $\Gamma(\alpha/n)$.

(b) Se φ_n é como acima, a v.a cuja f.c. é essa é uma v.a. $\Gamma(\alpha/n)$ com densidade

$$g(x) = \begin{cases} 0, & \text{se } x \leq 0, \\ \frac{1}{\Gamma(\alpha/n)} x^{\alpha/n-1} e^{-x}, & \text{se } x > 0. \end{cases}$$

(c) Temos que

$$G_n(u) = n \int_{-\infty}^{u} \frac{x^2}{1 + x^2} dF_n(x) =$$

$$\frac{n}{\Gamma(\alpha/n)} \int_0^u \frac{x^2}{1 + x^2} x^{\alpha/n-1} e^{-x} dx =$$

$$\frac{n}{\Gamma(\alpha/n)} \int_0^u \frac{x^{\alpha/n-1}}{1 + x^2} e^{-x} dx.$$

Agora,

$$\frac{n}{\Gamma(\alpha/n)} = \frac{\alpha}{(\alpha/n)\Gamma(\alpha/n)} = \frac{\alpha}{\Gamma(1 + \alpha/n)} \to \alpha$$

quando $n \to \infty$ e, portanto, pelo TCD, $G_n(u) \to G(u) = \alpha \int_0^u (x/(1 + x^2)e^{-x} dx$, de modo que a medida $dG(x)$ é dada pelo integrando, $dG(x) = \frac{x}{1+x^2} e^{-x} dx$.

9) Temos que provar que a f.c. de X é da forma $\varphi(t) = e^{-c|t|^\alpha}$ $c > 0$ e aqui $\alpha = 1/2$. Seja $\varphi_n(t)$ a f.c. de Y_n e $\psi(t)$ a f.c. de $\frac{\text{sinal}(X_{n_i})}{X_{ni}^2}$. Então,

$$\psi(t) = E\left[\exp\{it\frac{\text{sinal}(X_{n_i})}{X_{ni}^2}\}\right] = \int_{-n}^{n} \frac{\exp\{it\frac{\text{sinal}(x)}{x^2}\}}{2n}dx = \frac{1}{n}\int_0^n \cos(\frac{t}{x^2})dx.$$

Pela independência, $\varphi_n(t) = [\psi(t)]^n$, ou seja,

$$\varphi_n(t) = E\{e^{itY_n}\} = \left[\frac{1}{n}\int_0^n \cos(t/x^2)dx\right]^n = \left[1 - 1 + \frac{1}{n}\int_0^n \cos(t/x^2)dx\right]^n$$

$$= \left[1 - \frac{1}{n}\int_0^n [1 - \cos(t/x^2)dx\right]^n$$

$$= \left[1 - \frac{1}{n}\int_0^\infty [1 - \cos(t/x^2)]dx + o(n^{-1})\right]^n.$$

Para $n \to \infty$, $(1 - \lambda/n)^n \to e^{-\lambda}$ logo

$$\varphi_n(t) \to \varphi(t) = \exp\{-\int_0^\infty [1 - \cos(t/x^2)]dx\}.$$

Considere a transformação $t/x^2 = u$. Segue que

$$\int_0^\infty [1 - \cos(t/x^2)]dx = 2\int_0^\infty \frac{1 - \cos(tu)}{u^{1/2+1}}du.$$

Fazendo $tu = z$ obtemos que a integral é igual a $|t|^{1/2}[2\int_0^\infty (1 - \cos z)/(z^{3/2})dz]$ e a integral é uma constante $c > 0$. Logo $\varphi(t) = -e^{-c|t|^{1/2}}$ e o resultado segue.

