

Uma Iniciação aos Sistemas Dinâmicos Estocásticos

Publicações Matemáticas

**Uma Iniciação aos Sistemas
Dinâmicos Estocásticos**

Paulo Ruffino
UNICAMP



27^o Colóquio Brasileiro de Matemática

Copyright © 2009 by Paulo Ruffino
Direitos reservados, 2009 pela Associação Instituto
Nacional de Matemática Pura e Aplicada - IMPA
Estrada Dona Castorina, 110
22460-320 Rio de Janeiro, RJ
Impresso no Brasil / Printed in Brazil
Capa: Noni Geiger / Sérgio R. Vaz

27^a Colóquio Brasileiro de Matemática

- A Mathematical Introduction to Population Dynamics - Howard Weiss
- Algebraic Stacks and Moduli of Vector Bundles - Frank Neumann
- An Invitation to Web Geometry - Jorge Vitório Pereira e Luc Piro
- Bolhas Especulativas em Equilíbrio Geral - Rodrigo Novinski e Mário Rui Páscoa
- C^* -algebras and Dynamical Systems - Jean Renault
- Compressive Sensing - Adriana Schulz, Eduardo A. B. da Silva e Luiz Velho
- Differential Equations of Classical Geometry, a Qualitative Theory - Ronaldo Garcia e Jorge Sotomayor
- Dynamics of Partial Actions - Alexander Arbieto e Carlos Morales
- Introduction to Evolution Equations in Geometry - Bianca Santoro
- Introduction to Intersection Homology and Perverse Sheaves - Jean-Paul Brasselet
- Introdução à Análise Harmônica e Aplicações - Adán J. Corcho Fernandez e Marcos Petrucio de A. Cavalcante
- Introdução aos Métodos de Decomposição de Domínio - Juan Galvis
- Problema de Cauchy para Operadores Diferenciais Parciais - Marcelo Rempel Ebert e José Ruidival dos Santos Filho
- Simulação de Fluidos sem Malha: Uma Introdução ao Método SPH - Afonso Paiva, Fabiano Petronetto, Geovan Tavares e Thomas Lewiner
- Teoria Ergódica para Autômatos Celulares Algébricos - Marcelo Sobottka
- **Uma Iniciação aos Sistemas Dinâmicos Estocásticos - Paulo Ruffino**
- Uma Introdução à Geometria de Contato e Aplicações à Dinâmica Hamiltoniana - Umberto L. Hryniewicz e Pedro A. S. Salomão
- Viscosity Solutions of Hamilton-Jacobi Equations - Diogo Gomes

ISBN: 978-85-244-0304-0

Distribuição: IMPA
Estrada Dona Castorina, 110
22460-320 Rio de Janeiro, RJ
E-mail: ddic@impa.br
<http://www.impa.br>

Conteúdo

Apresentação	1
1 Teoria de probabilidade	3
1.1 Espaços de medida e de probabilidade	4
1.2 Variáveis aleatórias	6
1.3 Esperança de uma variável aleatória	9
1.4 Os espaços L^p	16
1.5 Esperança condicional	18
1.6 Independência	24
1.7 Convergências de variáveis aleatórias	29
1.8 Variáveis aleatórias gaussianas	30
2 Processos estocásticos	37
2.1 Processo canônico e continuidade	38
2.2 Filtração e tempos de parada.	40
2.3 Processos de Markov	42
2.4 Movimento Browniano	49
2.5 Martingales	53
2.5.1 Propriedades básicas	56
2.5.2 p -Variação de funções contínuas	57
2.5.3 Martingale local	63
3 Cálculo estocástico	65
3.1 Integral Estocástica de Itô	65
3.1.1 Aplicação a processos parados	66
3.1.2 Processos contínuos	68

3.2	Fórmula de Itô	74
3.3	Caracterização de Levy	78
3.4	Equações diferenciais estocásticas	81
3.5	Gerador infinitesimal	88
3.5.1	Fórmula de Dynkin	91
3.5.2	Aplicações	92
3.6	Integral de Stratonovich	95
4	Sistemas dinâmicos estocásticos	102
4.1	Trajетórias das soluções de uma EDE	102
4.2	Fluxos estocásticos e cociclos	104
4.3	Estabilidade: expoentes de Lyapunov	106
4.4	Conjugações de cociclo	107
4.5	Números de rotação em S^1	108
4.6	Outros exemplos e aplicações:	111
5	Apêndices	113
5.1	Independência e convolução de medidas	113
5.2	Critério de Kolmogorov	114
5.3	Gaussianas	118
5.4	Mais sobre o Movimento Browniano	119
5.4.1	Processo de Ornstein-Uhlenbeck	119
5.4.2	Outras construções	121
5.5	Teorema ergódico para processos de Markov	123
	Bibliografia	125
	Índice Remissivo	131

Apresentação

Sistemas dinâmicos estocásticos vem ganhando cada vez mais interesse nas últimas décadas pela diversidade de problemas cuja modelagem inclui algum aspecto probabilístico. Existem várias maneiras da aleatoriedade entrar em um sistema dinâmico contínuo ou discreto: por pulsos, por saltos limitados, ruído aditivo, multiplicativo, entre outras. Nossa abordagem da teoria de dinâmica estocástica é no sentido do ruído entrar no sistema como um semimartingale, o que inclui ruídos gaussianos em sistemas contínuos como movimento browniano por exemplo. Basicamente esta é a chamada teoria dinâmica das equações diferenciais estocásticas (EDE). Essa abordagem em particular tem sido extremamente frutífera nas últimas décadas, desde os trabalhos de Itô. Uma EDE permite, via aplicações diferenciáveis, transportar processos estocásticos de um espaço para outro, geometrizando portanto trajetórias não diferenciáveis. Além disso as propriedades probabilísticas dos martingales fornece uma ferramenta poderosa para abordar estabilidade, análise variacional, propriedades dinâmicas, geometria, finanças, EDP's entre outros.

Nossa intenção neste texto é fazer uma apresentação em tom expositório para alunos no final da graduação ou começo de pós-graduação na área de exatas. Tentamos fazer um estilo bastante livre, em algumas passagens até bastante informal. A leitura está dimensionada para ser mais leve do que os textos existentes atualmente. Isso para que o leitor entenda rapidamente as motivações, as técnicas e seja logo direcionado para os problemas e aplicações.

Foi feito um esforço para que os primeiros capítulos fossem suficientemente auto-contidos para uma primeira leitura. A expectativa é que o material seja útil para iniciantes terem um primeiro contato

com muitos dos conceitos e exemplos. No entanto, pela natureza do assunto, o leitor interessado vai perceber que o material exposto é somente um guia para um estudo a ser aprofundado posteriormente. Para se ter uma idéia de como os assuntos foram filtrados e resumidos, os 3 primeiros capítulos normalmente se abrem, cada um deles, em uma disciplina de um semestre na maioria dos melhores programas de graduação ou pós-graduação do país.

Incluimos cerca de quarenta exercícios de dificuldades variadas entremeados ao longo do texto. Alguns desses exercícios tem o enunciado que faz parte da teoria. Reservamos para acrescentar no apêndice alguns comentários mais técnicos, ou históricos ou até que são importantes num primeiro contato com a teoria, mas que não foram mencionados porque não estavam no caminho direto que nos propomos para o cálculo e sistemas dinâmicos estocásticos.

Por se tratar de uma primeira versão expositória neste nível, comentários, sugestões e correções dos vários erros que nos escaparam serão bem vindos. Finalmente, gostaríamos de agradecer as sugestões e correções dos alunos e pesquisadores do nosso grupo de estocástica na Unicamp.

Boa leitura!

Paulo Ruffino

Campinas, maio de 2009.

Capítulo 1

Noções de teoria da medida e probabilidade

Neste primeiro capítulo apresentaremos os conceitos da teoria de probabilidade sobre os quais irá se basear o modelo matemático da aleatoriedade dos sistemas dinâmicos estocásticos. Esse capítulo pretende, na medida do possível, dirigir-se a leitores sem familiaridade com a teoria básica de medida e integração. Apresentaremos assim, de uma maneira despretenciosa, uma versão simplificada desta teoria no sentido de, sempre que conveniente, as definições e demonstrações serão feitas para o caso finito ou discreto. Acreditamos que com essas idéias em mente, o leitor já terá uma idéia das motivações, e sobretudo, uma base razoável para compreender os capítulos seguintes. Aos enunciados ou exercícios que numa primeira leitura possam parecer mais avançados, sugiro ao leitor que pule e volte neles posteriormente. Alguns leitores talvez prefiram pensar que esse capítulo serve como um roteiro para se estudar posteriormente teoria da medida.

Sugerimos que o leitor use posteriormente, como referência paralela o texto em teoria de integração que mais lhe agradar. Quando for o caso, tentaremos indicar referências específicas de sessões ou capítulos de certos livros clássicos, por exemplo, Halmos [19], Bartle [3], Rudin [49], Royden [48], entre outros.

1.1 Espaços de medida e de probabilidade

Começamos pelo conceito fundamental da teoria que é a de um espaço de probabilidade. O conjunto abstrato base sobre o qual construiremos essa estrutura será denotado por Ω . Vamos considerar neste conjunto classes de subconjuntos, denominadas álgebras e σ -álgebras; logo mais se verá que esses conceitos correspondem a dizer que existe uma limitação nos eventos possíveis de ocorrerem em uma modelagem, ou ainda em outras palavras, uma limitação na possibilidade de medirmos ou atribuirmos a esses eventos uma probabilidade.

Definição 1.1. *Seja Ω um conjunto, uma classe \mathcal{F} de subconjuntos de Ω chama-se uma álgebra se:*

1. *O conjunto todo Ω e o vazio \emptyset pertencem a \mathcal{F} ;*
2. *Um subconjunto $A \in \mathcal{F}$ se e somente se seu complementar A^c também está em \mathcal{F} ;*
3. *para toda sequência $(A_i)_{1 \leq i \leq n} \in \mathcal{F}$, para algum $n \in \mathbf{N}$ temos que a união $\cup_{i=1}^n A_i \in \mathcal{F}$.*

A classe \mathcal{F} será chamada de uma σ -álgebra se a propriedade do item (3) acima valer para toda sequência infinita enumerável de elementos em \mathcal{F} .

Exercício 1.1. *Verifique que se \mathcal{F} e \mathcal{G} são σ -álgebras de um conjunto Ω então $\mathcal{F} \cap \mathcal{G}$ é também uma σ -álgebra.*

Pelo exercício acima, dada uma família qualquer \mathcal{A} de subconjuntos de Ω , é possível falarmos da menor σ -álgebra que contém essa família. A “menor” σ -álgebra, aqui, significa a intersecção de todas que contém a família \mathcal{A} , ou ainda em outras palavras, é o elemento minimal da ordem parcial nas σ -álgebras dada pela relação de continência. Note que as σ -álgebras que contem qualquer família de conjuntos é não vazia, já que o conjunto das partes $\mathcal{P}(\Omega)$ é uma σ -álgebra que contém toda família de subconjuntos. Denotaremos por $\sigma(\mathcal{A})$ essa σ -álgebra gerada pela família de subconjuntos \mathcal{A} . Um exemplo típico na teoria é quando tomamos a família \mathcal{A} como sendo os conjuntos abertos da reta (ou de qualquer outro espaço topológico): a menor σ -álgebra onde os conjuntos abertos (e portanto os fechados também) são mensuráveis é denominada a *σ -álgebra dos borelianos*.

Exercício 1.2. Dada uma família de subconjuntos A de Ω , descreva uma maneira construtiva de obtermos $\sigma(A)$, a σ -álgebra gerada por A em Ω .

A união de σ -álgebras pode não ser uma σ -álgebra. Denotamos por $\mathcal{F} \vee \mathcal{G}$ a menor σ -álgebra que contém $\mathcal{F} \cup \mathcal{G}$.

O par ordenado (Ω, \mathcal{F}) chama-se um *espaço mensurável*, e dizemos que um subconjunto $A \subseteq \Omega$ é mensurável (ou \mathcal{F} -mensurável, quando for preciso enfatizar a σ -álgebra), se $A \in \mathcal{F}$. Os conjuntos de \mathcal{F} são também chamados de *eventos*.

Definição 1.2. Uma medida μ sobre um espaço mensurável (Ω, \mathcal{F}) é uma função $\mu : \mathcal{F} \rightarrow \mathbf{R}_{\geq 0}$ tal que:

1. $\mu(\emptyset) = 0$;
2. (σ -aditiva), Se $(E_n)_{n \geq 1}$ é uma sequência de subconjuntos disjuntos \mathcal{F} -mensuráveis então $\mu(\bigcup_{n=1}^{\infty} E_n) = \sum_{n=1}^{\infty} \mu(E_n)$.

A medida μ é dita *finita* em Ω se $\mu(\Omega) \leq \infty$. Note que se μ_1 e μ_2 são medidas em (Ω, \mathcal{F}) , então $a_1\mu_1 + a_2\mu_2$, com $a_1, a_2 \geq 0$ também é uma medida. Não abordaremos aqui o espaço vetorial das medidas com sinal, que é o espaço onde poderíamos tomar quaisquer $a_1, a_2 \in \mathbf{R}$. Geometricamente, o conjunto de medidas forma um cone no espaço vetorial das medidas com sinal.

Estamos particularmente interessados naquelas medidas finitas μ onde a medida do espaço todo $\mu(\Omega) = 1$. Neste caso chamamos μ de uma *medida de probabilidade*, ou simplesmente uma *probabilidade* no espaço mensurável (Ω, \mathcal{F}) .

Pelo comentário no parágrafo acima, note que combinações convexas de medidas de probabilidades são também medidas de probabilidade. Um *espaço de probabilidade* é uma tripla $(\Omega, \mathcal{F}, \mathbf{P})$ onde Ω é um conjunto abstrato, \mathcal{F} é uma σ -álgebra de subconjuntos de Ω e \mathbf{P} é uma medida de probabilidade em \mathcal{F} .

São interessantes algumas relações que existem entre as topologias e as medidas que serão colocadas sobre os respectivos borelianos. Uma medida μ é dita *regular interior* se para todo boreliano B temos que

$$\mu(B) = \sup\{\mu(K) : K \subseteq B, K \text{ compacto}\}.$$

Analogamente, a medida é dita *regular exterior* se para todo boreliano B temos que

$$\mu(B) = \inf\{\mu(A) : B \subseteq A, A \text{ aberto}\}.$$

Uma *medidas de Radon* em um espaço topológico são medidas nos borelinos que são limitadas em compactos, regular exterior nos mensuráveis e regular interior nos abertos, ver, e.g. Folland [15]. A proposição seguinte garante que a maioria das medidas que trataremos aqui serão de Radon:

Proposição 1.1. *Se X é um espaço métrico completo e separável (também chamado de espaço polonês) então toda medida em X é regular interior e exterior.*

Voltaremos a falar de medidas de Radon no final da próxima seção.

1.2 Variáveis aleatórias

Dados dois espaços mensuráveis (X, \mathcal{F}) e (Y, \mathcal{G}) , dizemos que uma função $f : (X, \mathcal{F}) \rightarrow (Y, \mathcal{G})$ é *mensurável* se ela preservar a estrutura de mensurabilidade no sentido de imagens inversas de subconjuntos mensuráveis em Y são subconjuntos mensuráveis em X . Apesar do escopo de funções mensuráveis ser muito mais amplo, a título de primeira ilustração, note que, dados espaços topológicos X e Y , toda função contínua $f : X \rightarrow Y$ é mensurável nas σ -álgebras dos borelianos correspondentes.

Em teoria de probabilidade, funções mensuráveis recebem o nome especial de variáveis aleatórias, em outras palavras: dado um espaço de probabilidade $(\Omega, \mathcal{F}, \mathbf{P})$, uma função $X : \Omega \rightarrow Y$ é uma variável aleatória (v.a.) a valores no espaço mensurável (Y, \mathcal{G}) se $X^{-1}(A) \in \mathcal{F}$, para todo $A \in \mathcal{G}$. Neste texto o espaço Y será, na maioria das vezes, a reta real ou um espaço euclidiano \mathbf{R}^d . Mais adiante no estudo, iremos tomar variáveis aleatórias numa variedade diferenciável, no espaço de matrizes ou num grupo de Lie, enfim em qualquer espaço topológico; em todos esses casos a σ -álgebra será a dos borelianos.

Dada uma variável aleatória X em Ω , denotaremos por $\sigma(X)$ a σ -álgebra em Ω gerada por X , i.e., a menor σ -álgebra que faz com que X seja mensurável. Naturalmente $\sigma(X) \subseteq \mathcal{F}$, mas frequentemente

não vale a igualdade. Por exemplo, uma variável aleatória constante $X \equiv c$ gera a σ -álgebra trivial $\sigma(X) = \{\emptyset, \Omega\}$.

Exemplo 1.1 (Espaço de probabilidade finito).

Considere um espaço de probabilidade finito $\Omega = \{1, 2, \dots, n\}$. Atribua a σ -álgebra das partes de Ω , também chamada de σ -álgebra discreta. Assim, todos os subconjuntos unitários $\{i\}$ de Ω são mensuráveis. Qualquer elemento no simplexo $(x_1, x_2, \dots, x_n) \in \mathbf{R}^n$, com $x_j \geq 0$ para todo $1 \leq j \leq n$ e $\sum_{j=1}^n x_j = 1$ determina uma medida de probabilidade em Ω dada por $\mathbf{P}(i) = x_i$ para $1 \leq i \leq n$. Em particular, a medida de probabilidade é chamada de equiprovável se $\mathbf{P}\{i\} = \frac{1}{n}$ para todo $i \in \Omega$. É fácil ver que qualquer função $X : \Omega \rightarrow \mathbf{R}$ é uma variável aleatória. Alguém poderia modelar, por exemplo, o resultado do lançamento de um dado não viciado como a variável aleatória $X(i) = i$, com $n = 6$ e \mathbf{P} equiprovável.

Exemplo 1.2 (Medida de Lebesgue).

Seja $\Omega = \mathbf{R}$, com a σ -álgebra gerada pelos intervalos abertos de \mathbf{R} . Dado um intervalo $(a, b) \in [0, 1]$, atribua uma “pré-medida” $\mathbf{P}(a, b) = (b - a)$. A extensão dessa medida para toda a σ -álgebra dos borelianos é assegurada pelo teorema de extensão abaixo (ver e.g., entre muitas outras alternativas, Bartle [3, Thm. 9.8, p. 103]).

Teorema 1.2 (de extensão de Carathéodory-Hahn). *Dada uma medida (σ -finita) em uma álgebra \mathcal{A} , essa medida se estende unicamente para o completamento de $\sigma(\mathcal{A})$.*

O completamento de uma σ -álgebra \mathcal{F} , denotado por $\overline{\mathcal{F}}$ em relação a uma medida é a σ -álgebra $\sigma(\mathcal{F} \cup Z)$ onde Z é a classe de todos os subconjuntos de elementos em \mathcal{F} que possuem medida nula. Note portanto que o completamento depende da medida que colocamos na σ -álgebra inicial \mathcal{F} . Uma medida em uma álgebra é σ -finita se existem subconjuntos enumeráveis $A_1, A_2, \dots \subseteq \Omega$, todos de medida finita tais que $\Omega = \cup_{i \geq 1} A_i$.

A demonstração desse teorema é assunto clássico em referências de teoria da medida. Está baseado no interessante conceito de *medida exterior*.

O completamento dos borelianos na reta são chamados os conjuntos *Lebesgue mensuráveis*. A medida construída acima, quer esteja

aplicada nos Lebesgue-mensuráveis, quer esteja aplicada nos borelianos (classe menor. Por quê?), é chamada da medida de Lebesgue, que denotaremos por λ . Outra notação frequente que usaremos, sobretudo dentro dos sinais de integração é o clássico: dx .

A construção acima feita na reta se estende para medidas (volume) em hipercubos. Obtem-se assim a medida de Lebesgue nos espaços euclidianos \mathbf{R}^n . Deixamos os detalhes por conta do leitor. Ver também a construção via medida produto, na seção 1.6.

Nos interessa particularmente a medida de probabilidade dada pela medida de Lebesgue no intervalo $\Omega = [0, 1]$. Considere, por exemplo a função de Dirichlet $X : \Omega \rightarrow \mathbf{R}$ dada por $X(\omega) = 1$ se $\omega \in [0, 1]$ for irracional e $X(\omega) = 0$ se $\omega \in [0, 1]$ for racional. X é uma variável aleatória (por quê?).

□

Exemplo 1.3 (Distribuição de Dirac).

A *medida de Dirac*, denotada por δ_{x_0} em um espaço mensurável (Ω, \mathcal{F}) está concentrada no ponto x_0 , i.e. para um conjunto A que é \mathcal{F} -mensurável, $\delta_{x_0}(A) = 1$ se $x_0 \in A$, $\delta_{x_0}(A) = 0$ se $x_0 \notin A$.

□

A mensurabilidade de variáveis aleatórias $X : (\Omega, \mathcal{F}, \mathbf{P}) \rightarrow (Y, \mathcal{G})$ (i.e., a invariância da estrutura mensurável por “pull-back” X^{-1}) permite que empurremos (“push-forward”) a medida \mathbf{P} para o espaço mensurável (Y, \mathcal{G}) , chamada de *medida induzida*, *distribuição* ou ainda de *lei* da variável aleatória X , denotada por $X_*\mathbf{P}$, onde $X_*(B) = \mathbf{P}(X^{-1}(B))$ para todo $B \subseteq Y$ \mathcal{G} -mensurável.

Exemplo 1.4 (Distribuição uniforme).

Dizemos que uma variável aleatória $X : \Omega \rightarrow \mathbf{R}$ é *uniformemente distribuída*, se sua distribuição for a medida de Lebesgue normalizada em um certo intervalo, i.e. X é uniforme no intervalo $[a, b]$ se $X_*\mathbf{P} = 1/(b - a)\lambda$

Exemplo 1.5 (Outros exemplos clássicos de distribuições).

A *distribuição de Poisson* em \mathbf{N} , indexada pelo parâmetro $\lambda \in \mathbf{R}$ é dada por

$$P_\lambda(n) = \frac{\lambda^n e^{-\lambda}}{n!}.$$

A distribuição de poisson tem média e variância iguais a λ .

A *distribuição binomial*, também discreta, corresponde à probabilidade de, em N tentativas com probabilidade de sucesso p , a probabilidade de obter sucesso em k delas. Sua distribuição é dada por

$$P_{n,p}(k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Sua média é np e sua variância $np(1-p)$. Atirar moedas por exemplo, onde temos $p = 1/2$. A probabilidade de depois de se atirar N vezes, obtermos exatamente n caras e $(N-n)$ coroas é dado por $\binom{n}{k} 2^{-N}$.

Uma das distribuições mais importantes para nós é a *distribuição normal*, ou *gaussiana*, com média m e variância σ^2 é dada na reta por

$$P(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-m)^2}{2\sigma^2}\right\}.$$

Voltaremos a falar da distribuição gaussiana na Seção 1.8.

□

Exercício 1.3. Dado o espaço de probabilidade Ω como sendo o intervalo $[0, 1]$ da reta real, descreva a classe de variáveis aleatórias reais $X : \Omega \rightarrow \mathbf{R}$ que tem distribuição dada por $X_*\mathbf{P} = \sum_{i=1}^n \lambda_i \delta_{x_i}$, com $x_1, \dots, x_n \in \mathbf{R}$, $\lambda_1, \dots, \lambda_n \in \mathbf{R}$ e $\sum_{i=1}^n \lambda_i = 1$.

1.3 Esperança de uma variável aleatória

A *esperança* ou *média* de uma variável aleatória $X : (\Omega, \mathcal{F}, \mathbf{P}) \rightarrow \mathbf{R}$ é a integral

$$\mathbf{E}[X] = \int_{\Omega} X(\omega) d\mathbf{P}(\omega).$$

Esse valor também é chamado de *valor esperado* da variável aleatória X . Em certos contextos a notação m_X pode ser mais conveniente para a esperança. Note que a esperança representa, de fato, o sentido mais intuitivo que temos de média como “soma e divide pelo número total”. Só que aqui a “soma” é ponderada pelas medidas dos conjuntos mensuráveis, e o “número total” foi normalizado em um.

Para os leitores ainda não iniciados na teoria de integração de Lebesgue, pensem no caso discreto com $\Omega = \{1, 2, \dots, n\}$ finito e uma medida μ na σ -álgebra discreta de Ω . Dada uma função $f : \Omega \rightarrow \mathbf{R}$, então

$$\int_{\Omega} f(\omega) d\mu(\omega) = \sum_{i=1}^n f(i) \mu(\{i\}).$$

No caso de Ω não ser finito e f ser positiva e limitada, particione-se Ω numa quantidade finita de subconjuntos pela imagem inversa de uma partição de $\mathbf{R}_{\geq 0}$. Usa-se o mesmo somatório acima com os $f(i)$ sendo os limites inferiores dos intervalos da partição de \mathbf{R} . Esse procedimento é chamado de aproximação por funções simples. No caso geral, a integral de Lebesgue será dada pelo limite nessas partições da imagem. Mais precisamente, o que se encontra nos textos clássicos é a seguinte definição: f é uma *função simples* se existe um número finito de subconjuntos B_1, B_2, \dots, B_n mensuráveis em Ω tal que $f(\omega) = \sum_{i=1}^n \lambda_i 1_{B_i}(\omega)$, onde 1_{B_i} é a *função indicadora* ou *função característica* do conjunto B_i , i.e. $1_{B_i}(\omega) = 1$ se $\omega \in B_i$ e $1_{B_i}(\omega) = 0$ se $\omega \notin B_i$. Então a integral de Lebesgue para a função simples f é dada por

$$\int_{\Omega} f d\mu := \sum_{i=1}^n \lambda_i \mu(B_i). \quad (1.1)$$

Ficaremos neste texto somente com essa noção de integral como sendo a soma do produto do valor da função pelo “tamanho” do conjunto onde a função assume esse valor. Deixamos para o leitor verificar os detalhes dessa construção em outro texto. Essencialmente, o procedimento continua por linearidade (para funções limitadas), e limites, quando existir, para f não limitada. A classe de funções Lebesgue integráveis inclui e é muito maior do que a classe das funções Riemann integráveis.

Note que se f for uma função simples, sua representação como combinação linear de funções indicadoras não é única. Deixamos para o leitor:

Exercício 1.4. *Mostre que a definição da integral de Lebesgue como dada pela equação (1.1) não depende da representação que se toma da f como combinação linear de funções características.*

Dentre os teoremas de limites clássicos da integral de Lebesgue, enunciaremos aqui os que faremos referência futuramente. As técnicas de demonstração são clássicas e fáceis de se encontrar, omitiremos neste texto.

Dizemos que uma sequência de funções mensuráveis $f_n : (\Omega, \mathcal{F}, \mu) \rightarrow \mathbf{R}$ converge quase sempre (q.s., ou q.s. ponto a ponto) para uma função mensurável $g : \Omega \rightarrow \mathbf{R}$ se existe um conjunto mensurável E de medida nula tal que para todos os pontos ω em $\Omega \setminus E$, $f_n(\omega)$ converge para $g(\omega)$.

Teorema 1.3 (Teorema da convergência monótona de Beppo-Levy). *Dada uma sequência monótona crescente de funções $f_n : \Omega \rightarrow \mathbf{R}_{\geq 0}$, i.e. $f_n(x) \leq f_{n+1}(x)$ para x q.s. e para todo $n \in \mathbf{N}$ então*

$$\int \lim_{n \rightarrow \infty} f_n(x) \, d\mu(x) = \lim_{n \rightarrow \infty} \int f_n(x) \, d\mu(x).$$

no sentido de que se existir o limite de um lado, do outro lado também existe e é igual.

A demonstração segue praticamente direto da definição da integral de Lebesgue.

Teorema 1.4 (Lema de Fatou). *Dada uma sequência de funções mensuráveis não-negativas $f_n : \Omega \rightarrow \mathbf{R}$, então*

$$\int \liminf_{n \rightarrow \infty} f_n(x) \, d\mu(x) \leq \liminf_{n \rightarrow \infty} \int f_n(x) \, d\mu(x),$$

onde nos limites, fica entendido que é permitido termos valores infinitos.

Teorema 1.5 (Teorema da convergência dominada de Lebesgue). *Considere uma sequência de funções mensuráveis $f_n : \Omega \rightarrow \mathbf{R}$ que converge q.s. a uma função f . Se existir uma função integrável $g : \Omega \rightarrow \mathbf{R}$ tal que $|f_n| \leq g$ q.s. então não só o limite f é integrável como*

$$\int f(x) \, d\mu(x) = \lim_{n \rightarrow \infty} \int f_n(x) \, d\mu(x).$$

Os exercícios abaixo, que aparecem na introdução de [3], mostram como os teoremas de convergência para integral de Lebesgue são mais convenientes que os teoremas de convergência para integral de Riemann, que exigem convergência uniforme:

Exercício 1.5. *Mostre que dada a sequência de funções $f_n : (0, \infty) \rightarrow \mathbf{R}$ com $f_n = e^{-nx}/\sqrt{x}$, então $\lim_{x \rightarrow 0} f_n(x) = +\infty$, portanto f_n converge para 0 não uniformemente, apesar disso, o limite das integrais existe e*

$$\lim_{n \rightarrow \infty} \int_0^{+\infty} f_n dx = 0.$$

Exercício 1.6. *Se*

$$F(t) = \int_0^{+\infty} x^2 e^{-tx} dx,$$

então existe a derivada de F e é dada por

$$F(t) = - \int_0^{+\infty} x^3 e^{-tx} dx$$

Merece mais destaque para o entendimento das técnicas que usaremos a demonstração do seguinte resultado:

Teorema 1.6 (Teorema da medida induzida). *Seja $(\Omega, \mathcal{F}_1, m)$ um espaço de medida, (Y, \mathcal{F}_2) um espaço mensurável, $X : \Omega \rightarrow Y$ e $f : Y \rightarrow \mathbf{R}$ funções mensuráveis. Então*

$$\int_Y f dX_*m = \int_{\Omega} f \circ X dm.$$

Onde a igualdade significa que quando existe a integral de um lado, existe também a do outro lado e os dois números reais são iguais.

Demonstração: Mostramos para o caso de f ser uma função simples, com $f(y) = \sum_{i=1}^n \lambda_i 1_{B_i}(y)$, onde B_1, \dots, B_n são \mathcal{F}_2 -mensuráveis. O lado esquerdo é igual a

$$\sum_{i=1}^n \lambda_i X_*m(B_i)$$

que por definição é igual ao somatório $\sum_{i=1}^n \lambda_i m(X^{-1}(B_i))$, que é o lado direito. No caso geral tomamos aproximações de uma função por funções simples. \square

Exemplo 1.6 (Fórmula de mudança de variável).

Dados U e V abertos de \mathbf{R}^n , um difeomorfismo $g : U \rightarrow V$ e uma função real $f : V \rightarrow \mathbf{R}$ então a fórmula de mudança de variável no cálculo em várias variáveis é dada por:

$$\int_V f(y) |\det(d_y g^{-1})| dy = \int_U f \circ g(x) dx,$$

onde $|\det(d_y g^{-1})|$ é o módulo do determinante do Jacobiano de g^{-1} no ponto y . Comparando com o teorema da medida induzida, o que esta fórmula do cálculo está dizendo é que, quando se toma a medida de Lebesgue λ em U , a medida que o difeomorfismo g induz em sua imagem, i.e. $g_*\lambda$ é representada por $|\det(d_y g^{-1})| dy$, no sentido de que, dado um conjunto $B \subseteq V$ então

$$g_*\lambda(B) = \int_B |\det(d_y g^{-1})| dy = \int_V I_B(y) |\det(d_y g^{-1})| dy$$

\square

A *variância* de uma variável aleatória X é definida pela fórmula:

$$\text{var}(X) = \mathbf{E}[(X - m_X)^2].$$

A variância nos fornece uma primeira avaliação do quanto a variável aleatória se dispersa fora de sua média. Para p inteiro positivo, o p -ésimo momento de uma v.a. é dado por $\mathbf{E}|X|^p$. A variância portanto, é o segundo momento de $(X - \mathbf{E}X)$, chamado de *momento centrado*. Note que $\text{var}(X) = \mathbf{E}X^2 - (\mathbf{E}X)^2$.

A *covariância* entre duas variáveis aleatórias X e Y é definida por:

$$\text{cov}(X, Y) = \mathbf{E}[(X - \mathbf{E}X)(Y - \mathbf{E}Y)].$$

Note que $\text{var}(X + Y) = \text{var}(X) + \text{var}(Y) + 2\text{cov}(X, Y)$. Mais adiante veremos que a covariância é uma medida de quanto as variáveis X e Y são independentes.

Nas aplicações que se faz desta teoria em estatística, em alguns casos não é conveniente que a variância tenha sua unidade dimensional ao quadrado. Por exemplo, se a variável aleatória for de comprimento em metros (m), a variância aparece em metro ao quadrado (m^2). Assim, em aplicações práticas, para que se possa falar que um comprimento aleatório X está com um certo grau de segurança probabilística no intervalo $m_X \pm d(X)$, a grandeza $d(X)$ tem que estar na mesma unidade dimensional de X . Daí, nesses casos eles usam o parâmetro $d(X)$ como a raiz quadrada da variância, chamada de *desvio padrão*.

Exemplo 1.7 (Cálculo da esperança dada a distribuição).

É muito mais comum em problemas aplicados, sobretudo em estatística, conhecer a distribuição de uma variável aleatória do que a função mensurável que a modela propriamente. Usando o teorema da medida induzida, basta saber a distribuição desta v.a. para podermos calcularmos sua esperança (e muito mais coisa, como se verá mais a frente). De fato, dada nossa variável aleatória $X : \Omega \rightarrow \mathbf{R}$, considere a função identidade $id : \mathbf{R} \rightarrow \mathbf{R}$, $id(x) = x$ para todo $x \in \mathbf{R}$. Então, por esse teorema:

$$\int_{\mathbf{R}} id(x) dX_*\mathbf{P}(x) = \int_{\Omega} id \circ X(\omega) d\mathbf{P}(\omega)$$

e o lado direito é a esperança de X . Suponhamos, para fixar as ideias que a distribuição $X_*\mathbf{P}$ de X seja tal que num intervalo (a, b) da reta tenhamos $X_*\mathbf{P}(a, b) = \int_{(a,b)} h(x) dx$ para alguma função integrável h . Abusando da terminologia de medida, isto significa que $h(x) dx = dX_*\mathbf{P}(x)$, no sentido de que se integrarmos em relação à uma medida é o mesmo que se integrar em relação à outra. Aplicando na fórmula acima ficamos com :

$$\mathbf{E}[X] = \int_{\mathbf{R}} xh(x) dx.$$

A função h acima é chamada da densidade da medida $X_*\mathbf{P}$ em relação à medida de Lebesgue λ , que foi escrita acima como dx .

Para o cálculo dos momentos de uma variável aleatória, o argumento é o mesmo e basta calcularmos:

$$\mathbf{E}|X|^p = \int_{\mathbf{R}} x^p h(x) dx.$$

□

Essa idéia de densidade de uma medida em relação a outra é o conteúdo do teorema de Radon-Nikodym 1.9 nas próximas seções.

Lembramos que dado um espaço vetorial real normado V , seu dual (topológico) V^* é o espaço vetorial de aplicações lineares contínuas em \mathbf{R} , os elementos do dual são chamados de funcionais lineares contínuos. A norma de um funcional linear contínuo é dada por

$$\|F\|_{V^*} := \sup\{\|F(x)\| : \|x\|_V = 1\}$$

Se V tem dimensão finita então V^* também tem dimensão finita, por exemplo no espaço euclidiano $V = R^n$ é comum usarmos as bases canônicas $\{e_1, \dots, e_n\}$ de V e $\{dx_1, \dots, dx_n\}$ de V^* .

