

ÁLVARO RAPOSO GONÇALVES DE MELO JÚNIOR

Aplicação do Método de Euler-Maruyama na Resolução de Equações Diferenciais  
Estocásticas

São Paulo  
2013

ÁLVARO RAPOSO GONÇALVES DE MELO JÚNIOR

Aplicação do Método de Euler-Maruyama na Resolução de Equações Diferenciais  
Estocásticas

Dissertação apresentada à Escola  
Politécnica da Universidade de São  
Paulo para obtenção do título MBA-USP  
em Especialização em Engenharia  
Financeira

Área de Concentração: Especialização  
em Engenharia Financeira

Orientador: Prof. Dr. Nelson M. Kuhl

São Paulo  
2013

MBA/EF  
2013  
m/192 ap

M2013I

Esc Politécnica-Bib Eng Eletr



M2013BA

1. Equações diferenciais estocásticas.
2. Convergência para

M2013BA

[2511765]

## RESUMO

O principal objeto de estudo deste trabalho é um método numérico para resolução de equações diferenciais estocásticas (EDE), implementado a partir do modelo proposto por Euler-Maruyama. Para atestar sua validade, o que utilizamos como exemplo foi uma EDE cuja solução analítica já é conhecida. Foi também investigado neste trabalho a ordem de convergência forte e fraca deste método. Este método se mostrará de fácil implementação, apresentará rápida execução e bom nível de convergência.

Palavras-chave: Equações diferenciais estocásticas. Euler-Maruyama. Convergência forte. Convergência fraca.

## ABSTRACT

The main object of this work is a numerical method for solving stochastic differential equations (SDE), implemented in the model proposed by Euler-Maruyama. To prove its validity, what was used as an example was an SDE whose analytical solution is already known. It was also investigated in this work the order of weak and strong convergence of this method. This method will prove to be easy to implement, present fast execution and good level of convergence.

Keywords: Stochastic differential equations. Euler-Maruyama. Strong convergence. Weak convergence.

# Índice

Introdução.....	2
1. Conceitos Básicos de Probabilidade.....	3
1.1 Variável Aleatória.....	3
1.1.1 Probabilidade.....	3
1.1.2 Esperança, Variância e Momento.....	5
1.2 Vetores Aleatórios.....	7
1.2.1 Probabilidade, Distribuição e Função de Distribuição.....	7
1.2.2 Esperança, Variância e Covariância.....	9
1.2.3 Dependência e independência.....	10
1.3 Processos Estocásticos.....	11
1.3.1 Distribuição.....	14
1.3.2 Função de esperança e covariância.....	15
1.3.3 Estrutura de dependência.....	16
1.4 Movimento Browniano.....	17
1.4.1 Funções de distribuição, esperança e covariância.....	17
1.4.2 Propriedades do caminho.....	17
1.5 Esperança Condicional.....	21
1.6 $\sigma$ - álgebra.....	25
1.6.1 Esperança condicional geral.....	28
1.6.2 Regras Para o Cálculo de Esperança Condicional.....	30
1.6.3 Projeção da Esperança Condicional.....	34
1.7 Martingais.....	36
1.7.1 Interpretação como jogo justo.....	41
2. A Integral Estocástica e Simulação Numérica.....	43
2.1 A Integral de Riemann.....	43
2.2 Integral estocástica.....	44
2.3 O Método de Euler-Maruyama.....	47
2.4 Convergência forte e fraca do método de EM.....	50
2.5 O método de Milstein.....	53
2.6 Convergência forte do método de Milstein.....	55
Conclusão.....	57
Algoritmos.....	58
Bibliografia.....	64

## Introdução

Este trabalho tem como objetivo chegar a um método numérico simples para resolução de equações diferenciais estocásticas. É também objetivo deste trabalho verificar sua validade, e para isto são aplicados alguns testes de convergência sobre as soluções numéricas apresentadas. Aqui veremos também a teoria necessária para podermos abordar este assunto. A forma como abordaremos os tópicos apresentados aqui será com o mínimo de rigor matemático o possível, de forma que não fique uma leitura pesada.

Na primeira parte deste trabalho serão apresentados conceitos básicos sobre probabilidade, veremos variáveis aleatórias, como calcular suas variâncias, esperanças e momentos. A partir daí, seguiremos para vetores aleatórios. Depois vêm processos estocásticos, movimento browniano e esperança condicional. Neste ponto, os assuntos começam a ficar mais abstratos. Veremos aqui o conceito de  $\sigma$  – álgebra, e o de martingais. Este último é essencial para o entendimento da integral de Ito. Veremos ainda que a ideia por trás de um martingal é a de um jogo justo.

Na segunda parte do trabalho começaremos falando de integral de estocástica, apresentaremos as diferenças entre as integrais de Ito e a de Stratonovich, mostrando um exemplo para ilustrar essa diferença. Neste ponto começa o principal tópico deste trabalho. É neste ponto que podemos finalmente apresentar o método de Euler-Maruyama. Neste ponto aplicaremos este método em uma equação diferencial estocástica que é muito utilizada em modelos de precificação de ativos financeiros. A equação utilizada possui uma solução analítica. Desta forma será possível comparar os resultados do método com os da solução analítica. Por fim, falaremos um pouco sobre a convergência forte e fraca deste método.

# 1. Conceitos Básicos de Probabilidade

## 1.1 Variável Aleatória

Antes de definir o que é uma variável aleatória, devemos primeiramente definir o que é um espaço amostral.

O resultado de um jogo ou um experimento é aleatório. Por exemplo, quando lançamos uma moeda, o resultado deste lançamento pode ser cara ou coroa. Antes de lançarmos a moeda, sabemos os resultados possíveis, mas não sabemos qual é a saída que teremos antes de lançar a moeda.

O conjunto de todas as possíveis saídas para um determinado experimento é chamado de espaço amostral. Este conjunto será representado durante todo o trabalho pela letra  $\Omega$ .

Portanto, no exemplo do lançamento da moeda, temos:

$$\Omega = \{Cara, Coroa\}$$

Agora podemos definir variável aleatória.

Uma variável aleatória é uma função  $X$ , com domínio em  $\Omega$ , e contradomínio em  $\mathbb{R}$ . Ou seja, se tivermos:

$$\begin{aligned} X : \Omega &\rightarrow \mathbb{R} \\ \omega &\rightarrow X(\omega) \end{aligned}$$

Então  $X$  é uma variável aleatória.

Voltando novamente ao exemplo da moeda, podemos definir a seguinte variável aleatória:

$$\begin{aligned} X(Cara) &= 0 \\ X(Coroa) &= 1 \end{aligned}$$

### 1.1.1 Probabilidade

O próximo passo neste processo de abstração matemática é o de criar uma descrição probabilística para a variável aleatória  $X$ . A probabilidade mede a chance de um determinado evento ocorrer.

Ainda no exemplo da moeda, se a moeda sendo lançada for honesta, então devemos ter:

$$P(\{\omega : X(\omega) = 0\}) = P(\{\omega : X(\omega) = 1\}) = 0,5$$

Esta definição matemática é baseada em evidência empírica. Se lançarmos uma moeda honesta uma quantidade grande de vezes, iremos obter o resultado cara cerca de 50% das



vezes e coroa cerca de 50% das vezes também. Em teoria de probabilidade, a “Lei dos Grandes Números” dá uma justificativa teórica para este resultado.

Em geral, dados  $A, B \in \Omega$  dois eventos quaisquer, temos:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Se A e B forem disjuntos,

$$P(A \cup B) = P(A) + P(B)$$

Além disso, ainda temos:

$$P(A^c) = 1 - P(A) \quad , \quad P(\Omega) = 1 \quad , \quad P(\emptyset) = 0$$

A função

$$F_X(x) = P(X \leq x) = P(\{\omega : X(\omega) \leq x\}) \quad , \quad x \in \mathbb{R}$$

é conhecida como sendo a função de distribuição  $F_X$  de X.

Não é difícil mostrar que:

$$P(X \in [a, b]) = F_X(b) - F_X(a) \quad , \quad x \in \mathbb{R}$$

A função de distribuição pode ser contínua ou pode conter saltos. No caso em que ela possui saltos, temos:

$$F_X(x) = \sum_{k: x_k \leq x} p_k \quad , \quad 0 \leq p_k \leq 1 \quad \forall k \quad \text{e} \quad \sum_{k=1}^s p_k = 1$$

Esta função de distribuição é dita discreta. A variável aleatória com a distribuição acima é uma variável aleatória discreta.

Uma variável aleatória discreta possui apenas uma quantidade finita ou um conjunto enumerável de valores em seu espaço amostral. A sua função de distribuição possui saltos do tamanho  $p_k$  em  $x = x_k$ .

Um exemplo de distribuição discreta importante é a *distribuição binomial*,  $\text{Bin}(n, p)$ , com parâmetros  $n \in \mathbb{N}$  e  $p \in (0, 1)$  :

$$P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k=0,1,\dots,n$$

Outro exemplo de distribuição discreta importante é a distribuição de Poisson,  $Poi(\lambda)$ , com parâmetro  $\lambda > 0$ :

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

Ao contrário da função de distribuição discreta, a função de distribuição contínua não possui pulos, portanto  $P(X=x) = 0$  para todo  $x$ , ou seja, uma variável aleatória contínua assume um determinado valor com probabilidade zero.

Muitas distribuições contínuas de interesse possuem *densidade*  $f_X$ :

$$F_X(x) = \int_{-\infty}^x f_X(y) dy, \quad x \in \mathbb{R}, \text{ onde}$$

$$f_X(x) \geq 0 \text{ para todo } x \in \mathbb{R} \text{ e } \int_{-\infty}^{\infty} f_X(y) dy = 1.$$

Um exemplo importante de distribuição contínua é a *distribuição normal* ou *Gaussiana*,  $N(\mu, \sigma^2)$  com parâmetros  $\mu \in \mathbb{R}, \sigma^2 > 0$ . Sua densidade é dada por:

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

### 1.1.2 Esperança, Variância e Momento

Três características interessantes de uma variável aleatória são a *esperança*, a *variância* e o *momento*. Para uma variável aleatória contínua  $X$ , sua esperança  $E(X)$  é dada por:

$$\mu_X = E(X) = \int_{-\infty}^{\infty} x f_X(x) dx$$

Sua variância é definida como sendo:

$$\sigma_X^2 = var(X) = \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x) dx.$$

Seu  $n$ -ésimo momento,  $n \in \mathbb{N}$ :

$$E(X^n) = \int_{-\infty}^{\infty} x^n f_X(x) dx \quad .$$

Dada uma função  $g$  a valores reais, a esperança de  $g(X)$ , é dada por:

$$E_g(X) = \int_{-\infty}^{\infty} g(x) f_X(x) dx$$

Para uma variável aleatória discreta  $X$ , com probabilidade  $p_k = P(X = x_k)$ , sua esperança  $E(X)$  é dada por:

$$\mu_X = E(X) = \sum_{k=1}^{\infty} x_k p_k \quad .$$

Sua variância é definida como sendo:

$$\sigma_X^2 = \text{var}(X) = \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x) dx \quad .$$

Seu  $n$ -ésimo momento,  $n \in \mathbb{N}$  :

$$E(X^n) = \sum_{k=1}^{\infty} x_k^n p_k \quad .$$

Dada uma função  $g$  a valores reais, a esperança de  $g(X)$ , é dada por:

$$E_g(X) = \sum_{k=1}^{\infty} g(x_k) p_k \quad .$$

Podemos nos referir à esperança de uma variável aleatória como sendo o seu “centro de gravidade”, ou seja, valores aleatórios  $X(\omega)$  estão espalhados em torno de um valor não aleatório  $\mu_X$ . Este valor é o valor que se espera da variável aleatória.

O espalhamento ou dispersão em torno de  $\mu_X$  é descrito como sendo a variância de  $X(\omega)$ . Temos:

$$\begin{aligned} \sigma_X^2 &= \text{var}(X) = E((X - \mu_X)^2) \\ &= E(X^2 - 2\mu_X X + \mu_X^2) \\ &= E(X^2) - \mu_X^2 \end{aligned}$$

Um resultado interessante e conhecido é que para uma variável aleatória normal, temos:

$$P(\mu - 1,96\sigma < X < \mu + 1,96\sigma) = F_X(\mu + 1,96\sigma) - F_X(\mu - 1,96\sigma) = 0,95 \quad .$$

Portanto, com 95% de chance,  $X$  assumirá valores no intervalo  $[\mu - 1,96\sigma, \mu + 1,96\sigma]$  .

## 1.2 Vetores Aleatórios

Ao longo deste trabalho, faremos uso de estruturas de dimensão finita, ou até mesmo de dimensão infinita. Começaremos com dimensão finita, como um primeiro passo em direção à definição de processos estocásticos.

*Definição:* Dizemos que  $\mathbf{X} = (X_1, \dots, X_n)$  é um vetor aleatório de dimensão  $n \in \mathbb{N}$  se seus componentes forem variáveis aleatórias reais.

Se interpretarmos  $t = 1, \dots, n$  como instantes de tempo,  $X_t$  pode ser o resultado de um experimento no instante  $t$ . Esta sequência é uma série temporal, e ela pode representar, por exemplo, o preço de fechamento de uma determinada ação em  $n$  dias sucessivos.

De maneira análoga à variável aleatória, é possível introduzir os conceitos de função de distribuição, esperança, momento e matriz de covariância de um vetor aleatório, como uma tentativa de se descrever a sua distribuição e a sua estrutura de dependências. O único conceito novo até aqui é o de matriz de covariância, afinal, não faz sentido falarmos em covariância para uma variável aleatória.

### 1.2.1 Probabilidade, Distribuição e Função de Distribuição

Considere o experimento de se lançar uma moeda honesta duas vezes. Os possíveis resultados deste experimento seria (H, H), (H, T), (T, H) e (T, T), em que H representa cara e T representa coroa. Estes quatro pares de possíveis resultados formam o espaço amostral  $\Omega$ .

Se definirmos novamente a função como anteriormente, obtemos duas variáveis aleatórias,  $X_1$  e  $X_2$ , e  $\mathbf{X} = (X_1, X_2)$  um vetor aleatório bidimensional.

Note que:

$$\mathbf{X}(H, H) = (0, 0) \quad , \quad \mathbf{X}(T, H) = (1, 0) \quad , \quad \mathbf{X}(H, T) = (0, 1) \quad \text{e} \quad \mathbf{X}(T, T) = (1, 1) \quad .$$

Como a moeda é honesta, podemos então dizer que a probabilidade de cada um desses eventos ocorrer é de 0,25, ou seja:

$$P(\{\omega : \mathbf{X}(\omega) = (k, l)\}) = 0,25, \quad k, l \in \{0, 1\}$$

A coleção de probabilidades

$$F_X(\mathbf{x}) = P(X_1 \leq x_1, \dots, X_n \leq x_n) = P(\{\omega : X_1(\omega) \leq x_1, \dots, X_n(\omega) \leq x_n\}), \quad \mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$$

é a função de distribuição de  $F_X$  de  $X$ .

Com esta função é possível obter a probabilidade de um determinado evento  $X$  assumir valores em um determinado retângulo

$$(\mathbf{a}, \mathbf{b}) = \{\mathbf{x} : a_i < x_i \leq b_i, i=1, \dots, n\}$$

Por exemplo, se  $X$  é bidimensional,

$$P(X \in (\mathbf{a}, \mathbf{b})) = F_X(b_1, b_2) + F_X(a_1, a_2) - F_X(a_1, b_2) - F_X(b_1, a_2)$$

*Demonstração:*

$$F_X(b_1, b_2) = P(X_1 \leq b_1, X_2 \leq b_2)$$

$$F_X(b_1, b_2) = P[\{a_1 < X_1 \leq b_1, a_2 < X_2 \leq b_2\} \cup \{a_1 < X_1 \leq b_1, X_2 \leq a_2\} \cup \{X_1 \leq a_1, a_2 < X_2 \leq b_2\} \cup \{X_1 \leq a_1, X_2 \leq a_2\}]$$

$$F_X(b_1, b_2) = P[a_1 < X_1 \leq b_1, a_2 < X_2 \leq b_2] + P[a_1 < X_1 \leq b_1, X_2 \leq a_2] + P[X_1 \leq a_1, a_2 < X_2 \leq b_2] + P[X_1 \leq a_1, X_2 \leq a_2]$$

Note que:

$$P[X_1 \leq a_1, a_2 < X_2 \leq b_2] + P[X_1 \leq a_1, X_2 \leq a_2] = P[X_1 \leq a_1, X_2 \leq b_2]$$

E ainda:

$$P[a_1 < X_1 \leq b_1, X_2 \leq a_2] = P[X_1 \leq b_1, X_2 \leq a_2] - P[X_1 \leq a_1, X_2 \leq a_2]$$

Portanto:

$$F_X(b_1, b_2) = P[X \in (\mathbf{a}, \mathbf{b})] + F_X(b_1, a_2) - F_X(a_1, a_2) + F_X(a_1, b_2)$$

$$P[X \in (\mathbf{a}, \mathbf{b})] = F_X(b_1, b_2) + F_X(a_1, a_2) - F_X(b_1, a_2) - F_X(a_1, b_2)$$

Isto conclui a demonstração.

A coleção de probabilidades

$$P_X(B) = P(X \in B) = P(\{\omega : X(\omega) \in B\})$$

para subconjuntos  $B \subset \mathbb{R}^n$  “adequados”, constitui a distribuição de  $X$ . Por subconjuntos adequados de  $\mathbb{R}^n$ , entendam-se subconjuntos de Borel. Estes, por sua vez, são subconjuntos que podem ser obtidos através de uma quantidade enumerável de operações de  $\cap$ ,  $\cup$  ou  $^c$ . Mais adiante veremos uma definição mais precisa.

Matematicamente falando, a distribuição e a função de distribuição são noções equivalentes, pois com ambas é possível calcular a probabilidade de ocorrer um evento  $X \in B$ . Além do mais, conhecendo a distribuição de  $X$ , conhecemos também a distribuição das componentes  $X_i$ . Para isto, basta fazer  $x_j = \infty, j \neq i$ , na expressão da função de distribuição.

Analogamente às variáveis aleatórias, é possível definir vetores aleatórios discretos e contínuos, mas aqui iremos focar apenas no caso contínuo e suas distribuições.

Se a distribuição de um vetor aleatório  $X$  possui densidade  $f_X$ , então sua função de distribuição é dada por

$$F_X(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} f_X(y_1, \dots, y_n) dy_1 \dots dy_n, (x_1, \dots, x_n) \in \mathbb{R}^n$$

em que a densidade, é uma função que satisfaz:

- $f_X(x) \geq 0 \forall x \in \mathbb{R}^n$  ;
- $\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f_X(y_1, \dots, y_n) dy_1 \dots dy_n = 1$  .