10) (a) Devido à simetria dos valores possíveis de X_j, temos

$$\mathbb{E}[X_j] = 0.$$

e

$$\begin{aligned}\text{Var}(X_j) &= \mathbb{E}[X_j^2] \\ &= (j^2)^2 \cdot 2 \cdot \frac{1}{12j^2} + j^2 \cdot 2 \cdot \frac{1}{12} + 0 \\ &= j^4 \cdot \frac{2}{12j^2} + j^2 \cdot \frac{2}{12} \\ &= \frac{j^2}{6} + \frac{j^2}{6} = \frac{j^2}{3}.\end{aligned}$$

Portanto, como os X_j são independentes,

$$\mathrm{Var}(S_n) = \sum_{j=1}^{n} \mathrm{Var}(X_j) = \sum_{j=1}^{n} \frac{j^2}{3} \sim \frac{n^3}{9}.$$

A condição de Lindeberg requer que, para todo $\varepsilon > 0$,

$$\frac{1}{\mathrm{Var}(S_n)} \sum_{j=1}^{n} \mathbb{E}\left[X_j^2 \mathbf{1}_{\{|X_j| > \varepsilon\sqrt{\mathrm{Var}(S_n)}\}}\right] \to 0 \quad \text{quando } n \to \infty.$$

Observe que os termos X_j podem assumir valores grandes $\pm j^2$ com probabilidade $1/(12j^2)$, e:

$$\mathbb{E}\left[X_j^2 \mathbf{1}_{\{|X_j| = j^2\}}\right] = j^4 \cdot \frac{2}{12j^2} = \frac{j^2}{6}.$$

Para n grande, $\varepsilon\sqrt{\mathrm{Var}(S_n)} \sim \varepsilon\frac{n^{3/2}}{3}$. Somando os termos para os quais $j^2 > \varepsilon\sqrt{\mathrm{Var}(S_n)}$, ou seja, para $j > cn^{3/4}$, temos:

$$\sum_{j > cn^{3/4}} \mathbb{E}[X_j^2 \mathbf{1}_{|X_j| = j^2}] \sim \int_{cn^{3/4}}^{n} \frac{x^2}{6} dx \sim \frac{n^3}{18}.$$

O termo dominante é n^3, portanto

$$\frac{1}{\mathrm{Var}(S_n)} \sum_{j=1}^{n} \mathbb{E}[X_j^2 \mathbf{1}_{|X_j| > \varepsilon\sqrt{\mathrm{Var}(S_n)}}] \to \text{constante} \neq 0.$$

Logo, a condição de Lindeberg não é satisfeita.

Agora definimos

$$Y_j = X_j \mathbf{1}_{\{|X_j| \leq j\}}, \quad R_j = X_j - Y_j = X_j \mathbf{1}_{\{|X_j| = j^2\}}.$$

As Y_j's são variáveis truncadas que satisfazem a condição de Lindeberg. Portanto, obtemos que

$$\frac{3\sqrt{2}\sum_{j=1}^{n} Y_j}{n^{3/2}} \xrightarrow{d} \mathcal{N}(0,1).$$

Por outro lado, pelo lemma de Borel-Cantelli, obtemos que

$$\frac{\sum_{j=1}^{n} R_j}{n^{3/2}} \xrightarrow{P} 0.$$

Assim, conclui-se que
$$\frac{3\sqrt{2}S_n}{n^{3/2}} \xrightarrow{d} \mathcal{N}(0,1).$$

Capítulo 9

1) Observe que $|\pi_t(f)| = |f(t)| \leq \max_{t\in[0,1]} |f(t)| = \|f\|_\infty$.

3) Vamos começar com o cálculo da transformada de Laplace de X distribuída de acordo com a lei de Lévy, $L(t) = E(e^{-tX})$, $t \geq 0$.

$$\begin{aligned} L(t) &= \int_0^\infty \frac{a}{\sqrt{2\pi}} \frac{1}{x^{3/2}} \exp\Big(-\frac{a^2}{2x} - tx\Big)dx \\ &= \frac{2}{\sqrt{\pi}} \int_0^\infty \exp\Big(-u^2 - \frac{a^2 t}{2u^2}\Big)du \\ &= \frac{1}{\sqrt{\pi}} \int_{-\infty}^\infty \exp\left(-\left(u - \frac{a}{u}\sqrt{\frac{t}{2}}\right)^2 - a\sqrt{2t}\right)du \end{aligned}$$