O último resultado que deixaremos nesta seção mostra a importância das medidas de Radon em diferentes áreas da análise:

Teorema 1.7 (Teorema de Representação de Riesz). *Considere um espaço topológico localmente compacto X (dimensão finita, por exemplo). Seja $C_c(X)$ é o espaço vetorial das funções contínuas com suporte compacto, então todo funcional linear positivo F neste espaço pode ser representado por uma medida de Radon m_F em X . Além disso, a norma do operador $\|F\| = m_F(X)$.*

A representação acima é dada da seguinte maneira: dado $F : C_c(X) \rightarrow \mathbf{R}$, então existe uma (única) medida de Radon μ_F tal que, para toda função contínua $h : X \rightarrow \mathbf{R}$ temos que

$$F(h) = \int_X h(x)\mu(dx).$$

Esse funcional será positivo, i.e. se $h \geq 0$ então a integral em relação a μ também o será. Ver Folland [15] ou Halmos [19].

Se $C_c(X)$ tem dimensão finita, então o espaço X é um espaço finito com n pontos, a identificação entre um funcional $F = \alpha_1 dx_1 + \dots + \alpha_n dx_n$ e a medida de Radon $\mu_F = \alpha_1 \delta_1 + \dots + \alpha_n \delta_n$ é imediata.

Para o teorema de representação de Riesz acima se estender para todos os funcionais lineares, precisamos usar medidas com sinal. Deixamos para o leitor interessado ver na bibliografia recomendada.

1.4 Os espaços L^p

Os espaços vetoriais $L^p(\Omega, \mathcal{F}, \mu; \mathbf{R})$ são espaços de funções reais \mathcal{F} -mensuráveis (ou, como vimos, na linguagem de probabilidade, variáveis aleatórias) onde p determina uma espécie de grau de integrabilidade da função em relação à medida μ . Assim, definimos, para um espaço de medida $(\Omega, \mathcal{F}, \mu)$ e $1 \leq p < \infty$,

$$L^p(\Omega; \mathbf{R}) = \left\{ f : \Omega \rightarrow \mathbf{R} \text{ mensurável e } \int_{\Omega} |f|^p d\mu < \infty \right\}.$$

Para $p = \infty$, tomamos

$$L^\infty(\Omega; \mathbf{R}) = \{f : \Omega \rightarrow \mathbf{R} \text{ mensurável com } \text{supess}|f| < \infty\}.$$

onde o supremo essencial $\text{supess}|f| = c$ quando c é o ínfimo do conjunto $\{x : \mu\{\omega; f(\omega) \leq x\} = 0\}$.

Omitir ou não um ou mais dos parâmetros nos parêntesis da notação $L^p(\Omega, \mathcal{F}, \mu; \mathbf{R})$ depende de alguns desses parâmetros já estarem subentendidos ou se queremos enfatizar algum deles. Enunciamos aqui uma série de propriedades bem conhecidas que deixamos para o leitor verificar futuramente:

1. Os espaços L^p , com $1 \leq p \leq \infty$ são espaços vetoriais.
2. O operador em L^p dada por $f \rightarrow \|f\|_p = \left\{ \int_{\Omega} |f|^p d\mu < \infty \right\}^{1/p}$ é uma norma (pela desigualdade de Minkowsky). O mesmo vale para L^∞ .
3. Vale a chamada *desigualdade de Hölder*: Se $f \in L^p$, $g \in L^q$ com $1/p + 1/q = 1$ então fg é integrável (i.e., está em L^1) além disso $\|fg\|_1 \leq \|f\|_p \|g\|_q$
4. O espaço é completo, i.e., seqüências de Cauchy convergem para elementos dentro do próprio espaço.
5. Para $1 < p < \infty$, o dual topológico (de funcionais lineares contínuos) de L^p se representa em L^q , com $1/p + 1/q = 1$, da seguinte maneira: se $f \in L^p$, dado um funcional linear contínuo

$\alpha \in (L^p)^*$, então existe um elemento $g \in L^q$ tal que $\langle \alpha, f \rangle = \langle g, f \rangle$ onde

$$\langle g, f \rangle_{L^q, L^p} = \int fg \, d\mu.$$

As propriedades 1-4 acima se resumem em dizer que os espaços L^p são espaços de Banach, i.e. são espaços vetoriais normados e completos na métrica induzida pela norma. Para $p = q = 2$, a propriedade (5) diz que o espaço L^2 tem um produto interno dado pela integral do produto acima, portanto é um espaço de Hilbert, por definição deste (i.e., um espaço de Banach onde a norma provém de um produto interno). A propriedade (5) acima lê-se também como uma representação (também chamada de representação de Riesz) do dual de L^2 nele mesmo. Esse produto interno em L^2 torna possível falarmos de ortogonalidade neste espaço, portanto podemos dar algumas interpretações geométricas sobre propriedades de variáveis aleatórias, por exemplo, nas próximas seções veremos como que independência probabilística está relacionado com ortogonalidade neste espaço (ver Teorema 1.18). Observe que o conceito de covariância é justamente o produto interno neste espaço.

Observação 1.7 (L^p como quociente de relação de equivalência).

Rigorosamente falando, o operador $f \rightarrow \|f\|_p$, $1 \leq p \leq \infty$ determina uma norma nas classes de equivalência da relação $f \sim g$ se $f = g$, μ -q.s.. Alguns autores ainda distinguem o espaço \mathcal{L}^p de todas as funções f que são p -integráveis do espaço quociente onde de fato existe a norma $L^p = \mathcal{L}^p / \sim$. Por exemplo, se $\Omega = [0, 1]$ com a medida de Lebesgue, então $f \equiv 1$ e a função de Dirichlet dada por $g(x) = 1$ se x for irracional e $g(x) = 0$ se x for racional, então $f = g$ em $L^p(\Omega, \lambda)$, para todo $1 \leq p \leq \infty$.

□

Teorema 1.8 (Desigualdade de Tchebyshev). *Dada uma variável aleatória X então, para todo número real positivo $a > 0$ e para todo $1 \leq p \leq \infty$ temos que*

$$\mathbf{P}\{|X| > a\} \leq a^{-p} \mathbf{E}|X|^p.$$

Demonstração: Para todo $a > 0$ temos, pela definição de integral que

$$a\mathbf{P}\{|X| > a\} \leq \mathbf{E}|X|.$$

A desigualdade do enunciado agora segue para $1 \leq p \leq \infty$ considerando a v.a. $|X|^p$, o número real a^p e a observação de que $\mathbf{P}\{|X|^p > a^p\} = \mathbf{P}\{|X| > a\}$. □

Exercício 1.8. Verifique que para toda função $g : \mathbf{R} \rightarrow \mathbf{R}_{\geq 0}$ não-decrescente, e todo $a > 0$,

$$\mathbf{P}\{|X| > a\} \leq g(a)^{-1}\mathbf{E}g(X).$$

Exercício 1.9. Verifique que se o p -ésimo momento centrado $m_p = \mathbf{E}|X - \mathbf{E}X|^p$ da variável aleatória X existir, então podemos estimar quanto que a variável aleatória foge da sua média $\mathbf{E}X$ por:

$$\mathbf{P}\{|X - \mathbf{E}X| > km_p\} \leq \frac{1}{k^p}.$$

1.5 Esperança condicional

Dado um espaço mensurável (Ω, \mathcal{F}) e duas medidas μ_1 e μ_2 em \mathcal{F} , dizemos que μ_1 é *absolutamente contínua* com relação à μ_2 se sempre que para um dado conjunto mensurável $A \in \mathcal{F}$ acontecer de $\mu_2(A) = 0$ então $\mu_1(A) = 0$ também. Denotamos por $\mu_1 \ll \mu_2$. O teorema clássico abaixo formaliza as ideias que colocamos no Exemplo 1.7:

Teorema 1.9 (Teorema de Radon-Nikodym). *Dadas μ_1 e μ_2 medidas (σ -finitas) em (Ω, \mathcal{F}) , suponha que $\mu_1 \ll \mu_2$. Então existe uma função $h : \Omega \rightarrow \mathbf{R}_{\geq 0}$, \mathcal{F} -mensurável tal que $d\mu_1 = h(x)d\mu_2$ no sentido de:*

$$\mu_1(B) = \int_B h(x) d\mu_2,$$

para todo $B \in \mathcal{F}$.

A função h representa a densidade de μ_1 em relação à μ_2 e é chamada de derivada de Radon-Nikodym de μ_1 em relação à μ_2 . Frequentemente se escreve $h(x) = \frac{d\mu_1}{d\mu_2}(x)$. Verifique que h é única a

menos dos seus valores em conjuntos mensuráveis que tenham medida μ_2 nula.

Verifique que, uma outra interpretação para o significado da expressão $d\mu_1 = h(x)d\mu_2$ do enunciado do teorema é que para toda função mensurável e limitada f :

$$\int f d\mu_1 = \int fh d\mu_2.$$

Demonstração: Demonstraremos aqui somente para o caso discreto. Para $\Omega = \{1, 2, \dots, n\}$ finito com \mathcal{F} discreta, defina

$$h(j) = \frac{\mu_1\{j\}}{\mu_2\{j\}}$$

para $j \in \Omega$ com $\mu_2\{j\} \neq 0$. Como $\mu_2(\{i\}) = 0$ implica em $\mu_1(\{i\}) = 0$, então de fato não faz diferença qual a definição da função h onde $\mu_2(\{i\}) = 0$ (adote zero inicialmente, por exemplo, para fixar as ideias). Note que h é \mathcal{F} -mensurável e satisfaz a propriedade do enunciado.

Se Ω não for discreto, assuma que ambas μ_1 e μ_2 são finitas. Chame de C o conjunto de todas as funções mensuráveis f limitadas tais que

$$\int_B f d\mu_2 \leq \mu_1(B).$$

Note que C é não vazio (por quê?). Considere a ordem parcial induzida da reta nos elementos de C . Pelo teorema da convergência monôtona, toda sequência monôtona em C converge para um elemento de C . Veja que dadas f_1 e f_2 em C então, naturalmente $\max\{f_1, f_2\}$ também está em C . Assim, pelo lema de Zorn, existe elemento maximal, que pela propriedade anterior é o único elemento máximo que denotaremos por $g \in C$.

Por outro lado, suponha que exista um conjunto B tal que

$$\int_B g d\mu_2 < \mu_1(B),$$

isso vai contradizer a maximalidade da g (omitimos os detalhes dessa parte).

□

Exemplo 1.8 (Mudanças de coordenadas em \mathbf{R}^n).

No Exemplo (1.6) sobre fórmula de mudança de variável por um difeomorfismo $g : U \rightarrow V$ entre abertos de \mathbf{R}^n , a expressão da medida induzida como $|\det(d_y g^{-1})| dy$ significa formalmente que $g * \lambda$ é absolutamente contínua em relação à medida de Lebesgue original λ , além disso, a derivada de Radon-Nikodym é dada por

$$\frac{dg * \lambda}{d\lambda}(y) = |\det(d_y g^{-1})|.$$

para todo y em V .

□

Voltamos ao caso onde $(\Omega, \mathcal{F}, \mathbf{P})$ é um espaço de probabilidade.

Definição 1.3. *Dada uma variável aleatória real X em $L^1(\Omega, \mathcal{F}; \mathbf{R})$ e uma sub- σ -álgebra $\mathcal{G} \subseteq \mathcal{F}$, a esperança condicional $\mathbf{E}[X|\mathcal{G}]$ é uma variável aleatória \mathcal{G} -mensurável tal que*

$$\int_G X d\mathbf{P} = \int_G \mathbf{E}[X|\mathcal{G}] d\mathbf{P} \quad (1.2)$$

para todo conjunto $G \in \mathcal{G}$.

A ideia que se deve ter da esperança condicional é de um filtro ou transformador de uma função \mathcal{F} -mensurável em uma função \mathcal{G} -mensuráveis que não modifica as médias (integrais), quando olhamos nos subconjuntos de \mathcal{G} . Esse conceito é um dos mais importantes e mais ricos em teoria de probabilidade. Um martingale, por exemplo, nosso tipo favorito de processos estocásticos, é uma família de variáveis aleatórias integráveis $(X_t)_{t \in \mathbf{R}_{\geq 0}}$ tal que $\mathbf{E}[X_t|\mathcal{G}_s] = X_s$, onde $(\mathcal{G}_s)_{s \in \mathbf{R}_{\geq 0}}$ é uma família de σ -álgebras. Assim, todas as propriedades da esperança condicional passam para o processo (por exemplo: média constante, ver Teorema 1.11, item (b) abaixo). Voltaremos a falar de martingales no próximo capítulo.

Exemplo 1.9 (Espaço de probabilidade finito).

Suponha $\Omega = \{1, 2, \dots, 10\}$. Seja \mathcal{F} a σ -álgebra discreta, $\mathbf{P} = \sum_{i=1}^{10} \lambda_i \delta_i$, com $\sum_{i=1}^{10} \lambda_i = 1$ e \mathcal{G} a sub- σ -álgebra gerada pelos conjuntos $A_1 = \{1, 2\}$, $A_2 = \{3, 4, 5\}$, $A_3 = \{6\}$ e $A_4 = \{7, 8, 9, 10\}$.

Quando consideramos o espaço mensurável (Ω, \mathcal{G}) , esse se reduz, ou se preferir, fica isomorfo a um espaço mensurável finito com quatro elementos $\Omega' = \{a_1, a_2, a_3, a_4\}$ com a σ -álgebra discreta em Ω' . Toda função \mathcal{G} -mensurável em Ω é constante nos conjuntos A_j 's, com $j = 1, \dots, 4$ (por quê?). Logo, a esperança condicional $\mathbf{E}[\cdot | \mathcal{G}]$ é uma função constante nos A_j 's que preserva a propriedade de integração nos \mathcal{G} -mensuráveis. Dada uma variável aleatória X em Ω , com $X(i) = \alpha_i \in \mathbf{R}$, $i = 1, \dots, 10$ ficamos então com a seguinte descrição de sua esperança condicional:

$$\mathbf{E}[X|\mathcal{G}](i) = \sum_{i \in A_j} \alpha_i \lambda_i = \int_{A_j} X \, d\mathbf{P}$$

quando $i \in A_j$, com $i = 1, \dots, 10$ e $j = 1, \dots, 4$.

No caso geral temos o seguinte teorema de existência e unicidade q.s..

Teorema 1.10 (Existência da esperança condicional). *Dada uma variável aleatória integrável X e uma sub- σ -álgebra \mathcal{G} , existe e é única \mathbf{P} -q.s. a esperança condicional $\mathbf{E}[X|\mathcal{G}]$, que satisfaz a equação (1.2).*

Demonstração: Seja $X^+ = \max\{X, 0\}$ a parte positiva de X . Defina uma nova medida dada por

$$\mathbf{P}_{X^+ A} = \int_A X \, d\mu$$

sobre todos os conjuntos A que sejam \mathcal{G} -mensuráveis. Naturalmente $\mathbf{P}_{X^+} \ll \mathbf{P}|_{\mathcal{G}}$, portanto o teorema de Radon-Nikodym garante que existe uma função h^+ , \mathcal{G} -mensurável tal que $\mathbf{P}_{X^+} = h^+ \mathbf{P}$ sobre os \mathcal{G} -mensuráveis.

Repita o argumento para $X^- = \max\{-X, 0\}$ para obter uma função \mathcal{G} -mensurável h^- . Então a esperança condicional $\mathbf{E}[X|\mathcal{G}] = h^+ - h^-$. Para a unicidade, suponha que uma função g seja \mathcal{G} -mensurável e que também satisfaça a propriedade de integração (1.2). Então, para todo $G \in \mathcal{G}$

$$\int_G \mathbf{E}[X|\mathcal{G}] \, d\mathbf{P}|_{\mathcal{G}} = \int_G g \, d\mathbf{P}|_{\mathcal{G}}$$

se e somente se

$$\int_G (\mathbf{E}[X|\mathcal{G}] - g) d\mathbf{P}|_{\mathcal{G}} = 0$$

Integrando sobre os conjuntos G_1 onde o integrando é positivo e sobre G_2 onde o integrando é negativo, a expressão acima vale se e somente se $(\mathbf{E}[X|\mathcal{G}] - g) = 0$ em $\mathbf{P}|_{\mathcal{G}}$ quase todo ponto. \square

A *probabilidade condicional* vem diretamente da esperança condicional colocando, para todo $A \in \mathcal{F}$, $\mathbf{P}\{A|\mathcal{G}\} = \mathbf{E}[1_A|\mathcal{G}]$. As propriedades seguintes são consequências praticamente diretas da definição, suas demonstrações são bons exercícios para iniciantes:

Teorema 1.11. *Dadas duas variáveis aleatórias $X, Y \in L^1(\Omega)$:*

a) (*linearidade*) Para todos números reais a e b , temos

$$\mathbf{E}[aX + bY | \mathcal{G}] = a\mathbf{E}[X | \mathcal{G}] + b\mathbf{E}[Y | \mathcal{G}];$$

b) (*invariância da média*) $\mathbf{E}[\mathbf{E}[X | \mathcal{G}]] = \mathbf{E}[X]$;

c) (*propriedade de projeção*) $\mathbf{E}[X | \mathcal{G}] = X$ se X for \mathcal{G} -mensurável;

d) (*\mathcal{G} -mensuráveis agem como constantes*) $\mathbf{E}[YX|\mathcal{G}] = Y\mathbf{E}[X | \mathcal{G}]$ se Y for \mathcal{G} -mensurável;

e) (*filtração*) Dados $\mathcal{G}_1 \subseteq \mathcal{G}_2 \subseteq \mathcal{F}$ então

$$\mathbf{E}[X | \mathcal{G}_2 | \mathcal{G}_1] := \mathbf{E}[\mathbf{E}[X | \mathcal{G}_2] | \mathcal{G}_1] = \mathbf{E}[X | \mathcal{G}_1].$$

Exercício 1.10. *Verifique que a esperança condicional $\mathbf{E}[\cdot | \mathcal{G}] : L^1(\Omega, \mathcal{F}) \rightarrow L^1(\Omega, \mathcal{G})$ é um operador positivo, isto é, se $X, Y \in L^1(\Omega, \mathcal{F})$, com $X \leq Y$ então $\mathbf{E}[X | \mathcal{G}] \leq \mathbf{E}[Y | \mathcal{G}]$.*

Exercício 1.11. *Mostre que se $X \in L^2(\Omega, \mathcal{F})$ então a esperança condicional $\mathbf{E}[\cdot | \mathcal{G}] : L^2(\Omega, \mathcal{F}) \rightarrow L^2(\Omega, \mathcal{G})$ é a projeção ortogonal nesse subespaço.*

Desigualdade de Jensen:

Uma função $f : \mathbf{R} \rightarrow \mathbf{R}$ é *convexa* se para todo $x \in \mathbf{R}$, existe um coeficiente real α tal que $f(y) \geq f(x) + \alpha(y - x)$. Geometricamente

significa que seu gráfico tem a região superior convexa em \mathbf{R}^2 . Toda função convexa é contínua, além disso, se f for derivável então dado um ponto x , $\alpha = f'(x)$ é único. Mesmo que não se tenha a unicidade de α em relação a x , as possibilidades de escolha de α são sempre monótonas não decrescentes em relação à x .

Proposição 1.12 (Desigualdade de Jensen). *Dados X uma variável aleatória, f uma função convexa, e \mathcal{G} uma sub- σ -álgebra então:*

- i) $\mathbf{E}[f(X)] \geq f(\mathbf{E}[X]);$
- ii) $\mathbf{E}[f(X)|\mathcal{G}] \geq f(\mathbf{E}[X|\mathcal{G}]);$

Demonstração: Para (i), basta tomar $x = \mathbf{E}[X]$ e um $\alpha \in \mathbf{R}$ que satisfaz a propriedade da definição de convexidade da f em x , então

$$f(X) \geq f(x) + \alpha(X - x).$$

Tomando a esperança nos dois lados, o resultado segue por monotonicidade da esperança. Para (ii), fixemos uma função monótona $\alpha(x)$ para cada $x \in \mathbf{R}$ que satisfaz a propriedade de convexidade da f . Tome x como sendo a variável aleatória $x(\omega) = \mathbf{E}[X|\mathcal{G}]$. Considere a variável aleatória dada pela composição $\alpha(x(\omega))$. Note que $\alpha(x(\omega))$ também é \mathcal{G} -mensurável, já que a função α monótona é boreliano mensurável, portanto:

$$f(X) \geq f(x(\omega)) + \alpha(x(\omega))(X - x(\omega)).$$

Tomando a esperança condicional nos dois lados, o resultado segue por monotonicidade deste operador e pelo Teorema 1.11 [item (d)], que anula o segundo termo do lado direito.

□

Exercício 1.12. *Verifique que o operador linear $\mathbf{E}[\cdot | \mathcal{G}] : L^p(\Omega, \mathcal{F}) \rightarrow L^p(\Omega, \mathcal{G})$ é contínuo. Descreva o núcleo de $\mathbf{E}[\cdot | \mathcal{G}]$.*

Exercício 1.13. *Seja $h : (\Omega, \mathcal{F}, \mathbf{P}) \rightarrow (M, \mathcal{A})$ uma v. a. em um espaço mensurável (M, \mathcal{A}) que gera a sub- σ -álgebra $\sigma(h) \subseteq \mathcal{F}$. Dada uma v. a. real $f \in L^1(\Omega)$, mostre que existe uma função determinística $F : M \rightarrow \mathbf{R}$ integrável no sentido de $F \in L^1(M, \mathcal{A}, h_*\mathbf{P}; \mathbf{R})$ tal que*

$$\mathbf{E}[f|h] = F \circ h.$$

Isto significa que a melhor aproximação para a variável aleatória f , dada h é uma função F (determinística) de h . Sugestão: Defina $d\mathbf{P}_f = f d\mathbf{P}$. Verifique que $h_*(\mathbf{P}_f) \ll h_*(\mathbf{P})$ e que F dado pela derivada de Radon-Nikodym satisfaz a propriedade desejada.

1.6 Independência

Dados dois espaços de probabilidade $(\Omega_1, \mathcal{F}_1, \mathbf{P}_1)$ e $(\Omega_2, \mathcal{F}_2, \mathbf{P}_2)$, definimos um novo espaço de probabilidade chamado de *espaço produto* onde o conjunto base é $\Omega_1 \times \Omega_2$, a σ -álgebra produto, denotada por $\mathcal{F}_1 \otimes \mathcal{F}_2$, é aquela gerada pela álgebra dos retângulos $A \times B$ onde $A \in \mathcal{F}_1$ e $B \in \mathcal{F}_2$. A probabilidade produto $\mathbf{P}_1 \otimes \mathbf{P}_2$ é a extensão da medida na álgebra dos retângulos para a σ -álgebra $\mathcal{F}_1 \otimes \mathcal{F}_2$, onde a medida sobre os retângulos é dada pelo produto $\mathbf{P}_1 \otimes \mathbf{P}_2(A \times B) = \mathbf{P}_1(A) \cdot \mathbf{P}_2(B)$.

Dada uma quantidade infinita (enumerável ou não) de espaços de probabilidade $(\Omega_\lambda)_{\lambda \in \Lambda}$, com as respectivas σ -álgebras $(\mathcal{F}_\lambda)_{\lambda \in \Lambda}$ e medidas de probabilidade $(\mathbf{P}_\lambda)_{\lambda \in \Lambda}$, podemos também definir o espaço produto (infinito) da seguinte maneira: lembramos que um elemento ω desse espaço produto $\Omega = \prod_{\lambda \in \Lambda} \Omega_\lambda$ é uma função $\omega : \Lambda \rightarrow \prod_{\lambda \in \Lambda} \Omega_\lambda$ tal que a “coordenada” λ está em Ω_λ , i.e. $\omega(\lambda) \in \Omega_\lambda$. Tomamos a σ -álgebra $\mathcal{F} = \bigotimes_{\lambda \in \Lambda} \mathcal{F}_\lambda$ como aquela gerada pela álgebra das restrições em um número finito de coordenadas, os chamados *cilindros de dimensão finita* definidos da seguinte maneira: dados uma sequência finita de parâmetros $\lambda_1, \dots, \lambda_n$, uma sequência de subconjuntos mensuráveis nos respectivos espaços $A_1 \in \mathcal{F}_{\lambda_1}, \dots, A_n \in \mathcal{F}_{\lambda_n}$ então o subconjunto

$$C_{\lambda_1, \dots, \lambda_n}^{A_1, \dots, A_n} = \{\omega : \omega(\lambda_i) \in A_i, \text{ para todo } 1 \leq i \leq n\}$$

é o cilindro de dimensão finita determinados por $\lambda_1, \dots, \lambda_n$ e A_1, \dots, A_n .

Exercício 1.14. Verifique que os cilindros de dimensão finita $C_{\lambda_1, \dots, \lambda_n}^{A_1, \dots, A_n}$ formam uma álgebra.

Finalmente a medida de probabilidade \mathbf{P} no produto infinito é a extensão da probabilidade que na álgebra dos cilindros de dimensão

finita é dada por:

$$\mathbf{P}[C_{\lambda_1, \dots, \lambda_n}^{A_1, \dots, A_n}] = \prod_{i=1}^n \mathbf{P}_{\lambda_i}(A_i).$$

No caso de um produto finito de fatores, a construção que fizemos para espaços de probabilidades, vale igualmente para espaços de medida. O caso de quantidade infinita de fatores, se restringe a produtos de espaços de probabilidade.

Teorema 1.13 (Teorema de Tonelli). *Dada uma função mensurável não-negativa f no espaço produto $\Omega_1 \times \Omega_2$ então sua integral no espaço produto pode ser decomposta em integrais nos espaços de probabilidade originais, coordenada a coordenada e independente da ordem:*

$$\begin{aligned} \int_{\Omega_1 \times \Omega_2} f \, d(\mathbf{P}_1 \times \mathbf{P}_2) &= \int_{\Omega_1} \int_{\Omega_2} f(x, y) \, \mathbf{P}_2(dy) \, \mathbf{P}_1(dx) \\ &= \int_{\Omega_2} \int_{\Omega_1} f(x, y) \, \mathbf{P}_1(dx) \, \mathbf{P}_2(dy) \end{aligned}$$

no sentido de, se uma desses valores existirem, então os outros dois também existem e são iguais.

Na demonstração do teorema de Tonelli precisa ser mostrado, antes de tudo, que os integrandos $\int_{\Omega_2} f(x, y) \, \mathbf{P}_2(dy)$ e $\int_{\Omega_1} f(x, y) \, \mathbf{P}_1(dx)$ são funções mensuráveis em x e em y , respectivamente.

Teorema 1.14 (Teorema de Fubini). *Dada $f : \Omega_1 \times \Omega_2 \rightarrow \mathbf{R}$ uma função integrável, então sua integral também pode ser calculada integrando-se coordenada a coordenada como nas fórmulas do teorema de Tonelli 1.13 acima.*

Comparando esses últimos dois teoremas: de um lado o teorema de Tonelli é muito útil para verificar se uma função é integrável, simplesmente aplicando as integrações iteradas no seu módulo. Por outro, sabendo que ela é integrável, o teorema de Fubini nos dá um algoritmo para calcularmos sua integral.

Exercício 1.15. *Considere o seguinte exemplo de espaço produto, $\Omega_1 = \Omega_2 = \mathbf{N}$ com a medida da contagem dada pela série de deltas*

de Dirac sobre cada elemento de Ω_1 . Considere a seguinte função: $a : \mathbf{N} \times \mathbf{N} \rightarrow \mathbf{R}$ dada por $a_{n,n} = 1$, $a_{n,n+1} = -1$ e todos os outros valores são zero. Verifique que

$$\begin{aligned} \int_{\Omega_2} \int_{\Omega_1} a_{i,j} d\delta_i d\delta_j &= \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} a_{i,j} \\ &\neq \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} a_{i,j} \\ &= \int_{\Omega_1} \int_{\Omega_2} a_{i,j} d\delta_j d\delta_i. \end{aligned}$$

Isso contradiz os teoremas de Tonelli ou de Fubini?

Definição 1.4. Dado um espaço de probabilidade $(\Omega, \mathcal{F}, \mathbf{P})$, dizemos que dois eventos $A, B \in \mathcal{F}$ são independentes em relação à medida \mathbf{P} se $\mathbf{P}\{A \cap B\} = \mathbf{P}(A)\mathbf{P}(B)$.

Generalizando essa definição, dizemos que duas famílias de subconjuntos A_λ e B_α são independentes se $\mathbf{P}\{A_\lambda \cap B_\alpha\} = \mathbf{P}(A_\lambda)\mathbf{P}(B_\alpha)$ para todos os pares de elementos A_λ e B_α . Um conjunto finito de σ -álgebras $(\mathcal{F}_j)_{1 \leq j \leq n}$ é dito mutuamente independente (em relação à medida \mathbf{P}) se $\mathbf{P}(\bigcap_{j=1}^n A_j) = \prod_{j=1}^n \mathbf{P}(A_j)$, para todo $A_j \in \mathcal{F}_j$. Uma família de σ -álgebras é dita mutuamente independente se a propriedade do produto acima vale para todo subconjunto finito de índices dessa família. Dizemos então que uma família de variáveis aleatórias $(X_\lambda)_{\lambda \in \Lambda}$ são (mutuamente) independentes se as respectivas σ -álgebras $\sigma(X_\lambda)_{\lambda \in \Lambda}$ são (mutuamente) independentes (ver e.g. Halmos [19, §45]). Neste texto, usaremos a terminologia simplificada de “independência” para nos referir a independência mútua.

Exemplo 1.10 (Espaços produtos).

Tipicamente variáveis aleatórias independentes aparecem em espaços produtos. Se um espaço de probabilidade é dado por um produto, digamos $\Omega = \prod_{i=1}^{\infty} \Omega_i$; então, dadas variáveis aleatórias $X_i : \Omega \rightarrow \mathbf{R}$ onde cada X_i só depende da i -ésima coordenada, então X_1, X_2, \dots são variáveis aleatórias independentes

□

Exemplo 1.11 (V.a. independentes e identicamente distribuídas - i.i.d.).

No exemplo acima, se além da sequência X_1, X_2, \dots ser independente, as distribuições forem iguais, chamamos essa sequência de i.i.d.. Tipicamente as sequências i.i.d. são construídas como cópias de uma v.a. original. Dada uma variável aleatória X sobre um espaço de probabilidade $(\Omega, \mathcal{F}, \mathbf{P})$. Considere o espaço produto $\tilde{\Omega} = \prod_{n=1}^{\infty} \Omega$, com a correspondente σ -álgebra $\tilde{\mathcal{F}}$ e probabilidade $\tilde{\mathbf{P}}$. Os elementos de $\tilde{\Omega}$ são sequências de elementos de Ω , isto é, $\tilde{\omega} = (\omega_1, \omega_2, \dots)$. Defina em $\tilde{\Omega}$ a seguinte sequência de variáveis aleatórias: $X_i(\tilde{\omega}) = X(\omega_i)$. Assim, sobre o espaço estendido $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbf{P}})$ construímos uma infinidade de cópias da variável aleatória original X , além disso, essa sequência é i.i.d.. Esse modelo corresponde a nossa experiência com várias sequências de variáveis aleatórias bastante familiares como jogar moedas, atirar dados, roletas, etc.

□

Exercício 1.16. *Dê exemplo de um espaço de medida com três subconjuntos mensuráveis que são dois a dois independentes mas não são mutuamente independentes.*

Exercício 1.17. *Mostre que se uma variável aleatória X for independente da σ -álgebra \mathcal{G} , então a esperança condicional $\mathbf{E}[X|\mathcal{G}] = \mathbf{E}[X]$.*

Terminamos essa seção com resultados clássicos úteis na teoria. Lembre que o limite superior de uma sequência de conjuntos $(A_n)_{n \geq 1}$ é definido o conjunto de pontos que estão em infinitos A_n 's, i.e.:

$$\limsup_{n \rightarrow \infty} A_n = \bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n.$$

Analogamente,

$$\liminf_{n \rightarrow \infty} A_n = \bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} A_n.$$

que é o conjunto de pontos que a partir de um determinado $N \in \mathbf{N}$, aparece em todo A_n com $n > N$

Exercício 1.18. *Mostre que para toda seqüência de conjuntos A_n , $n \geq 1$, temos que*

$$\liminf A_n \subseteq \limsup A_n.$$

Lema 1.15 (Lema de Borel-Cantelli).

- 1) *Dada uma seqüência de conjuntos mensuráveis $(A_n)_{n \geq 1}$, se $\sum_{n=1}^{\infty} \mathbf{P}(A_n) < \infty$ então $\mathbf{P}(\limsup A_n) = 0$*
- 2) *Se os conjuntos mensuráveis da seqüência $(A_n)_{n \geq 1}$ forem independentes e $\sum_{n=1}^{\infty} \mathbf{P}(A_n) = \infty$ então $\mathbf{P}(\limsup A_n) = 1$*

Demonstração: Para (1), tome $F_k = \bigcup_{n=k}^{\infty} A_n$ então $0 \leq \mathbf{P}(F_k) \leq \sum_{n=k}^{\infty} \mathbf{P}(A_n)$ que vai para zero quando k vai para infinito. Como $\mathbf{P}(\limsup A_n) \leq \mathbf{P}(F_k)$ para todo $k \in \mathbf{N}$, então $\mathbf{P}(\limsup A_n) = 0$. Para (2), assumiremos que temos independência mútua e vamos mostrar que o complementar $\bigcup \bigcap A_n^c$ tem probabilidade nula. De fato, seja agora $F_k = \bigcap_{n=k}^{\infty} A_n^c$. Então:

$$\begin{aligned} \mathbf{P}(F_k) &= \prod_{n=k}^{\infty} \mathbf{P}(A_n^c) \\ &= \prod_{n=k}^{\infty} (1 - \mathbf{P}(A_n)) \\ &\leq \prod_{n=k}^{\infty} \exp\{-\mathbf{P}(A_n)\} \end{aligned}$$

já que $(1 - a) \leq e^{-a}$ para todo $a \in \mathbf{R}$. Portanto, pela divergência da série, o produtório vale zero. Assim, o complementar do $\limsup A_n$ é a união enumerável de conjuntos de probabilidade zero, portanto $\mathbf{P}(\limsup A_n) = 1$. □

Para ilustrar a primeira parte do lema, suponha que X_n é uma seqüência de v.a. discretas digamos a valores em $\{0, 1\}$ tal que $\mathbf{P}(X_n = 0) = 1/n^2$ para cada $n \geq 1$. Assim, a série $\sum_n \mathbf{P}(X_n = 0)$ converge e o lema de Borel-Cantelli assegura que a probabilidade de $X_n = 0$ ocorrer um número infinito de vezes é zero (aqui $A_n = X_n^{-1}(0)$). Dizendo ainda em outras palavras, observado os resultados

da sequência X_n , com probabilidade um ela só vale zero uma quantidade finita de vezes. Note que a convergência da série é crucial! Num jogo de apostas, essa convergência (i.e., essa integrabilidade em relação à medida infinita $\mu = \sum_i^\infty \delta_i$) pode fazer a diferença entre perder tudo ou sair milionário!

Sobre a segunda parte, a ideia é que, ou os conjuntos A_n já são grandes (probabilidade um), ou são pequenos, mas pela independência e divergência da série, é possível cobrir todo ponto de Ω (q.s.) um número infinito de vezes.