Se  $X$  possui densidade  $f_X$ , então todos os seus componentes  $X_i$ , seus pares  $(X_i, X_j)$ , suas triplas  $(X_i, X_j, X_k)$ , etc., possuem uma densidade. Elas são chamadas densidades marginais.

## 1.2.2 Esperança, Variância e Covariância

Da mesma forma que fizemos com variáveis aleatórias, podemos também definir esperança e variância para vetores aleatórios, e também podemos definir a matriz de covariância, que só faz sentido em ser falada no caso de vetores aleatórios.

A esperança de um vetor aleatório é dada por:

$$\mu_X = E(X) = (E(X_1), \dots, E(X_n))$$

Sua matriz de covariância é definida como sendo:

$$\Sigma_X = (\text{cov}(X_i, X_j); i, j = 1, \dots, n) \quad , \text{ onde}$$

$$\text{cov}(X_i, X_j) = E[(X_i - \mu_{X_i})(X_j - \mu_{X_j})] = E(X_i X_j) - \mu_{X_i} \mu_{X_j} \quad ,$$

é a covariância entre  $X_i$  e  $X_j$ .

Note que  $\text{cov}(X_i, X_i) = \sigma_{X_i}^2$ .

### 1.2.3 Dependência e independência

Intuitivamente falando, dois eventos são ditos independentes quando a probabilidade de que um deles aconteça não é alterada quando o outro ocorre ou deixa de acontecer, ou seja, não importa o que aconteça com o segundo evento, isso não afeta a probabilidade de o primeiro evento ocorrer ou não. Um exemplo muito comum desse fato é o do lançamento de duas moedas honestas. Se na primeira moeda o resultado do lançamento for, digamos, cara, em nada isso afeta o resultado do segundo lançamento.

Matematicamente falando, temos:

$$P(X_1 = k, X_2 = l) = P(X_1 = k) P(X_2 = l), \quad k, l \in \{0, 1\}.$$

Uma definição mais formal:

*Definição:* Dois eventos A e B são ditos independentes se:

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

Agora, duas variáveis aleatórias  $X_1$  e  $X_2$  são ditas independentes se:

$$P(X_1 \in B_1, X_2 \in B_2) = P(X_1 \in B_1) P(X_2 \in B_2).$$

Isto significa que os eventos  $\{X_1 \in B_1\}$  e  $\{X_2 \in B_2\}$  são independentes.

Alternativamente, é possível se estender a definição de independência a partir da função de distribuição.

Duas variáveis aleatórias  $X_1$  e  $X_2$  são ditas independentes se e somente se

$$F_{X_1, X_2}(x_1, x_2) = F_{X_1}(x_1) F_{X_2}(x_2), \quad x_1, x_2 \in \mathbb{R}$$

Se  $(X_1, X_2)$  possuir densidade  $f_{X_1, X_2}$  e densidades marginais  $f_{X_1}$  e  $f_{X_2}$ , então  $X_1$  e  $X_2$  são ditas independentes se e somente se

$$f_{X_1, X_2}(x_1, x_2) = f_{X_1}(x_1) f_{X_2}(x_2), \quad x_1, x_2 \in \mathbb{R}$$

A noção de independência pode ser estendida ainda para qualquer número finito de eventos e variáveis aleatórias. Note que a independência dos componentes de um vetor aleatório implica na independência de qualquer par de componentes, no entanto a recíproca não é verdadeira.

Os eventos  $A_1, \dots, A_n$  são ditos independentes se para qualquer que seja a escolha de índices  $1 \leq i_1 \leq \dots \leq i_k \leq n$  e inteiros  $1 \leq k \leq n$ ,

$$P(A_{i_1} \cap \dots \cap A_{i_k}) = P(A_{i_1}) \cap \dots \cap P(A_{i_k}).$$

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

$$P(X_1 \in B_1, \dots, X_n \in B_n) = P(X_1 \in B_1) \dots P(X_n \in B_n)$$

Isto significa que os eventos  $\{X_1 \in B_1\}, \dots, \{X_n \in B_n\}$  são independentes.

As variáveis aleatórias  $X_1, \dots, X_n$  são ditas independentes se e somente se sua função de distribuição acumulada puder ser escrita da seguinte forma:

$$F_{X_1, \dots, X_n}(x_1, \dots, x_n) = F_{X_1}(x_1) \dots F_{X_n}(x_n), \quad (x_1, \dots, x_n) \in \mathbb{R}^n$$

Se  $(X_1, \dots, X_n)$  possuir densidade  $f_{X_1, \dots, X_n}$  e densidades marginais  $f_{X_1}, \dots, f_{X_n}$ , então  $X_1, \dots, X_n$  são independentes se e somente se

$$f_{X_1, \dots, X_n}(x_1, \dots, x_n) = f_{X_1}(x_1) \dots f_{X_n}(x_n), \quad (x_1, \dots, x_n) \in \mathbb{R}^n$$

Isto completa a serie de definições se dependência e independência de variáveis aleatórias.

Exemplos:

1. Seja  $X = (X_1, \dots, X_n)$  um vetor gaussiano com média  $\mu = (\mu_1, \dots, \mu_n)$  e matriz de covariância  $\Sigma$ . Então temos:

$$f_X(x) = \frac{1}{(2\pi)^{n/2} (\det(\Sigma))^{1/2}} \exp\left\{-\frac{1}{2}(x-\mu)\Sigma^{-1}(x-\mu)\right\}$$

No caso em que as variáveis são descorrelacionadas, a matriz de covariância é diagonal, e temos:

$$f_X(x) = \prod_{i=1}^n \frac{1}{(2\pi)^{1/2} \sigma_i} \exp\left\{-\frac{(x_i - \mu_i)^2}{2\sigma_i^2}\right\} = f_{X_1}(x_1) \dots f_{X_n}(x_n)$$

ou seja, são independentes. Portanto, no caso gaussiano, descorrelação e independência são noções equivalentes. O próximo exemplo mostra que não-correlação não implica independência.

2. Seja  $X$  uma variável aleatória normal padrão. Como  $X$  é simétrica, isto é,  $X$  e  $-X$  possuem a mesma distribuição, então  $X^3$  também é, e logo, ambas possuem média zero. Assim:

$$\text{cov}(X, X^2) = E[XX^2] - E[X]E[X^2] = E[X^3] - E[X]E[X^2] = 0$$

Mas claramente  $X$  e  $X^2$  são dependentes, pois o evento  $\{X \in (-1, 1)\} = \{X^2 \in (0, 1)\}$ . Portanto, temos:

$$P(\{X \in (-1, 1)\}, \{X^2 \in (0, 1)\}) = P(\{X \in (-1, 1)\}) > P(\{X \in (-1, 1)\})P(\{X^2 \in (0, 1)\}) = P(\{X \in (-1, 1)\})^2$$

### 1.3 Processos Estocásticos

Suponha que a taxa de câmbio R\$/US\$ em qualquer instante entre as 9:00 da manhã e as 10:00 da manhã seja aleatória. Portanto, podemos imaginar que o caminho percorrido pelo câmbio neste intervalo seja uma realização  $X_t(\omega)$  dessa variável aleatória  $X_t$ . Para fazermos uma previsão sobre o valor de  $X_t$  as 11 da manhã é no mínimo razoável que



observemos a trajetória de  $X_t$  para  $t \in [9,10]$ . Um modelo matemático que descreve tal fenômeno é chamado de processo estocástico.

**Definição:** Um processo estocástico  $X$  é uma coleção de variáveis aleatórias  $(X_t, t \in T, \omega \in \Omega)$ , definida em um espaço  $\Omega$ .

Para nossos propósitos,  $T$  é frequentemente um intervalo. Neste caso,  $X$  é um processo contínuo no tempo. Um processo que não é contínuo no tempo é um processo discreto, e neste caso  $T$  é um conjunto finito, ou infinito e enumerável. O índice  $t$  da variável aleatória  $X_t$  normalmente refere-se ao tempo, e seguiremos esta convenção aqui.

Um processo estocástico  $X$  é uma função de duas variáveis. Para um instante de tempo  $t$  fixo,  $X_t$  é uma variável aleatória,

$$X_t = X_t(\omega), \omega \in \Omega.$$

Para uma dada trajetória  $\omega \in \Omega$ , trata-se de uma função do tempo,

$$X_t = X_t(\omega), t \in T.$$

Esta função é chamada de realização, trajetória ou caminho amostral do processo  $X$ .

A figura 1 abaixo ilustra bem este fato. Na primeira figura, cada um dos caminhos observados representa um  $\omega \in \Omega$  diferente. Já nas figuras do meio e de baixo, para cada  $t=0,1,\dots,0,9$ ,  $X_t$  representa uma variável aleatória.

As figuras foram retiradas do livro do Mikosch, mencionado na referência no final deste trabalho. As imagens contidas neste trabalho são todas deste livro.

**Exemplo:** Uma série temporal  $X_t, t=0,\pm 1,\pm 2,\dots$  é um processo discreto, com  $T=\mathbb{Z}$ . As séries temporais representam uma classe muito importante dos processos estocásticos. Elas são modelos relevantes na representação de muitos fenômenos da vida real, como por exemplo, a temperatura de um paciente em um hospital, o preço de fechamento de uma ação na bolsa, ou a quantidade de passageiros que voam de avião em um determinado mês do ano. Os modelos mais famosos de séries temporais os processos Auto-Regressivos com Média Móvel (ARMA, na sigla em Inglês). Estes modelos são dados por certas equações de diferenças, em que uma sequência  $(Z_t)$ , conhecida como ruído, é envolvida nestas equações. Por exemplo, um processo de média móvel de ordem  $q \geq 1$  é definido como:

$$X_t = Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q}, t \in \mathbb{Z}$$

Um processo auto-regressivo de ordem 1 é dado por:

$$X_t = \phi X_{t-1} + Z_t, t \in \mathbb{Z}$$

$\theta_1, \dots, \theta_q, \phi$  representam parâmetros reais dos modelos. A figura 1.3.2 abaixo representa dois exemplos de série temporais. Do lado esquerdo, vemos o retorno diário do índice S&P, e do lado direito, é uma simulação do processo auto-regressivo  $X_t = 0,5 X_{t-1} + Z_t, Z_t \sim N(0,1)$ .

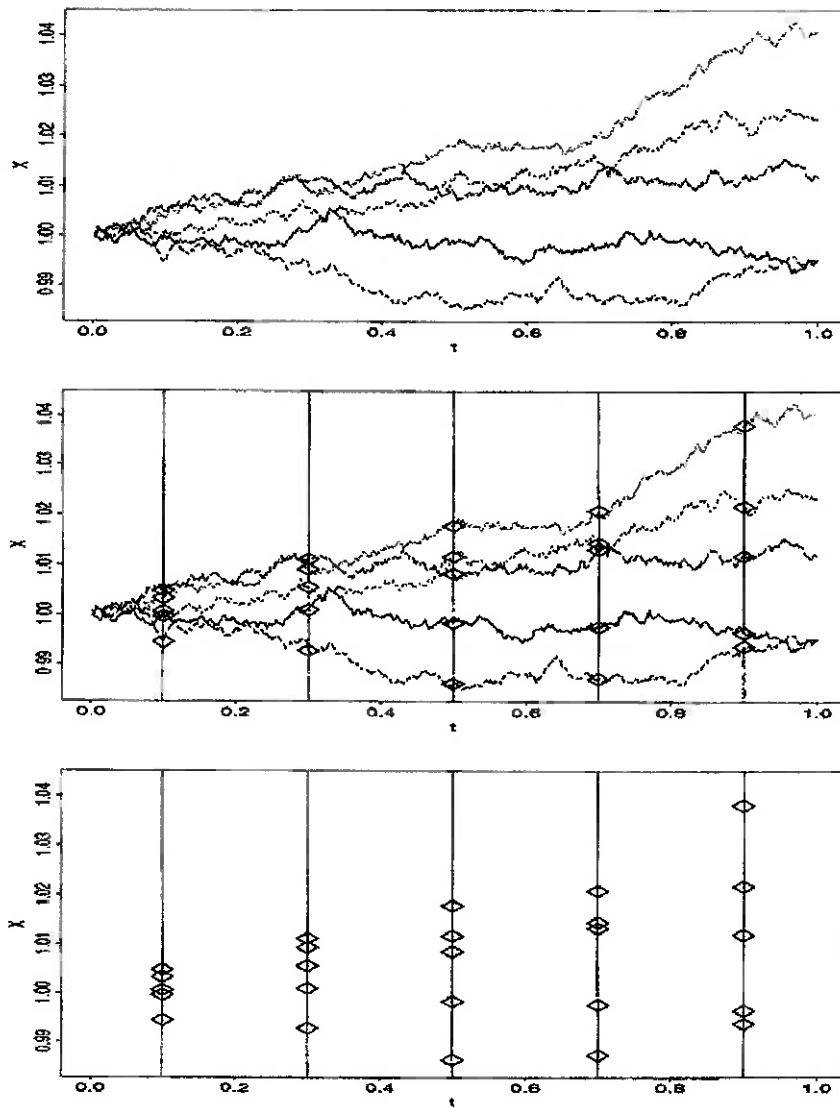


Figura 1.3.1: 5 caminhos aleatórios. Na primeira figura, cada um dos caminhos observados representa um  $\omega \in \Omega$  diferente. Já nas figuras do meio e de baixo, para cada  $t=0,1, \dots, 0,9$ ,  $X_t$  representa uma variável aleatória.

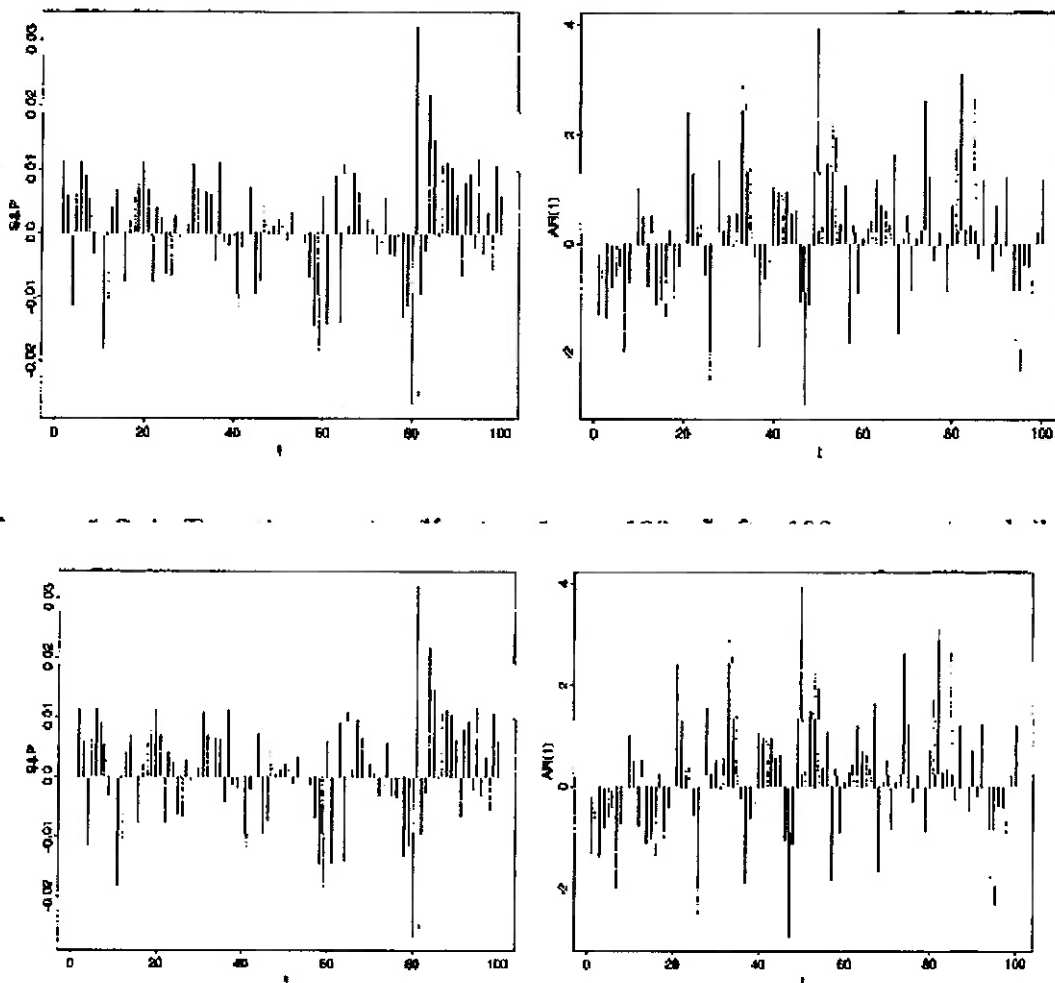


Figura 1.3.2: Duas séries temporais  $X_t$ ,  $t=1, \dots, 100$ . Do lado esquerdo temos 100 sucessivos log-retornos diários do índice S&P. Do lado direito, temos uma simulação de caminho aleatório de um processo auto-regressivo  $X_t = 0,5X_{t-1} + Z_t$ , onde os  $Z_t$  possuem distribuição  $N(0,1)$  e são independentes e identicamente distribuídas.

Observe que os conceitos de variável aleatória e de processos estocásticos não são tão diferentes assim. Ambos possuem realização, mas a realização de uma variável aleatória é um número, enquanto que a de um processo estocástico é uma função no tempo. Então, podemos imaginar um processo estocástico como sendo um elemento aleatório que toma funções como valores. Além disso, podemos interpretar uma variável aleatória ou um vetor aleatório como um caso especial de processo estocástico, com conjunto de índices  $T$  finito.

### 1.3.1 Distribuição

Em uma analogia com variável aleatória e com vetor aleatório, queremos introduzir a noção de distribuição, ou seja, as características não aleatórias a um processo estocástico, como sua distribuição, esperança, etc., e definir sua estrutura de dependências.

Uma forma de se fazer isto é interpretar um processo estocástico como uma coleção de vetores aleatórios. Portanto, temos a seguinte definição:

**Definição:** As distribuições de dimensão finita (*fidis*, na sigla em Inglês) de um processo estocástico  $X$  são todas as distribuições dos vetores de dimensão finita

$$(X_{t_1}, \dots, X_{t_n}), t_1, \dots, t_n \in T,$$

para todas as possíveis escolhas de tempos  $t_1, \dots, t_n \in T$  e para todo  $n \geq 1$ .