usando a mudança de variável $x = a^2/(2u^2)$. Usando o teorema mestre de Glasser obtemos que
$$L(t) = \exp(-a\sqrt{2t}).$$
Pelo teorema de extensão analítica, obtemos que $L(z) = \exp(-a\sqrt{2z})$ para z complexo tal que $\mathcal{R}e(z) \geq 0$. Finalmente, deduzimos que a f.c φ de X é dada por

$$\varphi(t) = \exp(-a\sqrt{-2it}), \quad t \in \mathbb{R}.$$

Usando Teorema 8.13, deduzimos que a lei de Lévy é estável com índice $\alpha = 1/2$.

7) Observe que $\max_{0\leq t\leq 1} X_n(t) = \max_{k\leq n} \frac{S_k}{\sigma\sqrt{n}}$. Por outro lado, a função $h : C([0,1]) \to \mathbb{R}$, $x \mapsto \max_{0\leq t\leq 1} x(t)$ é contínua pois

$$|h(x)| \leq \max_{0\leq t\leq 1} |x(t)| = \|x\|_\infty.$$

Usando o Teorema 9.4 e o Corolário 9.1, obtemos que

$$\max_{0\leq t\leq 1} X_n(t) \to \max_{0\leq t\leq 1} W(t)$$

em lei quando $n \to \infty$. Finalmente, para todo $x \geq 0$, temos que

$$P\Big\{\max_{0\leq t\leq 1} W(t) \leq x\Big\} = \frac{2}{\sqrt{2\pi}} \int_0^x e^{-t^2/2}dt.$$

9) (a) Temos que

$$W\Big\{\sup_{0\le t\le 1/2} x(t)\le 1\Big\}=\frac{2}{\sqrt{2\pi}}\int_0^{\sqrt{2}} e^{-t^2/2}dt.$$

(b) Em primeiro lugar, observe que

$$W\{0\le x(t)\le 1, t\in[0,1]\}\le W\{x(t)\ge 0, t\in[0,1]\}.$$

Agora, considere $\varepsilon>0$ e seja $A_\varepsilon=\{x(t)\ge -\varepsilon, t\in[0,1]\}$. Por simetria, temos que

$$W(A_\varepsilon)=W\Big\{\sup_{0\le t\le 1} x(t)\le \varepsilon\Big\}=\frac{2}{\sqrt{2\pi}}\int_0^{\varepsilon} e^{-t^2/2}dt.$$

Assim, obtemos que

$$W\{x(t)\ge 0, t\in[0,1]\}=\lim_{\varepsilon\to 0} W(A_\varepsilon)=0.$$

Capítulo 10

10) (a) Seja $\{X_n, n\ge 0\}$ uma C.M. Seja a matriz

$$\mathbf{P}=\begin{bmatrix} 0 & 1/2 & 1/2\\ 1/3 & 1/4 & 5/12\\ 2/3 & 1/4 & 1/12\end{bmatrix}$$

e seja $\mathbf{p}=(1/3,1/3,1/3)=(p_1,p_2,p_3)$, $p_k=P(X_0=k)$, $k\in\{1,2,3\}=I$.

Considere X_n com valores 1,2 e 3 e $f(X_n)$ com valores a,a,b, ou seja, $f(1)=f(2)\neq f(3)$.

Então, $\{f(X_n), n\ge 0\}$ não é uma C.M. porque, por exemplo,

$$P\{f(X_2)=b|f(X_0)=b, f(X_1)=a\}=$$

$$\frac{P(X_0=3,X_1=1,X_2=3)+P(X_0=3,X_1=2,X_2=3)}{P(X_0=3,X_1=1)+(P(X_0=3,X_1=2)}$$

$$=\frac{p_3\cdot p_{31}\cdot p_{13}+p_3\cdot p_{32}\cdot p_{23}}{p_3\cdot p_{31}+p_3\cdot p_{32}}=\frac{1/3\cdot 2/3\cdot 1/2+1/3\cdot 1/4\cdot 5/12}{1/3\cdot 2/3+1/3\cdot 1/4}=\frac{21}{44},$$

pois X_n é C.M.