1.7 Convergências de variáveis aleatórias

Segue abaixo um resumo das principais topologias (formas de convergência) no espaço de variáveis aleatórias que usaremos:

- 1) Uma sequência de v.a. $X_n : \Omega \rightarrow \mathbf{R}$ converge para uma certa v.a. X *quase sempre* (q.s.) ou \mathbf{P} -q.s. se existe um conjunto mensurável Ω' , com $\mathbf{P}(\Omega') = 1$ tal que a para todo $\omega \in \Omega'$, a sequência numérica $X_n(\omega)$ converge para $X(\omega)$.

Escrevemos $X = \lim X_n$ q.s..

- 2) A sequência X_n converge para X *em probabilidade* se para todo $\epsilon > 0$,

$$\lim_{n \rightarrow \infty} \mathbf{P}\{\omega; |X_n - X| > \epsilon\} = 0.$$

Escreveremos $X = \mathbf{P}\text{-lim } X_n$

- 3) A sequência X_n converge para X *em L^p* , $1 \leq p \leq \infty$ se $\lim_{n \rightarrow \infty} \|X_n - X\|_p = 0$

Escreveremos $X = L^p\text{-lim } X_n$.

Para os leitores ainda não familiarizados com a teoria, deixamos como exercício fazer os diagramas das implicações dos tipos de convergência (dominada e não dominada), baseados nos seguintes exercícios (ver, entre outros, Bartle [3]):

Exercício 1.19. 1. *Convergência q.s. implica em convergência em probabilidade. Convergência em probabilidade garante que*

existe uma subsequência que converge q.s.. Dê exemplos de variáveis aleatórias que convergem em probabilidade mas não convergem q.s..

2. *Convergência em L^p implica em convergência em probabilidade e que existe uma subsequência que converge q.s. Se a convergência for dominada então convergência em L^p é equivalente à convergência em probabilidade.*

Dentro desse contexto, o teorema da convergência dominada 1.5 garante que convergência q.s. de funções limitadas (“dominadas”) por uma função integrável então a convergência se dá também em L^p .

Dada uma sequência de medidas de probabilidade μ_n em um espaço métrico, normalmente completo e separável, $(M, (B))$, dizemos que a sequência μ_n converge fracamente para μ se μ_n convergir na topologia fraca-* para μ no dual de $C(M; \mathbb{R})$, isso é, se para toda função contínua com suporte compacto $f : M \rightarrow \mathbb{R}$ tivermos

$$\lim_{n \rightarrow \infty} \int_M f d\mu_n = \int_M f d\mu.$$

Dizemos que uma sequência de variáveis aleatórias X_n definidas em espaços de probabilidades (Ω_n, \mathbf{P}_n) converge em *distribuição* ou em *lei* para uma variável aleatória X se suas distribuições $(X_n)_* \mathbf{P}_n$ convergirem fracamente para a distribuição de X . Note que as v.a. acima não precisam necessariamente estar definidas no mesmo espaço de probabilidade.

1.8 Variáveis aleatórias gaussianas

Conforme vimos, frequentemente, sobretudo em modelagem estatística, a informação que temos de uma variável aleatória X é a sua distribuição $X_* \mathbf{P}$ em \mathbf{R} e não propriamente a função $X : \Omega \rightarrow \mathbf{R}$ que a define. O teorema da medida induzida já nos mostrou que isso é suficiente para calcularmos a média (esperança) e os momentos dessa variável aleatória. Nesta seção apresentaremos a distribuição gaussiana que é mais importante em teoria de probabilidade. A importância da gaussiana num primeiro momento se baseia em vários fatos: o primeiro deles é que são fechadas para limites q.s., limites em

probabilidade, limites em L^2 , convergência fraca. Outro fato essencial é que essas variáveis aleatórias formam um espaço vetorial dentro da variáveis aleatórias quadrado integráveis. (Ver mais comentários no Apêndice).

Definição 1.5. Dizemos que uma medida de probabilidade é gaussiana, ou normal $N(m, C)$, em \mathbb{R}^d se sua densidade é dada por:

$$P(x) = \frac{\sqrt{\det C^{-1}}}{(2\pi)^{d/2}} \exp\left\{-\frac{\langle C^{-1}(x - m), (x - m) \rangle}{2}\right\},$$

onde $m \in \mathbb{R}^d$ é chamado de média e C é uma matriz positiva definida chamada de matriz de covariância.

Analogamente, uma variável aleatória $X = (X_1, \dots, X_d)$ é chamada uma gaussiana $N(m, C)$ em \mathbb{R}^d se sua distribuição for uma medida gaussiana $N(m, C)$.

A terminologia para os parâmetros $m \in \mathbb{R}^d$ e a matriz positiva definida C vem do fato de que, se X é uma variável aleatória $N(m, C)$, então sua média é dada por $m = \mathbf{E}[X]$ e a matriz de covariâncias $C = \mathbf{E}[(X - m)(X - m)^t]$, onde “ t ” aqui significa transposta. Se a projeção ortogonal da v.a. X em uma certa direção $v \neq 0$ for constante, i.e. tiver variância nula, diremos que essa gaussiana é degenerada na direção v . Estamos assumindo que as medidas gaussianas são sempre não degeneradas, isso é, são absolutamente contínuas em relação à medida de Lebesgue ($\det C \neq 0$). Dizemos que uma variável aleatória $N(m, C)$ é gaussiana centrada se $m = 0$.

Exercício 1.20. Mostre que a distribuição $P(x)$ definida acima de fato corresponde a uma medida de probabilidade: $\int_{\mathbb{R}^d} P(x) dx = 1$ (sugestão: use mudança de variável para diagonalizar $C = Id_{d \times d}$, depois coordenadas polares em \mathbb{R}^2 para verificar que $\int \exp\{x_i^2/2\} dx_i = \sqrt{2\pi}$ para cada $i = 1, 2, \dots, d$).

A transformada de Fourier de uma medida μ em \mathbb{R}^d é a função complexa no dual de \mathbb{R}^d dada por:

$$\phi_\mu(u) = \int_{\mathbb{R}^d} e^{i\langle u, y \rangle} d\mu(y).$$

O integrando é limitado, portanto a integral sempre existe. A teoria de análise de Fourier garante que a transformada de Fourier acima caracteriza a medida μ (ver e.g. Folland [15]). Essa transformada também se aplica em medidas sobre espaços de Banach, neste caso, na fórmula acima, u é um elemento do dual e $\langle u, y \rangle$ é a avaliação $u(y)$.

Analogamente, para variáveis aleatórias, se X for uma v.a. em \mathbb{R}^d , sua transformada de Fourier é a transformada de Fourier de sua distribuição, isto é, pelo teorema da medida induzida:

$$\phi_X(u) = \mathbf{E}[e^{i\langle X, u \rangle}].$$

O exercício abaixo envolve somente calcular uma integral. A fórmula está dizendo que as distribuições das medidas gaussianas são autovetores do operador linear dado pela transformada de Fourier.

Exercício 1.21. *Verifique que uma medida gaussiana $N(m, C)$ tem transformada de Fourier dada por $\phi(u) = \exp\{i \langle u, m \rangle - \frac{1}{2} \langle Cu, u \rangle\}$. Conclua que as gaussianas centradas estão no autoespaço de autovetores associado ao autovalor 1 da transformada de Fourier.*

A fórmula do exercício também será útil para demonstrarmos os teoremas abaixo, que mostram mais propriedades básicas das gaussianas.

O primeiro enunciado é consequência direta das definições e da unicidade da transformada de Fourier:

Teorema 1.16. *Se uma sequência de variáveis gaussianas X_n converge em distribuição para a uma variável aleatória X , então X também é gaussiana.*

Demonstração: De fato, pela definição de convergência em distribuição temos que a transformada de Fourier $\phi_X(u)$ converge para o limite das transformadas:

$$\int_{\mathbb{R}^d} e^{i\langle u, x \rangle} P_X(dx) = \lim_{n \rightarrow \infty} \int_{\mathbb{R}^d} e^{i\langle u, x \rangle} P_{X_n}(dx).$$

Isto é, se cada X_n tem média m_n e matriz de covariância C_n então

$$\begin{aligned} \phi_X(u) &= \lim_{n \rightarrow \infty} \exp\{i \langle u, m_n \rangle - \frac{1}{2} \langle C_n, u \rangle\} \\ &= \exp\{i \langle u, \lim_{n \rightarrow \infty} m_n \rangle - \frac{1}{2} \langle \lim_{n \rightarrow \infty} C_n, u \rangle\}. \end{aligned}$$

Portanto, não só X é gaussiana como tem média e matriz de covariância dadas pelo limite desses valores.

□

Teorema 1.17. *Uma variável aleatória $X = (X_1, \dots, X_d)$ é gaussiana em \mathbb{R}^d se e somente se $Y = \lambda_1 X_1 + \dots + \lambda_d X_d$ é gaussiana para todo $\lambda_1, \dots, \lambda_d \in \mathbf{R}$.*

Demonstração: Se X for uma gaussiana em \mathbf{R}^d com média $m = (\mathbf{E}(X_1), \dots, \mathbf{E}(X_d))$, então para todo $u \in \mathbf{R}$, e todo $\bar{\lambda} = (\lambda_1, \dots, \lambda_d) \in \mathbf{R}^d$, sua transformada de Fourier satisfaz:

$$\begin{aligned} \phi_X(u\lambda_1, \dots, u\lambda_d) &= \mathbf{E} [e^{iu(\lambda_1 X_1 + \dots + \lambda_d X_d)}] \\ &= \exp\{i \langle u\bar{\lambda}, m \rangle - \frac{1}{2} \langle Cu\bar{\lambda}, u\bar{\lambda} \rangle\} \\ &= \exp\{iu \langle \bar{\lambda}, m \rangle - \frac{1}{2} u^2 \langle C\bar{\lambda}, \bar{\lambda} \rangle\} \\ &= \phi_Y(u) \end{aligned}$$

Portanto $Y = \lambda_1 X_1 + \dots + \lambda_d X_d$ é gaussiana com média $\langle \bar{\lambda}, m \rangle$ e variância $\langle C\bar{\lambda}, \bar{\lambda} \rangle$.

Reciprocamente, suponha que fixado um $\bar{\lambda} = (\lambda_1, \dots, \lambda_d) \in \mathbf{R}^d$, temos que Y seja normal, com média $\mathbf{E}Y = m_Y$ e variância $\mathbf{E}(Y - m_Y)^2 = \sigma_Y^2$. Então

$$\begin{aligned} \phi_Y(u) &= \mathbf{E} [e^{iu(\lambda_1 X_1 + \dots + \lambda_d X_d)}] \\ &= \exp\{iu(m_Y) - \frac{1}{2} u^2 \sigma_Y^2\}, \end{aligned}$$

com $m_Y = \langle \bar{\lambda}, m_X \rangle$ e

$$\begin{aligned} \sigma_Y^2 &= \mathbf{E} \left(\sum_{j=1}^d \lambda_j X_j - \sum_{j=1}^d \lambda_j m_j \right)^2 \\ &= \mathbf{E} \left(\sum_{j=1}^d \lambda_j (X_j - m_j) \right)^2 \\ &= \sum_{i,j=1}^d \lambda_i \lambda_j c_{ij}, \end{aligned}$$

isto é

$$\phi_Y(u) = \exp\{iu \langle \bar{\lambda}, m \rangle - \frac{1}{2}u^2 \langle C\bar{\lambda}, \bar{\lambda} \rangle\} = \phi_X(u\bar{\lambda})$$

portanto X é gaussiana. \square

Como consequência imediata deste último resultado temos que dadas duas variáveis aleatórias gaussianas X e Y independentes, considerando a matriz de covariância em \mathbf{R}^2 de (X, Y) , concluímos que $X + Y$ é gaussiana com média $m_X + m_Y$ e variância $\text{var}X + \text{var}Y$.

O próximo teorema mostra que vale a recíproca do teorema de integração de funções independentes desde que essas funções tenham distribuições gaussianas:

Teorema 1.18. *Dadas duas variáveis aleatórias gaussianas centradas X e Y , elas são independentes se e somente se $\mathbf{E}[XY] = 0$*

Demonstração: Se X e Y forem v.a. independentes, a teoria de integração garante que a integral do produto é o produto das integrais (ver, por exemplo Halmos [19, §45, Thm A]).

Para estabelecer a recíproca vamos provar que

$$P_1 := \mathbf{P}\{\omega : X(\omega) \in G_1 \text{ e } Y(\omega) \in G_2\}$$

é dado pelo produto $\mathbf{P}\{X \in G_1\}\mathbf{P}\{Y \in G_2\}$, para todo $G_1, G_2 \subseteq \mathbf{R}$ mensurável.

Considere a v.a. $Z(\omega) = (X, Y) \in \mathbf{R}^2$, com média $m = 0$ e matriz de covariância

$$C = \begin{pmatrix} \sigma_X^2 & 0 \\ 0 & \sigma_Y^2 \end{pmatrix}.$$

Então $P_1 = \mathbf{P}\{Z \in G_1 \times G_2\}$, portanto

$$\begin{aligned} P_1 &= \int_{G_1 \times G_2} \frac{\sqrt{\det C^{-1}}}{2\pi} \exp\left\{-\frac{1}{2} \left\langle C^{-1} \begin{pmatrix} x \\ y \end{pmatrix}, \begin{pmatrix} x \\ y \end{pmatrix} \right\rangle\right\} dx dy \\ &= \frac{1}{2\pi\sigma_X\sigma_Y} \int_{G_1 \times G_2} \exp\left\{-\frac{x^2}{2\sigma_X^2}\right\} \exp\left\{-\frac{y^2}{2\sigma_Y^2}\right\} dx dy \\ &= \left(\frac{1}{2\pi\sigma_X} \int_{G_1} \exp\left\{-\frac{x^2}{2\sigma_X^2}\right\} dx\right) \left(\frac{1}{2\pi\sigma_Y} \int_{G_2} \exp\left\{-\frac{y^2}{2\sigma_Y^2}\right\} dy\right) \\ &= \mathbf{P}\{X \in G_1\} \mathbf{P}\{Y \in G_2\} \end{aligned}$$

□

Teorema 1.19. *Se X_n é uma seqüência de v.a. gaussianas em \mathbb{R}^d e são tais que X_n converge em probabilidade para uma v.a. X , então X também é gaussiana.*

Demonstração: A métrica do ângulo no círculo é menor que a métrica euclídeana em \mathbb{R}^2 . Assim, temos que para todo $x, y, u \in \mathbb{R}^d$

$$\begin{aligned} |e^{i\langle u, x \rangle} - e^{i\langle u, y \rangle}| &\leq |\langle u, x \rangle - \langle u, y \rangle| = |\langle u, x - y \rangle| \\ &\leq |u| \cdot |x - y|. \end{aligned}$$

Assim, fixado $u \in \mathbb{R}^d \setminus \{0\}$ temos

$$\mathbf{P} \left\{ \omega : |e^{i\langle u, X_n \rangle} - e^{i\langle u, X \rangle}| > \varepsilon \right\} \leq \mathbf{P} \left\{ \omega : |X_n - X| > \frac{\varepsilon}{|u|} \right\}.$$

Portanto, pela hipótese, temos que $e^{i\langle u, X_n \rangle}$ também converge para $e^{i\langle u, X \rangle}$ em probabilidade. Como nestas v.a. as normas são limitadas (constantes iguais a um), pelo teorema da convergência dominada temos que $e^{i\langle u, X_n \rangle}$ converge para $e^{i\langle u, X \rangle}$ em L^1 , isto é

$$\lim_{n \rightarrow \infty} \mathbf{E} e^{i\langle u, X_n \rangle} = \mathbf{E} e^{i\langle u, X \rangle},$$

portanto em termos das transformadas de Fourier:

$$\begin{aligned} \phi_X(u) &= \lim_{n \rightarrow \infty} \exp\{i \langle u, m_n \rangle - \frac{1}{2} \langle C_n u, u \rangle\} \\ &= \exp\{i \langle u, \lim_{n \rightarrow \infty} m_n \rangle - \frac{1}{2} \langle \lim_{n \rightarrow \infty} C_n u, u \rangle\} \end{aligned}$$

portanto não só X é uma v.a. gaussiana como sua média e matriz de covariância são os limites da média e matriz de covariância de X_n . □

Observação 1.22 (Subespaço das gaussianas em L^2).

Como convergência em L^p implica em convergência em probabilidade, o resultado acima vale para convergência em L^p também. Em particular, considere o seguinte subespaço:

$$\mathcal{H}_1 = \{ \text{v.a. gaussianas de média } 0 \} \subset L^2(\Omega, \mathcal{F}, \mathbf{P}).$$

Pelo teorema acima, \mathcal{H}_1 é um subespaço fechado, portanto é um espaço de Hilbert. Além disso, pelo Teorema 1.18, temos que $f, g \in \mathcal{H}_1$ são ortogonais se e somente se são independentes. Voltaremos a esta observação no apêndice.

□

Capítulo 2

Processos estocásticos

Um processo estocástico é uma família de variáveis aleatórias em um espaço de estados mensurável (M, \mathcal{E}) que, tipicamente, é um espaço métrico com a σ -álgebra dos borelianos \mathcal{E} . Indexaremos essa família de v.a. por um conjunto abstrato T . Assim, um *processo estocástico* é uma aplicação $X : T \times \Omega \rightarrow M$, $(\omega, t) \mapsto X_t(\omega)$ tal que, fixado $t \in T$, $X_t(\omega)$ é mensurável em ω . Na grande maioria das vezes T representa o tempo: contínuo (\mathbb{R} ou $\mathbb{R}_{\geq 0}$) ou discreto (\mathbb{Z} ou \mathbb{N}).

Observação 2.1 (Sobre a generalidade do indexador T).

Embora essa interpretação do T sendo o tempo é a mais intuitiva e natural, a definição genérica acima para o parâmetro T pode ser útil em contextos que aparecerão posteriormente. Por exemplo, processos indexados em grupos. Para os iniciados vai aqui um exemplo interessante: fixado um ponto $x_0 \in M$ em uma variedade diferenciável, uma ação de um grupo de Lie G em M dada por $g \mapsto gx_0$ é um exemplo de um processo estocástico indexado em G onde Ω tem um único elemento. Para um Ω menos trivial, tome cada elemento deste espaço como sendo ações distintas de G em M . Outro exemplo mais ligado ao assunto principal deste texto: uma das construções do movimento browniano no Apêndice está baseada em um processo estocástico (gaussiano, como será visto) indexado em um espaço de Hilbert; tal processo será uma isometria entre elementos de T e gaussianas com a norma de $L^2(\Omega)$. Ainda um terceiro exemplo são os

chamados campos aleatórios na abordagem que faz Kunita [30] para uma demonstração do critério de Kolmogorov no espaço euclidiano onde T é um aberto de \mathbb{R}^d .

□

2.1 Processo canônico e continuidade

A menos que se faça menção explícita, em todo esse capítulo T denotará o tempo. Assim, podemos falar das trajetórias de um processo estocástico que é a dinâmica temporal quando o parâmetro aleatório ω é fixado, i.e., $X(\omega) : T \rightarrow M$. Se denotarmos por M^T o conjunto de todas as funções de T para M , as trajetórias de um processo é um subconjunto de M^T .

Para um subconjunto finito $\bar{t} = \{t_1, \dots, t_n\} \subseteq T$, denote por $\rho_{\bar{t}} : M^T \rightarrow M^{\bar{t}}$ a projeção dada pela restrição das funções em M^T nos pontos de \bar{t} . Em $M^{\bar{t}}$ temos a σ -álgebra natural $\mathcal{E}^{\bar{t}}$ que faz com que a bijeção dada pela avaliação $f \in M^{\bar{t}} \mapsto (f(t_1), \dots, f(t_n)) \in (M^n, \mathcal{E}^n)$ seja mensurável (em outras palavras, $\mathcal{E}^{\bar{t}}$ e \mathcal{E}^n são isomorfas). Os conjuntos $\rho_{\bar{t}}^{-1}(A) \in M^T$, com $A \in \mathcal{E}^{\bar{t}}$ são chamados de *cilindros de dimensão finita* em M^T . Verifique que, analogamente ao que fizemos na seção de espaços produtos, interseções e complementos desses cilindros são uniões finitas de cilindros, portanto eles formam uma álgebra de subconjuntos de M^T . Denotaremos por \mathcal{E}^T a σ -álgebra gerada pelos cilindros de dimensão finita. As probabilidades induzidas nas restrições finitas $M^{\bar{t}}$, i.e. as medidas $(\rho_{\bar{t}})_*(X_*\mathbf{P})$ são chamadas de distribuição de dimensão finita do processo X .

Note que, se tivermos dois subconjuntos finitos do tempo $\bar{t} \subseteq \bar{t}'$ e se $\pi_{\bar{t}', \bar{t}} : M^{\bar{t}'} \rightarrow M^{\bar{t}}$ for a projeção ou restrição de $M^{\bar{t}'}$ em $M^{\bar{t}}$ então, por composição, as medidas nos cilindros de dimensão finita satisfazem $(\pi_{\bar{t}', \bar{t}})_*(\rho_{\bar{t}'})_*(X_*\mathbf{P}) = (\rho_{\bar{t}})_*(X_*\mathbf{P})$. Essa compatibilidade das probabilidades nos cilindros de dimensão finita é condição necessária para, futuramente, fazermos o procedimento inverso, isto é, dadas probabilidades nos cilindros de dimensão finita, construirmos um processo. Ver condição de Chapman-Kolmogorov 2.1 mais a frente.

Todo processo estocástico tem uma representação canônica no espaço de suas próprias trajetórias. De fato, considere a variável

aleatória sobre o espaço de trajetórias $\tilde{X} : \Omega \rightarrow (M^T, \mathcal{E})$ dada por $\omega \mapsto (X_t(\omega) : T \rightarrow M)$. A mensurabilidade de X_t em ω para todo t , implica na mensurabilidade de \tilde{X} na álgebra dos cilindros de dimensão finita. Daí segue que \tilde{X} é mensurável em relação à \mathcal{E}^T . Considere sobre M^T a medida induzida $X_*(\mathbf{P})$.

Definição 2.1. *A representação canônica de um processo estocástico X é o processo \tilde{X} definido sobre o espaço de probabilidade nas trajetórias $(M^T, \mathcal{E}^T, X_*\mathbf{P})$, com $\tilde{X} : M^T \times T \rightarrow M$, dado pela avaliação $\tilde{X}_t(\omega) = \omega(t)$, onde $\omega \in M^T$.*

A ideia do processo canônico é podermos intuir sobre a dinâmica do processo, olhando as probabilidades diretamente nas trajetórias. Note que apesar de estarem definidos em espaços de probabilidade diferentes, os processos X e \tilde{X} tem as mesmas distribuições de dimensão finita, i.e. a mesma medida induzida nos cilindros de dimensão finita.

Nas próximas seções mostraremos uma construção que heurísticamente é o inverso do que fizemos aqui: ao invés de termos um processo e considerarmos as distribuições de dimensão finita, dada uma família de distribuições de dimensão finita desejada, construiremos um processo estocástico que satisfaça essas distribuições nos cilindros de dimensão finita. (Esse é o conteúdo do teorema de extensão de Kolmogorov 2.2, mais a frente).

Um processo estocástico X é chamado de um *processo gaussiano* se suas distribuições de dimensão finita para cada $\bar{t} = \{t_1, \dots, t_n\}$ são gaussianas em \mathbb{R}^n . O processo é chamado de *estacionário* se suas distribuições de dimensão finita não se alteram por translações no tempo, i.e., $(X_t)_{t \geq 0}$ e $(X_{t+h})_{t \geq 0}$ têm as mesmas distribuições de dimensão finita.

As classes de equivalências entre processos estocásticos que usaremos aqui estão em três níveis.

1. Dados dois processos $X : \Omega \times T \rightarrow M$ e $Y : \Omega' \times T \rightarrow M$, dizemos que X e Y são *versões* um do outro se suas distribuições de dimensão finita são iguais.
2. Dois processos X e Y definidos no mesmo espaço de probabilidade são ditos *modificação* um do outro se, para todo $t \in T$, $X_t = Y_t$ \mathbf{P} -q.s..

3. Por fim, a equivalência mais fina: X e Y são ditos *indistinguíveis* se $\mathbf{P}\{X_t = Y_t \text{ para todo } t \in T\} = 1$.

Naturalmente indistinguilidade implica em modificação, que por sua vez implica em versão.

Exemplo 2.1 (Modificação que não é indistinguível).

Seja $\Omega = [0, 1]$ com a medida de Lebesgue e $T = [0, 1]$. Defina os processos estocásticos $X_t(\omega) = \omega$ e $X'_t(\omega) = \omega$ se $t \neq \omega$, senão $X'_t(\omega) = 0$. Então $\mathbf{P}\{\omega : X_t(\omega) = X'_t(\omega)\} = \mathbf{P}\{[0, 1] \setminus \{t\}\} = 1$, portanto X e X' são modificações um do outro. Porém, $\mathbf{P}\{\omega : X_t(\omega) \neq X'_t(\omega) \text{ para todo } t \in T\} = 0$ portanto X e X' não são indistinguíveis. \square

Note que o processo canônico \tilde{X} associado a um processo X , como fizemos acima, é um exemplo de um versão de X , já que tem as mesmas leis, mesmo estando definido em espaços de probabilidades diferentes. Outra situação frequente onde se considera versões de processos é quando lidamos com um movimento browniano: estamos na realidade lidando com uma das (infinitas) versões desse processo, cuja caracterização está nas leis do processo. Exemplo onde aparecerão processos que são modificação um do outro está na próxima seção, na construção das trajetórias contínuas do movimento browniano. Finalmente, exemplos de processos indistinguíveis aparecem quando estudamos soluções de equações diferenciais estocásticas: se os campos forem lipschitziano, as soluções de uma equação diferencial estocástica é única a menos de indistinguilidade.

Definição 2.2. Se T e E são espaços topológicos então um processo $X : \Omega \times T \rightarrow E$ é contínuo q.s. se $\mathbf{P}\{\Omega : t \mapsto X_t(\omega) \text{ é contínua}\} = 1$.

Exercício 2.2. Dados T um espaço topológico separável e E um espaço de Hausdorff, se $X, X' : \Omega \times T \rightarrow E$ são modificações um do outro e q.s. contínuos então X e X' são indistinguíveis.

2.2 Filtração e tempos de parada.

Uma família de sub- σ -álgebras $(\mathcal{F}_t)_{t \geq 0} \subseteq \mathcal{F}$ é chamada de *filtração* no espaço mensurável (Ω, \mathcal{F}) se for crescente no sentido de $\mathcal{F}_s \subseteq \mathcal{F}_t$

quando $0 \leq s \leq t$. Quando necessário enfatizar, denotamos o espaço de probabilidade filtrado por $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbf{P})$. Um processo X é dito *adaptado* à uma filtração $(\mathcal{F}_t)_{t \geq 0}$ se X_t for \mathcal{F}_t -mensurável para todo $t \geq 0$. Todo processo é adaptado a sua filtração natural denotada $\overset{\circ}{\mathcal{F}}_t = \sigma\{X_s, 0 \leq s \leq t\}$. Limites laterais de filtrações são definidos, à esquerda, como $\mathcal{F}_{t-} = \bigvee_{s < t} \mathcal{F}_s$, i.e., a σ -álgebra $\sigma(\bigcup_{s < t} \mathcal{F}_s)$; e à direita, como $\mathcal{F}_{t+} = \bigcap_{s > t} \mathcal{F}_s$ e $\mathcal{F}_\infty = \bigvee_{t \geq 0} \mathcal{F}_t$. Uma filtração é dita *contínua à direita* se $\mathcal{F}_t = \mathcal{F}_{t+}$.

Um variável aleatória real $T \geq 0$ é um *tempo de parada* em relação à uma filtração (\mathcal{F}_t) se o conjunto $\{\omega : T(\omega) \leq t\}$ for \mathcal{F}_t -mensurável para todo $t \geq 0$.

Exercício 2.3. *Verifique que se a filtração \mathcal{F}_t for contínua à direita, então a definição acima é equivalente a dizermos que $\{T < t\}$ é \mathcal{F}_t -mensurável.*

Uma idéia intuitiva que ajuda a entender o conceito de tempo de parada começa com o seguinte: interprete uma σ -álgebra como sendo a possibilidade de se obter informações sobre as variáveis aleatórias mensuráveis em relação a ela. Assim, quanto maior for uma σ -álgebra, mais informações temos sobre uma família de variáveis aleatórias. Dado um processo adaptado em relação a uma filtração, conhecer um processo até um tempo t significa conhecer a σ -álgebra \mathcal{F}_t . Um tempo de parada T é uma variável aleatória que sinaliza algum acontecimento nas trajetórias (por exemplo entrada em um certo subconjunto no espaço de estados, como veremos na proposição abaixo), portanto, T só pode sinalizar que algo ocorre até o instante t se essa informação estiver em \mathcal{F}_t , caso contrário, a informação seria não-causal.

Dado um processo estocástico adaptado X em \mathbb{R}^d e $G \subseteq \mathbb{R}^d$, o tempo de entrada de X em G é dado pela variável aleatória $T_G(\omega) = \inf\{t \geq 0 : X_t(\omega) \in G\}$, onde convencionamos que $\inf\{\emptyset\} = \infty$.

Proposição 2.1. *Seja X um processo estocástico contínuo com \mathcal{F}_t sua filtração natural.*

- i) *Se G for fechado então T_G é um \mathcal{F}_t tempo de parada.*
- ii) *Se G for aberto então T_G é um \mathcal{F}_{t+} tempo de parada.*

Demonstração: Para (i), basta verificar que

$$\begin{aligned} \{T_G \leq t\} &= \{\omega : \inf_{s \leq t, s \in \mathbb{Q}} d(X_s(\omega), G) = 0\} \\ &= \bigcap_{n=1}^{\infty} \bigcup_{s \leq t, s \in \mathbb{Q}} \{\omega : d(X_s(\omega), G) < 1/n\} \in \mathcal{F}_t \end{aligned}$$

Para (ii):

$$\{T_G < t\} = \bigcup_{s \leq t, s \in \mathbb{Q}} \{X_s \in G\} \in \mathcal{F}_t.$$

Agora use o exercício 2.3 e o fato da filtração $(\mathcal{F}_{t+})_t$ ser contínua a direita.

□

O resultado acima pode ser melhorado observando que a filtração natural de um processo contínuo é contínua a direita.

Dado um \mathcal{F}_t -tempo de parada T , define-se a σ -álgebra \mathcal{F}_T formada pelos conjuntos $A \in \mathcal{F}_{\infty}$ tal que $A \cap \{T \leq t\}$ é \mathcal{F}_t -mensurável para todo $t \geq 0$.

Exercício 2.4. *Mostre que se T e S são tempos de parada então $S \vee T$ e $S \wedge T$ também são tempos de parada. Se $S \leq T$ então $\mathcal{F}_S \subseteq \mathcal{F}_T$.*

Exercício 2.5. *Dada uma sequência T_n de \mathcal{F}_t -tempos de parada, então $\sup_n T_n$ é tempo de parada. Se a filtração for contínua então $\inf_n T_n$, $\limsup_n T_n$ e $\liminf_n T_n$ são tempos de parada. Se $T_n \downarrow T$ então $\mathcal{F}_T = \bigcap_n \mathcal{F}_{T_n}$.*

2.3 Processos de Markov

Começamos apresentando o modelo discreto de processos de Markov, que também é chamado de *cadeia de Markov* finita. Dado um espaço de estados com número finito de elementos $M = \{e_1, e_2, \dots, e_n\}$, pensando no modelo de uma partícula saltando em tempos discretos entre esses estados, atribui-se a cada estado e_i uma probabilidade da partícula, estando ali, de saltar para o estado e_j . A essa probabilidade denotaremos por $P_{i,j}$, chamada de probabilidade de transição.

Naturalmente $\sum_j^n P_{i,j} = 1$ para todo $i \in [1, n]$. Outra observação importante aqui é que essas probabilidades de transição não se alteram com o tempo, neste caso dizemos que o processo (ou cadeia) de Markov é *homogêneo* no tempo .

Tecnicamente, um processo discreto $(X_n)_{n \in \mathbb{N}}$ é de Markov se a probabilidade condicional satisfizer

$$\mathbb{P}(X_{n+1} = x_{n+1} | X_0 = x_0, \dots, X_n = x_n) = \mathbb{P}(X_{n+1} = x_{n+1} | X_n = x_n)$$

para todo $n \geq 1$ e para toda sequência x_0, x_1, \dots, x_{n+1} de elementos no espaço de estados M . Essa condição significa que a incerteza do processo no tempo $t = n + 1$ (no futuro, digamos) só depende do estado imediatamente anterior $t = n$ (no presente, digamos), e não dos estados ou da trajetória nos tempos anteriores.

A intuição sobre as possíveis trajetórias e a evolução de um processo de Markov, pode ser pensada como se tivéssemos uma enorme quantidade de partículas distribuídas no espaço de estados onde cada uma obedece individualmente às leis de transição do parágrafo acima. Visto dessa maneira, naturalmente que as trajetórias e a evolução da distribuição dessas partículas no espaço de estados dependerá de como essas partículas foram distribuídas inicialmente.

Considere uma distribuição inicial $\mu_0 = \alpha_1 \delta_{e_1} + \dots + \alpha_n \delta_{e_n}$, com $\alpha_i \geq 0$, para todo $i \in [1, n]$ e $\sum_1^n \alpha_i = 1$. Podemos escrever essa distribuição como uma matriz linha dada por $\mu_0 = [\alpha_1 \dots \alpha_n]$. Com essa notação matricial, a distribuição do processo no tempo $n = 1$ será dada por

$$\mu_1 = \mu_0 P,$$

onde $P = (p_{i,j})_{1 \leq i,j \leq n}$ é a matrix das probabilidades de transição. Alguns autores preferem trabalhar com a transposta desta matrix de tal maneira que a ação sobre uma distribuição de probabilidade é feita à esquerda ao invés da direita, como fizemos acima. Preferimos nesta ordem porque no caso contínuo, dada uma distribuição inicial, a distribuição no tempo futuro será dada como uma ação adjunta (transposta, no caso finito dimensional como aqui). Na n -ésima iterada temos então que a distribuição é dada por

$$\mu_n = \mu_0 P^n.$$

Denotaremos as entradas da matriz P^n por $P(n)_{i,j}$ que corresponde à probabilidade de, saindo da posição i , chegar-se à posição j depois de n passos.

Esse fato pode ser verificado por indução, entrada por entrada. Ainda mais geral que isso:

Exercício 2.6. *Mostre que as matrizes com entradas não-negativas, com soma nas linhas (ou colunas) igual a um forma um semigrupo dentro do espaço das matrizes.*

As matrizes desse semigrupo não são necessariamente inversíveis, por exemplo:

$$P = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}.$$

Nem é comutativo, por exemplo:

$$\begin{pmatrix} 1/2 & 1/2 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1/2 & 1/2 \end{pmatrix} \neq \begin{pmatrix} 0 & 1 \\ 1/2 & 1/2 \end{pmatrix} \begin{pmatrix} 1/2 & 1/2 \\ 1 & 0 \end{pmatrix}.$$

É da propriedade de semigrupo acima que fica garantido que se pode compor matrizes de probabilidades de transição distintas, digamos, P e Q , para formarmos outros tipos de processos de Markov, só que agora, não mais homogêneos no tempo.