Um processo estocástico pode ser classificado por diferentes critérios. Um destes critérios é o tipo de *fidis*.

**Exemplo:** Processo gaussiano

Um processo estocástico é dito gaussiano se todas as suas *fidis* forem gaussianas multivariadas. Um vetor gaussiano pode ser determinado pela sua esperança e matriz de covariância. Logo, um processo estocástico gaussiano pode ser determinado pela sua coleção de esperanças e de matrizes de covariância.

Um processo gaussiano simples em  $T=[0,1]$  consiste em variáveis aleatórias independente e identicamente distribuídas (iid) com distribuição  $N(0,1)$ . Neste caso, as *fidis* são caracterizadas pelas seguintes funções de distribuição:

$$P(X_{t_1} \leq x_{t_1}, \dots, X_{t_n} \leq x_{t_n}) = P(X_{t_1} \leq x_{t_1}) \dots P(X_{t_n} \leq x_{t_n}) = \phi(x_{t_1}) \dots \phi(x_{t_n}),$$

em que  $\phi(x_1), \dots, \phi(x_n)$  representam as funções de distribuição da normal padrão calculadas em  $x_1, \dots, x_n$ .

### 1.3.2 Função de esperança e covariância

Para um vetor aleatório  $X=(X_1, \dots, X_n)$ , definimos sua esperança como sendo  $\mu_X=(E(X_1), \dots, E(X_n))$ , e sua matriz de covariância  $\Sigma_X=(cov(X_i, X_j); i, j=1, \dots, n)$ . Um processo estocástico  $X=(X_t, t \in T)$  pode ser considerado como uma coleção de vetores aleatórios  $(X_{t_1}, \dots, X_{t_n})$ . E para cada um deles, é possível calcular sua esperança e sua matriz de covariância. Alternativamente, podemos considerar essas quantidades como funções de  $t \in T$ . Portanto, considere a seguinte definição de esperança e covariância:

**Definição:** A função esperança de um processo estocástico  $X=(X_t, t \in T)$  é dada por:

$$\mu_X(t) = \mu_X = E(X_t), t \in T.$$

Sua função de covariância é dada por:

$$c_X(t, s) = cov(X_t, X_s) = E[(X_t - \mu_X(t))(X_s - \mu_X(s))], t \in T.$$

Por último, sua variância é dada por:

$$\sigma_X^2(t) = c_X(t, t) = var(X_t), t \in T.$$

Assim como no vetor aleatório, a função  $\mu_X(t)$  é determinística, e é tal que os caminhos aleatórios de  $X$  se concentram ao redor dela. A função  $c_X(t, s)$  representa uma medida de dependência no processo  $X$ . A função de variância  $\sigma_X^2(t)$  pode ser interpretada como uma medida do espalhamento dos caminhos amostrais de  $X$  em torno de  $\mu_X(t)$ .

Diferentemente do caso unidimensional, afirmações do tipo “95% dos caminhos amostrais estão entre  $\mu_X(t) \pm 2\sigma_X^2(t)$ ” são muito difíceis de serem demonstradas até mesmo para processos gaussianos, e em geral não são corretas.

### 1.3.3 Estrutura de dependência

Já introduzimos aqui o conceito de processo gaussiano especificando suas *fidis* como gaussianas multivariadas. Outra forma de classificar os processos estocásticos seria impor condições especiais à sua estrutura de dependências.

O processo  $X=(X_t, t \in T)$ ,  $T \subset \mathbb{R}$ , é dito estritamente estacionário se suas *fidis* são invariantes por deslocamentos no índice  $t$ :

$(X_{t_1}, \dots, X_{t_n}) \stackrel{d}{=} (X_{t_1+h}, \dots, X_{t_n+h})$  para todas as possíveis escolhas de índices  $t_1, \dots, t_n \in T, n \geq 1$  e  $h$  tal que  $t_1+h, \dots, t_n+h \in T$ . Note que não se trata de uma igualdade, mas se trata de distribuições idênticas, ou seja, quando afirmamos que  $A \stackrel{d}{=} B$ , então os elementos aleatórios (variáveis aleatórias, vetores aleatórios ou processos estocásticos)  $A$  e  $B$ , possuem a mesma distribuição, ou seja,  $P(A \in C) = P(B \in C)$ , qualquer que seja  $C$ , um conjunto “adequado”.

Estacionariedade também pode ser imposta nos incrementos de um processo. O processo em si não é necessariamente estacionário.

*Definição:* Seja  $X=(X_t, t \in T \subset \mathbb{R})$  um processo estocástico, e  $T$  um intervalo.  $X$  é dito estacionário por incrementos se:

$$X_t - X_s \stackrel{d}{=} X_{t+h} - X_{s+h}, \forall t, s, t+h, s+h \in T.$$

$X$  é dito independente por incrementos se para toda escolha de  $t_1 < \dots < t_n \in T, n \geq 1$ ,

$$X_{t_1} - X_{t_1}, \dots, X_{t_n} - X_{t_{n-1}}$$

forem variáveis aleatórias independentes.

*Exemplo:* Um processo estocástico  $(X_t, t \in [0, \infty))$  é dito um processo de Poisson homogêneo com taxa  $\lambda > 0$  se as seguintes condições forem satisfeitas:

- Começa em zero:  $X_0 = 0$ ;
- Possui incrementos independentes estacionários;
- Para todo  $t > 0, X_t \sim Poi(\lambda t)$ .

Note que, devido à estacionariedade dos incrementos,  $X_t - X_s, t > s$ , possui a mesma distribuição que  $X_{t-s} - X_0 = X_{t-s} \sim Poi(\lambda(t-s))$ .

## 1.4 Movimento Browniano

O movimento browniano possui um papel central em muitas áreas, como em probabilidade, em processos estocásticos, física, finanças, etc. É também muito importante neste trabalho. Começemos pela definição.

*Definição:* Um processo estocástico  $B=(B_t, t \in [0, \infty))$  é dito um processo de Wiener se as seguintes condições forem satisfeitas:

- Começa em zero:  $B_0=0$  ;
- Possui incrementos independentes estacionários;
- Para todo  $t>0, B_t \sim N(0, t)$  ;
- É contínua, não possui “saltos”.

### 1.4.1 Funções de distribuição, esperança e covariância

As *fidis* de um movimento browniano são gaussianas multivariadas, portanto  $B$  é dito um processo gaussiano. Para  $t>s$  temos que as variáveis aleatórias  $B_t - B_s$  e  $B_{t-s}$  possuem distribuição  $N(0, t-s)$ . Esta afirmação segue da estacionariedade por incrementos. De fato, temos que:

$$B_t - B_s \stackrel{d}{=} B_{t-s} - B_0 = B_{t-s} \sim N(0, t-s)$$

Assim, a variância é proporcional ao tamanho do intervalo  $[s, t]$ .

Da definição de processo browniano, é imediato que  $\mu_B(t) = E(B_t) = 0, t \geq 0$ , e como os incrementos  $B_s - B_0$  e  $B_t - B_s$  são independentes,  $t > s$ , a sua função de covariância fica:

$$c_B(t, s) = E[B_t B_s] = E[((B_t - B_s) + B_s) B_s] = E[(B_t - B_s)] E[B_s] + E[B_s^2] = 0 + s = s$$

Como um processo estocástico é definido por suas funções de esperança e covariância, então uma definição equivalente seria a seguinte:

*Definição:* Um movimento browniano é um processo gaussiano com:

$$\mu_B(t) = 0 \quad \text{e} \quad c_B(t, s) = \min(t, s) .$$

### 1.4.2 Propriedades do caminho

Neste tópico, fixaremos um caminho aleatório  $B_t(\omega), t \geq 0$ , e consideraremos as suas propriedades. Já sabemos da definição de movimento browniano que  $B_t(\omega)$  é contínuo. No entanto, observando alguns desses caminhos podemos ver que eles são extremamente

irregulares, apresentam muitas oscilações, e a razão disso é o fato de que  $B$  possui incrementos independentes.

De fato, os incrementos em intervalos adjacentes são independentes, não importa o tamanho destes intervalos.

Antes de continuarmos discutindo o quão irregular são os caminhos brownianos, faremos uma pequena pausa para a apresentação de uma definição contemplando uma classe de processos estocásticos, na qual o movimento browniano é um caso especial. Todos os membros dessa classe possuem caminhos aleatórios irregulares.

*Definição:* Um processo estocástico  $(X_t, t \in [0, \infty))$  é dito *H-auto-similar* para algum  $H > 0$  se suas *fidis* satisfizerem a seguinte condição:

$$(T^H B_{t_1}, \dots, T^H B_{t_n}) \stackrel{d}{=} (B_{Tt_1}, \dots, B_{Tt_n}) ,$$

para todo  $T > 0$ , qualquer que seja a escolha dos  $t_i \geq 0, i = 1, \dots, n, n \geq 1$ .

Note que auto-similaridade é uma propriedade da distribuição, e não do caminho, logo  $\stackrel{d}{=}$  não deve ser substituído por  $=$ .

De maneira grosseira, auto-similaridade implica que se mudarmos a escala de um caminho aleatório adequadamente, não importa se o intervalo de tempo é grande ou pequeno, estes caminhos serão semelhantes, mas não serão os mesmos.

A figura 1.4.2.1 a seguir ilustra bem este fato. Na medida em que vamos alterando a escala vemos que um mesmo caminho apresenta formatos semelhantes, mas não são iguais.

*Lema:* Os caminhos de processos auto-similares não são diferenciáveis em nenhum ponto.

Não demonstraremos este lema aqui. Apenas tomaremos como verdade. Agora, voltando ao caminho browniano, temos o seguinte teorema:

*Teorema:* O movimento browniano é 0,5-auto-similar, ou seja,

$$(T^{1/2} B_{t_1}, \dots, T^{1/2} B_{t_n}) \stackrel{d}{=} (B_{Tt_1}, \dots, B_{Tt_n}) ,$$

para todo  $T > 0$ , qualquer que seja a escolha dos  $t_i \geq 0, i = 1, \dots, n, n \geq 1$ .

*Demonstração:*

Como em ambos os lados da igualdade temos vetores gaussianos, então é suficiente mostrar que eles possuem a mesma esperança e a mesma matriz de covariância. Portanto:

$$E[T^{1/2} B_{t_i}] = T^{1/2} E[B_{t_i}] = T^{1/2} \cdot 0 = 0 = [B_{Tt_i}] .$$

$$\text{cov}(T^{1/2} B_{t_i}, T^{1/2} B_{t_j}) = E[T^{1/2} [(B_{t_i} - B_{t_j}) + B_{t_j}] B_{t_j} T^{1/2}]$$

$$\text{cov}(T^{1/2} B_{t_i}, T^{1/2} B_{t_j}) = T E[(B_{t_i} - B_{t_j}) E[B_{t_j}] + E[B_{t_j}^2]]$$

$$\text{cov}(T^{1/2} B_{t_i}, T^{1/2} B_{t_j}) = T t_j = c_B(T t_i, T t_j)$$

Na última passagem, assumi que  $t_i > t_j$ . Por fim, temos o seguinte corolário:

*Corolário:* Os caminhos de um movimento browniano não são diferenciáveis em nenhum ponto.

*Demonstração:* Este resultado sai da aplicação imediata do lema anterior.

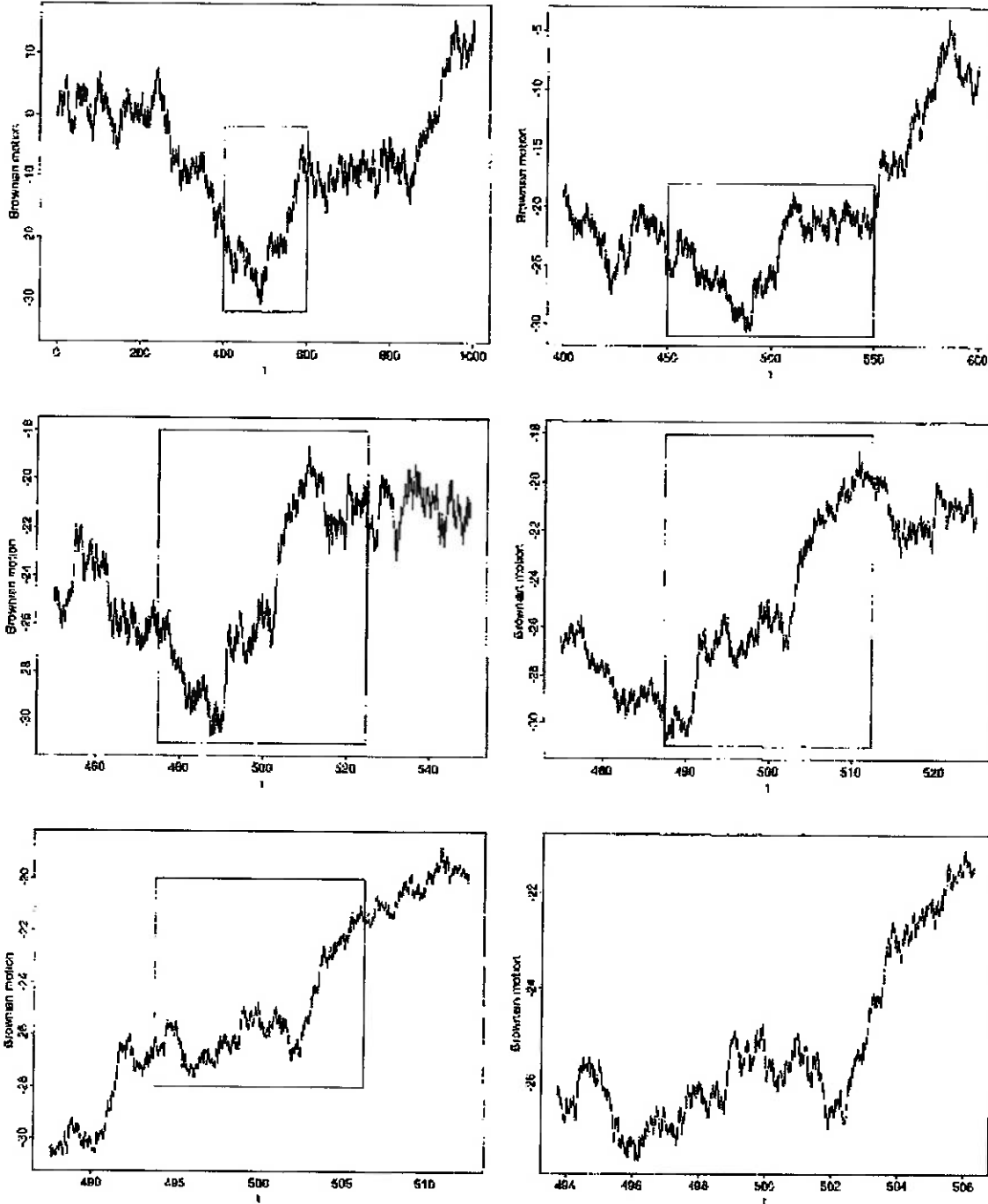


Figura 1.4.2.1: Auto-similaridade. Alterando a escala da figura obtemos figuras semelhantes, porém não são as mesmas.



A diferenciabilidade de uma função  $f$  implica que seu gráfico é “suave”, ou seja, localmente ela pode ser aproximada por uma reta. De fato, se o limite

$$\lim_{\Delta x \rightarrow 0} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x}$$

existir e for finito para algum  $x_0$  no domínio da  $f$ , então escrever para algum  $\Delta x$  pequeno:

$$f(x_0 + \Delta x) = f(x_0) + f'(x_0)\Delta x + h(x_0, \Delta x)\Delta x,$$

em que  $h(x_0, \Delta x) \rightarrow 0$  quando  $\Delta x \rightarrow 0$ . Portanto, em uma vizinhança pequena de  $x_0$ ,  $f$  é aproximadamente linear.

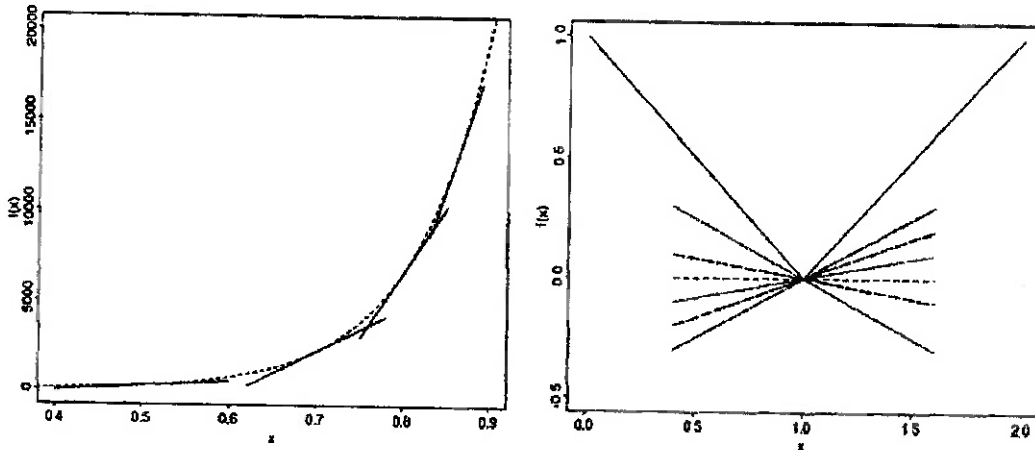


Figura 1.4.2.2: Do lado esquerdo temos uma função diferenciável. Qualquer ponto de seu gráfico pode ser aproximado por uma função linear, a única reta tangente neste ponto. Do lado direito  $f(x) = |x-1|$ . No ponto  $x=1$   $f$  não é diferenciável.

Na figura 1.4.2.2 acima, do lado esquerdo temos uma função que é suave, e em cada ponto de seu gráfico é possível se aproximar um vizinhança deste ponto por uma reta, que por sinal, é a única reta tangente neste ponto. Já no lado direito, vemos o gráfico da função  $f(x) = |x-1|$ . No ponto  $x=1$ , vemos que  $f$  não é diferenciável, ou seja, não há uma única reta tangente neste ponto.

Agora, se tentarmos imaginar uma função que não é diferenciável em nenhum ponto, chegaremos a conclusão de que isto é impossível. Esta função  $B$ , qualquer que seja o  $t_0$  em seu domínio, é impossível aproximar  $B_{t_0}$  por uma reta na vizinhança de  $t_0$ .

Uma outra indicação de que o movimento browniano é irregular é dada pelo seguinte fato:

**Teorema:** Um movimento browniano não tem variação limitada, qualquer que seja o intervalo de tempo  $[0, T]$ . Isto significa que:

$$\sup_{\tau} \sum_{i=1}^n |B_{t_i}(\omega) - B_{t_{i-1}}(\omega)| = \infty,$$

em que o supremo é tomado sobre todas as possíveis partições  $\tau: 0 = t_0 < \dots < t_n = T$  de  $[0, T]$ .