Por outro lado, se $\{f(X_n), n \geq 0\}$ fosse uma C.M. deveríamos ter

$$P\{f(X_2) = b|f(X_0) = b, f(X_1) = a\} = P\{f(X_2) = b|f(X_1) = a\} =$$

$$\frac{P(X_1 = 1, X_2 = 3) + P(X_1 = 2, X_2 = 3)}{P(X_1 = 1) + P(X_1 = 2)} = \frac{1/3 \cdot 1/2 + 1/3 \cdot 5/12}{1/3 + 1/3} = \frac{11}{24},$$

pois $P(X_1 = 1) = P(X_1 = 2) = 1/3$ (verifique!)

(b) Suponha que f seja 1-1. Então $f : \{1, 2, 3\} \to \{f(1), f(2), f(3)\} = \{a_1, a_2, a_3\}$, com a_i distintos. Logo

$$P(f(X_{n+1}) = j_{n+1}|f(X_0) = j_0, \ldots, f(X_n) = j_n\} = \quad (f : i_k \to j_k)$$

$$P(X_{n+1} = i_{n+1}|X_0 = i_0, \ldots, X_n = i_n\} = P(X_{n+1} = i_{n+1}|X_n = i_n\} =$$

$$P(f(X_{n+1}) = j_{n+1}|f(X_n) = j_n$$

e $\{f(X_n)\}$ é uma C.M.

12) Considere

$$P(X_{n+1} = j|X_n = i) = \frac{\sum P(X_{n+1} = j, X_n = i, \ldots, X_0 = i_0)}{\sum P(X_n = i, \ldots, X_0 = i_0)},$$

onde as somas são sobre $i_0, \ldots, i_{n-1}$. Agora, o termo genérico do numerador é igual a

$$\varphi(i, j)P(X_n = i, \ldots, X_0 = i_0) = \varphi(i, j)\varphi(i_{n-1}, i) \cdots \varphi(i_1, i_0)P(X_0 = i_0).$$

De modo similar, o termo genérico do denominador é igual a

$$\varphi(i_{n-1}, i)\varphi(i_{n-2}, i_{n-1}) \cdots \varphi(i_1, i_0)P(X_0 = i_0).$$

logo, $P(X_{n+1} = j|X_n = i) = \varphi(i, j)$ e X é uma C.M. Como, para cada n, isso é uma função de i, j, a cadeia tem transições estacionárias.

14) Defina $T = \inf\{n : X_n = 0\}$. Mostre que $P(T < \infty) = 1$. Para isso, seja $A_k = \{X_k = 0\}$ e escreva $P(T < \infty) = P(\cup_{k=1}^{\infty} A_k)$. Use o fato que $P(A_k) = \sum P(X_k = 0, X_{k-1} = i_{k-1}, \ldots, X_0 = i_0)$ e que X_k é uma C.M.

16) As probabilidades de transição a um passo da cadeia reversa são

$$Q_{ij} = P(X_n = j|X_{n+1} = i) = \frac{P(X_n = j, X_{n+1} = i)}{P(X_n = j)P(X_{n+1} = i)} P(X_n = j)$$

$$= \frac{P(X_n = j)}{P(X_{n+1} = i)} p_{ji}.$$

Se p_i é a distribuição estacionária, temos

$$p_j^{(n)} = P(X_n = j) = \sum_i p_i p_{ij}^{(n-1)} = p_j = \sum_i p_i p_{ij}.$$

Logo,,

$$Q_{ij} = \frac{p_j}{p_i} p_{ji}.$$

Mas, sabemos que $p_j = [\sum_i p_i]\pi_j = \pi_j$, logo

$$Q_{ij} = \frac{\pi_j}{\pi_i} p_{ji},$$

que é estacionária (não depende de n).