Dentre todas as distribuições por onde podemos inicializar um processo de Markov, existe (pelo menos) uma bastante interessante chamada de *medida invariante* que é tal que, ao longo do processo, permanece constante: $\mu_0 = \mu_0 P^n$, para todo $n \in \mathbf{N}$. A descrição matricial dessa medida corresponde a um autovetor associado ao autovalor 1 da ação de P à direita em \mathbf{R}^n . A existência da medida invariante, neste contexto é aplicação direta de álgebra linear e é conhecido como o teorema de Perron-Frobenius, que garante também que os autovalores tem módulo menores ou igual a 1. Uma outra maneira geométrica de se entender esse teorema é notando que a ação de P é uma isometria no cone de seqüências positivas de comprimento n com a norma l_1 , e existe um ponto neste cone que é atrator dos outros. Note ainda que, dada uma medida inicial qualquer, sua órbita pela ação de P , dada por $\mu_{n+1} = \mu_n P$ é uma seqüência infinita em um espaço compacto (o cone acima), portanto existe (pelo menos) uma subseqüência convergente para uma distribuição μ

Se essa medida invariante for única neste espaço, ela é chamada de *medida ergódica em relação ao processo de Markov* determinado por P . Um exemplo disso é quando a matriz P é irredutível. Se existir dois subespaços complementares onde a ação de P em cada um for irredutível, então, para cada um deles, existe uma medida invariante ergódica distinta, com suportes disjuntos em M .

Em outros textos o leitor encontrará também, ao invés de medida invariante, a denominação de *medida estacionária*, do processo de Markov, no sentido de que é para onde as distribuições do processo convergem quando o tempo vai para infinito, dada qualquer distribuição inicial. Isso está relacionado com o fato dos outros autovalores da matriz P serem menores do que 1, o que faz com que a distribuição invariante do parágrafo acima seja um ponto atrator de todas as outras distribuições.

No caso do espaço de estados M e do tempo t serem contínuos, uma das definições equivalentes de um processo de Markov é uma extensão das ideias dos parágrafos anterior. Considere um espaço métrico completo e separável (M, d) , uma *família* ou um *sistema de probabilidades de transição* $P_{s,t}(x, \cdot)$, é uma família de medidas de probabilidades sobre os borelianos de M indexados por $0 \leq s \leq t$ e $x \in M$. Interpretamos $P_{s,t}(x, A)$ como sendo a probabilidade de, a partir do ponto $x \in M$ no instante s , chegar-se no conjunto mensurável $A \subseteq M$, no instante $t \geq s$. Para viabilizar a existência de processos estocásticos satisfazendo essas probabilidades de transição, precisamos assumir que essa família satisfaz a seguinte *condição de compatibilidade*, conhecida como condição de Chapman-Kolmogorov:

$$P_{s,v}(x, A) = \int P_{s,t}(x, dy) P_{t,v}(y, A), \quad (2.1)$$

para todo $0 \leq s \leq t \leq v$, $x \in M$ e A boreliano de M . Como no caso discreto, estaremos interessados em sistemas de probabilidades de transição homogêneos no tempo, i.e. $P_{s,t}(x, A) = P_{0,t-s}(x, A)$. Escreveremos simplesmente $P_t(x, A)$ ao invés de $P_{s,s+t}(x, A)$.

Um processo $X_t : \Omega \times T \rightarrow M$ é de Markov (homogêneo) se ele obedece uma família de probabilidades de transição $P_t(x, dy)$. no sentido de $\mathbf{E}[1_{X_{s+t} \in A} | X_s] = P_t(X_s, A)$. Escrevendo de outra maneira,

se o processo X_t for \mathcal{F}_t -adaptado, então para toda função limitada $f \in C(M; \mathbf{R})$ temos

$$\mathbf{E}[f(X_{s+t})|\mathcal{F}_s] = (\mathbf{P}_t f)(X_s)$$

onde o operador $P_t f : C(M) \rightarrow C(M)$ é definido por $(P_t f)(x) = \int P_t(x, dy) f(y)$.

Uma medida μ é invariante se μ for invariante pela convolução com $P_t(x, \cdot)$ para todo $t > 0$, i.e. se

$$\mu(A) = \int_M P_t(x, A) \mu(dx)$$

para todo mensurável A . Combinação linear de medidas invariantes é uma medida invariante. Por outro lado, medidas invariantes minimais, i.e., que não podem ser decompostas em soma de outras medidas invariantes são chamadas *medidas ergódicas*. Todo espaço compacto tem pelo menos uma medida invariante (Teorema de Krylov-Bogoliubov). Ver Yosida [62], [61]. Todo espaço localmente compacto (de dimensão finita) tem uma única decomposição das medidas invariantes em medidas ergódicas, dada uma família de probabilidades de transição.

Exemplo 2.2 (Processos discretos).

Todo processo de Markov discreto em E finito satisfaz a condição de Chapman-Kolmogorov. De fato, dada a matriz de probabilidades de transição P , e uma função limitada $f : E \rightarrow \mathbf{R}$, fixado um ponto $i \in E$, então sua integral em relação à medida $P_1(i, j)$ é dada por

$$\int_E f(j) P_1(i, dj) = \sum_{j=1}^n f(j) p_{i,j}.$$

Da mesma maneira,

$$\int_E f(j) P_k(i, dj) = \sum_{j=1}^n f(j) p^{(k)}_{i,j}$$

para todo $k \in \mathbf{N}$. Assim, a condição de Chapman-Kolmogorov fica:

$$\begin{aligned} p(m+n)_{i,j} &= \int P^m(i, dy) P^n(y, j) \\ &= \sum_{k=1}^n p(m)_{ik} P(n)_{kj} \\ &= (P^m P^n)_{i,j} \end{aligned}$$

□

Exemplo 2.3 (Sequências i.i.d.).

Dada uma sequência de variáveis aleatórias i.i.d. X_1, X_2, \dots , analisar essas v.a. como um processo estocástico indexados em n X_n é um processo de Markov onde as probabilidades de transição são independentes do valor do processo no instante imediatamente anterior, i.e.

$$\mathbf{P}\{X_{n+1} \in A | \sigma(X_n)\} = \mathbf{P}\{X_1 \in A\}$$

para todo n .

□

Exemplo 2.4 (Sistemas determinísticos são processos de Markov).

Seja $f : M \rightarrow M$ uma dinâmica determinística no espaço métrico M . Então a sequência dada pela órbita de um ponto x_0 pela aplicação f é um processo de Markov a tempo discreto, sobre o espaço de probabilidade de um único ponto $\Omega = \{\omega\}$, que fica omitido na notação. As probabilidades de transição são dadas por medidas δ de Dirac $P_1(x, \cdot) = \delta_{f(x)}(\cdot)$

Se φ_t é um fluxo associado a um campo de vetores, então a dinâmica é também um processo de Markov a tempo contínuo, sobre o espaço de um espaço de probabilidade de um único ponto $\Omega = \{\omega\}$. As probabilidades de transição são dadas por $P_t(x, \cdot) = \delta_{\varphi_t(x)}(\cdot)$.

□

O problema que focaremos agora é o inverso dos exemplos que vimos: ao invés de ser dado um processo estocástico, nos é dada uma família de probabilidades de transição. Precisamos garantir a existência de um processo estocástico satisfazendo essas probabilidades de transição.

Denotamos por $\Omega = M^{\mathbb{R}^+}$ o conjunto de todas as funções de \mathbb{R}_+ em M , com a σ -álgebra \mathcal{F} gerada pelos cilindros de dimensão finita, i.e. tal que toda função avaliação $A_t : \Omega \rightarrow M$ dada por $A_t(\omega) = \omega(t)$, é mensurável. Temos então o seguinte teorema geral de existência de processos de Markov:

Teorema 2.2 (Teorema de extensão de Kolmogorov). *Dadas probabilidades de transição $P_t(\cdot, \cdot)$ e uma medida inicial ν em um espaço métrico completo e separável (M, d) , existe uma única medida de probabilidade \mathbf{P} em $(M^{\mathbb{R}^+}, \mathcal{F})$ tal que o processo canônico $X : M^{\mathbb{R}^+} \times \mathbb{R}_+ \rightarrow M$ dado por $X_t(\omega) = \omega(t)$ é um processo de Markov associado à $P_t(\cdot, \cdot)$ e tal que X_0 tem distribuição ν .*

Demonstração: Considere inicialmente a álgebra dos subconjuntos de $M^{\mathbb{R}^+}$ que são cilindros de dimensão finita definidos por:

$$C_{t_1, \dots, t_n}^{A_1, \dots, A_n} = \{ \omega \in M^{\mathbb{R}^+} : \omega(t_1) \in A_1, \dots, \omega(t_n) \in A_n, \text{ com} \\ 0 \leq t_1 \leq t_2 \leq \dots \leq t_n, \text{ e } A_1, \dots, A_n \in \mathcal{B}(\mathbb{R}) \}$$

Defina a seguinte medida sobre essa álgebra de cilindros:

$$\mathbf{P}(C_{t_1, \dots, t_n}^{A_1, \dots, A_n}) = \int_M \int_{A_1} \dots \int_{A_{n-1}} P_{t_{n-1}, t_n}(x_{n-1}, A_n) \cdot \\ P_{t_{n-2}, t_{n-1}}(x_{n-2}, dx_{n-1}) \dots \\ \dots P_{t_1, t_2}(x_1, dx_2) P_{t_1}(x_0, dx_1) d\nu(x_0).$$

A medida de probabilidade acima está bem definida já que pela equação de Chapman-Kolmogorov

$$\mathbf{P}(C_{t_1, \dots, t_i, \dots, t_n}^{A_1, \dots, A_i=M, \dots, A_n}) = \mathbf{P}(C_{t_1, \dots, \widehat{t}_i, \dots, t_n}^{A_1, \dots, \widehat{A}_i, \dots, A_n}).$$

O resultado do teorema agora segue pelo teorema de extensão de Carathéodory-Hahn 1.2 que estende unicamente medidas σ -finitas em álgebras para medidas na σ -álgebras gerada. \square

Verifique que a demonstração do teorema acima também assegura o seguinte: considere uma família de medidas de probabilidade $\mathbf{P}^{\vec{t}}$

nos espaços produto $\{E^{\bar{t}}\}_{\bar{t}}$, com $\bar{t} = \{0 \leq t_1 \leq t_2 \dots \leq t_n\} \subset \mathbb{R}_+$. Assuma a condição de compatibilidade dada por $(\pi_{\bar{t}_1, \bar{t}_2})_* \mathbf{P}^{\bar{t}_1} = \mathbf{P}^{\bar{t}_2}$, para todo $\bar{t}_2 \subseteq \bar{t}_1$ onde $\pi_{\bar{t}_1, \bar{t}_2} : E^{\bar{t}_1} \rightarrow E^{\bar{t}_2}$ é a projeção. Então existe uma única medida de probabilidade \mathbf{P} em $E^{\mathbb{R}_+}$ tal que $\mathbf{P}^{\bar{t}} = (\pi_{\mathbb{R}_+, \bar{t}})_* \mathbf{P}$ para todo \bar{t} finito.

2.4 Movimento Browniano

Dentre os processos estocásticos, o movimento browniano (abreviamos MB) é aquele que nos desperta mais interesse pela riqueza de propriedades, tanto do ponto de vista analítico quanto geométrico. A partir dessas propriedades, vê-se que o cálculo estocástico não serve somente como mais um modelo de ruído em sistemas dinâmicos, mas principalmente, porque o conjunto de propriedades dos processos envolvidos (martingales e semimartingales, como definiremos mais a frente), é uma ferramenta que traz informações sobre propriedades analíticas e geométricas do espaço onde trabalhamos. Inclusive informações determinísticas podem ser detectadas probabilisticamente, como veremos mais a frente.

Um *movimento browniano* na reta \mathbb{R} , ou MB linear, é um processo estocástico $B_t : \Omega \times \mathbf{R}_{\geq 0} \rightarrow \mathbf{R}$ tal que

1. tem trajetórias contínuas e inicializadas na origem: $B_0(\omega) = 0$ q.s.
2. Os incrementos no tempo $(B_t - B_s)_{0 \leq s \leq t}$ são gaussianas centradas de variância $(t - s)$, isto é, são $N(0, (t - s))$
3. Os incrementos em intervalos de tempo disjuntos são independentes: $(B_v - B_u)$ e $(B_t - B_s)$ são gaussianas independentes quando $0 \leq s \leq t \leq u \leq v$.

Neste primeiro momento apresentamos acima o movimento browniano linear em \mathbf{R} . Uma vez familiarizado com este processo, é fácil estendermos a definição para o espaço euclidiano \mathbf{R}^d , basta que no item (2) acima tenhamos $(B_t - B_s)_{0 \leq s \leq t} \in N(0, (t - s))$. Na prática, basta tomarmos d movimentos brownianos independentes na reta B_t^1, \dots, B_t^d que $B_t = (B_t^1, \dots, B_t^d)$ é um movimento browniano em \mathbf{R}^d . Veja mais adiante a caracterização de Levy, Teorema 3.9

Na literatura, existem várias maneiras equivalentes de se construir esse processo. Construiremos com detalhes uma dessas abordagens, via processos de Markov, que é a que consideramos a mais intuitiva do ponto de vista da dinâmica do fenômeno contínuo. Deixamos para apresentar no apêndice outras duas abordagens menos diretas (uma analítica funcional e outra via convergência de séries trigonométricas com coeficientes aleatórios) que também podem ser interessantes para explorar outras ideias no futuro (análise no espaço de trajetórias, simulações e outros). Em termos de simulação vale a pena mencionar que passeios aleatórios em uma grade (reticulado) discreta, ortogonal e regular em \mathbf{R}^d , com equiprobabilidade de saltar ao longo de cada uma das $2d$ arestas simulam a trajetória de um movimento browniano, embora não simule a velocidade de espalhamento. Um modelo ainda melhor, que simula também a velocidade é: cada salto é feito em um direções aleatórias uniformemente distribuída na esfera S^{d-1} e o comprimento é dado por uma variável aleatória de Poisson. Ver Pinsky [45].

Antes de continuarmos, verifique a seguinte equivalência:

Exercício 2.7. *Se X é um processo contínuo onde os incrementos são v.a. gaussianas centradas, mostre que X é um movimento browniano (incrementos independentes e $\mathbf{E}[X_t^2] = t$) se e somente se $\mathbf{E}[X_s X_t] = \inf\{s, t\}$.*

A independência entre os incrementos do processo querem dizer que temos um processo de Markov, além disso pelo fato das distribuições desses incrementos serem gaussianas, temos que fica bem definida as probabilidades nos cilindros de dimensão finita. A partir disso fazemos uma extensão da medida, da álgebra dos cilindros para a σ -álgebra gerada por elas, conforme se viu no Teorema de Extensão de Kolmogorov 2.2.

Finalmente, a continuidade das trajetórias ficará estabelecida, a menos de modificação do processo pelo seguinte teorema, cuja demonstração deixamos para o apêndice. Uma versão mais geral para campos aleatórios pode ser vista em Kunita [30].

Teorema 2.3 (Critério de Kolmogorov). *Dado um processo estocástico X , suponha que existam constantes $\alpha, \beta, K > 0$ tais que os*

incrementos do processo satisfaçam

$$\mathbf{E}[|X_{t+h} - X_t|^\alpha] \leq Kh^{1+\beta},$$

para todo $t, h \geq 0$. Então existe um processo contínuo \tilde{X}_t que é uma modificação de X .

Para o processo que estamos construindo, usamos as derivadas no zero da chamada função geradora de momentos $f(t) = \mathbf{E}[\exp\{t(B_{t+h} - B_t)\}]$, para verificar que $\mathbf{E}[(B_{t+h} - B_t)^4] = 3h^2$, portanto o critério de Kolmogorov se aplica e fica estabelecida a existência de uma modificação do processo gaussiano que construímos que tem trajetórias contínuas. A função geradora de momento acima é a *transformada de Laplace* de uma v.a. ou de uma medida, como acontece com a transformada de Fourier. É conhecido que a transformada de Laplace de uma gaussiana $N(m, \sigma^2)$ é dada por $\exp\{-mt + t^2\sigma^2/2\}$ (compare com o Exercício 1.21).

Espaço de Wiener. O espaço de probabilidade da versão canônica de um MB é chamado de *espaço de Wiener* $(W, \mathcal{F}, \mathbb{W})$, onde

$$W = \{\omega : \mathbf{R}_{\geq 0} \rightarrow \mathbf{R} \text{ contínua e } \omega(0) = 0\},$$

\mathcal{F} é a σ -álgebra gerada pelos cilindros de dimensão finita e \mathbb{W} é a medida induzida que garante que esse é o espaço do processo canônico. A medida \mathbb{W} é chamada de *medida de Wiener*. Esse espaço é um dos primeiros exemplos clássicos de uma medida gaussiana em um espaço de Banach.

Processo de Wiener e movimento browniano são sinônimos e conforme a conveniência da notação denotaremos esse processo por B_t ou por W_t .

Exercício 2.8. *Verifique que a σ -álgebra gerada pelos cilindros de dimensão finita coincide com a dos borelianos gerados pela métrica do supremo em $C([0, 1]; \mathbb{R})$.*

Observação 2.9 (“Medida” de ruído branco).

Seja $h \in L^2([0, 1])$ uma função simples subordinada a uma partição $0 = t_0 < t_1 < \dots < t_n = 1$ tal que $h(x) = \sum a_i \mathbf{1}_{[t_{i-1}, t_i]}$, com $a_i \in \mathbb{R}$.

Defina o operador linear contínuo $X : L^2([0, 1]) \rightarrow L^2(\Omega)$ que nas funções simples satisfaz:

$$X(h) = \sum a_i(B_{t_i} - B_{t_{i-1}})$$

$X(h)$ pode ser visto como uma integral $\int_0^t h(s) dB_s$, onde, ao invés da medida de Lebesgue nos intervalos, temos a medida $(B_{t_i} - B_{t_{i-1}})$ nos subintervalos da partição. De fato, essa integral existe para toda função $h \in L^2([0, 1])$, no sentido de limite em $L^2(\Omega)$, já que qualquer f pode ser aproximada em $L^2([0, 1])$ por funções simples. Mais adiante veremos que X é uma isometria chamada de *integral de Wiener*, e será estendida no próximo capítulo para a integral de Itô, onde o integrando é também um processo estocástico. A medida a valores em $L^2(\Omega)$ na reta dada por $(B_{t_i} - B_{t_{i-1}})$ é chamada de *ruído branco*, ver Apêndice ou Nualart [42].

□

A seguinte proposição mostra que, dado um movimento browniano (MB), podemos construir facilmente uma infinidade de outros a partir dele:

Proposição 2.4. *Seja B_t um movimento browniano. Então*

1. *Fixado $s > 0$, temos que $(B_{t-s} - B_s)_{t \geq 0}$ também é um MB independente de $\sigma\{B_u, u \leq s\}$.*
2. *$-B_t, t \geq 0$ é MB.*
3. *Para todo $c > 0$, $cB_{t/c^2}, t \geq 0$ é MB.*
4. *O processo X definido por $X_0 = 0$ e $X_t = tB_{1/t}, t > 0$ é MB.*

As demonstrações podem ser feitas como exercício. Note que o item (iv) significa que para todo MB B_t temos que

$$\lim_{t \rightarrow \infty} \frac{B_t}{t} = 0 \quad \mathbf{P}\text{-q.s..}$$

Observação 2.10 (Para uma segunda leitura). *A propriedade assintótica acima é válida para a maioria dos martingales que trataremos neste texto. Sendo mais preciso, se um martingale M_t satisfaz*

$$\int_0^\infty \frac{d[M, M]_{>t}}{(1+t)^2} < \infty$$

onde $[M, M]$ é a variação quadrática de M (ver Seção 2.5.2) então sua média assintótica é zero, ver Liptser [36].

2.5 Martingales

A teoria de martingales se estende enormemente, com várias intersecções interessantes com problemas geométricos, analíticos, além dos probabilísticos, naturalmente. A intenção desta seção é apresentar sucintamente a linguagem e as propriedades básicas úteis para a dinâmica estocástica.

Considere uma filtração em σ -álgebras $(\mathcal{F}_t)_{t \in T}$ em $(\Omega, \mathcal{F}, \mathbf{P})$ e X um processo \mathcal{F}_t -adaptado. O tempo T pode ser contínuo $\mathbf{R}_{\geq 0}$, \mathbf{R} ; ou discreto \mathbf{N} , \mathbf{Z} . Se $X_t \in L^1(\Omega)$ para todo $t \in T$, dizemos que o processo estocástico X é um *martingale* em relação à filtração \mathcal{F}_t e à medida \mathbf{P} se $\mathbf{E}[X_t | \mathcal{F}_s] = X_s$ q.s. para todo $s \leq t$. Note, portanto, que a esperança de um martingale é constante no tempo, e igual a $\mathbf{E}[X_0]$.

Observação 2.11 (Intuição sobre martingales).

Uma ideia intuitiva que se pode ter de um martingale é que dada a informação do processo no presente, representado pela informação “acessível” por \mathcal{F}_s , se tivermos que intuir sobre o estado futuro X_t , $t \geq s$, na média, os valores do processo permanecem no mesmo estado. Essa é uma ideia que na prática é bastante simples e nos habituamos com algumas situações assim: (no tempo t) quando buscamos um utensílio, um inseto ou uma pessoa, a busca começa, de fato, ao redor do último lugar onde foram vistos (\mathcal{F}_s). Ainda outra situação que aparece em modelagem estocástica em finanças: dado o fechamento da cotação de uma moeda estrangeira em um dia, idealmente, na média, a melhor aproximação que se pode fazer para o dia seguinte é o mesmo valor do fechamento; salvo, naturalmente, informações adicionais ou privilegiadas, que não correspondem ao que se considera “ideal”. Tecnicamente, em finanças essa situação ideal é denominado de princípio de não-arbitragem, ou mercado livre de arbitragem (ver e.g. Karatzas [26], Karatzas and Shreve [28] ou Sondermann [56]), que em palavras simples significa que não é possível obter lucro sem risco.

Para o leitor interessado em continuar explorando essas ideias intuitivas, recomendamos o texto clássico de divulgação sobre as origens (ver [10]) e uma interpretação física do martingale, escrito por um dos criadores dessa teoria, Doob [11].

□

Se para todo $0 \leq s \leq t$, tivermos $\mathbf{E}[X_t | \mathcal{F}_s] \geq X_s$ então o processo é chamado de *submartingale*. E se $\mathbf{E}[X_t | \mathcal{F}_s] \leq X_s$, chamamos X de *supermartingale*. No primeiro caso a esperança é crescente, no segundo, é decrescente. Se X for um martingale em L^p , pela desigualdade de Jensen temos que $|X|^p$ é um submartingale.

Exemplo 2.5 (Martingale regular).

Um dos exemplos de martingales mais simples de se construir é fixar uma variável aleatória X_∞ “no infinito” e assim, dada uma filtração \mathcal{F}_t , definimos o martingale como sendo $X_t = \mathbf{E}[X_\infty | \mathcal{F}_t]$. Na maioria das vezes teremos a situação inversa dessa: dado um martingale X , a pergunta que se faz é se ele converge para alguma variável aleatória X_∞ tal que $X_t = \mathbf{E}[X_\infty | \mathcal{F}_t]$. Martingales com essa propriedade de ser gerado por uma única variável aleatória no infinito (no sentido de supremo no tempo T) é chamado de *martingale regular*.

□

Exemplo 2.6 (Submartingales e Supermartingales).

Dado um martingale X , pela desigualdade de Jensen (Proposição 1.12), dada qualquer função convexa g , $g(X)$ é um submartingale. Em particular vai nos interessar o submartingale $|X|^p$.

Se X é um martingale e f é qualquer processo (ou função) crescente então $X + f$ é um submartingale e $X - f$ é supermartingale.

□

Exemplo 2.7 (Soma de variáveis aleatórias i.i.d.).

Dada uma sequência de variáveis aleatórias integráveis i.i.d. com média zero X_1, X_2, \dots , considere a filtração dada por

$$\mathcal{F}_n = \sigma\{X_1, \dots, X_n\} \subseteq \mathcal{F}.$$

Defina o processo S_n dado pela soma $S_n = \sum_{i=1}^n X_i$. Pelo exercício 1.17, a esperança condicional satisfaz

$$\mathbf{E}[S_n | \mathcal{F}_m] = S_{m \wedge n},$$

onde $m \wedge n = \min\{m, n\}$, que é equivalente à definição de martingale. Se as variáveis aleatórias X_1, X_2, \dots são gaussianas $N(0, 1)$, temos que S_n é uma aproximação discreta do movimento browniano. .

□

Exemplo 2.8 (Movimento browniano).

Denote por \mathcal{F}_t a filtração natural de um movimento browniano B_t . Para todo $0 \leq s \leq t$ temos que

$$\mathbf{E}[B_t | \mathcal{F}_s] = \mathbf{E}[(B_t - B_s) + B_s | \mathcal{F}_s] = B_s$$

pela independência entre a variável aleatória $(B_t - B_s)$ e a σ -álgebra \mathcal{F}_s , cf. Exercício 1.17.

O movimento browniano é o martingale contínuo mais rico em propriedades, inclusive porque pode gerar (quase) todos os outros por mudanças na escala de tempo. Nossas intenções neste texto são bastante modestas, portanto apresentamos somente as propriedades básicas, mas o leitor interessado precisa saber que existem dezenas de livros que tratam especificamente de muitas outras propriedades analíticas e geométricas. Só para mencionar algumas poucas referências, citamos por exemplo Revuz e Yor [47], Goldberg e Mueller [18], Emery [14], Hsu [22], Elworthy [12], [13], Hida [21] e as referências contidas em cada um desses textos.

□

Exercício 2.12. *Se B_t é um movimento browniano, usando a definição, verifique que $(B_t^2 - t)$ e $\exp\{\alpha B_t - 1/2\alpha^2 t\}$, $\alpha \in \mathbb{R}$ também são martingales.*

No próximo capítulo veremos que, pelo fato das integrais estocásticas de Itô serem geradores naturais de martingales, o exercício acima ficará trivializado, graças à fórmula de Itô. Para o leitor iniciante, no entanto, é interessante que verifique essa propriedade pela definição.

Outra propriedade que será útil é a seguinte igualdade:

Exercício 2.13. *Se M é um martingale então*

$$\mathbf{E}[M_t^2 - M_s^2] = \mathbf{E}(M_t - M_s)^2$$

para todo $s \leq t \in T$.

Um processo estocástico com tempo discreto A é dito *previsível*, se for mensurável uma unidade de tempo antes de sua realização, isto é A_n é \mathcal{F}_{n-1} -mensurável.

Teorema 2.5 (Decomposição de Doob). *Dado um processo estocástico discreto X_n integrável, \mathcal{F}_n adaptado, então existe uma decomposição $X = X_0 + M + A$ onde M é um \mathcal{F}_n -martingale e A um processo previsível. Com M e A inicializados no zero a decomposição é única.*

Demonstração: Defina indutivamente, $A_0 = 0$ e

$$A_n = A_{n-1} + \mathbf{E}[X_n - X_{n-1} \mid \mathcal{F}_{n-1}].$$

Assim, $M_0 = 0$ e

$$M_n = M_{n-1} + X_n - \mathbf{E}[X_n \mid \mathcal{F}_{n-1}]$$

Naturalmente A e M formam uma decomposição como a procurada. Para a unicidade, suponha que existe M' e A' com as mesmas propriedades da decomposição. Então $M - M' = A' - A$ é um martingale inicializado no zero e previsível. Então $(M - M')_n = \mathbf{E}[(M - M')_m \mid \mathcal{F}_n] = \mathbf{E}[(M - M')_m \mid \mathcal{F}_{n-1}]$, portanto, por indução finita temos que $(M - M') = 0$. Portanto a decomposição é única. \square

2.5.1 Propriedades básicas

O resultado abaixo é uma generalização da desigualdade de Tchebychev no caso do processo ser um martingale.

Teorema 2.6 (Desigualdade L^p de Doob). *Seja X_t um martingale contínuo com $t \in [0, b]$ ou $[0, \infty)$. Então, para $p \geq 1$:*

$$\lambda^p \mathbf{P}[\sup_t X_t \geq \lambda] \leq \sup_t \mathbf{E}|X_t|^p.$$

E para $p > 1$:

$$\|\sup_t X\|_p \leq \frac{p}{p-1} \sup_t \|X_t\|_p.$$

Exercício 2.14. Considere o processo $S_t = \sup_{s \leq t} B_s$ que aponta os máximos das trajetórias do movimento browniano até o tempo t . Mostre que a probabilidade de S_t crescer mais que linearmente decresce exponencialmente com o tempo, precisamente:

$$\mathbf{P}\{S_t \geq at\} \leq e^{-\frac{a^2 t}{2}}$$

para todo $a \geq 0$. (Sugestão: verifique que $\{S_t \geq at\} \subseteq \{\sup_{s \leq t} M_s^\alpha \geq e^{\alpha a - \frac{a^2}{2}}\}$, onde M_t^α é o martingale $e^{\alpha B_t - \frac{\alpha^2 t}{2}}$. Aplique a desigualdade de Doob em M_s^α , com $\lambda = e^{\alpha at - \frac{a^2 t}{2}}$, depois faça $\alpha = a$).

Uma família de variáveis aleatórias (X_t) integráveis é *uniformemente integrável* se satisfaz qualquer uma das definições equivalentes:

- i) $\lim_{k \rightarrow \infty} \sup_t \int_{|X_t| > k} |X_t| d\mathbf{P} = 0$.
- ii) O conjunto $\{X_t : t \in T\}$ é relativamente fracamente compacto em L^1 .
- iii) Se $\{X_t : t \in T\} \in L^p$ para $p > 1$, então é limitado nesse espaço.

Teorema 2.7. Seja X_t um martingale. Então são equivalentes:

- i) Existe uma variável aleatória X_∞ tal que $X_t = \mathbf{E}[X_\infty | \mathcal{F}_t]$;
- ii) X_t converge para uma v.a. X_∞ em L^1 ;
- iii) A família X_t é uniformemente integrável.

Se X_t for limitado em L^p então, não só os três itens acima valem, como a convergência também se dá em L^p .

2.5.2 p -Variação de funções contínuas

Seja $f : \mathbf{R}_{\geq 0} \rightarrow \mathbb{R}$ uma função contínua. Dizemos que f tem variação limitada, ou 1-variação limitada, ou ainda variação finita no intervalo se para todo $t \geq 0$ temos que, dadas partições $\Delta = \{t_0 = 0 < t_1 < \dots, t_n = t\}$ do intervalo $[0, t]$:

$$S_t := \sup_{\Delta} \sum_i |f(t_{i+1}) - f(t_i)| = \lim_{|\Delta| \rightarrow 0} \sum_i |f(t_{i+1}) - f(t_i)| < \infty.$$

A função crescente S_t definida acima é chamada de variação da f no intervalo $[0, t]$. Se f representa, por exemplo, a posição de uma partícula, S_t é o comprimento total percorrido entre os instantes 0 e t . Se f for de classe C^1 então $S_t = \int_0^t |f'(s)| ds$. Se f é crescente com $f(0) = 0$ então $S_t = f(t)$.

Se uma função f tem variação limitada, ela pode ser escrita como a diferença entre duas funções crescentes $f = f^+ - f^-$. De fato, basta tomar $f^+(t) = 1/2(S_t + f(t))$ e $f^-(t) = 1/2(S_t - f(t))$. Assim, como funções crescentes são deriváveis em quase todo ponto da reta, uma função de variação limitada também é derivável q.s..

Existe uma bijeção entre medidas de Radon em $[0, \infty)$ e funções de variação limitada dada pela seguinte aplicação: $\mu_f([a, b]) = f(b) - f(a)$. Essa bijeção também pode ser interpretada como uma conversão da integral de Stieltjes para a integral de Lebesgue. Recordamos que a integral de Stieltjes de uma função g em relação a uma função f é definido como:

$$\int_0^t g(s) df(s) = \lim_{|\Delta| \rightarrow 0} \sum_i g(t_i) (f(t_{i+1}) - f(t_i))$$

que por sua vez é igual à integral de Lebesgue $\int_{[0, t]} g d\mu_f$. Note que a integral $t \mapsto \int_0^t g df$ também tem variação limitada, se f o tiver.

Exercício 2.15. *Verifique que a função $f : [0, 1] \rightarrow \mathbf{R}$ dada por $f(x) = x^2 \sin\left(\frac{1}{x^2}\right)$ é derivável em $[0, 1]$. É possível calcular sua variação integrando o módulo da derivada? Por quê?*

A próxima proposição vai mostrar que as trajetórias dos martingales não tem variação limitada em geral, o que implica que não são deriváveis.

Proposição 2.8. *Seja M um martingale contínuo. Se suas trajetória tiverem variação limitada então M é constante (i.e. uma v.a. constante no tempo).*

Demonstração: Sem perda de generalidade, considere $M_0 = 0$. De fato se não o for, use $\bar{M}_t = M_t - M_0$. Seja V_t o processo estocástico que descreve a variação das trajetórias de M em $[0, t]$, então, dada

uma partição $0 = t_0 \leq t_1 \leq \dots \leq t_n = t$, temos, usando o resultado do Exercício 2.13 que :

$$\begin{aligned} \mathbf{E} [M_t^2] &= \mathbf{E} \left[\sum_{i=1}^n (M_{t_i}^2 - M_{t_{i-1}}^2) \right] \\ &= \mathbf{E} \left[\sum_{i=1}^n (M_{t_i} + M_{t_{i-1}})^2 \right] - 2 \sum_{i=1}^n \mathbf{E} (M_{t_i} M_{t_{i-1}}) \\ &= \mathbf{E} \left[\sum_{i=1}^n (M_{t_i} - M_{t_{i-1}})^2 \right] \end{aligned}$$

Portanto

$$\mathbf{E} [M_t^2] \leq \mathbf{E} \left[V_t \left(\sup_i |M_{t_i} - M_{t_{i-1}}| \right) \right]$$

que vai para zero pela continuidade de M . □

Uma maneira alternativa interessante de obter o mesmo resultado acima e ao mesmo tempo ter uma idéia comparativa da integral de Stieltjes em cada trajetória (fixado ω) com a integral estocástica de Itô, que será apresentada no próximo capítulo é o seguinte:

Exercício 2.16. *Suponha que M_t seja um martingale contínuo, limitado e de variação finita. Verifique, pela desigualdade de Jensen que se M não é constante então M_t^2 é um submartingale estrito. Por outro lado, conclua, usando integração em cada trajetória ω , que*

$$M_t^2 = M_0^2 + 2 \int_0^t M_s dM_s$$

é um martingale.

Para a integração estocástica propriamente dita que definiremos logo mais, a fórmula do exercício acima precisa de um termo de correção dado pela fórmula de Itô, veja também o Teorema 2.12 abaixo.

Generalizando a noção de variação limitada de funções contínuas, dado $p \geq 1$ e uma função $\sigma : [0, \infty) \rightarrow \mathbb{R}$, a p -variação de σ até o

instante t é dada por:

$$V_t(\sigma) = \lim_{|\Delta| \rightarrow 0} \sum_i |\sigma_{t_{i+1}} - \sigma_{t_i}|^p,$$

quando o limite existe (como refinamento para toda partição).

O seguinte lema, puramente analítico, mostra que cada função contínua tem uma única p -variação característica.

Lema 2.9. *Dada uma função contínua $\sigma : [0, \infty) \rightarrow \mathbb{R}$, se σ tem p -variação limitada (finita) e não nula em $0 \leq t < \infty$ então:*

- i) *Para $q < p$, σ tem q -variação infinita;*
- ii) *Para $p < q$, σ tem q -variação nula para todo $0 \leq t < \infty$.*

Para ver uma demonstração, ver por exemplo, o apêndice de Follmer em [56] ou Catuogno [7]. Para quem conhece dimensão de Hausdorff, compare a p -variação de uma função com a dimensão de Hausdorff do seu gráfico (pode não ser fácil!).

A p -variação de um processo estocástico X_t é o processo $[X]_t^p$ definido da seguinte maneira: dado uma sequência de partições Δ_n do intervalo $[0, t]$, com $|\Delta_n| \rightarrow 0$,

$$[X]_t^p := \mathbf{P}\text{-}\lim_{|\Delta_n| \rightarrow 0} \sum_i |X_{t_{i+1}} - X_{t_i}|^p,$$

quando o limite existe. O caso $p = 2$ é particularmente interessante para a teoria de martingales e movimento browniano, neste caso, chamamos de *variação quadrática* e denotaremos simplesmente por $[X]_t$.