Também não demonstraremos este fato aqui. O fato de não ser diferenciável em nenhum ponto, além de possuir variação ilimitada são as principais razões pela qual a integração pelos métodos tradicionais não funcionam.

### 1.5 Esperança Condicional

Nesta parte do trabalho falaremos sobre um conceito importante para que posteriormente possamos falar sobre  $\sigma$ -algebra .

Começaremos com a probabilidade condicional. Em cursos de teoria de probabilidade, aprendemos que a probabilidade de um certo evento  $A$  ocorrer dado que um outro evento  $B$  ocorreu é dada por:

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

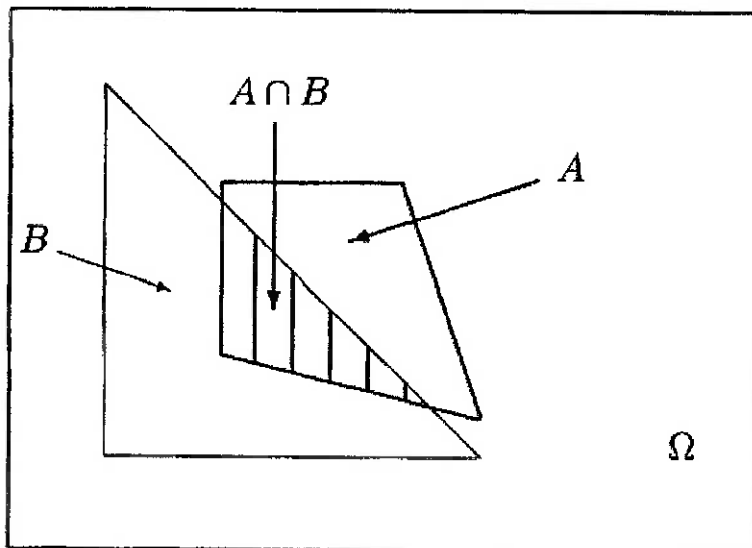


Figura 1.5.1: Dado que  $B$  ocorre, então  $A$  ocorre somente se  $A \cap B$  ocorrer.

Observando a figura 1.5.1 acima, vemos que basicamente o que acontece é que dado que  $B$  aconteceu, então  $A$  só pode ocorrer se  $A \cap B$  ocorrer. E dividimos por  $P(B)$  pois neste caso,  $B$  seria o novo espaço amostral, então precisamos normalizar o antigo valor de  $P(A \cap B)$  por  $P(B)$  .

Agora, dado que  $P(B) > 0$  , podemos definir a função de distribuição condicional de uma variável aleatória  $X$  dado  $B$ :

$$F_X(x|B) = \frac{P(X \leq x, B)}{P(B)}, x \in \mathbb{R} ,$$

e também a esperança condicional de  $X$  dado  $B$ :

$$E_X(x|B) = \frac{E(X I_B)}{P(B)}, x \in \mathbb{R}, \text{ em que}$$

$$I_B(\omega) = \begin{cases} 1, \omega \in B \\ 0, \omega \notin B \end{cases}$$

é a função indicadora do evento  $B$ . Apenas para efeito de ilustração, assumiremos que  $\Omega = \mathbb{R}$ . Se  $X$  é discreto, com valores  $x_1, x_2, \dots$ , então a esperança fica:

$$E[X|B] = \sum_{k=1}^{\infty} x_k \frac{P(\{\omega: X(\omega) = x_k\} \cap B)}{P(B)} = \sum_{k=1}^{\infty} x_k P(X = x_k|B).$$

Se  $X$  possui densidade  $f_X$ , então a esperança condicional fica:

$$E[X|B] = \frac{1}{P(B)} \int_{-\infty}^{\infty} x I_B f_X(x) dx = \frac{1}{P(B)} \int_B x f_X(x) dx$$

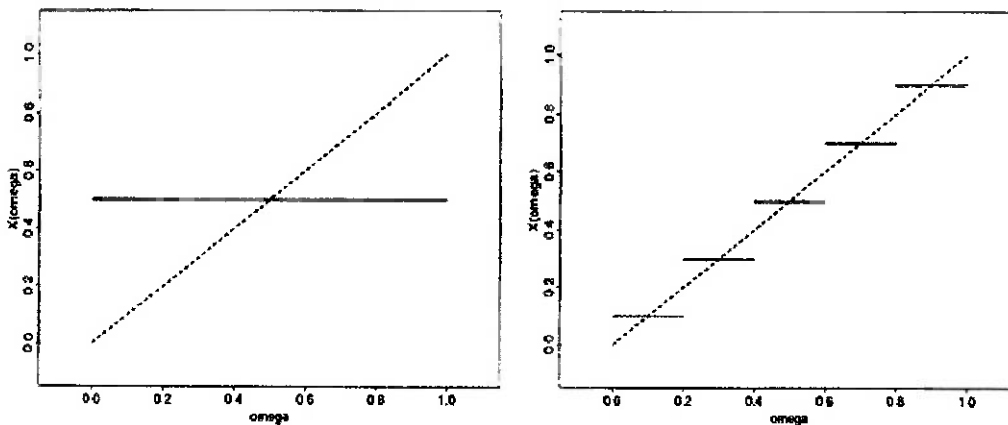


Figura 1.5.2: Do lado esquerdo temos uma variável uniforme  $Y \sim U(0,1)$ , representada pela linha pontilhada e sua esperança, representada pela linha contínua. Do lado direito também temos uma variável uniforme  $Y \sim U(0,1)$ , representada pela linha pontilhada e suas esperanças condicionais  $E[X|A_i]$ , onde  $A_i = [(i-1)/5, i/5]$ ,  $i=1, \dots, 5$ . Estas esperanças condicionais podem ser interpretadas como valores de uma variável aleatória discreta  $E[X|Y]$  com valores constantes distintos nos conjuntos  $A_i$ .

**Exemplo:** Considere a variável aleatória  $X(\omega) = \omega$ ,  $\Omega = (0,1]$ , munido da medida de probabilidade  $P((a,b)) = b-a$ ,  $(a,b) \subset \Omega$ .  $X$  possui distribuição uniforme em  $(0,1]$ . Sua função de distribuição é dada por:

$$F_X(x) = P(\{\omega: X(\omega) < x\}) = \begin{cases} P(\emptyset) = 0, x \leq 0 \\ P((0,x]) = x, x \in (0,1] \\ P((0,1]) = 1, x > 1 \end{cases}$$

Na figura 1.5.2 à esquerda, está representada a variável aleatória  $X$  e sua esperança  $E[X] = 0,5$ .

Agora, assumamos que  $A_i = \left[ \frac{i-1}{n}, \frac{i}{n} \right]$ ,  $i=1, \dots, n$ , ocorreu para algum  $i$ . Sabemos que  $f_X(x) = 1$  em  $[0, 1]$ , e  $P(A_i) = \frac{1}{n}$ . Então:

$$E[X|A_i] = \frac{1}{P(A_i)} \int_{A_i} x f_X(x) dx = n \int_{\frac{i-1}{n}}^{\frac{i}{n}} x dx = \frac{1}{2} \frac{2i-1}{n}.$$

A esperança condicional está ilustrada à direita. O valor de  $E[X|A_i]$  é atualizado no novo espaço  $A_i$ , dado que  $A_i$  aconteceu.

Agora, considere uma variável aleatória  $Y$  em  $\Omega$  que assume valores distintos  $y_i$  em conjuntos  $A_i$ , ou seja,

$$A_i = \{\omega : Y(\omega) = y_i\}, \quad i=1, 2, \dots$$

Então,  $A_i$  forma uma partição disjunta de  $\Omega$ , ou seja,  $A_i \cap A_j = \emptyset$ ,  $i \neq j$ , e  $\bigcup_{i=1}^{\infty} A_i = \Omega$ . Assumiremos por conveniência que  $P(A_i) > 0, \forall i$ .

Para uma variável aleatória  $X \in \Omega$ , com  $E[|X|] < \infty$ , definimos a esperança condicional de  $X$  dado  $Y$  como a variável aleatória discreta

$$E[X|Y](\omega) = E[X|A_i] = E[X|Y = y_i], \quad \omega \in A_i, \quad i=1, 2, \dots$$

Sabendo que  $A_i$  ocorreu, podemos nos restringir a  $\omega \in A_i$ . Neste caso, o  $E[X|Y](\omega)$  coincide com o caso clássico de  $E[X|A_i]$ .

*Exemplo: (Continuação)*

No exemplo anterior,  $E[X|A_i]$  pode ser interpretado como sendo valores de uma variável aleatória discreta  $E[X|Y]$ , onde  $Y$  é constante em cada um dos conjuntos

$A_i = \left[ \frac{i-1}{n}, \frac{i}{n} \right]$ ,  $i=1, \dots, n$ . Neste sentido,  $E[X|Y]$  pode ser considerado uma versão

grosseira da variável aleatória  $X$  original em cada um dos conjuntos  $A_i$ , isto é,  $E[X|Y]$  é uma aproximação de  $X$  dado que  $A_i$  ocorreu.

Mostraremos agora, algumas propriedades da variável aleatória  $E[X|Y]$ :

1) Ela é linear: para variáveis aleatórias  $X_1, X_2$  e constantes  $c_1, c_2$ :

$$E[c_1 X_1 + c_2 X_2 | Y] = c_1 E[X_1 | Y] + c_2 E[X_2 | Y].$$

*Demonstração:* Mostraremos apenas para o caso contínuo. O caso discreto é semelhante.

$f_{X_1}, f_{X_2}$  as respectivas densidades de  $X_1, X_2$ . Então:

$$E[c_1 X_1 + c_2 X_2 | Y] = \frac{1}{P(Y)} \int_Y (c_1 f_{X_1}(x) + c_2 f_{X_2}(x)) dx$$

$$E[c_1 X_1 + c_2 X_2 | Y] = c_1 \frac{1}{P(Y)} \int_Y f_{X_1}(x) dx + c_2 \frac{1}{P(Y)} \int_Y f_{X_2}(x) dx$$

$$E[c_1 X_1 + c_2 X_2 | Y] = c_1 E[X_1 | Y] + c_2 E[X_2 | Y]$$

2) As esperanças de  $X$  e de  $E[X|Y]$  são iguais:

$$E[X] = E[E[X|Y]] .$$

$$E[E[X|Y]] = \sum_{i=1}^{\infty} E[X|A_i] P(A_i) = \sum_{i=1}^{\infty} E[X I_{A_i}] = E[X \sum_{i=1}^{\infty} I_{A_i}] = E[X] .$$

Aqui, utilizamos o fato de que  $\sum_{i=1}^{\infty} I_{A_i} = I_{\bigcup_{i=1}^{\infty} A_i} = I_{\Omega} = 1$  .

3) Se  $X$  e  $Y$  são independentes, então  $E[X|Y] = E[X]$  .

Como já dissemos anteriormente, independência entre  $X$  e  $Y$  significa que

$$P(X \in A, Y = y_i) = P(X \in A) P(Y = y_i) = P(X \in A) P(A_i) .$$

Considere a variável aleatória  $I_{A_i}$ , e note que:

$$\{\omega : I_{A_i} = 1\} = A_i = \{\omega : Y(\omega) = y_i\}$$

Assim, podemos re-escrever a equação acima da seguinte forma:

$$P(X \in A, I_{A_i} = 1) = P(X \in A) P(I_{A_i} = 1) .$$

A relação é análoga com  $\{I_{A_i} = 0\}$  no lugar de  $\{I_{A_i} = 1\}$  . Portanto, as variáveis aleatórias  $X$  e  $I_{A_i}$  são independentes para  $\omega \in A_i$  . Temos:

$$E[X|Y](\omega) = E[X|A_i] = \frac{E[X I_{A_i}]}{P(A_i)} = \frac{E[X] E[I_{A_i}]}{P(A_i)} = E[X] , \text{ pois}$$

$$E[I_{A_i}] = 0 P(A_i^c) + 1 P(A_i) = P(A_i) .$$

Isto prova o item 3.

Resumo:

Até aqui aprendemos que:

- A esperança condicional  $E[X|Y]$  de  $X$  dado uma variável aleatória discreta  $Y$  é uma variável aleatória discreta;
- Ela coincide com o caso clássico de esperança condicional  $E[X|Y=y_i]$  nos conjuntos  $A_i = \{\omega : Y(\omega) = y_i\}$  ;
- Neste sentido, trata-se de uma aproximação grosseira de  $X$ . De fato, se  $Y$  for constante,  $E[X|Y] = E[X]$  , se  $Y$  assume  $n$  in  $\mathbb{N}^k$  valores distintos, então  $E[X|Y]$  também assume;
- $E[X|Y]$  não é função de  $X$ , mas sim de  $Y$ . A variável aleatória apenas define o tipo de função. De fato, podemos escrever:

$$E[X|Y] = g(Y) \text{ , onde } g(y) = \sum_{i=1}^n E[X|Y=y_i] I_{y_i}(y) \text{ .}$$

## 1.6 $\sigma$ - álgebra

Na seção anterior, vimos a definição de  $E[X|Y]$  onde  $X$  é uma variável aleatória e  $Y$  é uma variável aleatória discreta. Vimos também que não importa de fato quais valores  $Y$  assume, mas sim que  $Y$  assuma valores  $y_i$  distintos em subconjuntos  $A_i$  de  $\Omega$  . Assim,  $E[X|Y]$  pode ser interpretada como uma variável aleatória construída a partir de uma coleção  $\sigma(Y)$  de conjuntos de  $\Omega$  . Então, de maneira simbólica, podemos escrever:

$$E[X|Y] = E[X|\sigma(Y)] \text{ .}$$

A coleção  $\sigma(Y)$  nos fornece informações sobre a estrutura da variável aleatória  $Y(\omega)$  como função de  $\omega \in \Omega$  .

O nosso objetivo aqui é formalizar o significado de “coleção  $\sigma(Y)$  de conjuntos de  $\Omega$  ”. Chamaremos essa coleção  $\mathcal{F}$  de,  $\sigma$ -álgebra se forem satisfeitas as seguintes condições:

- A coleção  $\mathcal{F}$  é não vazia:  $\emptyset \in \mathcal{F}$  ,  $\Omega \in \mathcal{F}$  ;
- Se  $A \in \mathcal{F}$  , então  $A^c \in \mathcal{F}$  ;
- Se  $A_1, A_2, \dots \in \mathcal{F}$  , então  $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$  ,  $\bigcap_{i=1}^{\infty} A_i \in \mathcal{F}$  ;

*Exemplo:* Algumas  $\sigma$ -álgebra elementares

- $\mathcal{F}_1 = \{\emptyset, \Omega\}$  ;
- $\mathcal{F}_2 = \{\emptyset, A, A^c, \Omega\}$  ;

- $\mathcal{F}_3 = \{A: A \in \Omega\}$  ;

$\mathcal{F}_1$  é a maior  $\sigma$ -álgebra ,  $\mathcal{F}_3$  é a maior de todas.

Agora, suponha que  $C$  seja uma coleção de subconjuntos de  $\Omega$  , mas não necessariamente uma  $\sigma$ -álgebra . Poderíamos adicionar, por exemplo, o suficiente para formar o conjunto  $\mathcal{F}_3$  , mas é possível demonstrar que para uma dada coleção de subconjuntos de  $C$  de subconjuntos de  $\Omega$  , existe a menor  $\sigma$ -álgebra  $\sigma(C)$  em  $\Omega$  contendo  $C$ .

Chamaremos de  $\sigma(C)$  a  $\sigma$ -álgebra gerada por  $C$ .

*Exemplo:* (  $\sigma$ -álgebra gerada a partir de uma variável aleatória discreta)

Considere a variável aleatória  $Y$ , assumindo valores  $y_i$  e definindo os subconjuntos  $A_i = \{\omega: Y(\omega) = y_i\}$  que constituem uma partição disjunta de  $\Omega$  . Seja  $C = \{A_1, A_2, \dots\}$  . Como  $\sigma(C)$  é uma  $\sigma$ -álgebra , ela deve conter todos os conjuntos da forma:

$$A = \bigcup_{i \in I} A_i, I \subset \mathbb{N} = \{1, 2, \dots\} , \quad (1.6.1)$$

incluindo  $I = \emptyset (A = \emptyset)$  e  $I = \mathbb{N} (A = \Omega)$  . Os conjuntos da forma (1.6.1) formam uma  $\sigma$ -álgebra  $\sigma(Y)$  , pois:

- $I = \emptyset \Rightarrow \emptyset \in \sigma(Y)$  e  $I = \mathbb{N} \Rightarrow \Omega \in \sigma(Y)$  ;
- $A = \bigcup_{i \in I} A_i \in \sigma(Y)$  e  $A^c = \bigcup_{i \in I^c} A_i \in \sigma(Y)$  ;
- Dados  $I_1, I_2, \dots$  subconjuntos de  $\mathbb{N}$  então  $I = \bigcup_{j=1}^{\infty} I_j \subset \mathbb{N}$  e  $J = \bigcap_{j=1}^{\infty} I_j \subset \mathbb{N}$  , e portanto,  $A_1 = \bigcup_{i \in I} A_i \in \sigma(Y)$  e  $A_2 = \bigcup_{i \in J} A_i \in \sigma(Y)$  .

*Exemplo:* Os conjuntos de Borel

Considere  $\Omega = \mathbb{R}$  e  $C^{(1)} = \{[a, b]: -\infty < a < b < \infty\}$  . A  $\sigma$ -álgebra  $B_1 = \sigma(C^{(1)})$  contém uma grande variedade de subconjuntos de  $\mathbb{R}$  . Estes subconjuntos são conhecidos como conjuntos de Borel, e esta coleção como  $\sigma$ -álgebra de Borel. É muito difícil imaginar a variedade de conjuntos que  $B_1$  possui. Por exemplo, é fato que  $B_1 \subsetneq \mathcal{F}_3$  , no entanto, esta verificação não é fácil.