É fácil ver que

$$Q_{ij}^{(2)} = \frac{\pi_j}{\pi_i} p_{ji}^{(2)},$$

e, em geral, por indução,

$$Q_{ij}^{(n)} = \frac{\pi_j}{\pi_i} p_{ji}^{(n)}.$$

Portanto,

$$\sum_n Q_{ii}^{(n)} = \sum_n \frac{\pi_i}{\pi_i} p_{ii}^{(n)} = \sum_i p_{ii}^{(n)} = \infty,$$

logo a cadeia é recorrente.

Capítulo 11

6) Observe que se I é um intervalo de comprimento menor do que 1, então os intervalos $I, T^{-1}I, T^{-2}I, \ldots$ são disjuntos.

8) Pelo TEP, para todo $A \in \mathcal{F}$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \mathbf{1}_A(S^k\omega) \to P_1(A), \quad P_1 - \text{q.c},$$

e

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \mathbf{1}_A(S^k\omega) \to P_2(A), \quad P_2 - \text{q.c}.$$

Se P_1 e P_2 não forem mutualmente singulares, obtemos que $P_1(A) = P_2(A)$, ou seja uma contadição com o fato $P_1 \neq P_2$.

11) (a) No caso em que α é irracional, T é ergódica mas não é mixing. Para mostrar que T não é mixing, considere o caso de dois intervalos abertos disjuntos A e B e mostre que neste caso $P(A \cap T^{-n}B)$ não tem limite quando $n \to \infty$. No caso em que α é racional, T não é ergódica.
(b) Para $k \geq 2$, a transformação $T : \omega \mapsto k\omega \pmod 1$ é ergódica e mixing.

12) (a) Sim, usar a fórmula da probabilidade total.
(b) Não necessariamente, usar o Problema 8.

13) (a) Seja $X = \{X_n, n \geq 0\}$ um processo gaussiano tal que $\mathbb{E}(X_i) = 0$ para todo i, e função de covariância $R(i,j) := \text{Cov}(X_i, X_j)$.

Queremos mostrar que o processo X é estacionário se, e somente se, $R(i,j)$ depende apenas de $|i-j|$.

(⇒) Suponha que X é estacionário (em sentido estrito). Como o processo é gaussiano, a estacionariedade em sentido amplo implica estacionariedade estrita. Logo, para todos $i, j, h \in \mathbb{Z}_+$, temos:

$$\text{Cov}(X_i, X_j) = \text{Cov}(X_{i+h}, X_{j+h}).$$

Portanto,

$$R(i,j) = R(i+h, j+h),$$

o que mostra que $R(i,j)$ depende apenas da diferença $i-j$. Como $R(i,j) = R(j,i)$ (simetria da covariância), concluímos que existe uma função $r : \mathbb{Z}_+ \to \mathbb{R}$ tal que:

$$R(i,j) = r(|i-j|).$$

(⇐) Suponha agora que $R(i,j) = r(|i-j|)$. Queremos mostrar que X é estacionário.

Como o processo é gaussiano com média zero, sua distribuição é completamente determinada pela função de covariância R. Para qualquer $h \in \mathbb{Z}_+$, temos:

$$R(i+h, j+h) = r(|(i+h)-(j+h)|) = r(|i-j|) = R(i,j),$$

o que mostra que a função de covariância é invariante por translação. Assim, as distribuições conjuntas de vetores do tipo $(X_{t_1}, \ldots, X_{t_n})$ e $(X_{t_1+h}, \ldots, X_{t_n+h})$ coincidem para todo h, o que implica que X é estritamente estacionário.

Capítulo 12

3) É importante notar que, se $(X(t))_{t\geq 0}$ é um martingale local, a variável aleatória $X(t)$ não é necessariamente integrável. Em particular, não temos nenhuma informação sobre $X(0)$ além de ser $\mathcal{F}_0$-mensurável. Isto nos dá uma maneira simples de construir martingales locais que não são martingales. Considere por exemplo um martingale $(M(t))_{t\geq 0}$ e uma v.a. Z $\mathcal{F}_0$-mensurável tal que $E(|Z|) = \infty$. Então $X(t) = Z + M(t)$, $t \geq 0$, é um martingale local mas não é um martingale. Para verificar que $(X(t))_{t\geq 0}$ é um martingale local pode-se considerar a sequência localizante $\tau_n = n\mathbf{1}_{\{|Z|\leq n\}}$, $n \geq 1$.