Proposição 2.10. *Se B_t é um movimento browniano então sua variação quadrática $[B]_t = t$ q.s..*

Demonstração: Dada um partição $\Delta = \{0 = t_0 < t_1 < \dots < t_n =$

$t\}$ então

$$\begin{aligned}
& \left\| \sum_{i=1} (B_{t_i} - B_{t_{i-1}})^2 - t \right\|_{L^2}^2 \\
&= \mathbf{E} \left(\sum_{i=1} (B_{t_i} - B_{t_{i-1}})^2 - t \right)^2 \\
&= \mathbf{E} \left(\sum_{i=1} (B_{t_i} - B_{t_{i-1}})^2 \right)^2 - 2t \sum (t_i - t_{i-1}) + t^2 \\
&= 3 \sum (t_i - t_{i-1})^2 + 2 \sum_{i < j} (t_i - t_{i-1})(t_j - t_{j-1}) - t^2 \\
&= 3 \sum (t_i - t_{i-1})^2 + 2 \sum_{i=1} (t_i - t_{i-1})(t - t_i) - t^2 \\
&= 3 \sum (t_i - t_{i-1})^2 - 2 \sum_{i=1} t_i(t_i - t_{i-1}) + t^2
\end{aligned}$$

que converge para zero já que o segundo somatório converge para $\int_0^t s \, ds$ quando $n \rightarrow \infty$.

□

Uma generalização interessante da decomposição de Doob 2.5, no caso contínuo é a seguinte decomposição:

Teorema 2.11 (Decomposição de Doob-Meyer). *Dado um submartingale contínuo X , então existe uma decomposição $X = M + A$, onde M é um martingale e A é um processo crescente de variação finita.*

Um dos exemplos mais importantes que ilustram a decomposição de Doob é a seguinte caracterização da variação quadrática de um martingale:

Teorema 2.12. *Seja M um \mathcal{F}_t -martingale contínuo e limitado. Então M tem variação quadrática, além disso $[M]_t$ é o único processo adaptado crescente, com $[M]_0 \equiv 0$, \mathbf{P} -q.s. tal que $M^2 - [M]$ é um martingale.*

Demonstração: Seja $\Delta = \{0 = t_0 < t_1 < \dots < t_n = T\}$ uma partição de um intervalo $[0, T]$ tal que $0 \leq s \leq t < T$. Considere

então o processo $T_t^\Delta(M)$ que aproxima a variação quadrática $[M]$ dependendo de quão refinada é a partição Δ no seguinte sentido: supondo que $t_k \leq t < t_{k+1}$, então

$$T_t^\Delta(M) = \sum_{i=1}^k (M_{t_i} - M_{t_{i-1}})^2 + (M_t - M_{t_k})^2$$

Então, como acima, se $t_k \leq t < t_{k+1}$ e $t_j \leq s < t_{j+1}$ então

$$\begin{aligned} & \mathbf{E} [T_t^\Delta(M) - T_s^\Delta(M) | \mathcal{F}_s] \\ &= \mathbf{E} \left[\sum_{i=j}^k (M_{t_{i+1}} - M_{t_i})^2 + (M_t - M_{t_k})^2 + (M_{t_{j+1}} - M_s)^2 | \mathcal{F}_s \right] \\ &= \mathbf{E} [M_t^2 - M_s^2 | \mathcal{F}_s] \end{aligned}$$

Portanto $M^2 - T^\Delta(M)$ é um \mathcal{F}_t -martingale. A demonstração é concluída observando que $T_t^\Delta(M)$ é inicializado no zero, é contínuo, adaptado e converge para um processo crescente já que $T^\Delta(M)$ é crescente a menos da componente $(M_t - M_{t_k})^2$, que vai para zero quando se refina a partição.

Para a unicidade, note que se A e B são processos adaptados crescentes tais que $(M^2 - A)$ e $(M^2 - B)$ sejam martingales então a diferença entre esses processos $(A - B)$ é um martingale de variação finita portanto, pela Proposição 2.8 deve ser constante, logo $A = B$. \square

Dados dois martingales M e N , a *variação quadrática cruzada* $[M, N]$ é definido da seguinte maneira: dada uma sequência de partições Δ_n do intervalo $[0, t]$, com $|\Delta_n| \rightarrow 0$,

$$[M, N]_t := \mathbf{P}\text{-}\lim_{n \rightarrow \infty} \sum_i (M_{t_{i+1}} - M_{t_i}) (N_{t_{i+1}} - N_{t_i}).$$

Note que a variação quadrática de um martingale $[M] = [M, M]$. A variação quadrática cruzada é caracterizada pelo seguinte resultado:

Teorema 2.13. *Sejam M e N dois martingales, então a variação quadrática cruzada $[M, N]_t$ é o único processo com $[M, N]_0 = 0$ tal que $MN - [M, N]$ é um martingale.*

Demonstração: De fato, escreva $MN = \frac{1}{4}\{(M+N)^2 - (M-N)^2\}$. Portanto, pelo Teorema 2.12, basta verificar que

$$[M, N] = \frac{1}{4}\{[M+N] - [M-N]\}$$

□

A variação quadrática fornece uma família de produtos internos no espaço vetorial dos martingales quocientado por constantes, da seguinte maneira: cada produto interno é indexado no intervalo $[0, t]$, assim, é um operador bilinear, simétrico e positivo definido no sentido que $[M, M]_t = 0$ se e somente se M é constante no intervalo $[0, t]$.

2.5.3 Martingale local

Dado um processo estocástico X , e um \mathcal{F}_t -tempo de parada T , chamamos de *processo parado* X^T o processo que fica constante a partir de T :

$$(X^T)_t := X_{t \wedge T}$$

A localização em martingales é um ferramenta poderosa onde o termo localização aqui pode ser entendido tanto no sentido de espaço como de tempo, já que com tempos de parada adequados podemos restringir processos adaptados a certos conjuntos mensuráveis que nos convêm (abertos ou fechados), como foi visto na Proposição 2.1.

Um processo adaptado e contínuo X_t é um martingale local se existe uma sequência de tempos de parada $(T_n)_{n \geq 1}$ tal que

- i) T_n é crescente e $\lim_{n \rightarrow \infty} T_n = \infty$ q.s.;
- ii) para todo n , X^{T_n} é um martingale uniformemente integrável.

Note que para todo martingale M , é possível construir uma sequência de tempos de parada $(T_n)_{n \geq 1}$ que satisfaz as condições (i) e (ii), portanto todo martingale é um martingale local. A recíproca vale se o martingale local for uniformemente integrável. Não entraremos em maiores detalhes aqui, mas para satisfazer a curiosidade do leitor fechamos essa comparação com a seguinte caracterização:

Proposição 2.14. *Um martingale local X é um martingale se e somente se $\{X_T; T \text{ é tempo de parada}\}$ é uma família de v.a. uniformemente integrável.*

Os conceitos e resultados apresentados sobre martingales se estendem naturalmente para martingales locais. Por exemplo, o teorema seguinte tem demonstração muito simples, basta usar os tempos de parada da definição de martingale local e o teorema análogo visto anteriormente:

Teorema 2.15. *Seja M um martingale local contínuo, então existe um único processo adaptado crescente $[M, M]$ tal que $[M, M]_0 = 0$ e $M^2 - [M, M]$ é um martingale local.*

Também usando o resultado análogo anterior, temos:

Teorema 2.16. *Sejam M e N dois martingales locais, então a variação quadrática cruzada $[M, N]_t$ é o único processo com $[M, N]_0 = 0$ tal que $MN - [M, N]$ é um martingale local.*

Processos estocásticos que são soma de martingales (locais) com processos de variação limitada são chamados de *semimartingales*. O teorema da decomposição de Doob 2.11 por exemplo diz que submartingales e supermartingales são semimartingales especiais onde os processos de variação limitada são monótonos.

Capítulo 3

Cálculo estocástico

Neste capítulo mostraremos as ferramentas básicas do cálculo estocástico, enfatizando fórmulas e propriedades que mais se usa em sistemas dinâmicos.

3.1 Integral Estocástica de Itô

Introduzimos uma versão discreta do que virá a ser a integral de Itô. O que tem de interessante é que as propriedades mais importantes já são possíveis de se visualizar mesmo neste caso discreto.

Proposição 3.1. *Seja $(X_n)_{n \geq 0}$ um \mathcal{F}_n -martingale discreto e H_n um processo limitado \mathcal{F}_{n-1} -adaptado. Considere o processo $Y_n = X_0 + \sum_{j=1}^n H_j(X_j - X_{j-1})$, $n = 0, 1, 2, \dots$. Então Y_n é um \mathcal{F}_n -martingale.*

Demonstração: Inicialmente note que Y_n é obviamente \mathcal{F}_n -adaptado, uma vez que produtos e somas de funções \mathcal{F}_n -mensuráveis são \mathcal{F}_n -mensuráveis.

Para $m \leq n$ temos, por indução finita que

$$\mathbf{E}[Y_n | \mathcal{F}_m] = \mathbf{E}[Y_n | \mathcal{F}_{n-1} | \mathcal{F}_{n-2} | \dots | \mathcal{F}_m].$$

Portanto só precisamos provar que

$$\mathbf{E}[Y_n | \mathcal{F}_{n-1}] = Y_{n-1}$$

para todo $n \in \mathbf{N}$. Mas de fato:

$$\begin{aligned} \mathbf{E}[Y_{n-1} + H_n(X_n - X_{n-1}) | \mathcal{F}_{n-1}] &= Y_{n-1} + \\ &H_n \mathbf{E}[(X_n - X_{n-1}) | \mathcal{F}_{n-1}] \\ &= Y_{n-1}. \end{aligned}$$

□

Denotaremos o martingale $Y_n := \int_0^n H dX$ ou ainda, abreviando, por $Y = H \cdot X$. O processo H é chamado o *integrando* e X é chamado o *integrador*. O que a Proposição 3.1 garante é que se o integrador for um martingale, então o processo dado pela integral também o é. Observe também pela demonstração que se X for um submartingale (ou supermartingale) e o processo H for não-negativo então $Y = H \cdot X$ também será submartingale (ou supermartingale), e vice-versa se H for não-positivo.

Exemplo 3.1 (Integrador gaussiano).

Considere um processo gaussiano discreto com incrementos gaussianos e independentes: $X_0 \equiv 0$ e $X_n = \sum_{i=0}^{n-1} X_i + I^n(\omega)$, onde $(I^n)_{n \geq 1}$ são variáveis aleatórias independentes, gaussianas centradas $N(0, 1)$; note que assim, X_n é uma versão discreta do movimento browniano. Se H for determinístico, $Y = H \cdot X$ é uma soma de gaussianas independentes, portanto pelo Teorema 1.17, Y será uma v.a. gaussiana. No caso geral Y é um martingale com média zero e variância crescente com n , dada por

$$\mathbf{E} Y_n^2 = \sum_{i=1}^n \mathbf{E} H_i^2$$

(verifique como exercício). Essa fórmula da variância é a versão discreta da *isometria de Itô* que aparecerá logo mais a frente.

□

3.1.1 Aplicação a processos parados

Vamos ver as primeiras aplicações a processos parados:

Corolário 3.2. *Dado um martingale $(X_n)_{n \in \mathbf{N}}$ e um tempo de parada T então X^T também é um martingale.*

Demonstração: De fato, na proposição acima, basta tomar $H_n = 1_{\{n \leq T\}}$, isto é $H_n = 1 - 1_{\{T \leq n-1\}}$ que é \mathcal{F}_{n-1} -mensurável, e verificar que $X^T = \int H dX$. □

Dado um \mathcal{F}_n -tempo de parada T , a σ -álgebra \mathcal{F}_T associada a T , de eventos mensuráveis até o tempo T é definida por

$$\mathcal{F}_T = \{A \in \mathcal{F}_\infty; A \cap \{T \leq n\} \in \mathcal{F}_n, \text{ para todo } n \in \mathbf{N}\}.$$

Exercício 3.1. Verifique que \mathcal{F}_T definida acima é uma σ -álgebra. Depois conclua que se $S \leq T$ q.s. então $\mathcal{F}_S \subseteq \mathcal{F}_T$.

Teorema 3.3 (Tempo de parada opcional, versão discreta). *Seja (X_n) um processo estocástico integrável e \mathcal{F}_n -adaptado, então são equivalentes:*

- i) (X_n) é martingale;
- ii) $\mathbf{E}[X_T] = \mathbf{E}[X_S]$ para todo par de de tempos de parada limitados $0 \leq S \leq T$;
- iii) $X_S = \mathbf{E}[X_T | \mathcal{F}_S]$ q.s. para todo par de de tempos de parada limitados $0 \leq S \leq T$.

Demonstração: Assuma (1) e suponha que os tempos de parada limitados sejam tais que $S, T < K \in \mathbf{R}$. Considere o processo $H_n = 1_{n \leq T} - 1_{n \leq S}$. Então pela Proposição 3.1 $H \cdot X$ é martingale. Mas $(H \cdot X)_K - X_0 = X_T - X_S$. Como $\mathbf{E}H \cdot X = \mathbf{E}X_0$, segue que

$$\mathbf{E}X_S = \mathbf{E}X_T. \quad (3.1)$$

Repare agora que $S^B = S1_B + K1_{B^c}$ também é tempo de parada se $B \in \mathcal{F}_S$. De fato, basta ver que $\{S^B \leq t\} = B \cap \{S \leq t\}$, portanto S^B é tempo de parada.

Agora aplicando os tempos de parada $S^B = S1_B + K1_{B^c}$ e $T^B = T1_B + K1_{B^c}$ na equação 3.1, obtemos

$$\int_B X_T d\mathbf{P} = \int_B X_S d\mathbf{P},$$

para todo $B \in \mathcal{F}_S$, portanto $X_S = \mathbf{E}[X_T | \mathcal{F}_S]$.

De (iii) para (i) é óbvio da definição tomando tempos de parada determinísticos. □

O corolário abaixo é trivial quando se pensa na variação quadrática trajetória por trajetória, mas é menos óbvio quando aplicado ao processo $[M]$ que, por definição, é um limite em probabilidade.

Corolário 3.4. *Dado um martingale M e um tempo de parada T temos que $[M^T] = [M]^T$.*

Demonstração: De fato, $M^2 - [M]$ é martingale, portanto, pelo teorema do tempo de parada opcional

$$(M^2 - [M])^T = (M^T)^2 - [M]^T$$

é martingale. Logo, pela unicidade do Teorema 2.12, $[M^T] = [M]^T$. □

A condição de limitação do tempo de parada não pode ser removida. De fato, considere o seguinte:

Exemplo 3.2 (Tempo de parada não integrável).

Dado um movimento browniano canônico B_t , considere os tempo de parada

$$T = \inf\{t \geq 0, B_t \geq 1\},$$

e $S \equiv 0$. Então $\mathbf{E} B_T = 1$ mas $\mathbf{E} B_S = 0$. O que falha neste exemplo é que, apesar de $T \leq \infty$ q.s. como será verificado futuramente (Exemplo 3.8), T não é uniformemente limitado.

Sendo mais específico, o que ocorre aqui é que $\mathbf{E} T = \infty$. Chamamos a atenção para o fato dessa esperança ser equivalente a, no caso discreto termos variância no infinito $\sum_{i=1}^{\infty} \mathbf{E} H_n^2$ não convergente. Mais adiante vamos comparar esse exemplo com a classe de processos estocásticos integráveis. □

3.1.2 Processos contínuos

Dada uma filtração \mathcal{F}_s , com $s \in [0, t]$, os processos estocásticos *elementares* neste intervalo são processos \mathcal{F}_s -adaptados que são constantes por partes. Denotamos o espaço vetorial de processos elementares

neste intervalo por $\varepsilon([0, t])$. Sendo mais preciso: para cada processo $H_s \in \varepsilon([0, t])$, existe uma partição $\{0 = t_0 < t_1 < \dots < t_n = t\}$ tal que

$$H = H_0 1_{[0, t_1)} + H_1 1_{[t_1, t_2)} + \dots + H_{n-1} 1_{[t_{n-1}, t_n)}$$

onde cada H_j , $j = 0, 1, \dots, n-1$, são variáveis aleatórias limitadas \mathcal{F}_{t_j} -mensuráveis.

Considerando a filtração natural do movimento browniano \mathcal{F}_t , a *integral estocástica de Itô* de um processo elementar H em relação a um movimento browniano W_t é dado por

$$\int_0^t H_s dW_s = \sum_{i=1}^n H_{t_{(i-1)}} (W_{t_i} - W_{t_{(i-1)}}).$$

Variando $t \geq 0$, denotamos por Y_t o processo dado pela integral estocástica

$$Y_t = \int_0^t H_s dW_s.$$

Exercício 3.2. *A integral estocástica Y_t definido acima não depende da partição que torna o processo H_s constante por partes. Além disso, generalize para o caso contínuo a Proposição 3.1 que trata do caso discreto, i.e. verifique que Y_s é um \mathcal{F}_s -martingale com $s \in [0, t]$.*

Se o processo H_s for determinístico então Y_t é um processo gaussiano, como mostra o Exemplo 3.1 no caso discreto. No caso contínuo também será gaussiano, já que convergência de gaussianas é gaussiana, conforme resultados da Seção 1.8. Neste caso, como já foi dito, a integral $\int H dW$ é chamada de *integral estocástica de Wiener*.

Usaremos a seguinte proposição para obtermos a integral de Itô sobre processos adaptados mais gerais:

Proposição 3.5 (Isometria de Itô). *Dada uma integral de Itô $Y_t = \int_0^t H_s dW_s$, as variáveis aleatórias $Y_t \in L^2(\Omega)$ e $H_s \in L^2(\Omega \times [0, t])$ tem as mesmas normas nos respectivos espaços, isto é*

$$\mathbf{E} \left(\int_0^t h_t(\omega) dB_t \right)^2 = \mathbf{E} \left(\int_0^t h_s^2(\omega) dt \right)$$

Demonstração: De fato

$$\begin{aligned}
\mathbf{E}|Y_t|^2 &= \mathbf{E} \left(\sum_{i=1}^n H_{t_{i-1}} (W_{t_i} - W_{t_{(i-1)}}) \right)^2 \\
&= \sum_{i,j=1}^n \mathbf{E} [H_{t_{(i-1)}} (W_{t_i} - W_{t_{(i-1)}}) H_{t_{(j-1)}} (W_{t_j} - W_{t_{(j-1)}})] \\
&= \sum_{i=1}^n \mathbf{E} [H_{t_{(i-1)}}^2 (W_{t_i} - W_{t_{(i-1)}})^2] \\
&\quad + 2 \sum_{1=i < j}^n \mathbf{E} [H_{t_{(i-1)}} H_{t_{(j-1)}} (W_{t_i} - W_{t_{(i-1)}}) (W_{t_j} - W_{t_{(j-1)}})].
\end{aligned}$$

No primeiro somatório, exploramos a independência de $H_{t_{(i-1)}}$ com $(W_{t_i} - W_{t_{(i-1)}})$ para concluir que

$$\begin{aligned}
\mathbf{E} [H_{t_{(i-1)}}^2 (W_{t_i} - W_{t_{(i-1)}})^2] &= \mathbf{E} |H_{t_{(i-1)}}|^2 \mathbf{E} (W_{t_i} - W_{t_{(i-1)}})^2 \\
&= \mathbf{E} |H_{t_{(i-1)}}|^2 (t_i - t_{i-1}).
\end{aligned}$$

Portanto o primeiro somatório é igual a

$$\int_0^t \mathbf{E} |H_s|^2 ds = \mathbf{E} \int_0^t |H_s|^2 dt$$

pelos Teoremas de Tonelli 1.13 ou Fubini 1.14.

Agora basta mostrar que o segundo somatório se anula. De fato, usando a propriedade de invariância da média da esperança condicional, Teorema 1.11, temos que

$$\begin{aligned}
&\mathbf{E} [H_{t_{(i-1)}} H_{t_{(j-1)}} (W_{t_i} - W_{t_{(i-1)}}) (W_{t_j} - W_{t_{(j-1)}})] \\
&= \mathbf{E} \{ \mathbf{E} [H_{t_{(i-1)}} H_{t_{(j-1)}} (W_{t_i} - W_{t_{(i-1)}}) (W_{t_j} - W_{t_{(j-1)}}) | F_i] \} \\
&= \mathbf{E} \{ H_{t_{(i-1)}} H_{t_{(j-1)}} (W_{t_i} - W_{t_{(i-1)}}) \mathbf{E} [(W_{t_j} - W_{t_{(j-1)}}) | F_i] \},
\end{aligned}$$

porque $H_{t_{(i-1)}} H_{t_{(j-1)}} (W_{t_i} - W_{t_{(i-1)}})$ são \mathcal{F}_{t_i} -mensuráveis. Como $j > i$, segue que a esperança condicional da última linha é zero, portanto a esperança é zero, o que anula todos os termos do segundo somatório. \square

Estendemos naturalmente a definição da integral de Itô usando essa isometria.

Definição 3.1. Dizemos que um processo estocástico adaptado H_t é Itô integrável no intervalo $[0, t]$ se

$$\mathbf{E} \int_0^t H_s^2 ds < \infty. \quad (3.2)$$

Assim, dada uma sequência de processos elementares H^n convergindo para H na norma de $L^2(\Omega \times [0, t])$ então a integral de Itô de H é o limite

$$\int_0^t H_s dW_s = L^2\text{-}\lim_{n \rightarrow \infty} \int_0^t H_s^n dW_s.$$

Usando a isometria da Proposição 3.5 e o fato dos processos elementares serem densos em $L^2(\Omega \times [0, t])$, temos que a integral de Itô fica bem definida desta maneira.

Em termos práticos, inclusive para simulações numéricas, uma sequência canônica H^n que converge para H pode ser tomada simplesmente discretizando o processo H em partições do intervalo $[0, t]$ cujos subintervalos tem comprimento máximo tendendo para zero, isto é, dada uma partição $\Delta^n = \{0 = t_0 < t_1 < \dots < t_n = t\}$, sua malha (ou calibre) $|\Delta^n| = \max\{t_i - t_{i-1} : i = 1, 2, \dots, n\}$ vai para zero quando n vai para infinito. Os processos elementares associados a uma partição Δ^n são

$$H_s^n = H_0 1_{[0, t_1)} + H_1 1_{[t_1, t_2)} + \dots + H_{n-1} 1_{[t_{n-1}, t_n]},$$

que convergem para H .

Exercício 3.3. Mostre que os processos elementares são densos em $L^2(\Omega \times [0, t])$.

Como convergência em L^2 implica em convergência em probabilidade, uma vez que um processo H_t é Itô integrável, então essa integral pode ser escrita como um limite em probabilidade sobre as somas de Riemann sobre cada partição Δ do intervalo. Assim a integral de Itô

é dada também pelo limite em probabilidade das somas de riemann:

$$\int_0^t H_s dW_s = \mathbf{P}\text{-}\lim_{|\Delta| \rightarrow 0} \sum_{i=1}^n H_{t_{i-1}} (W_{t_i} - W_{t_{i-1}}). \quad (3.3)$$

Normalmente não é fácil calcular integrais estocásticas pela definição. Os melhores exemplos aparecem quando usamos a fórmula de Itô, que será apresentada na próxima seção. Resumimos as propriedades que vimos até agora da integral de Itô no seguinte:

Teorema 3.6 (Propriedades da Integral de Itô). *Dados G e H processos estocásticos integráveis então:*

1. (Linearidade) Para todo $a, b \in \mathbf{R}$,

$$\int (aG + bH) dW = a \int G dW + b \int H dW.$$

2. (Gerador de martingales) O processo

$$Y_t = \int_0^t H_s dW_s$$

é um martingale de média zero.

3. (Isometria)

$$\mathbf{E} \left(\int_0^t G_s dW_s \int_0^t H_s dW_s \right) = \int_0^t G_s H_s dt$$

4. (Variação quadrática)

$$\left\langle \int_0^t G_s dW_s, \int_0^t H_s dW_s \right\rangle = \int_0^t G_s H_s dt$$

Demonstração: O item (1) é consequência imediatas da construção. O item (2) é consequência da Proposição 3.1 e da continuidade da esperança condicional que garante que limite de uma sequência de martingales (integração de processos elementares) é um martingale. O item (3) segue da isometria da Proposição 3.5.

Mostraremos o item (4) para integrandos que sejam processos elementares. Neste caso, refinando as partições se necessário, tome uma partição $\{0 = t_0 < t_1 < \dots < t_n = t\}$ em relação a qual ambos G e H são v.a. constantes nos subintervalos.

Assim, em relação a essa partição, a variação quadrática pode ser calculada:

$$\begin{aligned} & \left\langle \int_0^t G_s dW_s, \int_0^t H_s dW_s \right\rangle = \\ & = \left\langle \sum_{i=1}^n G_{t_{(i-1)}} (W_{t_i} - W_{t_{(i-1)}}), \sum_{i=1}^n H_{t_{(i-1)}} (W_{t_i} - W_{t_{(i-1)}}) \right\rangle \\ & = \sum_{i=1}^n G_{t_{(i-1)}} H_{t_{(i-1)}} (t_i - t_{(i-1)}) \\ & = \int_0^t G_s H_s dt \end{aligned}$$

□

A teoria de integração feita nesta seção, considerando o movimento browniano como o integrador padrão, se estende facilmente para a integração em relação a qualquer semimartingale. No caso geral, naturalmente, as fórmulas de variação quadráticas são diferentes destas. O item (4) por exemplo, considerando integradores dados por semimartingales M_t^1 e M_t^2 , a fórmula se generaliza para

$$\left\langle \int_0^t G_s dM_s^1, \int_0^t H_s dM_s^2 \right\rangle = \int_0^t G_s H_s d \langle M^1, M^2 \rangle_s$$

No item (2), se substituirmos W_t por um martingale local, então Y_t também será um martingale local.

De qualquer maneira, o caso que fizemos não é tão restritivo se observarmos que a maioria dos martingales (locais) com que lidamos são frequentemente outras integrações em relação ao movimento browniano. Portanto, aplicando as mesmas propriedades do teorema acima, facilmente se calcula essas outras integrações e outras variações quadráticas mais gerais. Na notação da fórmula acima, por exemplo, normalmente $\langle M^1, M^2 \rangle$ será calculada usando o item (3) do teorema.

3.2 Fórmula de Itô

As integrações em relação a um processo de variação limitada, por exemplo em relação à variação cruzada $[X, Y]$, também tem variação limitada porque o processo original $[X, Y]$ pode ser escrito como a diferença de dois processos monótonos crescentes, portanto são deriváveis quase sempre. Daí que integrais assim são processos, mais que de variação limitada, são deriváveis $(\omega, t) \mathbf{P} \times \lambda$ -q.s..

O fato dos martingales terem trajetórias com variação quadrática (2-variação) vai implicar que o cálculo estocástico de Itô seja um cálculo de segunda ordem. A demonstração da fórmula de Itô abaixo vai deixar claro esse fato, partindo-se de uma série de Taylor:

Teorema 3.7 (Fórmula de Itô). *Seja $f : \mathbf{R}^n \subset \mathbf{R} \rightarrow \mathbf{R}$ uma função real de classe C^2 . Dado um semimartingale $X = (X^1, \dots, X^n)$ no domínio, temos que a composição $f(X)$ também é um semimartingale e:*

$$\begin{aligned} f(X_t) &= f(X_0) + \sum_{i=1}^n \int_0^t \frac{\partial f}{\partial x_i}(X_s) dX_s^i \\ &\quad + \frac{1}{2} \sum_{i,j=1}^n \int_0^t \frac{\partial^2 f}{\partial x_i \partial x_j}(X_s) d[X^i, X^j]_s. \end{aligned} \quad (3.4)$$

Observação 3.4 (Tempo de parada e forma infinitesimal).

1. A fórmula de Itô vale também se ao invés de integrarmos até t , fizermos a integração até um tempo de parada T . Note que neste caso, uma integração até T significa:

$$\int_0^T h_s dN_s := \int_0^M 1_{\{s < T(\omega)\}} h_s dN_s.$$

2. Outra maneira de enunciar a fórmula de Itô é na sua forma infinitesimal, na qual a maioria das EDE se apresentam:

$$\begin{aligned} df(X_t) &= \sum_{i=1}^n \frac{\partial f}{\partial x_i}(X_s) dX_s^i \\ &\quad + \frac{1}{2} \sum_{i,j=1}^n \frac{\partial^2 f}{\partial x_i \partial x_j}(X_s) d[X^i, X^j]_s. \end{aligned} \quad (3.5)$$

Significa simplesmente que a igualdade vale se tomarmos a integral estocástica nos dois lados da equação. Ainda uma terceira alternativa de notação, mais geométrica e mais curta é

$$df(X_t) = \langle \nabla f, dX_t \rangle + \frac{1}{2} Hf (dX, dX).$$

onde $\langle \cdot, \cdot \rangle$ é o produto interno em \mathbf{R}^n , ∇f é o gradiente da f , $Hf(\cdot, \cdot)$ é o hessiano da f , e fica subentendido que o operador bilinear é integrado em relação à variação quadrática cruzada. Em termos matriciais, usando o superíndice t para denotar a transposta, trata-se de $(dX)^t H(f) dX$, com a convenção do produto $dX^i dX^j = [X^i, X^j]$.

□

Antes da demonstração, vejamos ainda algumas aplicações fáceis mas ilustrativas:

Exemplo 3.3 (Potências do movimento browniano).

Tomando $n = 1$ no teorema acima, considere $f(x) = x^2$. Pela fórmula de Itô temos que

$$B_t^2 = 2 \int_0^t B_s dB_s + \int_0^t 1 ds,$$

e como a integral de Itô é martingale, temos uma descrição desta integral:

$$\int_0^t B_s dB_s = \frac{B_t^2 - t}{2}.$$

É interessante notar como a fórmula de Itô nos fornece facilmente uma enorme classe de exemplos de integrais de Itô e de martingales. Relembramos que a propriedade de martingale do lado direito da fórmula acima já tinha sido verificada pela definição no Exercício 2.12. Deixamos para o leitor verificar por indução que, para todo $n \geq 1$, os seguintes processos são martingales:

$$Y_t = 2B_t^{n+1} - \int_0^t B_s^{n-1} ds.$$

□

Exemplo 3.4 (Ainda voltando ao Exemplo 2.12).

Com $n = 2$, tome $f(x, y) = \exp\{\alpha x - \frac{\alpha^2}{2}y\}$. Segue pela fórmula de Itô, e pelo fato da variação quadrática $[t, \cdot]$ dar sempre zero que

$$f(B_t, t) = 1 + \alpha \int_0^t \exp\{\alpha B_s - \frac{\alpha^2}{2}s\} dB_s.$$

De onde segue, mais uma vez a propriedade de martingale do segundo processo do Exemplo 2.12.

Note ainda que o processo $Y_t = \alpha \int_0^t \exp\{\alpha B_s - \frac{\alpha^2}{2}s\} dB_s$ é portanto solução da seguinte equação estocástica:

$$dY_t = \alpha Y_t dB_t, \quad (3.6)$$

com condição inicial $Y_0 = 1$. O processo Y_t é chamado de *exponencial estocástica* do movimento browniano, ou ainda *movimento browniano geométrico*. Generalizando, dado um martingale M , a fórmula da exponencial de M é

$$\varepsilon^\alpha(M)_t = \exp\{\alpha M_t - \frac{\alpha^2}{2}[M]_t\} \quad (3.7)$$

que é a solução da EDE:

$$dx_t = \alpha x_t dM_t$$

□

Demonstração: (Da fórmula de Itô). Considere uma partição $\Delta = \{0 = t_0 < t_1, \dots < t_n = t\}$ do intervalo $[0, t]$. Pela fórmula de Taylor para funções de várias variáveis com resto de Lagrange, temos que para cada $\omega \in \Omega$,

$$\begin{aligned} f(X_{t_{k+1}}) &= f(X_{t_k}) + \sum_{i=1}^d \frac{\partial f}{\partial x_i}(X_{t_k})(X_{t_{k+1}}^i - X_{t_k}^i) \\ &\quad + \frac{1}{2} \sum_{i,j=1}^d \frac{\partial^2 f}{\partial x_i \partial x_j}(\xi_k)(X_{t_{k+1}}^i - X_{t_k}^i)(X_{t_{k+1}}^j - X_{t_k}^j) \end{aligned}$$

onde $\xi_k = X_{t_k} + \theta(X_{t_{k+1}} - X_{t_k})$ para algum $\theta \in [0, 1]$.

Assim, por cancelamento telescópico temos que

$$\begin{aligned} f(X_t) &= f(X_0) + \sum_{i=1}^d \sum_{k=0}^{n-1} \frac{\partial f}{\partial x_i}(X_{t_k})(X_{t_{k+1}}^i - X_{t_k}^i) \\ &\quad + \frac{1}{2} \sum_{i,j=1}^d \sum_{k=0}^{n-1} \frac{\partial^2 f}{\partial x_i \partial x_j}(\xi_k)(X_{t_{k+1}}^i - X_{t_k}^i)(X_{t_{k+1}}^j - X_{t_k}^j) \end{aligned}$$

Pela construção da integral de Itô, o primeiro somatório em k converge em probabilidade para

$$\int_0^t \frac{\partial f}{\partial x_i}(X_s) dX_s^i.$$

O segundo somatório em k converge para a integral de Riemann-Stieltjes

$$\int_0^t \frac{\partial^2 f}{\partial x_i \partial x_j}(X_s) d[X^i, X^j]_s$$

□

Corolário 3.8 (Integração por partes). *Dados semimartingales X e Y então*

$$(XY)_t = X_0 Y_0 + \int_0^t X_s dY_s + \int_0^t Y_s dX_s + [X, Y]_t,$$

em particular

$$X_t^2 = X_0^2 + 2 \int_0^t X_s dX_s + [X, Y]_t.$$

Demonstração: Exercício.

Observação 3.5.

1. Na demonstração da fórmula de Itô, usamos funções de classe C^2 justamente porque martingales contínuos tem trajetórias de p -variação característica com $p = 2$. Note que um cálculo envolvendo trajetórias contínuas de p -variação característica com $p > 2$ exigirá mais das derivadas da f . Esse é o caso da teoria chamada de “rough path” de T. Lyons. Ver por exemplo T. Lyons [38].

2. A fórmula de Itô é efetivamente a regra da cadeia, generalizada para composições com funções de 2-varição, independente do caráter probabilístico, i.e. independente de tomarmos ou não o limite em probabilidade.
3. Quanto ao aspecto probabilístico, a formula de Itô mostra que o espaço de semimartingales é invariante por composição com funções de classe C^2 .

□

3.3 Caracterização de Levy

O próximo teorema mostra uma caracterização do movimento browniano em termos da propriedade de martingale e sua variação quadrática.

Teorema 3.9 (Caracterização de Levy). *Seja X um processo estocástico em \mathbf{R}^d , contínuo, \mathcal{F}_t -adaptado com $X_0 = 0$. Então são equivalentes:*

- i) $X = (X^1, \dots, X^d)$ é MB em \mathbf{R}^d ;
- ii) X é martingale local contínuo com $[X^i, X^j] = \delta_{i,j} t$;
- iii) X é martingale local contínuo e para toda curva $f = (f_1, \dots, f_d)$, com $f_i \in L^2(\mathbf{R}_+)$, o processo

$$\varepsilon_t^{if} = \exp\left\{ i \sum_{k=1}^d \int_0^t f_k(s) dX_s^k + \frac{1}{2} \sum_{k=1}^d \int_0^t f_k^2(s) ds \right\}$$

é um martingale complexo.