É possível definir também a  $\sigma$ -álgebra de dimensão  $n$ ,  $B_n = \sigma(C^{(n)})$  e

$$C^{(n)} = \{[a, b]: -\infty < a_i < b_i < \infty, i = 1, \dots, n\} .$$

Os conjuntos  $[a, b]$  são chamados retângulos. Qualquer subconjunto de  $\mathbb{R}^n$  "razoável" é um conjunto de Borel. Por exemplo, bolas, esferas, curvas suaves, superfícies,

etc., são chamados de conjuntos de Borel. Também são conjuntos de Borel os conjuntos fechados e abertos. Para demonstrar que um determinado conjunto  $C \subset \mathbb{R}^n$  é de Borel basta obter C por uma quantidade enumerável de operações  $\cap$ ,  $\cup$  ou  $^c$ , agindo sobre retângulos. Por exemplo,  $(a, b), [a, b], (-\infty, a], (b, \infty)$  são Borel ( $n = 1$ ), pois:

- $(a, b) = \bigcup_{n \in \mathbb{N}} \left( a + \frac{1}{n}, b - \frac{1}{n} \right]$  ;
- $(a, b) = \bigcup_{n \in \mathbb{N}} \left( a + \frac{1}{n}, b - \frac{1}{n} \right]$  ;
- $(-\infty, a) = \bigcup_{n \in \mathbb{N}} \left( -n, a - \frac{1}{n} \right]$  ;
- $(b, \infty) = \left( \bigcup_{n \in \mathbb{N}} \left( -n, b + \frac{1}{n} \right] \right)^c$  .

Agora, retome o exemplo da  $\sigma$ -álgebra  $\sigma(Y)$  gerada a partir dos conjuntos  $A_i = \{\omega : Y(\omega) = y_i\}$ , para variável aleatória discreta  $Y$ , com valores  $y_i$ . Se  $Y$  for uma variável aleatória contínua, a  $\sigma$ -álgebra gerada a partir dos conjuntos  $\{\omega : Y(\omega) = y\}$ , do ponto de vista matemático, não é uma coleção rica o suficiente. Por exemplo,  $\{\omega : a < Y(\omega) \leq b\}$  não pertence a esta  $\sigma$ -álgebra. Portanto, assumiremos que pelo menos conjuntos dessa forma estarão em  $\sigma(Y)$  quando  $Y$  for uma variável aleatória contínua.

Como estaremos também interessados em vetor aleatório  $Y$ , segue a definição de  $\sigma(Y)$  para o caso multivariado.

**Definição:** Seja  $Y = (Y_1, \dots, Y_n)$  um vetor aleatório  $n$ -dimensional. A  $\sigma$ -álgebra  $\sigma(Y)$  é a menor  $\sigma$ -álgebra contendo todos os conjuntos da forma:

$$\{Y \in [a, b]\} = \{\omega : a_i < Y_i(\omega) \leq b_i, i = 1, \dots, n\}, -\infty < a_i < b_i < \infty .$$

A coleção  $\sigma(Y)$  é a  $\sigma$ -álgebra gerada pelo vetor aleatório  $Y$ . Um elemento em  $\sigma(Y)$  nos diz para qual  $\omega \in \Omega$  o vetor aleatório  $Y$  assume valores no retângulo  $[a, b]$ , ou em um conjunto de Borel, mais geral.

A  $\sigma$ -álgebra  $\sigma(Y)$  contém as informações essenciais sobre a estrutura do vetor aleatório  $Y$  como função de  $\omega \in \Omega$ . Ela contém todos os conjuntos da forma  $\{\omega : Y(\omega) \in C\}$ , para todo conjunto de Borel  $C \in \mathbb{R}^n$ .

Imaginar  $\sigma(Y)$  é muito difícil, e é mais difícil ainda imaginar no caso em que  $Y$  é um processo estocástico. Para o caso em que  $Y = (Y_t, t \in T, \omega \in \Omega)$ , a  $\sigma$ -álgebra  $\sigma(Y)$  é a menor  $\sigma$ -álgebra contendo todos os conjuntos da forma:

$$\{\omega : \text{o caminho amostral } (Y_t(\omega), t \in T, \omega \in \Omega) \text{ pertence a } C\} ,$$

para todo conjunto "adequado"  $C$  de funções em  $T$ .

Não foi dada uma definição mais formal devido às nossas ferramentas ainda serem limitadas, mas teremos uma ideia melhor sobre  $\sigma(Y)$  depois desse exemplo abaixo:



*Exemplo:* Seja  $B=(B_s, s \leq t)$  um movimento browniano em  $[0, t]$ . A coleção:

$$\mathcal{F}_t = \sigma(B) = \sigma(B_s, s \leq t),$$

é a menor  $\sigma$ -álgebra contendo a informação essencial sobre a estrutura do processo  $B$  em  $[0, t]$ . É possível demonstrar que esta  $\sigma$ -álgebra é gerada por todos os conjuntos da forma:

$$A_{t_1, \dots, t_n}(C) = \{\omega : (B_{t_1}(\omega), \dots, B_{t_n}(\omega)) \in C\},$$

para qualquer conjunto de Borel  $C$  de dimensão  $n$ , e qualquer escolha de  $t_i \in [0, t], n \geq 1$ .

Podemos ver que  $\sigma(Y)$  é um objeto não-trivial. Daqui em diante, iremos utilizar a seguinte regra para tentar imaginar os conjuntos  $\sigma(Y)$ :

Para uma variável aleatória, vetor aleatório ou processo estocástico  $Y \in \Omega$ , a  $\sigma$ -álgebra  $\sigma(Y)$  gerada por  $Y$  contém a informação essencial sobre a estrutura de  $Y$  como função de  $\omega \in \Omega$ . Ela consiste de todos os subconjuntos  $\{\omega : Y(\omega) \in C\}$  para conjuntos adequados  $C$ .

Como  $Y$  gera  $\sigma(Y)$ , também dizemos que  $Y$  contém informação representada por  $\sigma(Y)$  ou  $Y$  carrega a informação  $\sigma(Y)$ .

Concluimos a seguir com a seguinte regra útil:

Seja  $f$  uma função agindo em  $Y$ , e considere o conjunto  $\{\omega : f(Y(\omega)) \in C\}$ , para conjuntos adequados  $C$ . Para funções "bem comportadas",  $\sigma(f(Y)) \subset \sigma(Y)$ . Isto significa que  $f$  não fornece nenhuma informação nova sobre a estrutura de  $Y$ . Dizemos que a informação nova gerada por  $\sigma(f(Y))$  está contida em  $\sigma(Y)$ .

Segue um exemplo simples para mostrar este fato:

*Exemplo:* Seja  $B$  um movimento browniano, e defina a  $\sigma$ -álgebra  $\mathcal{F}_t = \sigma(B_s, s \leq t), t \geq 0$ . Considere  $f(B) = B_t$  para um  $t$  fixo. Dado que conhecemos a estrutura de  $B_t$ , temos que  $\sigma(f(B)) \subset \mathcal{F}_t$ . A recíproca é claramente falsa. Se conhecermos somente  $B_t$ , não é possível construir todo o processo  $(B_s, s \leq t)$  a partir de  $B_t$ .

### 1.6.1 Esperança condicional geral

Vimos até aqui a definição de  $E[X|Y]$  em que  $Y$  é uma variável aleatória discreta. Vimos que não é necessário saber os valores de  $Y$  explicitamente, mas a sua definição dependia da partição que fazíamos do espaço amostral ( $A_i = \{\omega : Y(\omega) = y_i\}$ ). Vimos que estes conjuntos  $A_i$  geram uma  $\sigma$ -álgebra  $\sigma(Y)$ . Este é o ponto de partida da definição de  $E[X|\mathcal{F}]$ ,  $\mathcal{F}$  é uma  $\sigma$ -álgebra em  $\Omega$ . Em aplicações sempre iremos escolher  $\mathcal{F} = \sigma(Y)$  em que  $Y$  é uma variável aleatória, um vetor aleatório ou um processo estocástico.

Na seção anterior vimos que a estrutura essencial de  $Y$  está contida em  $\sigma(Y)$ . Neste sentido, dizemos que  $Y$  carrega a informação de  $\sigma(Y)$ .

Sejam  $Y, Y_1, Y_2$  elementos aleatórios (variáveis aleatórias, vetores aleatórios ou processos estocásticos) em  $\Omega$  e  $\mathcal{F}$  uma  $\sigma$ -álgebra em  $\Omega$ .

Dizemos que:

- A informação de  $Y$  está contida em  $\mathcal{F}$ , ou  $Y$  não contém mais informação que a contida em  $\mathcal{F}$  se  $\sigma(Y) \subset \mathcal{F}$ ;
- $Y_2$  contém mais informação que  $Y_1$  se  $\sigma(Y_1) \subset \sigma(Y_2)$ .

Agora, podemos apresentar uma definição mais rigorosa de  $E(X|\mathcal{F})$  sob uma  $\sigma$ -álgebra  $\mathcal{F}$  abstrata:

**Definição:** Uma variável aleatória  $Z$  é chamada esperança condicional de  $X$  dada uma  $\sigma$ -álgebra  $\mathcal{F}$ , e escrevemos  $Z = E(X|\mathcal{F})$ , se:

- $Z$  não contém mais informação que a contida em  $\mathcal{F}$ ;
- $Z$  satisfaz a relação:

$$E(XI_A) = E(ZI_A), \forall A \in \mathcal{F}. \quad (I)$$

Esta segunda condição nos diz que as variáveis aleatórias  $X$  e  $E(X|\mathcal{F})$  são “próximas”, mas não no sentido de coincidir para qualquer que seja  $\omega \in \Omega$ , mas no sentido de médias, em certos conjuntos  $A$  de  $\mathcal{F}$ . Ou seja, esta definição é uma versão formal da ideia que tínhamos obtido anteriormente com a esperança condicional, de que  $E(X|\mathcal{F})$  é uma versão grosseira da variável aleatória  $X$ .

A segunda condição ainda nos dá certa liberdade de escolha na construção da variável aleatória  $Z = E(X|\mathcal{F})$ . Isso significa que pode existir uma variável aleatória  $Z' \neq Z$  em conjuntos que possuem probabilidade zero. Por esta razão,  $E(X|\mathcal{F})$  deveria ser interpretado no sentido de *quase sempre* (q.s.). Mas neste trabalho esta sigla será suprimida.

**Exemplo:** Esperança condicional sob condições discretas

Queremos mostrar aqui que  $E(X|Y)$  sob condições discretas nos leva a um caso especial de  $E(X|\mathcal{F})$  quando  $\mathcal{F} = \sigma(Y)$ . Retome o exemplo passado em que todo elemento de  $A \in \mathcal{F}$  é da forma:

$$A = \bigcup_{i \in I} A_i = \bigcup_{i \in I} \{\omega : Y(\omega) = y_i\}, I \subset \mathbb{N} = \{1, 2, \dots\}.$$

Definimos  $Z(\omega) = E(X|A_i), \omega \in A_i$ . Vimos que neste caso,  $Z$  é meramente uma função de  $Y$  e não de  $X$ . Logo  $\sigma(Z) \subset \sigma(Y)$ . Além disso, dado um  $A \in \mathcal{F}$ ,

$$E(XI_A) = E\left(X \sum_{i \in I} I_{A_i}\right) = \sum_{i \in I} E(XI_{A_i}).$$

Mas por outro lado,  $ZI_A$  é uma variável aleatória discreta com esperança

$$E(ZI_A) = \sum_{i \in I} E(X|A_i) P(A_i) = \sum_{i \in I} E(XI_{A_i}).$$

Assim,  $Z$  satisfaz a condição de esperança condicional, e não contém mais informação que  $Y$ . Logo,  $Z = E(X|\sigma(Y))$ .

Portanto, no caso discreto  $E(X|Y)$  e  $E(X|\sigma(Y))$  representam a mesma variável aleatória. Isto sugere a seguinte definição:

*Definição:* Seja  $Y$  um elemento aleatório em  $\Omega$  e  $\sigma(Y)$  a  $\sigma$ -álgebra gerada por  $Y$ . Então, a esperança condicional de uma variável aleatória  $X$  dado  $Y$  é definida por:

$$E(X|Y) = E(X|\sigma(Y)) .$$

*Exemplo:* A probabilidade condicional e a esperança condicional clássicas

A probabilidade condicional e a esperança condicional clássicas são casos especiais da noção geral da esperança condicional que acabamos de definir. De fato, seja  $B$  tal que  $P(B) > 0$  e  $P(B^c) > 0$ , e defina  $\mathcal{F}_B = \sigma(\{B\})$ . Portanto, temos que:

$$\mathcal{F}_B = \{\emptyset, \Omega, B, B^c\} .$$

Do exemplo anterior, sabemos que:

$$E(X|\mathcal{F}_B)(\omega) = E(X|B), \omega \in B .$$

Esta é a noção clássica de esperança condicional, pois se especificarmos  $X = I_A$  para algum evento  $A$ , obtemos para  $\omega \in B$ :

$$E(I_A|\mathcal{F}_B)(\omega) = E(I_A|B) = \frac{P(A \cap B)}{P(B)} .$$

## 1.6.2 Regras Para o Cálculo de Esperança Condicional

Devido à forma como definimos a esperança condicional a uma  $\sigma$ -álgebra, é em geral muito difícil, se não impossível, efetuar o seu cálculo. No caso em que a  $\sigma$ -álgebra é gerada a partir de uma variável aleatória discreta é exceção. Portanto, listaremos a seguir algumas regras para lidarmos com o cálculo dessas esperanças condicionais.

Mas antes de apresentarmos essas regras, enunciaremos a seguir um teorema que será útil mais a frente.

*Teorema:* Seja  $[\Omega, \mathcal{F}, P]$  um espaço de probabilidade, ou seja,  $\mathcal{F}$  uma  $\sigma$ -álgebra em  $\Omega$  e  $P$  uma medida de probabilidade em  $\mathcal{F}$ . Considere a  $\sigma$ -álgebra  $\mathcal{F}' \subset \mathcal{F}$ , e denote por  $P'$  a medida de probabilidade  $P$  restrita à  $\mathcal{F}'$ . Seja ainda  $X$  uma variável aleatória em  $\Omega$ . Se  $E(|X|) < \infty$  então existe uma variável aleatória  $Z$  tal que:

- $\sigma(Z) \subset \mathcal{F}'$  ;
- Para todo  $A \in \mathcal{F}'$ ,

$$\int_A Z(\omega) dP(\omega) = \int_A X(\omega) dP, A \in \mathcal{F}.$$

Se essas duas condições forem satisfeitas para outra variável aleatória  $Z'$ , então:

$$P(Z \neq Z') = 0.$$

Basicamente, o que podemos tirar de importante deste teorema, é que se  $E(|X|) < \infty$ , então  $E(X|\mathcal{F})$  existe e é único.

Agora, vamos às regras:

1º) A esperança condicional é linear:

$$E([c_1 X_1 + c_2 X_2]|\mathcal{F}) = c_1 E(X_1|\mathcal{F}) + c_2 E(X_2|\mathcal{F})$$

2º) Esperança de  $X$  e de  $E(X|\mathcal{F})$  são a mesma:

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

3º) Se  $X$  e a  $\sigma$ -álgebra  $\mathcal{F}$  são independentes,  $E(X|\mathcal{F}) = E(X)$ . Em particular, se  $X$  e  $Y$  são independentes  $E(X|Y) = E(X)$ .

Uma breve discussão sobre esta regra:

$X$  e  $\mathcal{F}$  independentes significa que se conhecemos  $\mathcal{F}$  não sabemos nenhuma informação sobre  $X$  e vice-versa. De maneira formal, as variáveis aleatórias  $X$  e  $I_A$  são independentes para todo  $A \in \mathcal{F}$ . Logo:

$$E(X I_A) = E(X) E(I_A) = E(E(X) I_A), \forall A \in \mathcal{F}.$$

Comparando com (I) acima, vemos que  $Z = E(X)$  é a esperança condicional  $E(X|\mathcal{F})$ . Isto prova a regra 3.

4º) Se a  $\sigma$ -álgebra  $\sigma(X)$ , gerada pela variável aleatória  $X$ , está contida em  $\mathcal{F}$ , então  $E(X|\mathcal{F}) = X$ .

Em particular, se  $X$  for função de  $Y$ ,  $\sigma(X) \subset \sigma(Y)$ , então  $E(X|Y) = X$ .

Para demonstrar esta regra, utilizamos o seguinte argumento:

$\sigma(X) \subset \mathcal{F} \Rightarrow \mathcal{F}$  contém toda a informação sobre a variável aleatória  $X$ . Neste caso, podemos tratar  $X$  como não aleatório. Logo:

$$E(X|\mathcal{F})(\omega) = E(X(\omega)|\mathcal{F}) = X(\omega) E(1|\mathcal{F}) = X(\omega).$$

Podemos estender esta regra para situações mais gerais:

5º)  $\sigma(X) \subset \mathcal{F}$ ,  $G$  uma variável aleatória:

$$E(XG|\mathcal{F}) = XE(G|\mathcal{F}) .$$

Em particular, se  $X$  for função de  $Y \Rightarrow \sigma(X) \subset \sigma(Y) \Rightarrow E(XG|Y) = XE(G|Y) .$

6º) Se  $\mathcal{F}$  e  $\mathcal{F}'$  são duas  $\sigma$ -álgebras,  $\mathcal{F} \subset \mathcal{F}'$ , então:

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

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

A segunda afirmação é aplicação direta da regra 4.

A primeira pode ser demonstrada a partir da propriedade (I), que diz que  $A \in \mathcal{F}$  e  $Z = E(X|\mathcal{F})$ ,  $E(ZI_A) = E(XI_A)$ .

Mas pela regra 5, como  $A \in \mathcal{F} \subset \mathcal{F}'$ ,

$$E[E(X|\mathcal{F}')|\mathcal{F}]I_A = E(E(X|\mathcal{F}')I_A|\mathcal{F}) = E(E(XI_A|\mathcal{F}')|\mathcal{F}) \quad (II)$$

Agora, se calcularmos a esperança de (II), aplicando a regra 2, temos:

$$E[E(E(XI_A|\mathcal{F}')|\mathcal{F})] = E(E(XI_A|\mathcal{F}')) = E(XI_A)$$

Assim,  $Z = E(E(XI_A|\mathcal{F}')|\mathcal{F})$  também satisfaz (I). Mas pela unicidade de  $E(X|\mathcal{F}) \Rightarrow Z = Z'$ . Isto prova a regra 6.

7º) Se  $X$  independe de  $\mathcal{F}$ , e a informação carregada pelo elemento aleatório  $G$  está contida em  $\mathcal{F}$ , então para qualquer função  $h(x, y)$ ,

$$E(h(X, G)|\mathcal{F}) = E(E_X(h(X, G))|\mathcal{F}) ,$$

onde  $E_X(h(X, G))$  significa que fixamos  $G$  e tomamos a esperança com respeito a  $X$ .