4) Ao multiplicar por e^{-bt} a desigualdade (12.14), obtemos

$$\frac{d}{dt}\Big(e^{-bt}\int_0^t f(s)ds\Big) \leq ae^{-bt}.$$

Integrando a desigualdade acima deduzimos que

$$e^{-bt}\int_0^t f(s)ds \leq \frac{a}{b}(1-e^{-bt}).$$

Finalmente, usando esta última desigualdade em (12.14) obtemos o resultado desejado.

5) (i) Como $X_n \geq 0$ temos que $X \geq 0$, q.c e portanto

$$(X - X_n)^+ \leq X^+ = X, \text{ q.c.}$$

Como $X \in L^1$, pelo teorema de convergência dominada, obtemos que $E((X - X_n)^+) \to 0$ quando $n \to \infty$. Por outro lado,

$$|X_n - X| = 2(X - X_n)^+ - (X - X_n).$$

Tomando a esperança desta última desiguadade e usando fato que por hipótese $E(X - X_n) \to 0$, as $n \to \infty$, obtemos que $E(|X - X_n|) \to 0$ também.
(ii) Como $X_n \to X$ em probabilidade, existe uma subsequência $(X_{n_k})_k$ tal que $X_{n_k} \to X$, q.c. Por outro lado, usando o lema de Fatou, obtemos que

$$E(X^2) \leq \liminf_{k\to\infty} E(X_{n_k}^2) \leq \sup_n E((X - X_n)^2) < \infty.$$

Agora, para todo $a > 0$, usando as desigualdades de Cauchy-Schwarz e Markov, obtemos que

$$E(|X - X_n|) \leq E(|X - X_n| \wedge a) + a^{-1}E(|X - X_n|^2).$$

Devido a convergência em probabilidade de X_n para X, o primeiro termo da desigualdade acima tende a 0 quando $n \to \infty$, para todo $a > 0$. O segundo termo pode se tornar arbitrariamente pequeno escolhando a grande o suficiente. Assim, concluimos que $E(|X - X_n|) \to 0$ quando $n \to \infty$.

Índice Remissivo

PEDRO ALBERTO MORETTIN é graduado em Matemática pela Universidade de São Paulo (USP), com mestrado e doutorado em Estatística pela University of California, Berkeley, nos Estados Unidos. É autor de diversos livros e inúmeros artigos, publicados no Brasil e no exterior. Recebeu os prêmios ABE, da Associação Brasileira de Estatística (ABE), e o Mahalanobis Award, conferido pelo Governo da Índia. Foi presidente da ABE e do Interamerican Statistical Institute e vice-presidente do International Statistical Institute. É Professor Emérito do Departamento de Estatística da USP e Pesquisador Emérito do CNPq.

CHRISTOPHE GALLESCO foi aluno da École Nationale Supérieure de Physique de Grenoble, na França. É mestre em Física pela Universidade Joseph Fourier de Grenoble e doutor em Estatística pela Universidade de São Paulo. Atualmente, é professor do Departamento de Estatística da Universidade Estadual de Campinas (UNICAMP). É autor de diversos artigos científicos na área de Probabilidade. Suas linhas de pesquisa incluem processos markovianos, processos com memória infinita, passeios aleatórios em meio aleatório e modelos de percolação.

A obra apresenta tópicos da Teoria de Probabilidade em nível avançado, indicada para um ou dois semestres de programas de mestrado e doutorado em Probabilidade e Estatística. Pode ser útil, também, em programas de pós-graduação em Economia, Matemática e Engenharia. Além dos tópicos usuais, apresenta um capítulo sobre Cálculo Estocástico, necessário para se entender a formulação da integral estocástica de Itô, que tem aplicações, por exemplo, na obtenção da fórmula de Black-Scholes, usada em finanças para se calcular o preço teórico de uma opção.

ISBN: 978-65-01-91821-1

9 786501 918211