Demonstração: Assuma (i), isto é que $X = (B^1, \dots, B^n)$ é um MB em \mathbf{R}^n . Temos que para todo t , X_t é uma gaussiana em \mathbf{R}^n , portanto os B^j 's são independentes, portanto são martingales e portanto martingales locais. Porém, pelo Teorema 2.13, sabemos que $B^i B^j - [B^i, B^j]$ é martingale local. Mas

$$\begin{aligned} \mathbf{E} [B_t^i B_t^j \mid \mathcal{F}_s] &= \mathbf{E} [B_t^i \mid \mathcal{F}_s] \mathbf{E} [B_t^j \mid \mathcal{F}_s] \\ &= B_t^i B_t^j \end{aligned}$$

porque são independentes se $1 \leq i \neq j \leq d$. Portanto B_t^i, B_t^j é martingale, o que implica que $[B_t^i, B_t^j]$ é martingale, de variação limitada e inicializado no zero, portanto só pode ser zero q.s.. O caso $1 \leq i = j \leq d$ é óbvio. Fica então estabelecido o item (ii).

Assumindo (ii), considere

$$M_t = \sum_{k=1}^d \int_0^t f_k(s) dX_s^k.$$

Então, por hipótese e pelas propriedades da integral de Itô (ver comentários depois do Teorema 3.6) sua variação quadrática será dada por

$$\begin{aligned} [M]_t &= \left[\sum_{k=1}^d \int_0^t f_k(s) dX_s^k \right]_t \\ &= \sum_{k,l=1}^d \int_0^t f_k f_l d[X^k, X^l]_s \\ &= \sum_k \int_0^t f_k^2 d[X^k]_s \\ &= \sum_k \int_0^t f_k^2 ds. \end{aligned}$$

Portanto, pela fórmula da exponencial estocástica (3.7), se tomarmos $\alpha = i$ então ε_t^{if} é um martingale local. A hipótese de que as f_k 's são quadrado integráveis implica que M_t é limitado em L^2 , portanto ε_t^{if} também o será. Assim, concluímos que de fato ε_t^{if} é um martingale, que é o item (iii).

Finalmente, assuma (iii), e tome um vetor $\xi \in \mathbf{R}^d$. Fixe dois números reais $0 < s < t$ e considere a função $f = \xi 1_{[s,t]}$. Então ε^{if} é um martingale, com $\varepsilon_s^{if} = 1$ e

$$\varepsilon_t^{if} = \exp\{i \langle \xi, X_t - X_s \rangle + \frac{1}{2} \|\xi\|^2 (t - s)\}$$

Seja $A \in \mathcal{F}_s$, então, usando esperança condicional temos que

$$\mathbf{E}1_A \varepsilon_t^{if} = \mathbf{E}1_A \varepsilon_s^{if}$$

que é igual a $\mathbf{P}(A)$. Portanto

$$\mathbf{E}1_A e^{i\langle \xi, X_t - X_s \rangle} = \mathbf{P}(A) e^{-\frac{1}{2} \|\xi\|^2 (t-s)},$$

o que implica que a transformada de Fourier

$$\mathbf{E} [e^{i\langle \xi, X_t - X_s \rangle}] = e^{-\frac{1}{2} \|\xi\|^2 (t-s)}.$$

Portanto, como a transformada de Fourier caracteriza a medida, $(X_t - X_s)$ é gaussiana com matriz de covariância $C = (t - s)1d$. Somado ao fato de $(X_t - X_s)$ ser independente de \mathcal{F}_s , concluímos que $X_t = (X^1, \dots, X^d)$ é um MB.

□

Corolário 3.10. *Movimento browniano é o único martingale local contínuo com variação quadrática t .*

A próxima proposição é outra aplicação interessante da fórmula de Itô. Note que ela faz uma caracterização probabilística de uma propriedade determinística.

Proposição 3.11. *Seja $f : \mathbf{R}^n \rightarrow \mathbf{R}$ uma função de classe C^2 em \mathbf{R}^n . A função f é harmônica se e somente se para todo movimento browniano B_t , sua imagem $f(B_t)$ é um martingale em \mathbf{R} .*

Demonstração: Dada f de classe C^2 e um movimento browniano $B_t = (B^1, \dots, B^n)$ em \mathbf{R}^n , temos, pela fórmula de Itô (3.7) e pelo Teorema de caracterização de Levy 3.9, que

$$f(B_t) = f(B_0) + \sum_{i=1}^n \int_0^t \frac{\partial f}{\partial x_i}(B_s) dB_s^i + \frac{1}{2} \sum_{i=1}^n \int_0^t \frac{\partial^2 f}{\partial x_i^2}(B_s) ds.$$

já que os outros termos se anulam.

Assim, se f for harmônica, o último somatório se anula, portanto só restam os termos que são martingales.

Reciprocamente, suponha que $f(B_t)$ seja um martingale. Suponha por absurdo que existe um ponto $x_0 \in \mathbf{R}^n$, tal que o laplaciano

$\Delta f(x_0) > 0$. Seja $U \subset \mathbf{R}^n$ um aberto conexo contendo x_0 tal que neste conjunto $\Delta f > 0$. Seja T o tempo de parada dado pelo tempo de saída de U do movimento browniano inicializado em x_0 . Então

$$[f(B_t)]^T = f(B_0) + \sum_{i=1}^n \int_0^T \frac{\partial f}{\partial x_i}(B_s) dB_s^i + \frac{1}{2} \int_0^T \Delta f(B_s) ds.$$

Isso implica que $[f(B_t)]^T$ não é um martingale visto que a última integração se faz com um integrando estritamente positivo, o que é uma contradição. Segue que Δf é identicamente nulo. \square

Esse tipo de caracterização via imagem de movimento browniano se estende para aplicações harmônicas entre variedades riemannianas, ver por exemplo Catuogno e Ruffino [6].

3.4 Equações diferenciais estocásticas

O resultado principal desta seção é um teorema de existência e unicidade de solução de equações diferenciais estocásticas (EDE) em \mathbf{R}^n com condição inicial:

$$\begin{cases} dx_t = f_0(t, x_t) dt + \sum_{i=1}^m f_i(t, x_t) dW_t^i, \\ x_0 = x(0) \in \mathbf{R}^n \end{cases} \quad (3.8)$$

onde W^1, \dots, W^m são movimentos brownianos independentes e os campos f_0, f_1, \dots, f_m são contínuos e de Lipschitz na variável x , i.e. existe um $K > 0$ tal que

$$|f_i(t, x) - f_i(t, y)| < K|x - y|.$$

para todo $i = 0, 1, \dots, m$, $t \geq 0$ e $x, y \in \mathbf{R}^n$. O campo de vetores f_0 associado a integração determinística é frequentemente chamado de “drift” da equação; enquanto que os outros campos f_1, \dots, f_m são chamados de *coeficientes* ou *campos de difusão*.

A equação (3.8) tem um significado de uma equação integral no sentido de que um processo x_t será solução dessa EDE se quando

integrarmos nos dois lados verificarmos a seguinte igualdade:

$$x_t - x_0 = \int_0^t f_0(s, x_s) ds + \sum_{i=1}^m \int_0^t f_i(s, x_s) dW_s^i.$$

O leitor familiarizado com a demonstração determinística de existência e unicidade de soluções de EDO vai notar que ambas as demonstrações, i.e. tanto para EDO quanto para EDE, estão baseadas na mesma ideia de construir um operador que seja contração em um espaço adequado. Relembramos aqui uma das versões do

Teorema 3.12 (Teorema do ponto fixo de Banach). *Seja (M, d) um espaço métrico completo. Seja $S : M \rightarrow M$ uma aplicação tal que existe $\alpha \in (0, 1)$ e um inteiro $n \geq 1$ tal que $d(S^n(x), S^n(y)) \leq \alpha d(x, y)$. Então existe um único $x \in M$ tal que $S(x) = x$.*

O que precisamos introduzir de diferente do caso de EDO é justamente o espaço de candidatos à solução onde acontecerá a contração. Para um t fixado, considere a seguinte norma no espaço de semimartingales contínuos e adaptados:

$$|U|_E = \left\{ \mathbf{E} \sup_{s \leq t} |U_s - V_s|^2 \right\}^{1/2}$$

Por ser composição de duas normas, $|\cdot|_E$ também é uma norma. Usaremos que

Lema 3.13. *O espaço E de semimartingales contínuos adaptados com a norma $|\cdot|_E$ é um espaço de Banach.*

Agora vamos à demonstração do

Teorema 3.14 (Existência e unicidade de solução de EDE). *Existe um único processo estocástico x_t , que satisfaz a EDE (3.8) a menos de indistinguibilidade.*

Demonstração: Para simplificar a notação, consideremos uma única componente de ruído na nossa equação, i.e. $m = 1$. O caso geral é extensão imediata, com fórmulas mais longas.

Considere o operador $\mathcal{S} : E \rightarrow E$ definido por

$$(\mathcal{S}U)_t = x_0 + \int_0^t f_0(s, U_s) ds + \int_0^t f_1(s, U_s) dW_s.$$

Estamos procurando um ponto fixo para esse operador. Vamos usar a seguinte métrica, para um t fixado:

$$\Phi_t(U, V) = \mathbf{E} \left[\sup_{s \leq t} |U_s - V_s|^2 \right],$$

que corresponde a $|U - V|_{E_t}^2$. Usando que para todo $h, k \in \mathbf{R}$, vale a desigualdade $(h + k)^2 \leq 2(h^2 + k^2)$, temos que

$$\begin{aligned} \Phi_t(\mathcal{S}U, \mathcal{S}V) &= \mathbf{E} \left[\sup_{s \leq t} \left| \int_0^s f_0(s, U_s) ds + \int_0^s f_1(s, U_s) dW_s \right. \right. \\ &\quad \left. \left. - \int_0^s f_0(s, V_s) ds - \int_0^s f_1(s, V_s) dW_s \right|^2 \right] \\ &\leq 2\mathbf{E} \left[\sup_{s \leq t} \left(\int_0^s f_0(s, U_s) - f_0(s, V_s) ds \right)^2 \right. \\ &\quad \left. + \sup_{s \leq t} \left(\int_0^s f_1(s, U_s) - f_1(s, V_s) dW_s \right)^2 \right] \end{aligned}$$

Usando as desigualdades L^p de Doob (Teorema 2.6, com $p = 2$) e de Cauchy-Schwartz para o produto interno em L^2 , ficamos com

$$\begin{aligned} \Phi_t(\mathcal{S}U, \mathcal{S}V) &\leq 8\mathbf{E} \left(\int_0^t f_1(s, U_s) - f_1(s, V_s) dW_s \right)^2 \\ &\quad + \mathbf{E} \left[\left(\int_0^t 1 ds \right) \left(\int_0^t |f_0(s, U_s) - f_0(s, V_s)|^2 ds \right) \right] \end{aligned}$$

que, pela isometria de Itô (Teorema 3.6), fica

$$\begin{aligned}
&\leq 8\mathbf{E} \int_0^t |f_1(s, U_s) - f_1(s, V_s)|^2 ds \\
&\quad + 2t \mathbf{E} \int_0^t |f_0(s, U_s) - f_0(s, V_s)|^2 ds \\
&\leq 2K^2(t+4) \mathbf{E} \int_0^t \sup_{s \leq r} |U_s - V_s|^2 dr \\
&= 2K^2(t+4) \int_0^t \Phi_r(U, V) dr. \tag{3.9}
\end{aligned}$$

Para um $0 < T < \infty$ fixado grande o suficiente, vamos abreviar as constantes da formula acima e colocar $C := 2K^2(T+4)$. Mostraremos que para todo $t < T$ teremos

$$\Phi_t(\mathcal{S}^n U, \mathcal{S}^n V) \leq \frac{C^n t^n}{n!} \Phi_T(U, V),$$

para todo $n \in \mathbf{N}$. De fato, é óbvio para $n = 0$; assuma verdadeira para n então, pela desigualdade (3.9) teremos:

$$\begin{aligned}
\Phi_t(\mathcal{S}^{n+1} U, \mathcal{S}^{n+1} V) &\leq C \int_0^t \frac{C^n r^n}{n!} \Phi_T(U, V) dr \\
&= \frac{C^{n+1} t^{n+1}}{(n+1)!} \Phi_T(U, V).
\end{aligned}$$

Note que essa desigualdade mostra também que a imagem de \mathcal{S} está de fato em E . O argumento de existência finaliza com o fato de existir um $n \in \mathbf{N}$ tal que \mathcal{S}^n é uma contração uniforme em relação à métrica Φ_t . Portanto, pelo teorema do ponto fixo de Banach, temos que existe um semimartingale que é solução da EDE 3.8.

Quanto à unicidade da solução, note que a norma $|\cdot|_E$ que definimos no espaço E na verdade é uma seminorma no sentido de podermos definir uma relação de equivalência entre dois semimartingales neste espaço $X \sim Y$ dada por $|X - Y|_E = 0$. Isso significa que

$$\mathbf{E} \sup_{s \leq t} |U_s - V_s|^2 = 0,$$

o que implica que existe um conjunto $\Omega' \subset \Omega$ de probabilidade total $\mathbf{P}(\Omega') = 1$ tal que se $\omega \in \Omega'$ então $X_s(\omega) = Y_s(\omega)$ para todo $s \leq t$. Portanto a unicidade está no espaço de Banach quocientado pela relação de equivalência E/\sim , que é o mesmo que dizer que a unicidade se dá a menos de indistinguibilidade. \square

Observação 3.6 (Sobre soluções de EDE).

1. EDE's são geradores naturais de semimartingales: a componente de variação limitada da solução da equação (3.8) está na primeira parcela do lado direito e a componente de martingale são as outras parcelas que correspondem às integrações em relação ao movimento browniano.
2. Uma EDE em geral pode ter como integradores outros martingales, note que em relação a essa direção de generalização, o crucial nesta demonstração é que a variação quadrática desses integradores sejam limitadas por um λt quando aplicamos a isometria de Itô.
3. Da mesma maneira que vimos para a fórmula de Itô, aqui o tempo t pode ser trocado por um tempo de parada T .
4. Normalmente, uma solução de uma EDE é verificada usando-se a fórmula de Itô. Nem sempre é fácil achar uma solução explícita. Para quem trabalha com aplicações diretas à física, engenharia ou economia, métodos de simulações numéricos podem ser bem vindos. Em Kloeden e Platen [29], além de apresentar vários algoritmos para soluções numéricas, temos uma fonte grande de exemplos numa tabela de EDE's com as respectivas soluções explícitas.
5. No caso dos campos de vetores serem somente localmente Lipschitz, o teorema acima vale localmente. Pode acontecer de termos tempo de explosão finito, i.e. x_t indo para o infinito ou para um ponto fora do domínio. Neste caso o tempo de explosão, que depende das trajetórias para cada ω , é um tempo de parada.

\square

Exemplo 3.5 (Equações lineares).

Seja A uma matriz $n \times n$ e $b_i(t)$, $i = 1, \dots, m$ curvas contínuas (ou mensuráveis) em \mathbf{R}^n . Considere a seguinte EDE linear não homogênea:

$$\begin{cases} dx_t = A x_t dt + \sum_{i=1}^m b_i(t) dW_t^i, \\ x_0 = a \in \mathbf{R}^n \end{cases}$$

Tome $\phi_t = e^{At}$ a solução fundamental da equação linear $\dot{x} = Ax_t$. Então temos que a solução é dada pela fórmula de variação de parâmetros

$$x_t = \phi_t(a) + \phi_t \sum_{i=1}^m \int_0^t \phi_s^{-1} b_i(s) dW_s^i.$$

De fato, usando a fórmula de Itô na forma infinitesimal, temos que

$$\begin{aligned} dx_t &= d\phi_t(a) + d\left(\phi_t \sum_{i=1}^m \int_0^t \phi_s^{-1} b_i(s) dW_s^i\right) \\ &= (\dot{\phi}_t)(a) dt + \left[\dot{\phi}_t \sum_{i=1}^m \int_0^t \phi_s^{-1} b_i(s) dW_s^i\right] dt \\ &\quad + \phi_t \sum_{i=1}^m \phi_t^{-1} b_i(s) dW_t^i \\ &= \left(A\phi_t(a) + A\phi_t \sum_{i=1}^m \int_0^t \phi_s^{-1} b_i(s) dW_s^i\right) dt \\ &\quad + \sum_{i=1}^m b_i(s) dW_t^i \\ &= A\left(\phi_t(a) + \phi_t \sum_{i=1}^m \int_0^t \phi_s^{-1} b_i(s) dW_s^i\right) dt \\ &\quad + \sum_{i=1}^m b_i(s) dW_t^i \\ &= Ax_t dt + \sum_{i=1}^m b_i(s) dW_t^i \end{aligned}$$

De onde segue que de fato x_t é a solução da equação. □

Exemplo 3.6 (Outros casos).

Vimos no Exemplo 2.12 o exemplo da EDE (3.6) que é linear no ruído:

$$dx_t = \lambda x_t dB_t,$$

cujas solução é a exponencial estocástica $\varepsilon^\alpha(B)_t = \exp\{\alpha B_t - \frac{\alpha^2 t}{2}\}$.

Uma aplicação direta da fórmula de Itô é acharmos a EDE que um dado semimartingale satisfaz. Por exemplo, verifique que $x_t = \cos(B_t)$ é solução de

$$dx_t = -\sqrt{1-x_t^2} dB_t - \frac{1}{2}x_t dt$$

com condição inicial $x_0 = 1$. □

Observação 3.7 (Sobre soluções fortes e fracas).

Dizemos que uma solução da EDE (3.8) é *forte* quando ela puder ser construída a partir dos movimentos brownianos estipulados nesta mesma equação. A demonstração de existência e unicidade de soluções de EDE que fizemos acima é para soluções fortes. Em contraste, uma solução é dita *fraca* se ela for escrita em relação a outro movimento browniano que não os integradores originais. Um exemplo simples para ilustrarmos é o seguinte. Considere a EDE:

$$dx_t = K_t dW_t, \tag{3.10}$$

onde $K_t \in O(n)$ é uma curva no grupo de matrizes ortogonais e W_t é um movimento browniano em \mathbf{R}^n . Na próxima seção veremos que $\tilde{B}_t = K_t B_t$ também é um movimento browniano. Assim, a equação pode ser reescrita em relação a esse novo processo como

$$dx_t = d\tilde{W}_t, \tag{3.11}$$

cujas solução é $x_t = x_0 + \tilde{W}_t$. Assim, essa solução é uma solução fraca da equação (3.10). Uma solução forte de (3.10) em geral pode ser muito difícil de se calcular. Outro exemplo menos trivial é o processo de Bessel (ver Oksendal [43]) que dá o raio de um MB no plano

$R = \sqrt{(B^1)^2 + (B^2)^2}$. Usando a fórmula de Itô, verificamos que esse processo satisfaz a EDE

$$dR = \frac{1}{2R} dt + d\tilde{B}_t,$$

onde o MB \tilde{B}_t satisfaz:

$$\tilde{B}_t = \int_0^t B^1((B^1)^2 + (B^2)^2)^{-1/2} dB^1 + \int_0^t B^2((B^1)^2 + (B^2)^2)^{-1/2} dB^2. \quad (3.12)$$

Também por resultado da próxima seção podemos verificar que \tilde{B}_t é um movimento browniano.

3.5 Gerador infinitesimal

Quando os campos de vetores envolvidos em uma EDE são autônomos (i.e. não dependem da variável t), então as soluções são processos de Markov contínuos em \mathbf{R}^n , também chamados de *processos de difusão* ou *difusão de Itô*. De fato, neste caso, dado que o processo se encontra em um certo ponto $y \in \mathbf{R}^n$, se pensarmos na discretização da solução no tempo como um processo elementar, as probabilidades de transição a partir de y , para essa discretização, ficam determinadas pelas direções dos campos f_0, f_1, \dots, f_m . A demonstração deste teorema, usando a independência dos integradores em incrementos disjuntos pode ser encontrado em qualquer um dos livros clássicos de cálculo estocástico: Oksendal [43], Revuz e Yor [47], Kunita [30], Karatzas e Shreve [27] ou o clássico Ikeda e Watanabe [23], entre muitos outros.

As probabilidades de transição de um processo de difusão de Itô, i.e. probabilidades de transição de soluções de EDE homogêneas no tempo, determinam as distribuições finito dimensionais do processo. Para que essas distribuições fiquem efetivamente bem definidas, só nos falta fixar a condição inicial x_0 em \mathbf{R}^n , onde o processo será disparado. É o mesmo que dizer que, disparando o processo em x_0 , tomando $X_t^{x_0}$ a solução da EDE com esta condição inicial, então dado um conjunto aberto A e um tempo $t \geq 0$, a probabilidade de transição de x_0 para A neste tempo é:

$$P_t(x_0, A) = \mathbf{P}\{\omega : X_t^{x_0} \in A\}.$$

Um *semigrupo de operadores de Feller* $(T_t)_{t \geq 0} : C_0 \rightarrow C_0$ agindo no espaço das funções contínuas com suporte compacto C_0 satisfaz as propriedades:

- i) (contínuo e positivo em C_0) $T_0 = 1d$, $\|T_t\| \leq 1$ e é operador positivo para todo $t \geq 0$;
- ii) (semigrupo) $T_{t+s} = T_t \circ T_s$, para todo $s, t \geq 0$;
- iii) (continuidade no tempo) $\lim_{t \rightarrow 0} \|T_t(f) - f\| = 0$ para toda $f \in C_0$.

As probabilidades de transição mencionadas acima geram o seguinte semigrupo de operadores Feller $T_t : C_0(\mathbf{R}^n) \rightarrow C_0(\mathbf{R}^n)$, para $t \geq 0$:

$$(T_t f)(x) = \int_{\mathbf{R}^n} f(y) P_t(x, dy).$$

As propriedades de semigrupo Feller podem ser facilmente verificadas; em particular, a propriedade (ii) acima segue da condição de Chapman-Kolmogorov. Pelo teorema da medida induzida 1.6 também podemos escrever esse semigrupo como

$$(T_t f)(x) = \mathbf{E}[f(X_t)].$$

O *gerador infinitesimal* \mathcal{A} do semigrupo T_t é a derivada em t deste semigrupo em alguma função f . Assim, para um ponto $x_0 \in \mathbf{R}^n$ temos que

$$(\mathcal{A}f)(x_0) = \lim_{t \searrow 0} \frac{T_t f(x_0) - f(x_0)}{t},$$

quando o limite existir.

Usando a fórmula de Itô podemos calcular o gerador infinitesimal associado a uma EDE, ou melhor dizendo, associado a suas probabilidades de transição. Verificaremos que trata-se de um operador de segunda ordem, portanto funções de classe C^2 estão no seu domínio.

Proposição 3.15. *Dada a EDE*

$$dx_t = f_0(x_t) dt + \sum_{i=1}^m f_i(x_t) dW_t^i,$$

o gerador infinitesimal do processo de Markov associado é dado pelo operador de segunda ordem, que aplicado a $h \in C^2$ é dado por

$$\mathcal{A}(h)(x) = (f_0)h(x) + \frac{1}{2} \sum_{i=1}^m (f_i)^t H(h) f_i(x)$$

onde $H(h)$ é a hessiana de h e f_i^t é a transposta do vetor f_i representado matricialmente.

Demonstração: De fato, seja h uma função de classe C^2 . Então, pela fórmula de Itô,

$$h(x_t) = h(x_0) + \int_0^t \langle \nabla h(x_s), dx_s \rangle + \frac{1}{2} \int_0^t H(h)(x_s) [dx_s^t, dx_s].$$

Onde a expressão $H(h)(x_s) [dx_s^t, dx_s]$ significa, a variação quadrática de $dx_s^t H(h) dx_s$ com a convenção de que $dW^i dW^j = \delta_{ij} ds$.

Tomando a esperança, as componentes que são martingales desaparecem. Além disso a última integral pode ser reescrita como:

$$\begin{aligned} \mathbf{E}h(x_t) &= h(x_0) + \int_0^t \langle \nabla f(x_s), f_0(x_s) \rangle ds \\ &\quad + \frac{1}{2} \sum_{i=1}^m \int_0^t f_i^t H(h)(x_s) f_i ds. \end{aligned}$$

Derivando na variável t obtemos o resultado. □

A proposição acima também pode ser vista como corolário imediato do seguinte lema mais geral, ver, entre outros [43, Lemma 7.8]:

Lema 3.16. *Se um processo $Y_t \in \mathbf{R}^n$ for uma integral*

$$Y_t = \int_0^t u(s, \omega) ds + \int_0^t v(s, \omega) dB_s,$$

com u e v processos integráveis. Então, para toda $f : \mathbf{R}^n \rightarrow \mathbf{R}$ de classe C^2 temos que

$$\begin{aligned} \mathbf{E}[f(Y_t)] &= f(0) + \sum_{i=1}^n \mathbf{E} \int_0^t \langle \nabla f, du_s \rangle \\ &\quad + \frac{1}{2} \sum_{i,j=1}^n \int_0^t (vv^t)_{i,j} \frac{\partial^2 f}{\partial x_i \partial x_j}(Y_s) ds. \end{aligned}$$

O tempo t acima pode ser substituído por um tempo de parada.

Outra caracterização do MB em termos de processo de Markov contínuo é a seguinte: Um processo de difusão X_t é um MB se seu gerador infinitesimal é dado por $\frac{1}{2}\Delta$. Essa caracterização vai ser útil em geometria estocástica para definir MB em uma variedade riemanniana: um processo ali será um MB se for uma difusão com gerador infinitesimal dado por $\frac{1}{2}\Delta$, onde o Laplaciano aqui ($\Delta = \text{div}\nabla$) é o operador de Laplace-Beltrami.

Exercício 3.8. *Use o Lema 3.16 para mostrar que os processos \tilde{W}_t e \tilde{B}_t das fórmulas (3.11) e (3.11) são de fato movimentos brownianos.*

3.5.1 Fórmula de Dynkin

Combinando a fórmula de Itô com um tempo de parada finito q.s., e o truque de tomar a esperança para eliminar a componente martingale, temos a seguinte

Teorema 3.17 (Fórmula de Dynkin). *Dada uma função $f \in C^2$ e um tempo de parada $\tau < M$ q.s. para algum $M > 0$ então:*

$$\mathbf{E}[f(X_\tau)] = f(X_0) + \mathbf{E} \int_0^\tau \mathcal{A}f(x_s) ds.$$

Se chamarmos $u_t = \mathbf{E}[f(X_t)]$, a fórmula de Dynkin está dizendo que

$$\frac{\partial u}{\partial t} = \mathbf{E}(\mathcal{A}f)(X_t).$$

A famosa equação reversa de Kolmogorov vai garantir que o gerador infinitesimal comuta com a esperança, i.e.

Teorema 3.18 (Equação reversa de Kolmogorov). *Para uma função $f \in C^2$ e $u_t = \mathbf{E}[f(X_t)]$ então temos:*

$$\frac{\partial u}{\partial t} = \mathcal{A}u.$$

3.5.2 Aplicações

Nos próximos exemplos exploraremos a fórmula de Dynkin. Os dois primeiros exemplos fornecem informações sobre a dinâmica do MB (ver Oksendal [43]). No terceiro exemplo resolveremos o problema clássico de Dirichlet de função harmônica com condição de fronteira.

Exemplo 3.7 (tempo de saída de conjunto limitado).

Suponha que temos um MB X_t inicializado em $a \in \mathbf{R}^n$, cujo gerador infinitesimal sabemos que é $\frac{1}{2}\Delta$. Dado uma esfera de raio R , maior que $|a|$, considere o tempo de parada T dado pelo tempo de saída de X_t da bola $B_0(R)$. Em princípio não sabemos nem se esse tempo de parada é finito, já que muitas trajetórias ficam eternamente dentro desta bola. Assim, para garantir que nas fórmulas estaremos usando um tempo de parada limitado, considere a sequência de tempos de parada $\tau_k = \tau \wedge k$.

Considere a função diferenciável $f(x) = |x|^2$, cujo laplaciano é constante $\Delta f = 2n$, em todo ponto. Assim, aplicando a fórmula de Dynkin temos:

$$\begin{aligned} \mathbf{E}f(X_{\tau_k}) &= |a|^2 + \mathbf{E} \int_0^{\tau_k} n \, ds & (3.13) \\ &= |a|^2 + n\mathbf{E}\tau_k. \end{aligned}$$

Portanto, como $f(X_\tau) \equiv R^2$ temos que para todo $k \in \mathbf{N}$

$$\mathbf{E} [\tau_k] \leq \frac{1}{n}(R^2 - |a|^2).$$

E quando k tende para infinito, τ_k tende para τ , portanto, pela equação (3.13), temos :

$$\mathbf{E} [\tau] = \frac{1}{n}(R^2 - |a|^2).$$

Concluimos que o tempo de saída do MB de um conjunto limitado é finito com probabilidade um, e ainda mais, esse tempo é integrável. \square

Exemplo 3.8 (Recorrência do MB).

Considere o problema de inicializarmos o MB em um ponto b fora da bola $B_R(0)$ centrada na origem com raio R , e o tempo que o processo leva para atingir a bola, se é que a atinge!

Suponha inicialmente que temos um MB B_t inicializado em $b \in \mathbf{R}^2$. Considere o anel aberto em \mathbf{R}^2 dado por $A_k = \{x : R < |x| < 2^k R\}$. Para um ponto b neste anel, chamemos de τ_k o tempo de saída do conjunto A_k . Considere uma função $f : \{|x| > R - \varepsilon\} \rightarrow \mathbf{R}$ tal que nos anéis A_k temos $f(x) = -\log|x|$. Como $\Delta f = 0$ em A_k , então, pela fórmula de Dynkin,

$$\mathbf{E}f(B_{\tau_k}) = f(b)$$

para todo $k \geq 2$. Essa média acima está sendo feita sobre trajetórias que param no círculo interno de raio R , com probabilidade $p_k = \mathbf{P}\{|B_{\tau_k}| = R\}$, e com as que param no círculo externo de raio $2^k R$, com probabilidade $q_k = \mathbf{P}\{|B_{\tau_k}| = 2^k R\}$, com $p_k + q_k = 1$. A fórmula acima quer dizer que

$$-(\log R)p_k - (\log R + k \log 2)q_k = -\log|b|$$

para todo k . Ou ainda:

$$p_k = 1 - \frac{\log(|b|/R)}{k \log 2}$$

Portanto p_k tende a 1 quando k vai para infinito. Isto significa que com probabilidade 1 as trajetórias fora da bola, entram nela em um tempo finito. Ou ainda, inicializando fora de um conjunto aberto, esse conjunto aberto será visitado por todas (q.s.) as trajetórias, e mais, um número infinito de vezes. Essa propriedade é chamada de *recorrência* do MB no plano.

Essa recorrência em \mathbf{R}^2 implica, em particular, olhando-se em cada coordenada euclideana, em recorrência do MB linear. Esse fato implica que o tempo de parada do Exemplo 3.2, de fato é finito q.s. embora, como foi dito, não é integrável.

Considere agora as mesmas construções de A_k , τ_k , p_k e q_k feitas anteriormente, mas no \mathbf{R}^n , com $n \geq 3$. Tome a função harmônica $f(x) = |x|^{2-n}$. Então ficamos com

$$R^{2-n}p_k + (2^k R)^{2-n}q_k = |b|^{2-n}$$

para todo k . Ou ainda:

$$p_k = \frac{|b|^{2-n} - (2^k R)^{2-n}}{R^{2-n} - (2^k R)^{2-n}}.$$

Assim, o fato de p_k tender a $\left(\frac{|b|}{R}\right)^{2-n} < 1$ quando k cresce para infinito significa que com probabilidade estritamente positiva, trajetórias não entrarão na bola de raio R . Quanto maior a dimensão, menos trajetórias entrarão nesta bola. Este fenômeno é chamado de *transiência* do MB em dimensão $n \geq 3$. □

Exemplo 3.9 (Problema de Dirichlet).

Nessa aplicação da fórmula de Dynkin, mostraremos um argumento probabilístico para se resolver o problema clássico de Dirichlet (determinístico). Suponha que temos um domínio aberto, limitado e conexo $D \in \mathbf{R}^n$, com fronteira ∂D mensurável. O problema de Dirichlet consiste em se achar uma função harmônica em D que satisfaça uma condição de fronteira dada por uma função mensurável $g : \partial D \rightarrow \mathbf{R}$. O problema pode então ser equacionado como

$$\begin{cases} \Delta u = 0 & \text{em } D, \\ u|_{\partial D} = g \end{cases}$$

Fixamos um ponto $a \in D$. Disparamos um movimento browniano B_t em a e seja τ o tempo de parada quando esse processo toca na fronteira. Quando aplicamos a fórmula de Dynkin na função harmônica u que estamos procurando ficamos com:

$$\mathbf{E} [u(B_\tau)] = u(a).$$

A variável aleatória $B_\tau : \Omega \rightarrow \partial D$ induz uma medida μ_a em ∂D . Ao mesmo tempo u em ∂D coincide com g . Assim, pelo teorema da medida induzida podemos calcular $u(a)$ como a integral:

$$u(a) = \int_{\partial D} g(x) \mu_a(dx).$$

Essa procedimento probabilístico para se calcular a função harmônica pode servir para aproximações por simulações. Por exemplo,

construe-se um reticulado cúbico de pontos para simular um movimento browniano neste reticulado. Cada trajetória que toca na fronteira, considera-se o valor de g naquele ponto. Com alguns milhares de trajetórias simuladas em um computador, tomando a média delas podemos obter uma boa aproximação da função harmônica u . Esse procedimento de achar a média sobre uma quantidade grande de simulações é chamado de *método Monte Carlo*. □

Observação 3.9 (Fronteira de Poisson).

Geometricamente falando, *espaço ou fronteira de Poisson* Π de uma (sub)variedade com estrutura de grupo e uma medida de Borel μ é um espaço topológico compacto onde podemos representar todas as funções μ -harmônicas da variedade nas funções mensuráveis reais em Π . Mostrar que o problema de Dirichlet acima tem solução, significa mostrar que a fronteira topológica ∂D é o espaço de Poisson de domínios abertos limitados D no \mathbf{R}^n . A medida μ que toma-se aqui que torna equivalente as definições de μ -harmônica e a definição por núcleo do laplaciano é uma medida gaussiana dada pela probabilidade de transição do movimento browniano $P_t(x, dy)$, para qualquer $t > 0$. Ver detalhes sobre fronteira de Poisson em grupos de Lie em Furstenberg [17] ou em Lopez, Ruffino e San Martín [37]. □

3.6 Integral de Stratonovich

Voltando à definição da integral de Itô na equação (3.3) repare que nas somas de Riemann, o integrando é avaliado no extremo esquerdo de cada subintervalo da partição. Isso garante que a integral será um martingale.

De fato, a título de exemplo, considere a integral $\int_0^t B_s dB_s$ e uma partição $\Delta = \{0 = t_0 < t_1 < \dots < t_n = t\}$. Calculando as esperanças das somas de Riemann, com os integrandos avaliados como na integral de Itô temos:

$$\mathbf{E} \sum_{i=0}^{n-1} B_{t_i} (B_{t_{i+1}} - B_{t_i}) = \sum_{i=0}^{n-1} \mathbf{E} [B_{t_i} (B_{t_{i+1}} - B_{t_i})] = 0$$

para toda partição Δ .

No entanto, se avaliarmos os integrandos no extremo direito dos subintervalos teremos:

$$\begin{aligned} \mathbf{E} \sum_{i=0}^{n-1} B_{t_{i+1}} (B_{t_{i+1}} - B_{t_i}) &= \sum_{i=0}^{n-1} \mathbf{E} [(B_{t_{i+1}} - B_{t_i})^2] \\ &= \sum_{i=0}^{n-1} t_{i+1} - t_i \\ &= t \end{aligned}$$

para toda partição Δ .