Ilustraremos esta regra com o seguinte exemplo:

*Exemplo:* Sejam  $X$  e  $Y$  duas variáveis aleatórias independentes. Portanto, pelas regras 7 e 5, temos:

$$E(XY|Y) = E(E_X(XY)|Y) = E(YE(X)|Y) = YE(X)$$

$$E(X+Y|Y) = E(E_X(X+Y)|Y) = E(E(X)+Y|Y) = E(X)+Y$$

Segue agora, dois exemplos para ilustrar a utilização das regras:

*Exemplo:* Movimento browniano

$B = (B_t, t \geq 0)$  um movimento browniano.

$\mathcal{F}_s = \sigma(B_x, x \leq s)$ . Calcularemos  $E(B_t|\mathcal{F}_s)$ .

$$E(B_t | \mathcal{F}_s) = E(B_t | \mathcal{F}_x, x \leq s), s \geq 0 .$$

Se  $s \geq t$  então  $\mathcal{F}_t \subset \mathcal{F}_s$  e pela regra 4  $E(B_t | \mathcal{F}_s) = B_t$  .

Se  $s < t$  , então pela regra 1:

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

Mas  $B_t - B_s$  e  $\mathcal{F}_s$  são independentes, e pela regra três, temos:

$$E(B_t - B_s | \mathcal{F}_s) = E(B_t - B_s) = 0 .$$

Além disso,  $\sigma(B_s) \subset \sigma(B_x, x \leq s) = \mathcal{F}_s$  , logo  $E(B_s | \mathcal{F}_s) = B_s$  .

Portanto,  $E(B_t | \mathcal{F}_s) = B_{\min(t, s)}$  .

**Exemplo:** Movimento browniano quadrático

$B = (B_t, t \geq 0)$  um movimento browniano.

$X_t = B_t^2 - t, t \geq 0$  Um processo estocástico.

Pelos mesmos argumentos do exemplo anterior, temos que  $E(X_t | \mathcal{F}_s) = X_t, s \geq t$  .

Agora, para  $s < t$  , temos:

$$B_t^2 - t = [(B_t - B_s) + B_s]^2 - t = (B_t - B_s)^2 + B_s^2 + 2(B_t - B_s)B_s - t .$$

Tomando a esperança condicional, temos:

$$E(X_t | \mathcal{F}_s) = E[(B_t - B_s)^2 | \mathcal{F}_s] + E[B_s^2 | \mathcal{F}_s] + 2E[(B_t - B_s)B_s | \mathcal{F}_s] - t$$

Mas  $(B_t - B_s)$  e  $(B_t - B_s)^2$  são independentes de  $\mathcal{F}_s$  , e  $\sigma(B_s^2) \subset \sigma(B_s) \subset \sigma(\mathcal{F}_s)$  .

Aplicando as regras 3, 4 e 5, temos:

$$E(X_t | \mathcal{F}_s) = E[(B_t - B_s)^2] + B_s^2 + 2B_s E[B_t - B_s] - t = t - s + B_s^2 + 0 - t = B_s^2 - s = X_s .$$

Portanto,  $E(X_t | \mathcal{F}_s) = X_{\min(t, s)}$  .

### 1.6.3 Projeção da Esperança Condicional

Nesta seção, consideraremos  $\mathcal{F}$  uma  $\sigma$ -álgebra, e denotaremos por  $L^2(\mathcal{F})$  a coleção de variáveis aleatórias  $Z$  em  $\Omega$ , satisfazendo as seguintes condições:

- $Z$  tem 2º momento finito:  $E(Z^2) < \infty$  ;
- A informação carregada por  $Z$  está contida em  $\mathcal{F}$  :  
 $\sigma(Z) \subset \mathcal{F}$ . Se  $\mathcal{F} = \sigma(Y) \Rightarrow Z = f(Y)$  .

A variável aleatória  $E(X|\mathcal{F})$  pode ser entendida como uma versão atualizada da esperança de  $X$  dada a informação  $\mathcal{F}$ . A esperança condicional tem certa propriedade de otimalidade na classe  $L^2(\mathcal{F})$ , é a chamada propriedade da projeção.

*Propriedade da projeção:* Seja  $X$  uma variável aleatória com  $E(X^2) < \infty$ . A esperança condicional  $E(X|\mathcal{F})$  é a variável aleatória em  $L^2(\mathcal{F})$  que é a mais próxima de  $X$  no sentido quadrático médio. Isto significa que:

$$E([X - E(X|\mathcal{F})]^2) = \min_{Z \in L^2(\mathcal{F})} E[(X - Z)^2] . \quad (1.6.3.1)$$

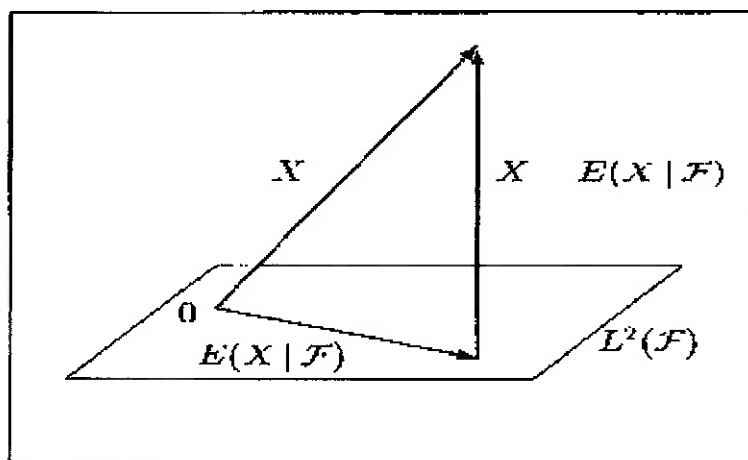


Figura 1.6.3.1: Ilustração da propriedade da projeção da esperança condicional  $E[X|\mathcal{F}]$ . Se  $E[X^2] < \infty$ , e  $E[Y^2] < \infty$ , então  $\langle X, Y \rangle = E[XY]$  define um produto interno, e  $\|X - Y\| = \sqrt{\langle X - Y, X - Y \rangle}$  é a distância entre  $X$  e  $Y$ . Assim como no espaço Euclidiano, dizemos que  $X$  e  $Y$  são ortogonais se  $\langle X, Y \rangle = 0$ . Neste sentido,  $E[X|\mathcal{F}]$  é a projeção ortogonal de  $X$  em  $L^2(\mathcal{F})$ .  $\langle X - E[X|\mathcal{F}], Z \rangle = 0$  para todo  $Z \in L^2(\mathcal{F})$ , e  $\langle X - Z, X - Z \rangle$  é mínimo para  $Z = E[X|\mathcal{F}]$ .

Na figura 1.6.3.1 acima verificamos esta propriedade. Neste sentido,  $E(X|\mathcal{F})$  é a projeção da variável aleatória  $X$  em  $L^2(\mathcal{F})$ .

Se  $\mathcal{F} = \sigma(Y)$ ,  $E(X|Y)$  é a função de  $Y$  que tem o segundo momento finito e que é mais próxima de  $X$  no sentido quadrático médio.

As vezes nos referimos a  $E(X|\mathcal{F})$  como a melhor previsão de  $X$  dado  $\mathcal{F}$ . Para

que a palavra *predição* possua algum sentido, reconsidere os exemplos da seção anterior. Provamos que para  $s \leq t$ , as melhores predições para os futuros de  $B_t$  e  $B_t^2 - t$  dadas as informações sobre o movimento browniano até o momento eram  $B_s$  e  $B_s^2 - s$ . Esta propriedade representa toda uma classe de “*martingal*” com segundo momento finito. A melhor predição para o momento futuro é o momento presente.

Faremos aqui uma pequena pausa, para falar sobre a *desigualdade de Jensen* e a *desigualdade de Cauchy-Schwarz* que utilizaremos mais a seguir. Não demonstraremos aqui estes resultados.

### 1) Desigualdade de Jensen:

Seja  $f$  uma função convexa em  $\mathbb{R}$ . Se  $E(|X|)$  e  $E(|f(X)|)$  são finitos, então:

$$f(E(X)) \leq E(f(X)) .$$

Em particular,

$$E(|X|^q)^{1/q} \leq E(|X|^p)^{1/p}, 0 < q < p .$$

Esta é a desigualdade de *Lyapunov*.

A desigualdade de Jensen continua válida para o caso de esperança condicional. Portanto, se  $\mathcal{F}$  for uma  $\sigma$ -álgebra em  $\Omega$ :

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

Em particular:

$$|E(X|\mathcal{F})| \leq E(|X|\mathcal{F}) \quad \text{e} \quad [E(X|\mathcal{F})]^2 \leq E(X^2|\mathcal{F}) .$$

### 2) Desigualdade de Cauchy-Schwarz:

Dadas duas variáveis aleatórias  $X$  e  $Y$ , temos:

$$E(|XY|) \leq [E(X^2)]^{1/2} [E(Y^2)]^{1/2} .$$

Agora, veremos a prova da propriedade (1.6.3.1). Calcularemos a esperança utilizando somente as regras da seção anterior. Primeiramente, note que  $Z' = E(X|\mathcal{F}) \in L^2(\mathcal{F})$ , pois carrega informação sobre  $\mathcal{F}$  somente e possui segundo momento finito, pois pela *desigualdade de Jensen*, juntamente com a *regra 2*:

$$E[(E(X|\mathcal{F}))^2] \leq E[(E(X^2|\mathcal{F}))] = E[X^2] < \infty$$

Agora, seja  $Z \in L^2(\mathcal{F})$  qualquer. Então:

$$E[(X-Z)^2] = E[((X-Z') + (Z'-Z))^2] = E[(X-Z')^2] + E[(Z'-Z)^2] + 2E[(X-Z')(Z'-Z)]$$



Mas  $Z, Z' \in L^2(\mathcal{F}) \Rightarrow Z - Z' \in L^2(\mathcal{F})$ . Pela regra 5:

$$E[(X - Z')(Z' - Z) | \mathcal{F}] = (Z' - Z)E[(X - Z') | \mathcal{F}]$$

Pelas regras 1 e 4:

$$E(X - Z' | \mathcal{F}) = E(X | \mathcal{F}) - E(Z' | \mathcal{F}) = Z' - Z' = 0.$$

Portanto, temos:

$$E[(X - Z)^2] = E[(X - Z')^2] + E[(Z' - Z)^2] \geq E[(X - Z')^2]$$

Portanto, a igualdade só é atingida quando  $Z = Z'$ . Logo,  $Z' = E(X | \mathcal{F})$  realmente representa o elemento de  $L^2(\mathcal{F})$ , em que o mínimo de  $E[(X - Z)^2]$  é atingido.

## 1.7 Martingais

A noção de martingal é essencial para o entendimento da integral de Ito. A integral estocástica de Ito é construída constituindo martingais. A ideia por trás de um martingal é a de um jogo justo, em que o ganho líquido é calculado via esperança condicional.

Assuma que  $(\mathcal{F}_t, t \geq 0)$  seja uma coleção de  $\sigma$ -álgebras em um mesmo espaço  $\Omega$  e que toda  $\mathcal{F}_t$  é subconjunto de uma  $\sigma$ -álgebra  $\mathcal{F}$  em  $\Omega$ .

**Definição:** Dizemos que a coleção  $(\mathcal{F}_t, t \geq 0)$  de  $\sigma$ -álgebras é uma filtração se  $\mathcal{F}_s \subset \mathcal{F}_t$  para todo  $0 \leq s \leq t$ .

Assim, a filtração é um fluxo crescente de informações. Se  $(\mathcal{F}_n, n = 0, 1, \dots)$  é uma sequência de  $\sigma$ -álgebras em  $\Omega$  e  $\mathcal{F}_n \subset \mathcal{F}_{n+1}$  para todo  $n$ , chamamos então  $(\mathcal{F}_n)$  de filtração também.

Para nossas aplicações, uma filtração está sempre ligada a um processo estocástico.

**Definição:** O processo estocástico  $Y = (Y_t, t \geq 0)$  é dito adaptado à filtração  $(\mathcal{F}_t, t \geq 0)$  se  $\sigma(Y_t) \subset \mathcal{F}_t, \forall t \geq 0$ .

Um processo estocástico  $Y = (Y_t, t \geq 0)$  é sempre adaptado à filtração natural gerada por  $Y$ ,  $\mathcal{F}_t = \sigma(Y_s, s \leq t)$ .

Assim, quando dizemos que um processo estocástico  $Y$  é adaptado a uma filtração  $(\mathcal{F}_t, t \geq 0)$ , então isto significa que  $Y$  não carrega mais informação que  $\mathcal{F}_t$ .

Se  $Y = (Y_n, n = 0, 1, \dots)$  for um processo estocástico discreto, então definimos a sua

adaptabilidade de maneira análoga. Para uma filtração  $(\mathcal{F}_n, n=0, 1, \dots)$  é necessário que  $\sigma(Y_n) \subset \mathcal{F}_n$ .

Alguns exemplos de processos adaptados:

*Exemplo:* Seja  $(B_t, t \geq 0)$  um movimento browniano e  $\mathcal{F}_t$  sua filtração natural. Um processo estocástico da forma  $X_t = f(t, B_t), t \geq 0$ , e  $f$  é uma função de duas variáveis é adaptado à  $\mathcal{F}_t, t \geq 0$ . Alguns exemplos de  $f$ :

- $X_t^{(1)} = B_t$  ;
- $X_t^{(2)} = B_t^2$  ;
- $X_t^{(3)} = B_t^2 - t$  ;
- $X_t^{(4)} = B_t^3$  ;
- $X_t^{(5)} = B_t^4$  .

Processos que dependem de todo o passado do movimento browniano também podem ser adaptados:

- $X_t^{(6)} = \max_{0 \leq s \leq t} B_s$  ;
- $X_t^{(7)} = \min_{0 \leq s \leq t} B_s^2$  ;

Se um processo estocástico  $Y$  é adaptado à filtração natural browniana  $(\mathcal{F}_t, t \geq 0)$ , diremos então que  $Y$  é adaptado ao movimento browniano. Isto significa que  $Y_t$  é função de  $B_s, s \leq t$ . Os exemplos a seguir não são adaptados ao movimento browniano:

- $X_t^{(8)} = B_{t+1}$  ;
- $X_t^{(9)} = B_T - B_t, T > 0$  fixo ;
- $X_t^{(10)} = B_T + B_t, T > 0$  fixo .

Nestes 3 últimos exemplos seria necessário ter informações sobre o processo estocástico em momentos futuros.

*Exemplo:* Aumentando a filtração

Considere o movimento browniano  $B = (B_t, t \geq 0)$  e a sua correspondente filtração natural  $\mathcal{F}_t = \sigma(B_s, s \leq t)$ . O processo estocástico  $X_t = B_t^2$  gera a filtração natural:

$$\mathcal{F}_t' = \sigma(B_s^2, s \leq t) ,$$

que é menor que  $\mathcal{F}_t$ , pois a partir de  $B_t^2$  o máximo que podemos fazer é reconstruir toda a informação de  $|B_t|$ , mas não de  $B_t$ .

Então,  $\mathcal{F}_t$  também é uma filtração natural de  $(B_t^2)$ . Logo, podemos trabalhar com diferentes filtrações para um processo estocástico.

Agora, considere um processo estocástico  $X=(X_t, t \geq 0)$  em  $\Omega$  e suponha que tenhamos a informação  $\mathcal{F}_s$  no instante  $s$  presente. Como esta informação influencia nosso conhecimento sobre o comportamento de  $X$  no futuro? Se  $\mathcal{F}_s$  e  $X$  são dependentes, é de se esperar que o conhecimento sobre  $\mathcal{F}_s$  reduza a incerteza sobre o comportamento de  $X$  no futuro. Se soubermos que certos eventos ocorreram no passado podemos incluir este conhecimento nos cálculos. Assim, o fato de conhecermos  $\mathcal{F}_s$  implica que podemos prever  $X_t$  melhor do que sem  $\mathcal{F}_s$ . A ferramenta matemática para descrever este ganho de informação é a esperança condicional:  $E(X_t | \mathcal{F}_s), 0 \leq s \leq t$ .

Nas seções anteriores, vimos que  $E(X_t | \mathcal{F}_s)$  é a melhor previsão para  $X_t$  dado  $\mathcal{F}_s$ . Além disso, dos exemplos na seção 1.6.2 (movimento browniano e movimento browniano quadrático), as melhores previsões para  $X_t = B_t$  e  $X_t = B_t^2 - t$  satisfazem  $E(X_t | \mathcal{F}_s) = X_s$ . Para estes processos, a melhor previsão é o seu valor presente. Claramente, isso pode ser mudado se a filtração for alterada.

Vejamos agora a definição de martingal:

*Definição:* O processo estocástico  $X=(X_t, t \geq 0)$  é chamado martingal contínuo no tempo com respeito à filtração  $(\mathcal{F}_t, t \geq 0)$ , e escrevemos  $(X, \mathcal{F}_t)$ , se:

- $E(|X_t|) < \infty$  ;
- $X$  é adaptado à  $(\mathcal{F}_t)$  ;
- $E(X_t | \mathcal{F}_s) = X_s, 0 \leq s \leq t$ , ou seja,  $X_s$  é a melhor previsão de  $X_t$  dado  $\mathcal{F}_s$ .

Também é possível descrever um martingal discreto no tempo  $X=(X_n, n=0, 1, \dots)$ . Neste caso, adaptamos a última propriedade da definição anterior:

- $E(X_{n+k} | \mathcal{F}_n) = X_n, k \geq 0$ .

Note que neste caso, é suficiente que esta condição seja satisfeita para  $k=1$ , pois pela *regra 6*, temos:

$$\begin{aligned} E(X_{n+1} | \mathcal{F}_n) &= E(E(X_{n+1} | \mathcal{F}_{n+1}) | \mathcal{F}_n) = E(X_{n+1} | \mathcal{F}_n) \\ &= E(E(X_{n+2} | \mathcal{F}_{n+2}) | \mathcal{F}_n) = E(X_{n+2} | \mathcal{F}_n) \\ &= \dots = E(X_{n+k} | \mathcal{F}_n) \end{aligned}$$

Assim, podemos agora definir martingal em tempo discreto:

**Definição:** O processo estocástico  $X=(X_n, n=0,1,\dots)$  é chamado martingal de tempo discreto com respeito à filtração  $(\mathcal{F}_n, n=0,1,\dots)$ , e escrevemos  $(X, \mathcal{F}_n)$ , se:

- $E(|X_n|) < \infty$  ;
- $X$  é adaptado à  $(\mathcal{F}_n)$  ;
- $E(X_{n+1}|\mathcal{F}_n) = X_n, \forall n=0,1,\dots$ , ou seja,  $X_n$  é a melhor predição de  $X_{n+1}$  dado  $\mathcal{F}_n$ .

Note que  $E(Y_{n+1}|\mathcal{F}_n) = 0$ , onde  $Y_{n+1} = X_{n+1} - X_n, n=0,1,\dots$ , é outra forma de escrever a última propriedade da definição anterior.

A sequência  $(Y_n)$  é chamada sequência das diferenças de martingal com respeito à filtração  $(\mathcal{F}_n)$ . Frequentemente omitiremos a filtração na apresentação de um martingal, pois isso será óbvio pelo contexto.

Um martingal possui a propriedade de que sua esperança é constante. De fato,  $E(X_t|\mathcal{F}_s) = X_s, s < t$ . Pela regra 2, obtemos:

$$E(X_s) = E(E(X_t|\mathcal{F}_s)) = E(X_t), \forall s, t.$$

Isto fornece uma maneira fácil de se mostrar que um processo estocástico não é um martingal. Por exemplo, se  $B$  é um movimento browniano,  $E(B_t^2) = t, \forall t$ . Portanto,  $(B_t^2)$  não é um martingal. A recíproca, no entanto, não é verdadeira, pois  $E(B_t^3) = 0, \forall t$ , mas  $(B_t^3)$  não é um martingal como veremos mais adiante. Vejamos alguns exemplos:

**Exemplo:** Soma parcial de variáveis aleatórias constitui um martingal

Seja  $(Z_n)$  uma sequência de variáveis aleatórias independentes, com esperança finita, e  $Z_0 = 0$ . Considere a soma parcial

$$R_n = \sum_{i=1}^n Z_i, n \geq 0$$

e sua filtração natural correspondente  $\mathcal{F}_n = \sigma(R_0, \dots, R_n)$ . Note que  $\mathcal{F}_n = \sigma(Z_0, \dots, Z_n)$ .

De fato,  $(R_0, \dots, R_n)$  e  $(Z_0, \dots, Z_n)$  contém a mesma informação, pois  $R_i = Z_1 + \dots + Z_i$  e  $Z_i = R_i - R_{i-1}$ . Aplicando as regras 1, 3 e 4, obtemos:

$$E(R_{n+1}|\mathcal{F}_n) = E(R_n|\mathcal{F}_n) + E(Z_{n+1}|\mathcal{F}_n) = R_n + E(Z_{n+1})$$

Logo, se  $E(Z_n) = 0$  para todo  $n$ , então  $(R_n, n=0,1,\dots)$  é um martingal com respeito à  $(\mathcal{F}_n, n=0,1,\dots)$ .

*Exemplo:* Coletando informações sobre uma variável aleatória

Seja  $Z$  uma variável aleatória em  $\Omega$  com  $E(|Z|) < \infty$  e  $(\mathcal{F}_t, t \geq 0)$  uma filtração em  $\Omega$ . Defina o processo estocástico  $X$  a seguir:

$$X_t = E(Z | \mathcal{F}_t), t \geq 0.$$

Como  $\mathcal{F}_t$  é crescente com o tempo,  $X_t$  fornece cada vez mais informação sobre  $Z$ . Em particular, se  $\sigma(Z) \subset \mathcal{F}_t$  para algum  $t$ , então  $X_t = Z$ .

Vamos mostrar que  $X$  é um martingal. Pela desigualdade de Jensen, e pela regra 2, temos:

$$E(|X_t|) = E(|E(Z | \mathcal{F}_t)|) \leq E(E(|Z| | \mathcal{F}_t)) = E(|Z|) < \infty$$

Além disso,  $X_t$  é obtido a partir da informação  $\mathcal{F}_t$ , então  $\sigma(X_t) \subset \mathcal{F}_t$ .

Por último,  $E(X_t | \mathcal{F}_s) = E(E(Z | \mathcal{F}_t) | \mathcal{F}_s) = E(Z | \mathcal{F}_s) = X_s$ .

Assim,  $X$  obedece às propriedades que definem um martingal de tempo contínuo.

*Exemplo:* O movimento browniano é um martingal

Seja  $B = (B_t, t \geq 0)$  um movimento browniano. Dos exemplos vistos na seção 1.6.2,  $(B_t)$  e  $(B_t^2 - t)$  são martingais com respeito à filtração natural  $\mathcal{F}_t = \sigma(B_s, s \leq t)$ . Agora, considere  $X_t = B_t^3 - 3t B_t$ .

$$E(X_t | \mathcal{F}_s) = E(B_t^3 | \mathcal{F}_s) - 3E(t B_t | \mathcal{F}_s)$$

$E(B_t^3) = E([(B_t - B_s) + B_s]^3) = E((B_t - B_s)^3) + E(B_s^3) + 3E((B_t - B_s)^2 B_s) + 3E((B_t - B_s) B_s^2)$ . Aqui, omitimos a condição das esperanças, para que a expressão ficasse “mais limpa”, mas na verdade todas essas esperanças são calculadas dado  $\mathcal{F}_s$ .

Mas  $(B_t - B_s)$ ,  $(B_t - B_s)^2$  e  $(B_t - B_s)^3$  são independentes de  $\mathcal{F}_s$ , portanto:

$$E(B_t^3 | \mathcal{F}_s) = 0 + B_s^3 + 3B_s(t-s) + 3B_s \cdot 0 = B_s^3 + 3B_s(t-s)$$

Daí vemos que  $B_t^3$  não é um martingal, mas:

$$E(X_t | \mathcal{F}_s) = 0 + B_s^3 + 3B_s(t-s) + 3B_s \cdot 0 - 3t B_s = B_s^3 + 3s B_s = X_s.$$

Portanto,  $(X_t, (\mathcal{F}_t))$  é um martingal.

*Exemplo:* Transformada de um martingal

Seja  $Y=(Y_n, n=0, 1, \dots)$  uma sequência de diferenças de martingais com respeito à filtração  $(\mathcal{F}_n, n=0, 1, \dots)$ . Considere o processo estocástico  $C=(C_n, n=0, 1, \dots)$  e assumamos que para todo  $n$ , a informação carregada por  $C_n$  está contida em  $\mathcal{F}_{n-1}$ , isto é,  $\sigma(C_n) \subset \mathcal{F}_{n-1}$ . Isto significa que dado  $\mathcal{F}_{n-1}$ , conhecemos  $C_n$  completamente no instante  $n-1$ . A sequência  $C_n$  é dita previsível ou predizível com respeito à  $(\mathcal{F}_n)$ .

Agora, defina o processo estocástico:

$$X_0=0, \quad X_n=\sum_{i=1}^n C_i Y_i, \quad n \geq 1$$

Por razões óbvias, o processo  $X$  é denotado por  $C \circ Y$ . Ela é chamada transformada de martingal de  $Y$  por  $C$ .

Esta transformada é martingal se  $E(C_n^2) < \infty$  e  $E(Y_n^2) < \infty$  para todo  $n$ . De fato, pela desigualdade de Cauchy-Schwarz:

- $E(|X_n|) \leq \sum_{i=1}^n E(|C_i Y_i|) \leq \sum_{i=1}^n [E(C_i^2)E(Y_i^2)]^{1/2} < \infty$  ;
- $X_n$  é claramente, adaptado a  $\mathcal{F}_n$ , pois  $Y_i$  não carrega mais informação que  $\mathcal{F}_n$ , e  $(C_1, \dots, C_n)$  é previsível;
- Por último, aplicando a *regra 5*, lembrando que  $C_n$  é previsível:

$$E(X_n - X_{n-1} | \mathcal{F}_{n-1}) = E(C_n Y_n | \mathcal{F}_{n-1}) = C_n E(Y_n | \mathcal{F}_{n-1}) = 0.$$

Aqui neste último passo, utilizamos o fato de que a última propriedade de um martingal pode ser escrita como uma sequência de diferenças de um martingal.

Portanto,  $(X_n, (\mathcal{F}_n))$  é um martingal.

### 1.7.1 Interpretação como jogo justo

Suponha que estamos em um jogo de tempo contínuo, e que a cada instante de tempo  $t$ , tenhamos  $X_t$  como o valor do jogo. Suponha também que  $(X_t)$  é adaptado à filtração  $(\mathcal{F}_t)$ . Então  $(X_t - X_s)$  pode ser pensado como ganho obtido no intervalo  $[s, t]$  do jogo. Portanto, a predição que pode ser feita em  $s$  sobre o ganho que teremos em  $t$  dada a informação  $\mathcal{F}_s$ , é dada por:

$$E(X_t - X_s | \mathcal{F}_s) = E(X_t | \mathcal{F}_s) - X_s$$

Se  $(X, (\mathcal{F}_t))$  é um martingal, o lado direito da equação se anula, Isto significa que a melhor predição para os ganhos futuros é zero. Ganho zero é exatamente o que se espera de um jogo justo, ou seja, espera-se que não haja arbitragem.

## 2. A Integral Estocástica e Simulação Numérica

Neste capítulo o objetivo é apresentar a integral estocástica de Ito. Também iremos apresentar aqui a parte de simulação numérica. Como a simulação é o foco deste trabalho, não daremos muita importância para a formalização dos conceitos por trás de toda a teoria de integral estocástica, apenas daremos uma ideia do que se trata, ou seja, será apresentado aqui o mínimo possível de teoria.

### 2.1 A Integral de Riemann

Aqui iremos apenas dar uma noção da integral de Riemann. Suponha por simplicidade que  $f$  seja uma função a valores reais definida em  $[0, 1]$ . Poderíamos considerar um intervalo  $[a, b]$  ao invés disso. Considere a seguinte partição do intervalo  $[0, 1]$ :

$$\tau_n: 0 = t_0 < t_1 < \dots < t_{n-1} < t_n = 1,$$

e defina

$$\Delta_i = t_i - t_{i-1}, \quad i = 1, \dots, n.$$

Uma partição intermediária,  $\sigma_n$  de  $\tau_n$  pode ser construída tomando qualquer valor  $y_i$  tal que  $t_{i-1} < y_i < t_i$ . Dadas as partições  $\sigma_n$  e  $\tau_n$  podemos definir a soma de Riemann

$$S_n = S_n(\tau_n, \sigma_n) = \sum_{i=1}^n f(y_i)(t_i - t_{i-1}) = \sum_{i=1}^n f(y_i)\Delta_i.$$

Basicamente, essa soma representa uma aproximação da área entre o gráfico da função e o eixo  $x$ , desde que  $f$  seja não negativa.

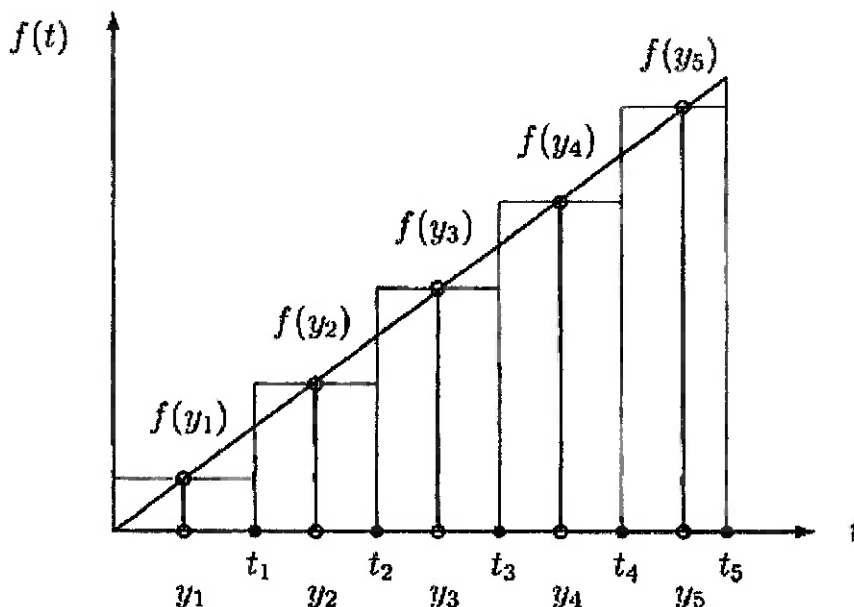


Figura 2.1.1: Uma ilustração da soma de Riemann, com partição  $(t_i)$  e partição intermediária  $(y_i)$ . A soma das áreas dos retângulos aproxima a área entre o gráfico e o eixo  $t$ .



A figura 2.1.1 acima ilustra este fato. Agora, definindo

$$\Delta = \max_{i=1, \dots, n} \Delta_i,$$

podemos tomar o limite para  $\Delta \rightarrow 0$ . Claramente, quando  $\Delta \rightarrow 0$  então  $n \rightarrow \infty$ .

Se o limite

$$S = \lim_{n \rightarrow \infty} S_n = \lim_{n \rightarrow \infty} \sum_{i=1}^n f(y_i) \Delta_i,$$

existir e o limite for independente da escolha das partições  $\tau$  e  $\sigma_n$  e  $\tau_n$ , então  $S$  é chamada a integral ordinária ou integral de Riemann de  $f$  em  $[0,1]$ .

## 2.2 Integral estocástica

Em teoria de probabilidades, é comum denotarmos a esperança de uma variável aleatória  $X$  por

$$E(X) = \int_{-\infty}^{\infty} t dF_X(t),$$

onde  $F_X$  é a função de distribuição de  $X$ . Isto significa que  $E(X)$  definida como uma integral de Riemann-Stieltjes, ou integral de Lebesgue-Stieltjes. De maneira grosseira,

$$\int_{-\infty}^{\infty} t dF_X(t) \approx \sum_i y_i [F_X(t_i) - F_X(t_{i-1})],$$

para uma partição  $(t_i)$  de  $\mathbb{R}$ , e uma partição  $(y_i)$  intermediária correspondente.

Além disso, dos cursos de cálculo elementar, é conhecida a definição da integral de uma função  $f$  em relação a outra  $g$ . A saber:

$$\int_0^1 f(t) dg(t) = \int_0^1 f(t) g'(t) dt,$$

desde que  $g'(t)$  exista.

Ambas as integrais acima são exemplos de como é possível encarar o problema de se integrar uma função com respeito a outra. O objetivo aqui neste capítulo é sugerir um método de se calcular

$$\int_0^1 f(t) dB_t(\omega), \quad (2.2.1)$$

onde  $f$  é uma função ou um processo estocástico, e  $B_t(\omega)$  é um movimento browniano. Mais do que isso, iremos também implementar um algoritmo para avaliar este tipo de

integral.

Então, da mesma forma como fizemos na seção anterior com a integral de Riemann, dada uma partição  $(t_i)$  de  $[0, T]$  e uma partição intermediária  $(y_i)$  correspondente, podemos então desta forma aproximar a integral

$$\int_0^T f(t) dB_t(\omega)$$

por uma soma da forma

$$\sum_i y_i [B_{t_i}(\omega) - B_{t_{i-1}}(\omega)] .$$

O problema deste tipo de soma é que quando  $f$  também for um processo estocástico, como por exemplo, o próprio  $B_t(\omega)$ . Na forma como a integral de Riemann é definida, não importa a partição que tomamos, nem a sua correspondente partição intermediária, no limite para  $\max \Delta t \rightarrow 0$  a integral converge sempre para o mesmo valor. Mas esse não é o caso na integral estocástica de Riemann-Stieltjes. Dependendo da partição que tomamos, e de sua correspondente partição intermediária, o valor da integral muda.

Para ilustrar este fato, tomaremos 2 partições intermediárias  $(y_i)$  diferentes. Serão elas:

- $y_i = t_i$  (Canto esquerdo de  $[t_i, t_{i+1}]$ ),  $i=0, 1, \dots, N-1$ , e
- $y_i = \frac{t_i + t_{i+1}}{2}$  (Centro de  $[t_i, t_{i+1}]$ ),  $i=0, 1, \dots, N-1$ .

No primeiro caso, obtemos:

$$\begin{aligned} \sum_{i=0}^{N-1} B_{y_i}(\omega) (B_{t_{i+1}}(\omega) - B_{t_i}(\omega)) &= \frac{1}{2} \sum_{i=0}^{N-1} [B_{t_{i+1}}(\omega)^2 - B_{t_i}(\omega)^2 - (B_{t_{i+1}}(\omega) - B_{t_i}(\omega))^2] \\ &= \frac{1}{2} \left[ B_T(\omega)^2 - B_0(\omega)^2 - \underbrace{\sum_{i=0}^{N-1} (B_{t_{i+1}}(\omega) - B_{t_i}(\omega))^2}_{(I)} \right] . \end{aligned}$$

Pode ser demonstrado que o termo  $(I)$  destacado acima tem valor esperado  $T$  e variância da ordem do tamanho do passo da partição. Portanto,

$$(I) \xrightarrow{N \rightarrow \infty} T \Rightarrow \int_0^T B_t(\omega) dB_t(\omega) = \frac{1}{2} B_T(\omega)^2 - \frac{1}{2} T . \quad (2.2.2)$$

Agora, no segundo caso:

Para  $y_i = \frac{1}{2}[t_i + t_{i+1}]$  significa que precisamos avaliar  $B_t(\omega)$  em  $\frac{1}{2}[t_i + t_{i+1}]$ .

Pode ser demonstrado que se calcularmos  $y_i = \frac{1}{2}[B_{t_i}(\omega) + B_{t_{i+1}}(\omega)]$  e somarmos a isso um incremento independente com distribuição  $N(0, \Delta t/4)$  obtemos um valor para  $B_{y_i}(\omega)$  que mantém todas as condições da definição de movimento browniano visto anteriormente na seção 1.4. Portanto, temos:

$$\begin{aligned} \sum_{i=0}^{N-1} B_{y_i}(\omega)(B_{t_{i+1}}(\omega) - B_{t_i}(\omega)) &= \sum_{i=0}^{N-1} \left[ \frac{B_{t_{i+1}}(\omega) + B_{t_i}(\omega)}{2} + \Delta Z_i \right] (B_{t_{i+1}}(\omega) - B_{t_i}(\omega)) \\ &= \frac{1}{2}(B_T(\omega)^2 - B_0(\omega)^2) - \underbrace{\left[ \sum_{i=0}^{N-1} \Delta Z_i (B_{t_{i+1}}(\omega) - B_{t_i}(\omega)) \right]}_{(II)}, \end{aligned}$$

onde  $\Delta Z_i \sim N(0, \Delta t/4)$ . Pode ser demonstrado que o termo (II) destacado acima tem valor esperado 0 e variância da ordem do tamanho do passo da partição. Portanto, para esta partição,

$$(II) \xrightarrow{N \rightarrow \infty} 0 \Rightarrow \int_0^T B_t(\omega) dB_t(\omega) = \frac{1}{2} B_T(\omega)^2. \quad (2.2.3)$$

No primeiro caso, o que temos é o que conhecemos como a integral de Ito, onde a partição intermediária é tomada no canto esquerdo de cada subintervalo, e no segundo caso, temos a integral de Stratonovich, com a partição intermediária tomada no centro de cada subintervalo.