A integral de Stratonovich é uma média entre essas duas situações vistas acima. A vantagem dela em relação à integral de Itô vai aparecer na simplificação da fórmula de Itô para ela, que vai voltar a ser uma fórmula de primeira ordem, como no cálculo tradicional. Além disso, modela com mais simplicidade uma classe muito grande de sistemas dinâmicos e processos estocásticos em variedades diferenciáveis.

Definição 3.2. *Dado um semimartingale contínuo X_t e f_t um processo $\sigma(X_t)$ -adaptado, a integral de Stratonovich é definida como o seguinte limite em probabilidade:*

$$\int_0^t f_s \circ dX_s = \mathbf{P} - \lim_{|\Delta| \rightarrow 0} \sum_{i=1}^n \left(\frac{f_{t_{i+1}} + f_{t_i}}{2} \right) (X_{t_{i+1}} - X_{t_i}).$$

(quando o limite existe).

O seguinte teorema vai dizer que a classe de processos integráveis segundo Stratonovich é a mesma que para a integral de Itô.

Teorema 3.19 (Fórmula de conversão Itô-Stratonovich). *Se $(f_t)_{t \geq 0}$ é um semimartingale então:*

$$\int_0^t f_s \circ dX_s = \int_0^t f_s dX_s + \frac{1}{2} [f, X]_t,$$

no sentido de que se um lado existir, o outro também existe e são iguais.

Demonstração: Dada uma partição, a soma de Riemann para a integral de Stratonovich se decompõe como:

$$\begin{aligned} \sum_{i=0}^{n-1} \frac{1}{2} (f_{t_{k+1}} + f_{t_k})(X_{t_{k+1}} - X_{t_k}) &= \\ \sum_{i=0}^{n-1} f_{t_k}(X_{t_{k+1}} - X_{t_k}) + \frac{1}{2} \sum_{i=0}^{n-1} (f_{t_{k+1}} - f_{t_k})(X_{t_{k+1}} - X_{t_k}). \end{aligned}$$

O resultado segue diretamente das definições em termos de convergência em probabilidade quando a malha da partição vai para zero. □

O resultado abaixo é uma fórmula para converter EDE's de Itô em EDE's de Stratonovich.

Teorema 3.20. *Seja $X, Y_1, \dots, Y_m : \mathbf{R}_{\geq 0} \times \mathbf{R}^n \rightarrow \mathbf{R}^n$ campos vetoriais com X contínuo e Lipschitz na segunda coordenada e Y_j contínuos e C^2 na segunda coordenada. Então a solução da EDE de Stratonovich*

$$dx_t = X(t, x_t) dt + \sum_{j=1}^m Y_j(t, x_t) \circ dB_t^j,$$

com $x_0 = x(0)$ é a solução da EDE de Itô:

$$dx_t = \tilde{X}(t, x_t) dt + \sum_{j=1}^m Y_j(t, x_t) dB_t^j,$$

onde

$$\begin{aligned} \tilde{X}(t, x_t) &= X(t, x) + \frac{1}{2} \sum_{j=1}^m \sum_{i=1}^n \left(Y_j^i \frac{\partial Y_j}{\partial x_i}(t, x) \right) \\ &= X(t, x) + \frac{1}{2} \sum_{j=1}^m d(Y_j)_{(t,x)}(Y_j), \end{aligned}$$

onde $d(Y_j)_{(t,x)}(Y_j)$ é a diferencial (jacobiano) de cada campo Y_j avaliado no ponto (t, x) , na direção do próprio campo Y_j .

Demonstração: Temos que as integrações de Stratonovich, pela fórmula de conversão, em termos de integrais de Itô ficam:

$$\int_0^t Y_j(s, x_s) \circ dB_t^j = \int_0^t Y_j(s, x_s) dB_t^j + \frac{1}{2}[Y_j(t, x_t), B_t^j]. \quad (3.14)$$

Precisamos calcular a variação quadrática cruzada do último termo acima. Pela fórmula de Itô,

$$\begin{aligned} Y_j(t, x_t) &= Y_j(0, x_0) + \int_0^t \frac{\partial Y_j}{\partial t}(s, x_s) ds + \sum_{i=1}^n \int_0^t \frac{\partial Y_j}{\partial x_i}(s, x_s) dx^i \\ &\quad + \frac{1}{2} \sum_{i,k} \int_0^t \frac{\partial^2 Y_j}{\partial x^i \partial x^k} d[x^i, x^k]_s, \end{aligned}$$

onde $x_t = (x_t^1, \dots, x_t^n)$. A parte cuja variação quadrática é diferente de zero, portanto a única parte relevante para a variação cruzada da equação (3.14) é a que contém a componente martingale no primeiro somatório. De fato:

$$\begin{aligned} \sum_{i=1}^n \int_0^t \frac{\partial Y_j}{\partial x_i}(s, x_s) dx^i &= \sum_{i=1}^n \int_0^t \frac{\partial Y_j}{\partial x_i}(s, x_s) (\tilde{X}^i(s, x_s)) ds \\ &\quad + \sum_{i=1}^n \int_0^t \frac{\partial Y_j}{\partial x_i}(s, x_s) \left(\sum_{k=1}^m Y_k^i dB_t^k \right) \end{aligned}$$

onde, apesar de \tilde{X} ser justamente o que queremos calcular, neste ponto ele não é relevante já que a componente associada a ele terá variação quadrática zero. Assim, aplicando a propriedade da variação quadrática da integral de Itô (Teorema 3.6, item (4)), ficamos com:

$$\begin{aligned} [Y_j(t, x_t), B_t^j] &= \left[\sum_{i=1}^n \sum_{k=1}^m \int_0^t \frac{\partial Y_j}{\partial x_i}(s, x_s) Y^i dB_t^k, B_t^j \right] \\ &= \sum_{i=1}^n \sum_{k=1}^m \int_0^t \frac{\partial Y_j}{\partial x_i}(s, x_s) Y^i d[B_t^k, B_t^j] \\ &= \int_0^t d(Y_j)_{(s, x_s)}(Y_j) ds. \end{aligned}$$

O que completa a demonstração. □

O que torna a integração de Stratonovich tão interessante é o fato da sua fórmula de Itô (de composição com semimartingales) ser simples como no caso do cálculo de primeira ordem, o que facilita contas e cálculos de soluções de EDE:

Teorema 3.21 (Fórmula de Itô para integral de Stratonovich). *Seja $F : \mathbf{R}^d \rightarrow \mathbf{R}$ uma função de classe C^3 e $X = (X^1, \dots, X^d)$ um semimartingale d -dimensional. Então:*

$$F(X_t) = F(X_0) + \sum_{i=1}^d \int_0^t \frac{\partial F}{\partial x_i}(X_s) \circ dX_s^i$$

O somatório do lado direito também será escrito como $\int_0^t \langle \nabla F, dX_s \rangle$, onde ∇F é o gradiente da F .

Demonstração: Convertendo a soma de integrais de Stratonovich para para Itô, temos:

$$\begin{aligned} \sum_{i=1}^d \int_0^t \frac{\partial F}{\partial x_i}(X_s) \circ dX_s^i &= \sum_{i=1}^d \int_0^t \frac{\partial F}{\partial x_i}(X_s) dX_s^i \\ &\quad + \frac{1}{2} \sum_i^d \left[\frac{\partial F}{\partial x_i}(X_s), dX_s^i \right]. \end{aligned}$$

Para calcularmos a variação cruzada no último somatório usamos a fórmula de Itô para identificar a componente martingale de

$$\begin{aligned} \frac{\partial F}{\partial x_i}(X_t) &= \frac{\partial F}{\partial x_i}(X_0) + \sum_{j=1}^d \int_0^t \frac{\partial^2 F}{\partial x_j \partial x_i}(X_s) dX_s^j \\ &\quad + \frac{1}{2} \sum_{j,k=1}^d \int_0^t \frac{\partial^3 F}{\partial x_j \partial x_k \partial x_i}(X_s) d[X_k, X^j]_s. \end{aligned}$$

Portanto, a componente martingale local de $\frac{\partial F}{\partial x_i}(X_t)$ está contido no processo

$$\sum_{j=1}^d \int_0^t \frac{\partial^2 F}{\partial x_j \partial x_i}(X_s) dX_s^j.$$

Portanto, voltando à variação quadrática da primeira equação desta demonstração, temos que

$$\left[\frac{\partial F}{\partial x_i}(X_s), dX_s^i \right] = \sum_{j=1}^d \int_0^t \int_0^t \frac{\partial^2 F}{\partial x_j \partial x_i}(X_s) d[X^j, X^i]_s.$$

Substituindo esse processo naquela equação, temos que a igualdade do teorema segue pela fórmula de Itô para integral de Itô (3.4). \square

Exemplo 3.10 (Aplicações: integração e solução de EDE).

Como foi dito, o fato da composição ter uma regra da cadeia de primeira ordem como no cálculo com funções diferenciáveis, as integrações e soluções de EDE ficam muito mais fáceis de serem procuradas quando estão na forma de uma integral de Stratonovich. Uma vez achada a solução, podemos voltar ao caso da integral de Itô pela fórmula de conversão. Considere os casos simples:

$$\int_0^t B_s \circ dB_s = \frac{1}{2} B_t^2,$$

que não é um martingale, conforme visto no Exemplo 3.3. Ou ainda

$$\int_0^t \cos^2(B_s) \circ dB_s = \frac{1}{2} [B_t - \sin(B_t) \cos(B_t)],$$

(verifique). Solução de uma equação linear, com A uma matriz $d \times d$ e B_t um MB linear:

$$dx_t = Ax_t dB_t$$

é dada simplesmente por $x_t = \exp\{AB_t\} x_0$. Ou ainda, com A^1 e A^2 matrizes $d \times d$ e (B_t^1, B_t^2) MB's independentes, então

$$dx_t = A^1 x_t dB_t^1 + A^2 x_t dB_t^2$$

tem solução $x_t = \exp\{A^1 B_t^1 + A^2 B_t^2\} x_0$, independente de A^1 e A^2 comutarem ou não. \square

Exercício 3.10. Use a fórmula de conversão Itô-Stratonovich para verificar que dada uma difusão (sistema autônomo) escrito como uma EDE de Stratonovich:

$$dx_t = X(x_t) dt + \sum_{j=1}^m Y_j(x_t) \circ dB_t^j,$$

essa difusão tem gerador infinitesimal dado por

$$\mathcal{A} = X + \frac{1}{2} \sum_{j=1}^m (Y_j)^2.$$

onde $(Y_j^2 f) = Y_j(Y_j f)$.

Capítulo 4

Sistemas dinâmicos estocásticos

A intenção deste capítulo é, além de mostrar as ferramentas apresentadas nos capítulos anteriores sendo usadas na análise da dinâmica estocástica, também apresentar ao leitor uma variedade de problemas onde temos alguma experiência, daí ter um certo apelo pessoal. O pesquisador, neste ponto, está convidado, ao ler esses tópicos, propor generalizações e novos problemas relacionados.

4.1 Trajetórias das soluções de uma EDE

Na maioria das aplicações consideramos equações diferenciais estocásticas autônomas (difusões). Além de gerarem processos de Markov, tecnicamente não perdemos generalidade pois com o mesmo truque do caso determinístico, podemos transformar um sistema não autônomo em um autônomo aumentando uma dimensão (a temporal $\mathbf{R}_{\geq 0}$) no espaço de estados. Considere então uma EDE:

$$dx_t = X(x_t) dt + \sum_{j=1}^m Y_j(x_t) \circ dB_t^j, \quad (4.1)$$

onde X, Y_1, \dots, Y_m são campos vetoriais de classe C^1 e $\{B_t^1, \dots, B_t^m\}$ são MB lineares independentes. Se assumirmos que as derivadas dos campos (jacobianos) são limitadas então o tempo de explosão para todas as condições iniciais é infinito, ver Kunita [30].

Uma maneira interessante de intuirmos sobre a solução é o teorema de aproximação de Wong-Zakai, em [60]. Antes disso, vamos definir o que é uma *aproximação poligonal do movimento browniano*. Considere um MB B_t e uma partição, digamos com sub-intervalos de comprimentos iguais a Δt . Então uma aproximação poligonal de B_t é o processo contínuo, linear por partes u_t tal que nos múltiplos de Δt , $u_{n(\Delta t)} = B_{n(\Delta t)}$, para $n \in \mathbf{N}$. A vantagem desse processo é que ele é derivável em todo ponto exceto nos múltiplos de Δt .

Para cada movimento browniano linear B^j , $j = \{1, \dots, m\}$ da equação (4.1), considere uma aproximação poligonal u_t^j

Teorema 4.1 (Aproximação de Wong-Zakai). *As trajetórias da solução da equação (4.1) podem ser aproximadas pelas trajetórias da EDO (em cada ω):*

$$dx_t^u = X(x_t^u) dt + \sum_{j=1}^m Y_j(x_t^u) du_t^j,$$

ou, na forma diferencial:

$$\dot{x}_t^u = X(x_t^u) + \sum_{j=1}^m Y_j(x_t^u) \dot{u}_t^j,$$

A convergência nas trajetórias se dá da seguinte maneira: para todo $\epsilon > 0$:

$$\lim_{(\Delta t) \rightarrow 0} \mathbf{P} \left\{ \sup_{s \in [0, t]} |x_s - x_s^u| < \epsilon \right\} = 0$$

(ver, entre outros [58]). A convergência acima é chamada *uniforme em probabilidade nos compactos* (ucp).

Um *sistema de controle*, ou *poli-sistema dinâmico* ou ainda, *sistema de controle afim* com os mesmos campos de vetores da equação

(4.1), é uma equação onde temos funções no tempo que podem controlar o impacto dos campos vetoriais nas trajetórias:

$$\dot{x}_t = X(x_t^u) + \sum_{j=1}^m v^j(t) Y_j(x_t^u). \quad (4.2)$$

A função vetorial $v(t) = (v^1(t), \dots, v^m(t))$ é chamada de *função de controle* ou *controlador*. A relação entre as possíveis trajetórias da equação de controle (4.2) de acordo com $v(t)$, e as trajetórias da EDE (4.1) são muito estreitas: é como dizer que numa EDE, o controlador é uma função aleatória (“derivada” de um processo gaussiano).

O controle pode ser feito dentro do subespaço afim de direções gerado por X somado ao subespaço gerado pelos campos de difusão; naturalmente, não se tem dinâmica em outras direções ortogonais aos campos. Essa relação estreita entre os pontos atingíveis pela equação de controle e a trajetória realizada por alguma das trajetórias guiadas pelo MB em (4.1) é dada pelo seguinte:

Teorema 4.2 (Teorema do Suporte). *O suporte das v.a. das pela solução da (4.1) é igual ao fecho dos pontos atingíveis pela equação de controle (4.2).*

Uma demonstração pode ser encontrada no livro clássico Ikeda e Watanabe [23]. Note que a relação aqui é como se o controle fosse dado, formalmente por $u_t^j = \frac{dB_t^j}{dt}$, uma derivada que sabemos que não existe, daí a necessidade da equação estocástica ser escrita na forma integral.

4.2 Fluxos estocásticos e cociclos

Voltamos à comparação entre nossa equação original (4.1) e a equação de controle (4.2). Para a equação (4.2), existe um fluxo de controle que depende da função controle $u(t)$. Isto é, existe uma família de difeomorfismos (locais) $\phi_t : \mathcal{C} \times \mathbf{R}^d \rightarrow \mathbf{R}^d$, onde \mathcal{C} é a família de todos os controles possíveis (digamos, contínuos ou mensuráveis) tais que dada uma condição inicial x_0 , então a solução para um certo controle $u(t)$, é dada por $x_t(u) = \phi_t(u)(x_0)$.

Repare que a condição de semigrupo de difeomorfismos de solução passa a valer no seguinte sentido, para todo $t, s \geq 0$:

$$\phi_{(t+s)}(u)(x_0) = \phi_t(u(s + \cdot)) \circ \phi_s(u)(x_0)$$

onde a translação no tempo (que chamamos de “shift”) do controle $u(s + \cdot)$ é necessária porque depois que o sistema saiu do ponto x_0 , seguindo as direções dadas pelo controle u , quando ele atinge o ponto $\phi_s(u)(x_0)$, as instruções que ele recebe daí pra frente para que a igualdade da equação acima se mantenha, são as instruções de u a partir do tempo s . Esse shift então é um novo controle $v(t) = u(s + t) \in \mathcal{C}$. Esse shift (ou atualização nos parâmetros do semigrupo solução) sempre aparece quando falamos de equações controladas por funções no tempo.

Pela mesma razão, no caso de fluxo estocástico para a equação (4.1), precisamos fazer o shift do movimento browniano. Como a relação entre u e B_t se faz na sua derivada, para mantermos a imagem do shift no espaço de “controle” (espaço de Wiener, i.e. contínuas inicializadas no zero), o shift será representado no espaço de probabilidade canônico: no espaço de Wiener. Assim o shift será dado por $\theta_s : \Omega \rightarrow \Omega$ e satisfaz

$$B_t(\theta_s(\omega)) = B_{t+s} - B_s.$$

O fluxo estocástico então satisfaz as seguintes propriedades chamada de *cociclo* :

$$\phi_{t+s}(\omega)(x_0) = \phi_t(\theta_s \omega)(\cdot) \circ \phi_s(\omega)(x_0).$$

Ver também L. Arnold [1].

Dadas as condições de diferenciabilidade (C^1) nos campos de vetores da equação, existe uma dependência diferenciável no espaço \mathbf{R}^d do fluxo estocástico em relação a condição inicial. Isto é, fixado um t e um ω (q.s.), podemos falar da diferencial $d\phi_t(x_0)$ que é uma transformação do espaço tangente a x_0 ao espaço tangente a $d\phi_t(x_0)$, que neste caso se identificam com o próprio \mathbf{R}^d .

Resumindo, temos as seguintes propriedades do fluxo estocástico:

Teorema 4.3. *Para campos diferenciáveis C^∞ na equação (4.1) temos que o correspondente fluxo estocástico $\phi : \mathbf{R}_{\geq 0} \times \Omega \times \mathbf{R}^d \rightarrow \mathbf{R}^d$ satisfaz:*

- i) (*Identidade em $t = 0$*) $\phi_0 \equiv Id$, **P**-q.s.
- ii) (*Cociclo*) $\phi_{t+s}(\omega)(x_0) = \phi_t(\theta_s\omega)(\cdot) \circ \phi_s(\omega)(x_0)$.
- iii) (*Diferenciabilidade no espaço*) *Diferenciável na variável x_0 .*

4.3 Estabilidade: expoentes de Lyapunov

Exploramos a diferenciabilidade do fluxo estocástico no espaço. Suponha o caso de termos um ponto fixo x_0 , i.e. onde todos os vetores da equação (4.1) se anulam. A linearização do fluxo neste ponto é a diferencial do fluxo estocástico $d\phi_t(x_0)$, que preserva o ponto, mas não sua vizinhança. A linearização (diferencial) corresponde a procurar propriedades deste fluxo no espaço tangente a x_0 , já que em x_0 a dinâmica se trivializou. Chamando de $v_t = d\phi_t(x_0)(v_0)$ essa dinâmica no espaço tangente, temos que v_t satisfaz a seguinte EDE:

$$dv_t = dX_{(x_t)}(v_t) dt + \sum_{j=1}^m dY_{(x_t)}^j(v_t) dB_t^j$$

Dada uma solução X_t de uma EDE em \mathbf{R}^d , para cada medida ergódica deste processo temos o seguinte resultado de existência dos expoentes de Lyapunov, que são parâmetros que medem crescimento ou decrescimento exponencial de distâncias entre pontos relativamente pertos, em direções específicas. Trata-se de uma extensão do teorema ergódico de Oselede para fluxos estocásticos. Ver, e.g. Carverhill [5], Elworthy [12], Arnold [1] entre outros.

Teorema 4.4 (Teorema ergódico multiplicativo). *Dada uma medida de probabilidade ergódica μ para o fluxo estocástico ϕ_t da EDE (4.1) Para $(\omega, x) \mathbf{P} \times \mu$ -q.s. existem os expoentes de Lyapunov $\lambda_d \leq \dots \leq \lambda_1$ (determinísticos) e uma filtração aleatória em subespaços*

$$\{0\} = V^{d+1} \subseteq V^{d+1}(\omega, x) \subseteq \dots \subseteq V^1(\omega, x) = \mathbf{R}^d$$

Tal que, se $v_0 \in V^k(\omega, x) \setminus V^{k-1}(\omega, x)$ então:

$$\lim_{t \rightarrow \infty} \frac{1}{t} \log |d\phi_t(\omega, x)(v_0)| = \lambda_k.$$

Além disso os subespaços da filtração, chamados de subespaços de Oseletets são invariantes pelo fluxo:

$$d\phi_t(x)(V^k(\omega, x)) = V^k(\theta_t\omega, \phi_t(\omega)x).$$

Os números λ_k , $k = 1, \dots, d$, são chamados os *expoentes de Lyapunov* do sistema dinâmico estocástico dado pela equação (4.1).

Esse teorema é estrutural no estudo de dinâmica estocástica, trata-se essencialmente de uma álgebra linear para esses sistemas. A demonstração da versão enunciado acima é uma aplicação da versão apresentada em Ruelle [50] para produto de matrizes aleatórias.

O enunciado do teorema ergódico acima foi feito com tempo positivo; neste caso, temos que tangente a cada subespaço de Oseletets associado a um expoente de Lyapunov negativo, existe uma subvariedade invariante (aleatória) que contrai na direção da trajetória $\phi_t(\omega, x)$; é a chamada *subvariedade estável* associada ao expoente λ_k . A estabilidade do sistema está associada com o maior expoente de Lyapunov: se $\lambda_1 < 0$ então a subvariedade estável contém uma vizinhança de x , cuja dinâmica é contração exponencial na direção de $\phi_t(\omega, x)$, que é a situação em que chamamos de estabilidade exponencialmente assintótica. Ver Arnold [1] e as referências contidas nele. Como foi dito, por ser estrutural, expoentes de Lyapunov aparecem em várias centenas de artigos nas melhores revistas especializadas da área nas últimas 2 ou 3 décadas.

4.4 Conjugações de cociclo

O fato do fluxo estocástico ser um cociclo, significa que ele é uma componente do fluxo num espaço maior, que é o espaço produto $\Omega \times \mathbf{R}^d$, onde o fluxo, chamado de *skew-product* é dado por $\Phi_t : \Omega \times \mathbf{R}^d \rightarrow \Omega \times \mathbf{R}^d$ com $\Phi_t(\omega, x) = (\theta_t(\omega), \phi_t(\omega)(x))$.

Exercício 4.1. Verifique que Φ_t é um fluxo em $\Omega \times \mathbf{R}^d$, isto é $\Phi_0 = Id$ e $\Phi_{t+s} = \Phi_t \circ \Phi_s$.

Quando falamos de conjugações de fluxos estocásticos precisamos respeitar a estrutura de cociclo. Isto significa que para termos essa conjugação, digamos entre dois fluxos estocásticos ϕ_t e ξ_t , devemos

mostrar uma conjugação do fluxo skew-product ou uma conjugação por um homeomorfismo aleatório, digamos $H(\omega)$ que satisfaça:

$$\xi_t(x_0) = H^{-1}(\theta_t\omega) \circ \phi_t(\omega) \circ H(\omega)(x_0).$$

Exercício 4.2. *Verifique que dada a conjugação aleatória $H(\omega)$ acima, $(\omega, x) \mapsto (\theta_t\omega, H(\omega)(x))$ é uma conjugação entre os fluxos skew-product associados a ϕ_t e a ξ_t .*

Todos os teoremas de conjugação na teoria de sistemas dinâmicos diferenciáveis clássica pode ser recolocada aqui com a questão da existência da conjugação aleatória mensurável que respeita o cociclo como acima. P. Imkeller chama essa conjugação de *cohomologia de cociclo*, ver [24].

Exemplo 4.1 (Hartman-Grobman estocásticos).

Exemplos dessas conjugações aparecem nas versões estocásticas dos teoremas de Hartman-Grobman, ver Coayla-Teran, Ruffino e Mohammed [9], [8], ou, com outra abordagem, Imkeller [24].

Se o sistema tiver um ponto fixo x_0 cuja vizinhança só apresentar direções de contração ou repulsão (tecnicamente, significa ser *hiperbólico*, i.e. ter os expoentes de Lyapunov diferentes de zero), então localmente existe um homeomorfismo aleatório $H(\omega)$, com domínio aleatório tal que no domínio apropriado o fluxo estocástico e seu linearizado são conjugados:

$$\phi_t(x) = H^{-1}(\theta_t\omega)(\cdot) \circ d(\phi_t)_{x_0}(\omega)(\cdot) \circ H(\omega)(x).$$

4.5 Números de rotação em S^1

Incluimos aqui um outro exemplo em dimensão baixa de aplicação da teoria de dinâmica estocástica no sentido de semimartingales. Dado um espaço de probabilidade $(\Omega, \mathcal{F}, \mathbf{P})$, considere θ uma transformação ergódica em Ω . Dado um homeomorfismo aleatório $f : \Omega \times S^1 \rightarrow S^1$ que preserva a orientação em S^1 , considere a sequência de homeomorfismos aleatórios $f_n(\omega, \cdot) = f(\theta^{n-1}\omega, \cdot)$. Em particular, tomando Ω como sendo espaço produto, sequências i.i.d. podem ser vistas dessa maneira.

Essa sequência de homeomorfismos aleatórios geram um processo de Markov em S^1 dado por

$$\phi_n(\omega, x) = f(\theta^{n-1}, \cdot) \circ \dots \circ f(\theta\omega, \cdot) \circ f(\omega, x).$$

Seja μ uma medida invariante em S^1 para esse processo de Markov induzido.

Para cada f_n , considere o levantamento $F_n : \mathbf{R} \rightarrow \mathbf{R}$, i.e. tal que $\pi \circ F_n = f_n \circ \pi$ onde $\pi : \mathbf{R} \rightarrow S^1$ dado pelo ângulo no círculo. Esses levantamentos não são únicos. Tomaremos em particular aqueles onde $F_n(0) \in (-1/2, 1/2]$, portanto dados da seguinte forma:

$$F_n(\omega, x) = x + \beta_n(\omega, \pi(x)),$$

com $\beta_n(\omega, \cdot) : S^1 \rightarrow (-1/2, 3/2]$.

Temos então o seguinte teorema ergódico para existência do número de rotação da sequência f_1, f_2, \dots agindo em S^1 :

Teorema 4.5 (Teorema ergódico para número de rotação). *O número de rotação do sistema dinâmico ergódico dado pelos homeomorfismos aleatórios f_1, f_2, \dots em S^1*

$$\rho(f, \theta) = \lim_{n \rightarrow \infty} \frac{F_n \circ \dots \circ F_1(x) - x}{n} \pmod{1}$$

existe, é independente de x e é dado por

$$\int_{S^1} \mathbf{E}\beta(\omega, x) d\mu(x) \pmod{1}$$

P-q.s.

A demonstração deste teorema é uma aplicação direta do teorema ergódico de Birkhoff 5.7. Ver Ruffino [52]. Neste mesmo artigo mostramos que no caso estocástico também podemos reconstruir o número de rotação de um sistema contínuo se fizermos uma amostragem em um sistema estocástico linear com frequência grande o suficiente:

Considere a seguinte EDE linear em \mathbf{R}^2 :

$$dx_t = Ax_t dt + \sum_{i=1}^m B^i x_t \circ dW_t^i, \quad (4.3)$$

onde A, B^1, \dots, B^m são matrizes 2×2 , $(W_t^1, \dots, W_t^m)_{t \geq 0}$ é um MB em \mathbf{R}^m . Denotemos por $\varphi(t, \omega)$ o fluxo solução.

A coordenada angular na forma de uma função contínua $\alpha_t \in \mathbf{R}$ é a solução da seguinte EDE em termos da integral de Itô:

$$d\alpha_t = [\langle As_t, v_t \rangle + \sum_{i=1}^m \left(\frac{1}{2} \langle (B^i)^2 s_t, v_t \rangle - \langle B^i s_t, s_t \rangle \langle B^i s_t, v_t \rangle \right)] dt + \sum_{i=1}^m \langle B^i s_t, v_t \rangle dW_t^i,$$

onde (s_t, v_t) é um processo nas bases ortonormais, tal que s_t descreve $x_t/|x_t|$, ou ainda, o que é o mesmo, $s_t = \pi(\alpha_t)$. O número de rotação do sistema contínuo é facilmente calculado neste caso lembrando que a componente martingale tem média assintótica nula, s_t é um processo de Markov, então, usando o teorema ergódico para processos de Markov (Teorema 5.8):

$$\begin{aligned} \rho(\varphi) &= \lim_{t \rightarrow \infty} \frac{\alpha_t}{t} \\ &= \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t [\langle As_t, v_t \rangle + \sum_{i=1}^m \left(\frac{1}{2} \langle (B^i)^2 s_t, v_t \rangle - \langle B^i s_t, s_t \rangle \langle B^i s_t, v_t \rangle \right)] dt \\ &= \int_{S^1} f(x) \mu(dx) \quad \mathbf{P} \times \mu\text{-q.s.} \end{aligned}$$

onde x se identifica com (s, v) e

$$f(s, v) = \langle As, v \rangle + \sum_{i=1}^m \left(\frac{1}{2} \langle (B^i)^2 s, v \rangle - \langle B^i s, s \rangle \langle B^i s, v \rangle \right).$$

Teorema 4.6 (Teorema da amostragem estocástico). *Amostrando no tempo, com intervalo $T > 0$ o fluxo linear ϕ_t que é solução do sistema (4.3), obtemos uma sequência de matrizes aleatórias i.i.d. que, tomando seu número de rotação reescalado no tempo, converge para $\rho(\varphi)$ q.s.:*

$$\lim_{T \rightarrow 0} \frac{1}{T} \rho(\varphi(T, \omega), \theta_T) = \rho(\varphi) \quad a.s..$$

4.6 Outros exemplos e aplicações:

Concluiremos este texto listando uma série de outros tópicos de interesse em dinâmica estocástica. Não pretendemos ser exaustivo. Muitos deles ainda contém problemas em aberto:

Decomposição de fluxos estocásticos

Um assunto interessante é estudar a difusão que ocorre dentro do grupo de difeomorfismos de \mathbf{R}^d (ou de uma variedade). Quando nos voltamos a um dos subgrupos desse grupo de dimensão infinita, podemos nos perguntar se podemos fatorar um fluxo estocástico com uma componente naquele subgrupo composta com outro processo em outro subgrupo de interesse. Algumas respostas parciais foram conseguidas com o grupo de isometrias de onde, neste caso, obtemos as médias assintóticas das rotações e no “resto” da decomposição, os expoentes de Lyapunov, ver Liao [35] e Ruffino [53]. Outras decomposições foram feitas posteriormente em Silva [55].

Sistemas Hamiltonianos estocásticos

Sistemas hamiltonianos estocásticos ainda não tem um conceito bem estabelecido do ponto de vista estocástico. Note que a existência de campos em direções diferentes do campo Hamiltoniano implicaria que o sistema não estaria contido no mesmo nível de energia. Existem no entanto várias abordagens interessantes. Uma delas de Cami-Lazaro e Ortega [32], onde eles definem sistemas hamiltonianos em variedades de Poisson. As difusões hamiltonianas ali definidas ficam contidas em folheações simpléticas e são pontos críticos de um funcional energia $S(\gamma)$ adequado. Outra abordagem aparece em Xue-Mei Li [34].

Bifurcação para processos de Markov

Várias abordagens estão sendo feitas para uma teoria de bifurcação em difusões. Mas ainda não se conhece uma abordagem unificada que aglutine todas as definições. Possivelmente uma abordagem geral para processos de Markov possa ser feita usando a teoria da dinâmica de 2-pontos, ver e.g. Kunita [31].

Geometria estocástica

As equações de Itô ou Stratonovich em variedades diferenciais vem ganhando bastante interesse, não só por generalizar propriedades conhecidas, mas porque apontam para propriedades geométricas e topológicas. O movimento browniano em uma variedade riemanniana carrega informações sobre a geometria desta, a título de exemplo, ver [6]. Em particular, um grande impulso foi dado quando Itô fez o levantamento horizontal no fibrado das bases de um processo na variedade. A partir do transporte paralelo estocástico, integração de 1-formas, teoria de Hodge-de Rham, análise estocástica no espaço de laços e vários outros tópicos foram ganhando interesse. Em particular, recentemente, difusões em folheações, ver Ledesma [33] e referências dali.

Capítulo 5

Apêndices

Reservamos este espaço no apêndice para acrescentarmos alguns comentários ou algumas demonstrações que julgamos relevantes. Alguns desses itens não foram mencionados antes neste texto porque não estavam no caminho direto que nos propomos entre cálculo estocástico e sistemas dinâmicos estocásticos, mas que são fundamentais num curso mais completo. Ainda assim, não pretendemos esgotar tudo que deve cobrir um curso de introdução no assunto.

5.1 Independência e convolução de medidas

Proposição 5.1. *A soma de variáveis aleatórias independentes tem distribuição dada pela convolução das duas distribuições.*

Em uma fórmula, a proposição diz que se X e Y são variáveis aleatórias independentes em um espaço vetorial V , então

$$(X + Y)_* \mathbf{P} = 1.(x + y)X_* \mathbf{P}(dx) \times Y_* \mathbf{P}(dy)$$

no sentido de

$$\begin{aligned} (X + Y)_* \mathbf{P}(A) &= \iint_{V \times V} 1_A(x + y)X_* \mathbf{P}(dx)Y_* \mathbf{P}(dy) \\ &= (X_* \mathbf{P}) * (Y_* \mathbf{P})(A) \end{aligned}$$

onde $(X_*\mathbf{P}) * (Y_*\mathbf{P})$ é a convolução das duas medidas.

Proposição 5.2 (Lei fraca dos grandes números). *Dada uma sequência $(X_n)_{n \geq 1}$ de variáveis aleatórias i.i.d. que seja integrável, então*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n X_i = \mathbf{E}X_1$$

Uma demonstração simples e direta pode ser encontrada no Folland [15]. Esta propriedade também é corolário do teorema ergódico de Birkhoff (ver Teorema 5.7).

5.2 Critério de Kolmogorov

A demonstração do critério de Kolmogorov é bastante técnica, no entanto, trata-se de um resultado que é imprescindível para mostrarmos a existência do movimento browniano com trajetórias contínuas. Provaremos a seguinte versão em dimensão um deste teorema:

Teorema 5.3 (Critério de Kolmogorov). *Dado um processo estocástico X , suponha que existam constantes $\alpha, \beta, K > 0$ tais que os incrementos do processo satisfaçam*

$$\mathbf{E} |X_{t+h} - X_t|^\alpha \leq Kh^{1+\beta},$$

para todo $t, h \geq 0$. Então existe um processo contínuo \tilde{X}_t que é uma modificação de X .

Provaremos o teorema no intervalo considerando, sem perda de generalidade, que t está no intervalo $[0, 1]$. Considere o subconjunto Δ dos números diádicos, i.e. que podem ser escritos em sistema binário como $0, a_1 a_2 \dots a_k$, com cada algarismo $a_j \in \{0, 1\}$, $j = 1, \dots, k < \infty$, i.e.

$$0, a_1 a_2 \dots a_k = \sum_{j=1}^k a_j 2^{-j}.$$

Seja Δ_m os números que na representação binária tem exatamente m algarismos depois da vírgula, i.e. $\Delta_m = \{0, a_1 a_2 \dots a_m\}$. Naturalmente $\Delta = \cup^\infty \Delta_m$.

Demonstração: Tome $0 < \gamma < \beta\alpha^{-1}$. Sejam $s, t \in \Delta_m$ elementos consecutivos, i.e. $t - s = 2^{-m}$. Então, pela desigualdade de Tchebychev,

$$\begin{aligned} \mathbf{P}\{|X_t - X_s| > 2^{-m\gamma}\} &\leq 2^{m\gamma\alpha} \mathbf{E} |X_t - X_s|^\alpha \\ &\leq K 2^{m\gamma\alpha} 2^{-m(1+\beta)} \\ &= K 2^{m(\gamma\alpha - \beta - 1)}. \end{aligned}$$

Seja

$$A_m = \left\{ \omega : |X_t(\omega) - X_s(\omega)| > 2^{-m\gamma} \text{ para algum par } t, s \in \Delta_m, \right. \\ \left. \text{tal que } |t - s| = 2^{-m} \right\}.$$

Como o número de elementos consecutivos em Δ_m é 2^m então

$$\mathbf{P}(A_m) \leq 2^m K 2^{m(\gamma\alpha - \beta - 1)} = K 2^{m(\gamma\alpha - \beta)}.$$

Note que $(\gamma\alpha - \beta) < 0$. Assim,

$$\sum_{m=1}^{\infty} \mathbf{P}(A_m) < \infty,$$

portanto, pelo lema de Borel-Cantelli temos que:

$$\mathbf{P}(\liminf A_m^c) = 1.$$

Tome $\omega \in \liminf A_m^c$, então existe $m_0(\omega) \in \mathbf{N}$ tal que para todo $m \geq m_0(\omega)$,

$$|X_t(\omega) - X_s(\omega)| \leq 2^{-\gamma m},$$

para todo $t, s \in \Delta_m$ consecutivos.

Mostraremos agora que $X_t|_\Delta$ é contínuo com esta restrição de domínio, para todo ω fixado em $\liminf A_m^c$.

De fato, para esse ω fixado, considere $t \in \Delta$. Dado um $\varepsilon > 0$ tome

$$m = \min \left\{ m : \frac{2^{-\gamma m}}{2^\gamma - 1} < \varepsilon, t \in \Delta_m \right\}.$$

Afirmamos que se $|s - t| < 2^{-m}$ para $s \in \Delta$ então $|X_s - X_t| < \varepsilon$. De fato, suponha que $s \in \Delta_n$, com $n > m$ e tome uma seqüência

finita s_m, s_{m+1}, \dots, s_n em Δ de tal forma que: o primeiro elemento $s_m = t$, o último elemento $s_n = s$; que respeita a decomposição em diádicos $s_j \in \Delta_j$ para $j = m, m+1, \dots, n$; e que cada par (s_{j-1}, s_j) apresente elementos consecutivos em Δ_j , com $j = m+1, m+2, \dots, n$.

Então,

$$\begin{aligned} |X_s - X_t| &\leq \sum_{j=m+1}^n |X_{s_{j-1}} - X_{s_j}| \leq \sum_{j=m+1}^n 2^{-\gamma j} \\ &\leq \frac{2^{-\gamma(m+1)}}{1 - 2^{-\gamma}} \leq \frac{2^{-\gamma m}}{2^\gamma - 1} < \varepsilon. \end{aligned}$$

Como os diádicos Δ são densos em $[0, 1]$ então existe uma única extensão contínua $\tilde{X}_t(\omega)$ tal que $\tilde{X}_t(\omega) = X_t(\omega)$ se $t \in \Delta$ e ω em $\liminf A_m^c$.

Vamos mostrar agora que $\tilde{X}_t(\omega)$ é modificação de X_t . e fato, fixado um $t \in [0, 1]$:

1) Se $t \in \Delta$ então não tem nada a ser feito. Aliás, note que restrito a Δ , os processos $\tilde{X}_t|_\Delta$ e $X_t|_\Delta$ são indistinguíveis porque

$$\mathbf{P}\{\omega : \tilde{X}_t(\omega) = X_t(\omega), \text{ para todo } t \in \Delta\} = \mathbf{P}\{\liminf A_m^c\} = 1.$$

2) Considere o caso $t \notin \Delta$. Queremos provar que existe um subconjunto que pode depender de t , $\Omega_t \subset \Omega$, de probabilidade um tal que se $\omega \in \Omega_t$ então $\tilde{X}_t(\omega) = X_t(\omega)$. Considere então a sequência de subconjuntos mensuráveis:

$$\Omega_{n,t} = \{\omega : |\tilde{X}_t(\omega) - X_t(\omega)| \leq \frac{1}{n}\},$$

para cada $n \in \mathbf{N}$. Seja δ a probabilidade

$$\delta := \mathbf{P}\{\Omega_{n,t}^c\} = \mathbf{P}\{\omega : |\tilde{X}_t(\omega) - X_t(\omega)| > \frac{1}{n}\}$$

Vamos mostrar que $\delta = 0$ para todo $n \in \mathbf{N}$. Tome uma sequência $s_i \rightarrow t$, com $s_i \in \Delta$, para todo $i = 1, 2, \dots$. A continuidade de \tilde{X}_t e o fato das restrições coincidirem $\tilde{X}_t|_\Delta = X_t|_\Delta$ implicam que para todo $\omega \in \Omega_{n,t}^c$ temos:

$$|\lim_{i \rightarrow \infty} X_{s_i}(\omega) - X_t(\omega)| > \frac{1}{n}.$$

Para cada $\omega \in \Omega_{n,t}^c$, considere a função mensurável $N : \Omega_{n,t}^c \rightarrow \mathbf{N}$, com $N(\omega)$ sendo o menor inteiro positivo tal que se $i \geq N(\omega)$ então $|X_{s_i}(\omega) - X_t(\omega)| > \frac{1}{n}$. Temos então que

$$\Omega_{n,t}^c = \bigcup_{i=1}^{\infty} N^{-1}(i).$$

Logo, existe um k , inteiro positivo tal que

$$\mathbf{P}\left\{\bigcup_{i=1}^k N^{-1}(i)\right\} \geq \delta/2.$$

Portanto, para todo $i > k$,

$$\mathbf{P}\{\omega : |X_{s_i} - X_t| > 1/n\} \geq \delta/2.$$

Agora, pela desigualdade de Tchebyshev:

$$\mathbf{P}\{\omega : |X_{s_i} - X_t| > 1/n\} \left(\frac{1}{n}\right)^\alpha \geq \mathbf{E} |X_{s_i} - X_t|^\alpha,$$

isto é,

$$\frac{\delta}{2} \left(\frac{1}{n}\right)^\alpha \geq \mathbf{E} |X_{s_i} - X_t|^\alpha,$$

para todo $i > k$, o que implica que $\delta = 0$ pois, por hipótese

$$\mathbf{E} |X_{s_i} - X_t|^\alpha \leq K(t - s_i)^{1+\beta},$$

para todo $i > k$. Portanto $\mathbf{P}\{\Omega_{n,t}\} = 1$ e

$$\Omega_t = \bigcap_{n=1}^{\infty} \Omega_{n,t}$$

é o conjunto de probabilidade total procurado. □

A demonstração do teorema de Kolmogorov (também chamado de Kolmogorov-Tchentsov) acima garante mais do que o enunciado.

Na verdade, a modificação \tilde{X} do processo X que construímos é localmente γ -Hölder contínua para todo $\gamma \in (0, \beta/\alpha)$. Mais precisamente, existe uma v.a. $h(\omega)$ positiva q.s. e uma constante δ tal que

$$\mathbf{P} \left\{ \omega : \sup_{0 \leq t-s \leq h(\omega)} \frac{|\tilde{X}_t - \tilde{X}_s|}{|t-s|^\gamma} < \delta \right\} = 1.$$

Para mais detalhes, inclusive aspectos históricos, ver por exemplo Karatzas e Shreve [27, pp. 53 e 126] ou Kunita [30].

5.3 Gaussianas

Um fato interessante, que mostra como que as gaussianas tem papel estrutural na teoria, é a possibilidade de decompor o espaço $L^2(\Omega)$ como polinômios dessas variáveis aleatórias. Na Observação 1.22, verificamos que como convergência em $L^p(\Omega)$ implica em convergência em probabilidade, então o Teorema 1.19 garante que o subespaço de gaussianas centradas \mathcal{H}_1 contido em $L^2(\Omega)$ seja fechado. Esse subespaço corresponde ao primeiro subespaço \mathcal{H}_1 da *decomposição em caos de Wiener*, dado por polinômios de Hermite compostos com gaussianas. Obtem-se assim uma decomposição de L^2 em somas diretas ortogonais:

$$L^2(\Omega) = \bigoplus_{i=0}^{\infty} \mathcal{H}_i$$

onde \mathcal{H}_i , $i = 0, 1, \dots$ são subespaços de funções obtidas pela composição do i -ésimo polinômio de Hermite h_i com v.a. gaussianas. Os polinômios de Hermite são definidos por:

$$h_n(x) = (-1)^n e^{x^2/2} \frac{d^n}{dx^n} e^{-x^2/2}.$$

Os primeiros polinômios desta série são $h_0(x) = 1$, $h_1(x) = x$, $h_2(x) = x^2 - 1$, $h_3(x) = x^3 - 3x$, etc. Esses polinômios também são ortogonais quando na reta tomamos a medida gaussiana $e^{-x^2/2}$. Para ver mais sobre caos de Wiener recomendamos, por exemplo Nualart [42].

Outro resultado que mostra que as gaussianas ocupam um papel estrutural na teoria é o seguinte teorema que mostra que normalizando soma de v.a. independentes obtemos uma gaussiana (ver e.g. Folland [15]):

Teorema 5.4 (Teorema do limite central). *Considere uma sequência X_1, X_2, \dots de variáveis aleatórias independentes com médias m_1, m_2, \dots e variâncias $\sigma_1^2, \sigma_2^2, \dots$. Então*

$$\mathbf{P} \left\{ \frac{\sum_{i=1}^n X_i - \sum_{i=1}^n m_i}{\sqrt{\sum_{i=1}^n \sigma_i^2}} < k \right\} = \Phi(k)$$

onde

$$\Phi(k) = \int_{-\infty}^k \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx.$$

5.4 Mais sobre o Movimento Browniano

Do ponto de vista histórico, esse nome vem da observação do botânico inglês chamado Robert Brown, que, observando grãos de pólen e outras diminutas partículas em suspensão na água, observou o fenômeno de vibração ininterrupta em todas as direções. Na época essa descoberta causou um debate interessante sobre criacionismo de matéria orgânica (viva) e possibilidade ou não de movimentos intrínsecos de seres inanimados. Em suas próprias palavras, no seu primeiro artigo onde menciona esse fenômeno, em 1828: “not only organic tissues, but also inorganic substances, consist of what [he] calls animated or irritable particles.” A discussão sobre criacionismo incluía a pergunta se o movimento correspondia a uma espécie de “vida” da partícula. Posteriormente outros modelos apareceram para mostrar que na verdade a vibração vinha do impacto da partícula observada com milhões de outras partículas do líquido. O primeiro modelo físico, de A. Einstein e Smoluchowski só apareceu no começo do século XX. Ver mais sobre essas e outras informações em E. Nelson [41]

5.4.1 Processo de Ornstein-Uhlenbeck

Outro aspecto histórico interessante foi a tentativa de modelar o movimento browniano como diferenciáveis no tempo. Esse enfoque

teve início com as idéias de Langevin e culminou em 1930 com os resultados de Ornstein-Uhlenbeck [57]. O aspecto principal da abordagem sugerida por Langevin é a tentativa de aproximar o modelo matemático do movimento browniano (ainda se inspirando nas partículas de grãos de pólen) das equações newtonianas que regem a dinâmica de uma partícula. Essencialmente, ao invés de introduzirmos o ruído branco na velocidade da partícula, como no modelo clássico (Einstein-Smoluchowski):

$$\begin{aligned} dx_t &= dB_t \\ x_0 &= x(0) \end{aligned}$$

com $x_t \in \mathbf{R}$; assumimos que a velocidade $v_t = dx_t/dt$ da partícula existe, de tal modo que agora o ruído é introduzido na aceleração, de acordo com a equação de Langevin :

$$dv_t = -\beta v_t dt + \sigma dB_t \quad (5.1)$$

Nessa equação o parâmetro β é uma constante de amortecimento, tem dimensão de frequência (inverso do tempo). Em outras palavras, nesse contexto teríamos que é a aceleração que agora está perturbada pelo ruído, *i.e.* formalmente :

$$\frac{d^2 x_t}{dt^2} = -\beta v + \sigma \frac{dB_t}{dt}$$

O resultado principal dessa abordagem é que, com as trajetórias x_t deriváveis (!) podemos ajustar os parâmetros β e σ para obter uma aproximação, em termos de distribuição de dimensão finita, do processo de Wiener (que representa o modelo probabilístico, não derivável, de Einstein-Smoluchowski).

Dadas as condições iniciais $x_0 = x(0)$ e $v_0 = v(0)$ a solução da equação linear de Langevin (5.1) é dada pelo processo gaussiano

$$v_t = e^{-\beta t} v_0 + e^{-\beta t} \int_0^t e^{\beta s} dB_s$$

(lembre que integral de Wiener fornece variáveis gaussianas); o processo x_t então será dado por

$$x_t = x_0 + \int_0^t v_s ds.$$

Repare que a integral de uma curva no espaço das gaussianas também é uma gaussiana, portanto o processo x_t também é um processo gaussiano. O processo x_t é conhecido como *o processo de Ornstein-Uhlenbeck*.

5.4.2 Outras construções

Ainda sobre o movimento browniano, em que uma variedade Riemanniana, a distribuição do MB é dada também pela solução da equação do calor, isto é, o chamado núcleo do calor $P_t(x)$ na variedade. Seu gerador infinitesimal, como semigrupo agindo em funções duas vezes diferenciáveis $C^2(\mathbb{R})$ é dado por $T_t f(x) = d/dt\{\mathbf{E}[f(B_t^x)]\}$ que é o operador de Laplace- Beltrami $\Delta/2$, onde aqui, B_t^x é o MB inicializado em x . Pelo aspecto geométrico, apesar de não estar no escopo deste texto, o MB "enxerga" a métrica riemanniana e se espalha, a partir de um certo ponto, uniformemente em todas as direções, com "espalhamento" (variância) crescente no tempo. Dentre outras interpretações equivalentes, a grosso modo isso significaria que se a variedade riemanniana fosse um terreno, as trajetórias desse processo têm maior probabilidade de contornar uma elevação (regiões de grande curvatura), com difusão ao longo de geodésicas, do que probabilidade de subir e descer a elevação. Ver artigo de divulgação de [45].

Uma segunda abordagem para a construção do MB que gostaríamos de mencionar, mais analítico funcional, é a seguinte, envolvendo um processo indexado em um espaço de Hilbert.

Teorema 5.5. *Seja H um espaço de Hilbert separável. Então existe um espaço de probabilidade $(\Omega, \mathcal{F}, \mathbf{P})$ e uma transformação linear $X : H \rightarrow L^2(\Omega)$ isométrica tal que $X(h)$ é gaussiana centrada.*

A demonstração se baseia no seguinte:

Teorema 5.6. *Dada uma medida de probabilidade ν em \mathbb{R} , existe um espaço de probabilidade $(\Omega, \mathcal{F}, \mathbf{P})$ e uma sequência de variáveis aleatórias i.i.d. X_n definidas em Ω com distribuição ν .*

De fato, basta tomar o espaço produto $\Omega = (\mathbb{R}, \nu)^\infty$, e para cada $\omega = (x_1, x_2, \dots)$ em Ω , defina $X_n(\omega) = x_n$. Vamos agora à

Demonstração: (do Teorema 5.5).

Tome uma base ortonormal $\{e_n\}$ de H e considere o espaço de probabilidade Ω definido no teorema acima onde se define uma sequência g_n de variáveis gaussianas centradas independentes. Dado um elemento $h \in H$, defina $X(h)$ como sendo a série $\sum \langle h, e_n \rangle g_n$. \square

Enfatizamos que a isometria do Teorema 5.5 significa que $X(h)$ tem variância $\|h\|_H^2$, i.e.

$$\mathbf{E} \left[X(h)^2 \right] = \|h\|_H^2.$$

Tomaremos o espaço de Hilbert do Teorema 5.5 como sendo o $L^2(\mathbf{R}_+)$. O movimento browniano neste contexto aparece como sendo o processo que é imagem da isometria X das funções características do intervalo $[0, t]$, para $t \geq 0$, i.e. $B_t = X(1d_{[0,t]})$. O processo B_t definido assim é gaussiano, $\mathbf{E}[B_t^2] = t$, para todo $t \geq 0$ e tem incrementos independentes (i.e. ortogonais pela isometria). Novamente, trajetórias contínuas são obtidas aplicando o critério de Kolmogorov. Em geral, para $f \in L^2(\mathbf{R}_+)$, sua imagem $X(f)$ corresponde à integral de Wiener:

$$X(f) = \int_0^\infty f(s) dW_s,$$

como no Capítulo 3.

Finalmente, a terceira e última abordagem que gostaríamos de mencionar aqui, que é útil do ponto de vista computacional ou para análise de Fourier das trajetórias é a seguinte. Dada uma sequência $(g_n)_{n \geq 0}$ de gaussianas $N(0, 1)$ definidas no espaço de probabilidade Ω (ver Teorema 5.6), o processo $W : \mathbb{R}_{\geq 0} \times \Omega \rightarrow \mathbb{R}$ definido por

$$W_t = g_0 t + \sum_{n=1}^{\infty} \sum_{k=2^{n-1}}^{2^n-1} g_k \sqrt{2} \frac{\sin \pi k t}{k \pi}$$

é um movimento browniano. Essa abordagem foi idealizada inicialmente por Wiener. Atualmente, esse é um caso particular do chamado “modelo numérico” de Malliavin, ver por exemplo [39, I.4.3 e Prop.(X) 1.3.1]. Outras propriedades dessa decomposição trigonométrica das

trajetórias podem ser vistas em de Andrade e Ruffino [2]. Na fórmula acima, é sabido que a convergência é uniforme (no parâmetro n) para as trajetórias do movimento browniano, veja também Kac [25, Sec. 2].

5.5 Teorema ergódico para processos de Markov

Teorema 5.7 (Teorema ergódico de Birkhoff). *Seja $\theta : \Omega \rightarrow \Omega$ uma transformação no espaço de medida $(\Omega, \mathcal{F}, \mathbf{P})$ que preserva a medida. Dado $f \in L^1(\Omega)$ então a média temporal ao longo da órbita de θ existe \mathbf{P} -q.s. e é dada pela média espacial:*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n f(\theta^{i-1}\omega) = \mathbf{E}[f|\mathcal{G}],$$

onde $\mathcal{G} \subseteq \mathcal{F}$ é a sub- σ -álgebra gerada pelos conjuntos invariantes de θ

Uma transformação θ é dita *ergódica* em relação à medida \mathbf{P} se todos os conjuntos invariantes tiverem medidas zero ou um. Assim, se θ for ergódica (como no caso i.i.d. onde a transformação ergódica é o *shift* no espaço de probabilidade produto), então a esperança condicional é a própria esperança $\mathbf{E}[f]$. Demonstrações desse teorema podem ser encontradas em todos os livros introdutórios em teoria ergódica, por exemplo Halmos [20], Billingsley [4], Walters [59], Mañé [40], Pollicot [46] e vários outros.

Esse teorema vale em um contexto ainda mais geral, quando temos uma medida ergódica de um processo de Markov.

Teorema 5.8 (Teorema ergódico para processos de Markov). *Seja X_t um processo em M inicializado em x , de Markov, com uma medida de probabilidade invariante ergódica μ . Dada uma função integrável $f : M \rightarrow \mathbf{R}$ então a média temporal da f ao longo do processo é igual a μ -média da f em M . Isto é*

$$\lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t f(X_s) ds = \int_M f(x) d\mu(x) \quad \mathbf{P} \times \mu\text{-q.s.}$$

Bibliografia

- [1] L. Arnold - *Random Dynamical Systems*. Springer-Verlag, 1998.
- [2] A. de Andrade and Ruffino, P. R. C. - Wiener Integral in the space of sequences of real numbers. *Archivum Mathematicum (BRNO)* Vol. 36, no. 2 (2000) pp. 95-101 .
- [3] R. G. Bartle - *The elements of integration and Lebesgue measure*. John Wiley & Sons, Inc. (1966). Reprinted in 1995.
- [4] P. Billingsley – *Ergodic theory and information*. John Wiley and Sons, Inc., New York 1965.
- [5] A. Carverhill – Flows of stochastic dynamical systems: ergodic theory. *Stochastics*, **14**, 273–317, 1985.
- [6] P. J. Catuogno and Ruffino, P. R. C. – Product of harmonic maps is harmonic: an stochastic approach. *Seminaire de Probabilité XL*. Lect. Notes on Mathematics 1899, Springer-Verlag, p. 227-234, 2007.
- [7] P. J. Catuogno - *Notas de aula*. Departamento de Matemática, UNICAMP, 2008.
- [8] E. A. Coayla-Teran, Mohammed, S.-E. A. and Ruffino, P.R.C. – Hartman-Grobman theorems along hyperbolic stationary trajectories. *Discrete and Continuous Dyn. Systems (série A)* Vol. 17 (2), p. 281-292, 2007.

- [9] E. A. Coayla-Teran and Ruffino, P. R. C. – Stochastic versions of Hartman-Grobman theorems. *Stochastics and Dynamics*, Vol. 4, no.4, 571-594 (2004).
- [10] J. L. Doob – *Stochastic processes*. John Wiley, 1953.
- [11] J. L. Doob – What is a martingale? *The American Mathematical Monthly*, Vol. 78, No. 5 (May, 1971), pp. 451-463.
- [12] D.Elworthy - *Geometric aspects of difusion on manifolds. École d'été de probabilités de Saint-Flour, XV-XVII—1985-1987*, 277–426. Lect. Notes on Maths. 1362. Springer-Verlag (1988).
- [13] D. Elworthy - *Stochastic calculus on Manifolds*. London Mathematical Society. Lecture Notes Series 70. Cambridge University Press, 1983.
- [14] M. Emery - *Stochastic calculus in manifolds*. Universitext, Springer Verlag. 1990.
- [15] G. B. Folland – *Real analysis. Modern techniques and their applications*. John Wiley & Sons, Inc., New York, 1984.
- [16] H Föllmer – Calcul d'Itô sans probabilité. *Seminaire de Probabilité XV*, Springer Lect. Notes in Maths 850, (1981) 143–150. Reeditado em inglês: *Itô calculus without probability*, em apêndice de Sondermann [56].
- [17] H. Furstenberg – A Poisson formula for semi-simple Lie groups, *Ann. of Math.* (2) 77 (1963), 335–386.
- [18] Goldberg, S. I. and C. Mueller - Brownian Motion, Geometry, and Generalizations of Picard's Little Theorem. *Ann. Probab.* Vol. 11(4) (1983), 833-846.
- [19] P. R. Halmos – *Measure Theory*. Springer Verlag, 1974.
- [20] P. R. Halmos – *Lectures on Ergodic Theory*. Chelsea, 1956.
- [21] T. Hida – *Brownian Motion*. Springer Verlag 1980.

- [22] E. P. Hsu – *Stochastic analysis on manifolds*. Graduate Studies in Mathematics, 38. American Mathematical Society, Providence, RI, 2002.
- [23] N. Ikeda e Watanabe, S. - *Stochastic Differential Equations and Diffusion processes*. North-Holland, 2^a ed., 1989.
- [24] P. Imkeller and Lederer, C – The cohomology of stochastic and random differential equations, and local linearization of stochastic flows, *Stochastics and Dynamics*, **2**(2) (2002) 131-159.
- [25] M. Kac – *Integration in function space and some of its applications*. Accad. Nazionale Lincei Scuola Normale Sup. Pisa, 1980.
- [26] I. Karatzas – *Lectures on the Mathematics of Finance*. CRM Monographs Series vol.8. 1996.
- [27] I. Karatzas and Shreve, S. E. – *Brownian motion and stochastic calculus*. Springer-Verlag, NY (1988).
- [28] I. Karatzas and Shreve, S. E. – *Methods of Mathematical Finance*. Springer-Verlag, NY, second printing 1999.
- [29] P. E. Kloeden, Platen, E. – *Numerical solution of stochastic differential equations*, Berlin, New York: Springer-Verlag, (1992).
- [30] H. Kunita – Stochastic differential equations and stochastic flows of diffeomorphisms. *École d'été de probabilités de Saint-Flour*, XII—1982, 143–303, Lecture Notes in Math., 1097, Springer, Berlin, 1984.
- [31] H. Kunita – Stochastic flows and stochastic differential equations. (Reprint of the 1990 original) Cambridge Studies in Advanced Mathematics, 24. Cambridge University Press, 1997.
- [32] J.A Lázaro-Camí e Ortega, J.-P. Stochastic Hamiltonian dynamical systems. *Rep. Math. Phys.* **61** (1), 65-122 (2008).
- [33] S. Ledesma – Difusões em Folheações. Tese de doutorado. IMECC- UNICAMP, 2009.

- [34] X.-M. Li - An averaging principle for a completely integrable stochastic Hamiltonian system. *Nonlinearity* **21** (2008) 803?–822.
- [35] M. Liao – Decomposition of stochastic flows and Lyapunov exponents, *Probab. Theory Rel. Fields* **117** (2000) 589-607.
- [36] R. S. Liptser – A strong law of large numbers for local martingales. *Stochastics*, vol.**3**, pp. 217 – 228, 1980.
- [37] J. N. Lopez, Ruffino, P. R. C and San Martin, L. A. B. – Poisson spaces for proper semigroups of semi-simple Lie groups. *Stochastics and Dynamics*, **7** (2007), 273–298.
- [38] T. Lyons and Qian, Z. – *System control and rough paths*. Oxford Mathematical Monographs. Oxford University Press, Oxford, 2002.
- [39] P. Malliavin – *Stochastic analysis*. Springer-Verlag, 1997.
- [40] R. Mañe – *Teoria Ergódica*. Projetos Euclides, Publicação IMPA, 1983.
- [41] E. Nelson – *Dynamical Theories of Brownian Motion*. (First edition) Princeton University Press (1967). Second edition, August 2001. Posted on the Web at <http://www.math.princeton.edu/~nelson/books.html>
- [42] D. Nualart – *The Malliavin calculus and related topics*. Springer-Verlag, 1995.
- [43] B. Oksendal - *Stochastic Differential Equations, an Introduction with Applications*. 6th edition. Universitext, Springer-Verlag, 2003.
- [44] W. Parry – *Topics in Ergodic Theory*. Cambridge Tracts in Mathematics. Cambridge, UK, 1981.
- [45] M. A. Pinsky - Can you feel the shape of a manifold with Brownian motion? *Topics in Contemporary Probability and Its Applications*, 1995

- [46] M. Pollicot and Yuri. M. – *Dynamical Systems and Ergodic Theory*. London Mathematical Society, 1998.
- [47] D. Revuz e Yor, M. - *Continuous Martingale and Brownian Motion*. Springer-Verlag 1991.
- [48] H. L. Royden – *Real analysis*. Third edition. Macmillan Publishing Company, New York, 1988.
- [49] W. Rudin – *Principles of Mathematical Analysis*, McGraw-Hill, 1989.
- [50] D. Ruelle – Ergodic theory of differentiable dynamical systems, Publications of I.H.E.S.. Bures-sur-Yvette, France, 1979.
- [51] P.R.C. Ruffino – Rotation numbers for stochastic dynamical systems. *Stochastics and Stochastics Reports*, vol. **60**, pp. 289-318, 1997.
- [52] P. R. C. Ruffino – A sampling theorem for rotation numbers of linear processes in R^2 . *Random Operators and Stochastic Equations*, **8** (2), 2000.
- [53] P. R. C. Ruffino – Decomposition of Stochastic Flows and Rotation Matrix, *Stochastics and Dynamics*, **2**(1), (2002) 93-107.
- [54] L.A.B. San Martin e Marques, M.S.F. - *Cálculo Estocástico*. Monografia do 18º Colóquio Brasileiro de Matemática, 1991.
- [55] F. B. da Silva – Decomposições de fluxos estocásticos. Tese de Doutorado, IMECC-UNICAMP, 2009.
- [56] D. Sondermann – *Introduction to Stochastic Calculus for Finance: A New Didactic Approach*. Springer-Verlag, Berlin, Heidelberg, 2006.
- [57] G. E. Uhlenbeck e L. S. Ornstein – On the theory of Brownian motion, *Physical Review* **36**(1930), 823–841. Republicado por Nelson Wax, “Selected Papers on Noise and Stochastic Processes”, Dover, 1954.
- [58] K. Twardowska – Wong-Zakai approximation for stochastic differential equations. *Acta Applic Math.* (**43**), (1996), 317-359.

- [59] P. Walters – *An introduction to ergodic theory*. Springer-Verlag, New York, 1982.
- [60] E. Wong and M. Zakai – Riemann-Stieltjes approximations of stochastic integrals. *Z. Wahr. verw.Geb.* **12** (1969), 87–97.
- [61] K. Yosida – *Functional Analysis*. (6th edition). Grundlehren der Math. Wissenschaften 123. Berlin, Heidelberg, New York: Springer-Verlag, 1968.
- [62] K. Yosida – The Markoff Process with a stable distribution. *Proc. Imp. Acad. Tokyo*, **16** (1940), 43–54.

Índice

- σ -álgebra, 4
 - completamento de, 7
 - discreta, 7
 - dos borelianos, 4
 - gerada, 4
 - produto, 24
 - trivial, 7
- Álgebra, 4
- Aproximação poligonal do MB, 103
- Cadeia de Markov, 42
- Campos de difusão, 81
- Caracterização de Levy, 49
- Cilindro de dimensão finita, 24, 38
- Cociclo, 105
- Coefficientes de difusão, 81
- Cohomologia de cociclo, 108
- Condição
 - de Chapman-Kolmogorov, 45
- Condição de compatibilidade, 45
- Conjunto
 - Lebesgue mensurável, 7
 - mensurável, 5
- Continuidade absoluta, 18
- Convergência
 - em L^p , 29
 - em distribuição, 30
 - em lei, 30
 - em probabilidade, 29
 - fraca de medidas, 30
 - q.s., 29
- Convergência uniforme em probabilidade nos compactos (ucp), 103
- Conversão Itô-Stratonovich, 96, 97
- Convolução de medidas, 46, 113
- Covariância, 13
- Decomposição em caos de Wiener, 118
- Desigualdade
 - L^p de Doob, 56
 - de Hölder, 16
 - de Jensen, 22
 - de Minkowsky, 16
- Desvio padrão, 14
- Distribuição, 8
 - binomial, 9
 - de Poisson, 8
 - gaussiana, 9
 - normal, 9
- Drift de uma EDE, 81
- EDE linear, 86
- Equação reversa de Kolmogorov, 91

- Equacao de Langevin, 120
- Espaço
- de Banach, 17
 - de Hilbert, 17
 - de probabilidade, 5
 - finito, 7
 - de Wiener, 51
 - mensurável, 5
 - polonês, 6
 - produto, 24
- Espaço das gaussianas em L^2 , 35
- Esperança, 9
 - condicional, 20
- Eventos, 5
- Expoentes de Lyapunov, 107
- Exponencial estocástica, 76
- Fórmula de Dynkin, 91
- Fórmula de Itô, 74
 - para integral de Stratonovich, 99
- Filtração
 - contínua à direita, 41
- Filtração em σ -álgebras, 40
- Fronteira de Poisson, 95
- Função
 - característica, 10
 - convexa, 22
 - de Dirichlet, 17
 - harmônica, 80
 - indicadora, 10
 - mensurável, 6
 - simples, 10
- Função de Controle ou controlador, 104
- Função geradora de momento, 51
- Gerador infinitesimal, 89
- Grupos de Lie, 95
- Independência, 26
- Independência mútua, 26
- Integrabilidade uniforme, 57
- Integral
 - de Itô, 69
 - de Wiener, 52, 122
- Integral de Lebesgue, 10
- Integral estocástica
 - de Itô (discreta), 65
 - de Wiener, 69
 - de Itô, 72
- Integrando, integrador, 66
- Isometria de Itô, 66
- Laplaciano, 91
- Lema
 - de Borel-Cantelli, 28
 - de Fatou, 11
- Lim inf de conjuntos, 27
- Lim sup de conjuntos, 27
- Média, 9
- Método Monte Carlo, 95
- Martingale, 53
 - regular, 54
 - sub-, 54
 - super-, 54
- Medida, 5
 - σ -aditiva, 5
 - σ -finita, 7
 - com sinal, 5
 - da contagem, 25
 - de Dirac, 8
 - de Lebesgue, 7
 - de probabilidade, 5
 - de Radon, 6

- de Wiener, 51
- ergódica, 46
- exterior, 7
- finita, 5
- gaussiana, 31
- induzida, 8
- invariante, 44
- produto, 24
- regular exterior, 6
- regular interior, 5
- Medida invariante, 46
- Momento centrado, 13
- Movimento browniano, 49
 - aproximação discreta, 55
 - geométrico, 76
 - gerador infinitesimal do, 91
 - recorrência em \mathbf{R} e \mathbf{R}^2 , 93
 - transiência em \mathbf{R}^n , $n \geq 3$, 94
- Movimento browniano linear, 49
- Operador de Laplace-Beltrami, 91
- Operador positivo, 22
- Princípio de não-arbitragem, 53
- Probabilidade, 5
 - condicional, 22
 - de transição, 45
- Problema de Dirichlet, 94
- Processo
 - de Ornstein-Uhlenbeck, 121
- Processo de difusão, 88
- Processo de Markov, 42
 - homogêneo, 43, 45
 - medida estacionária, 45
- Processo estocástico, 37
 - adaptado, 41
 - distribuição de dimensão finita, 39
- elementar, 68
- estacionário, 39
- gaussiano, 39
- indistinguibilidade de, 40
- Itô integrável, 71
- modificações de, 39
- parado X^T , 63
- previzível, 56
- trajetórias de um, 38
- versões de, 39
- Representação canônica, 38
- Ruído branco, 52
- Semigrupo de Feller, 89
- Semimartingales, 64
- Shift, translação no tempo, 105
- Sistema de controle afim, 103
- Sistema estocástico hiperbólico, 108
- Solução forte e fraca, 87
- Subespaços de Oseledets, 107
- Subvariedade estável, 107
- Tempo de entrada, 41
- Tempo de explosão, 85
- Tempo de parada, 41
 - σ -álgebra \mathcal{F}_T associada, 67
- Teorema
 - da convergência dominada - de Lebesgue, 11
 - da convergência monótona - de Beppo-Levy, 11
 - da extensão de Carathéodory-Hahn, 7
 - da medida induzida, 12
 - da representação de Riesz, 15
 - de aproximação de Wong-Zakai, 103

- de extensão de Kolmogorov, 48
- de Hartman-Grobman estocásticos, 108
- de Krylov-Boguliubov, 46
- de Perron-Frobenius, 44
- de Radon-Nikodym, 18
- do limite central, 119
- do suporte, 104
- ergódico de Birkhoff, 123
- ergódico para processos de Markov, 123
- existência e unicidade de solução de EDE, 82
- ponto fixo de Banach, 82
- Transformada de Fourier de medida, 31
- Transformação ergódica, 123
- Transformada de Laplace de variável aleatória, 51
- Valor esperado, 9
- Variável aleatória, 6
 - p -momento, 13
 - gaussianas, 30
 - distribuição de, 8
 - lei de, 8
 - uniformemente distribuída, 8
- Variância, 13
- Varição
 - p -, 59
 - finita ou limitada, 57
 - quadrática, 60
 - quadrática cruzada, 62
- Variiedade de Poisson, 111