O algoritmo *script1.py* apresentado no final deste trabalho, foi implementado em Python e foi criado para ilustrar esta diferença. Na tabela a seguir, representamos os valores obtidos com o cálculo da integral (2.2.1) pelos métodos de Ito e de Stratonovich, com  $f(t) = B_t(\omega)$ . Neste exemplo usamos  $T = 1$ ,  $N = 1000$  e  $\Delta t = 1/N$ .

Ito	Ito - erro	Stratonovich	Strat - erro
0,176	0,035	0,726	0,016
-0,124	0,022	0,333	0,022
-0,085	0,011	0,430	0,004
0,557	0,009	1,056	0,008
-0,204	0,035	0,251	0,010
0,264	0,022	0,765	0,023
-0,344	0,026	0,139	0,008

-0,430	0,006	0,087	0,023
-0,443	0,006	0,047	0,016
0,669	0,039	1,126	0,004

Tabela 1: Diferença entre integrais de Ito e Stratonovich

Na tabela podemos observar o valor obtido pelo *script1.py* para a integral (2.2.1) quando tomamos uma partição com partição intermediária no canto esquerdo de cada subintervalo  $[t_i, t_{i+1}]$ , e quando tomamos uma partição intermediária no centro de  $[t_i, t_{i+1}]$ . Como sabemos o valor exato da integral (2.2.1) tanto pelo método de Ito quanto por Stratonovich, criamos assim, as colunas de erro acima.

O método para gerarmos a tabela acima foi o seguinte:

1. Foi gerado  $N = 1000$  valores  $dw_i \sim N(0, dt), i=1, \dots, 1000$ ;
2. A partir disto, foi gerado um vetor  $w$ , em que cada posição  $k=0, \dots, 999$  deste vetor é igual a soma de todos os  $dw[i], i=1, \dots, k$ , ou seja,  $w[0] = 0, w[k] = dw[0] + \dots + dw[k]$ ;
3. Este vetor  $w$  representa uma simulação de caminho browniano. Então o valor exato da integral (2.2.1) no caso em que  $f$  é o próprio caminho browniano é conhecido tanto para partição de Ito quanto para partição de Stratonovich, e seus valores são (2.2.2) e (2.2.3), respectivamente.
4. A partir disto, o que foi feito foi simplesmente fazer a soma:

$$\sum_{i=0}^{N-1} B_{y_i}(\omega) dB_t$$

A partir deste ponto o que foi feito foi comparar os valores obtidos na soma com o que seria o valor exato. E assim obtivemos as colunas de erros acima. Para  $N = 1000$  passos no intervalo  $[0, 1]$  pudemos observar que os erros se encontram na ordem de  $10^{-2}$ .

## 2.3 O Método de Euler-Maruyama

Uma equação diferencial estocástica escalar autônoma pode ser escrita na forma integral. Temos:

$$X(t) = X_0 + \int_0^t f(X(s)) ds + \int_0^t g(X(s)) dB_s(\omega) \quad (2.3.1)$$

Aqui,  $f$  e  $g$  são funções escalares e a condição inicial  $X_0$  é uma variável aleatória. A segunda integral do lado direito de (2.3.1) é calculada com respeito a um movimento

browniano, e daqui em diante estamos considerando a versão de Ito.

É comum escrever a equação (2.3.1) na sua forma diferencial. Assim, obtemos:

$$dX(t) = f(X(t))dt + g(X(t))dB_t(\omega), X(0) = X_0, 0 \leq t \leq T \quad (2.3.2)$$

Isto é apenas outra forma de dizer que  $X$  resolve (2.3.1).

Note que se  $g \equiv 0$  e  $X_0$  for constante, o problema se torna determinístico, e (2.3.2) se reduz a uma equação diferencial ordinária, a saber:

$$\frac{d}{dt}X(t) = f(X(t)), X(0) = X_0, 0 \leq t \leq T$$

Para podermos aplicar um método numérico em (2.3.2) sobre  $[0, T]$  primeiramente discretizamos o intervalo. Seja então  $\Delta t = T/L$  e  $\tau_j = j\Delta t$ . Nosso objetivo aqui é estimar  $X(\tau_j)$ , que será representado por  $X_j$ . O método de Euler-Maruyama é da forma

$$X_j = X_{j-1} + f(X_{j-1})\Delta t + g(X_{j-1})(B_{\tau_j}(\omega) - B_{\tau_{j-1}}(\omega)), j = 1, \dots, L \quad (2.3.3)$$

Para entendermos de onde vem (2.3.3), note que da forma integral em (2.3.1), temos:

$$X(\tau_j) = X(\tau_{j-1}) + \int_{\tau_{j-1}}^{\tau_j} f(X(s))ds + \int_{\tau_{j-1}}^{\tau_j} g(X(s))dB_s(\omega) \quad (2.3.4)$$

Cada termo do lado direito de (2.3.3) aproxima o seu correspondente em (2.3.4). Vale notar que no caso determinístico ( $g \equiv 0$  e  $X_0$  constante), (2.3.3) se reduz à forma do método de Euler.

Para ilustrar o método de Euler-Maruyama, o aplicaremos aqui em uma equação diferencial estocástica bem conhecida:

$$dX(t) = \lambda X(t)dt + \mu X(t)dB_t(\omega), X(0) = X_0 \quad (2.3.5)$$

onde  $\lambda$  e  $\mu$  são constantes reais. Então temos:

$$f(X) = \lambda X \quad \text{e} \quad g(X) = \mu X$$

Esta equação diferencial estocástica surge, por exemplo, em modelos de precificação de um ativo financeiro. Não demonstraremos aqui, mas a solução exata desta equação diferencial é:

$$X(t) = X_0 \exp\left\{\left(\lambda - \frac{1}{2}\mu^2\right)t + \mu B_t(\omega)\right\} \quad (2.3.6)$$

No algoritmo *script2.py* apresentado no final deste trabalho foi implementado o método

de Euler-Maruyama, e neste método utilizamos na equação (2.3.5) os valores  $\lambda=2$ ,  $\mu=1$ , e  $X_0=1$ . Foi criado o caminho browniano com no intervalo  $[0,1]$   $\delta t=2^{-8}$ , e calculamos a solução pela fórmula apresentada em (2.3.6). A seguir, aplicamos o método de Euler-Maruyama com  $\Delta t=R\delta t$  com  $R=4$  ( $R>1$  inteiro garante que os pontos estimados pelo Euler-Maruyama serão também calculados na solução exata (2.3.6)). Note ainda que:

$$B_{\tau_j} - B_{\tau_{j-1}} = B_{jR\Delta t} - B_{(j-1)R\Delta t} = \sum_{k=(j-1)R+1}^{jR} dB_k,$$

então, apenas a título de curiosidade, para calcular o próximo ponto  $X(\tau_j)$  a partir de um determinado ponto  $X(\tau_{j-1})$  estimado pelo método de Euler-Maruyama no algoritmo, basta somar todas as variações de  $B(\omega)$  entre os termos  $B_{\tau_{j-1}}$  e  $B_{\tau_j}$ . Como o algoritmo gera os termos  $dB_k$ , então a forma mais fácil de se fazer essa conta seria calcular de fato essa soma, já que em Python temos uma função pronta para fazer esse cálculo.

Agora, abaixo temos um exemplo de uma execução do algoritmo:

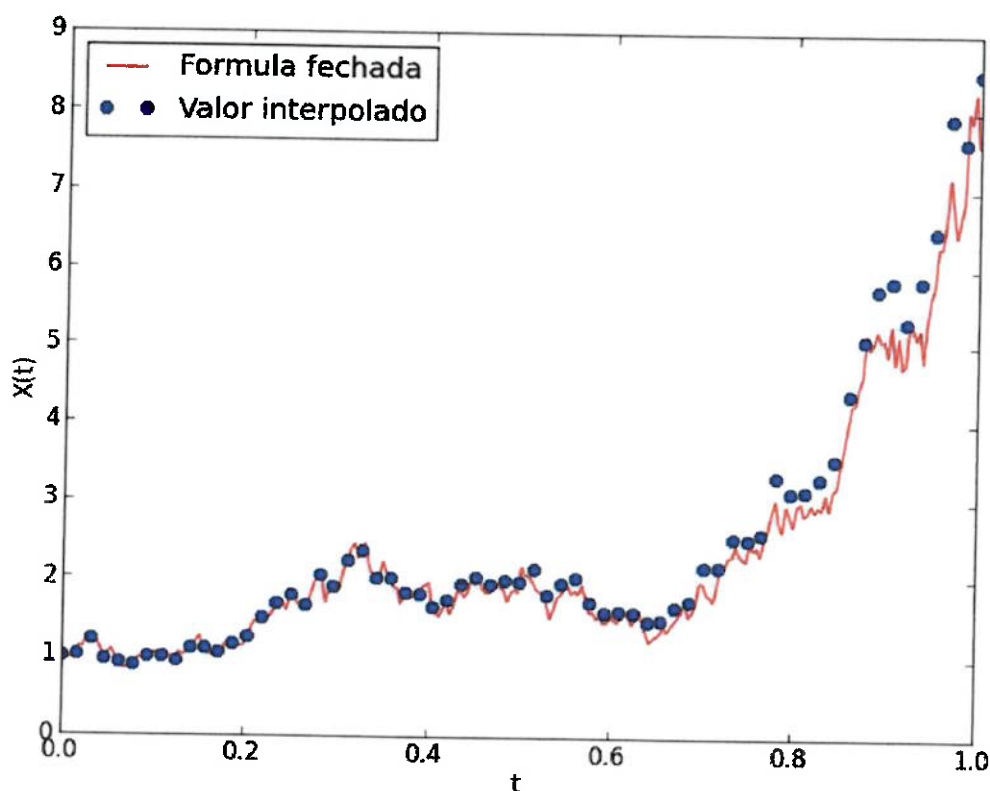


Gráfico 2.3.1. O método de Euler-Maruyama.

A linha contínua vermelha representa o gráfico da solução exata  $X$  da equação (2.3.6), e os pontos azuis representam a aproximação de  $X$  calculada nestes pontos. Podemos observar no gráfico acima que a aproximação pelo método é bastante boa, dada a sua simplicidade na implementação.

## 2.4 Convergência forte e fraca do método de EM

No *script2.py* mencionado na seção anterior a solução de EM se aproxima da solução real à medida que  $\Delta t$  decresce, ou seja, o método de EM converge para a solução real. Lembrando que tanto  $X_n$  quanto  $X(\tau_n)$  são variáveis aleatórias, para que possamos ter uma noção mais precisa da convergência, precisamos decidir uma forma de se medir a diferença entre elas. Uma forma de se encarar este problema seria calculando  $E(|X_n - X(\tau_n)|)$ . Um método dito possuir uma *ordem de convergência forte* igual a  $\gamma$  se existir uma constante  $C$  tal que

$$E(|X_n - X(\tau_n)|) \leq C \Delta t^\gamma \quad (2.4.1)$$

para qualquer  $\tau = n\Delta t$  fixo em  $[0, T]$  e  $\Delta t$  suficientemente pequeno.

Nos nossos testes numéricos, focaremos no erro no final do caminho, ou seja, em  $t = T$ , pois se a desigualdade (2.4.1) é válida para qualquer ponto do caminho, então é válido também para o fim do caminho. Portanto, definimos:

$$e_{\Delta t}^F := E(|X_L - X(T)|), \quad L\Delta t = T \quad (2.4.2)$$

Esta é a definição de erro no ponto final no sentido de convergência forte. Se a desigualdade (2.4.1) vale para  $\gamma = \frac{1}{2}$  em qualquer ponto de  $[0, T]$ , então certamente vale também para o ponto final. Então para  $\Delta t$  suficientemente pequeno, temos:

$$e_{\Delta t}^F \leq C \Delta t^{\frac{1}{2}} \quad (2.4.3)$$

O algoritmo *script3.py* apresentado no final deste trabalho mostra a convergência forte do método de Euler-Maruyama aplicado em (2.3.5). Aqui utilizamos os mesmos valores para  $\mu$ ,  $\tau$  e  $X_0$  que utilizamos em *script2.py*. O processo seguido foi o seguinte:

1. Primeiro geramos 1000 caminhos aleatórios diferentes em  $[0, 1]$ , com  $\delta t = 2^{-9}$ ;
2. Para cada um desses caminhos foi aplicado o método de Euler-Maruyama, visto na seção anterior, com 5 tamanhos de passos distintos:  $\Delta t = 2^{p-1}$ ,  $p = 1, \dots, 5$ ;
3. Armazenamos o ponto final gerado pelo método no  $s$ -ésimo caminho e  $p$ -ésimo tamanho de passo em uma matriz, obtendo assim, uma matriz de tamanho  $1000 \times 5$ ;
4. Calculamos a média sobre todas as linhas desta matriz, obtendo assim um vetor com 5 entradas. Cada entrada é uma aproximação do erro pela média para cada tamanho de passo, ou seja, o  $p$ -ésimo elemento desse vetor é uma aproximação para  $e_{\Delta t}^F$  para  $\Delta t = 2^{p-1} \delta t$ .

Se a desigualdade (2.4.3) for próximo da igualdade, então tomando o logaritmo, temos:

$$\log e_{\Delta t}^F \approx \log C + \gamma \log \Delta t \quad (2.4.4)$$

Portanto, o próximo passo foi plotar o gráfico de  $\log e_{\Delta t}^F$  versus  $\log \Delta t$ . Plotamos ainda em listrado, uma linha reta com coeficiente angular igual a 0,5 apenas para efeito de

comparação.

Feito isso, obtivemos o coeficiente angular da reta ajustada pelo método dos mínimos quadrados. Neste gráfico obtivemos um  $\gamma = 0,5080578$ , com resíduo igual a 0,00023157. Nos vários testes que foi feito com este algoritmo o que se observou foi  $\gamma$  variando em torno de 0,5. Isto reforça o fato de que este método possui ordem de convergência forte  $\gamma = 0,5$ .

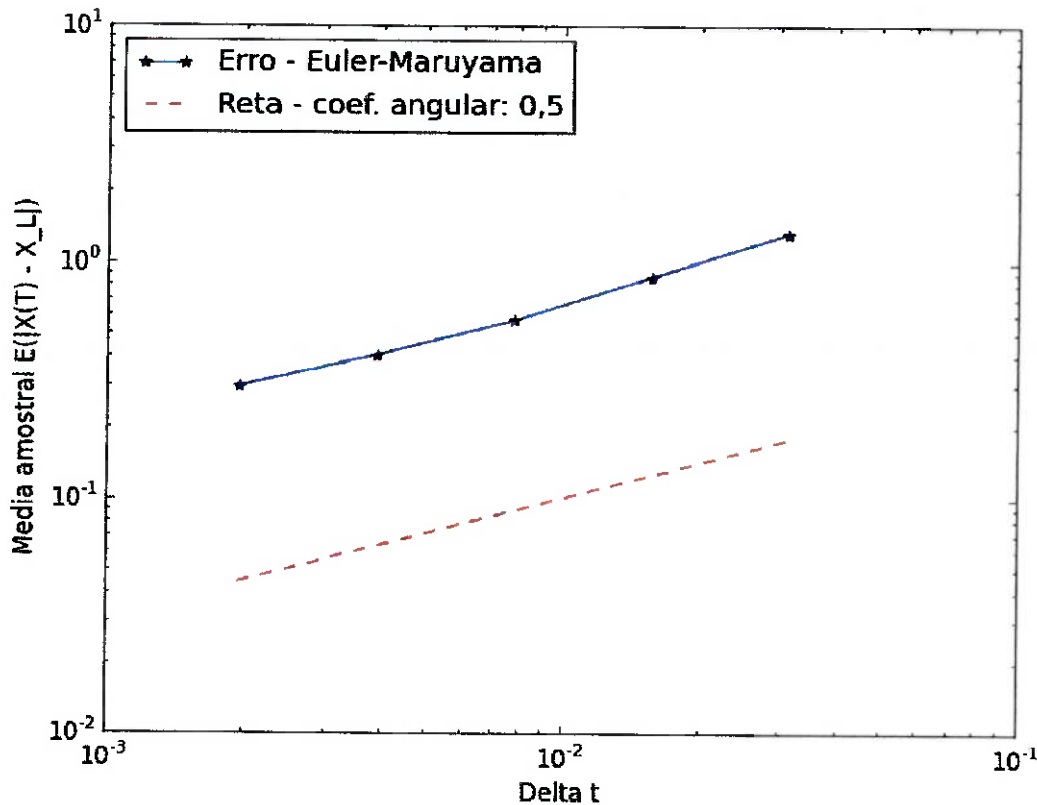


Figura 2.4.1: Convergência forte. Linha vermelha tracejada como referência, e azul com asteriscos são os erros para cada tamanho de passo.

Nessa simulação de ordem de convergência forte estamos negligenciando aqui algumas fontes de erros, como por exemplo:

1. *Erro amostral*: Erro devido à aproximação de valor esperado pela média amostral;
2. *Vício de número aleatório*: Erro devido ao gerador de números aleatórios;
3. *Erro de arredondamento*: Erro computacional.

Destes 3 erros mencionados, para uma simulação computacional o principal deles é provavelmente o primeiro. Na referência [2] mencionada no fim deste trabalho, na hora de criar estes exemplos foram feitos experimentos para se encontrar uma quantidade de caminhos aleatórios gerados o suficiente e tamanho de passos pequenos o suficiente para que pudéssemos observar a ordem de convergência. Aqui, apenas utilizei os mesmos valores iniciais utilizados neste artigo.



Apesar da definição (2.4.1) de convergência forte envolver valor esperado ela possui implicações para simulações individuais. A desigualdade de Markov diz que se uma variável aleatória  $X$  tem valor esperado finito, então para qualquer  $a > 0$  temos:

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

Logo, tomando  $a = \Delta t^{1/4}$ , temos:

$$P(|X_n| - X(\tau) \geq \Delta t^{1/4}) \leq C \Delta t^{1/4} ,$$

ou equivalentemente,

$$P(|X_n| - X(\tau) < \Delta t^{1/4}) \geq 1 - C \Delta t^{1/4} .$$

Isso mostra que em um ponto fixo qualquer de  $[0, T]$  o erro é pequeno com probabilidade próxima de 1.

A ordem de convergência forte mede com que taxa a “média dos erros” cai à medida que  $\Delta t \rightarrow 0$ . Uma alternativa menos exigente seria medir a taxa com que o “erro das médias” cai. Isso nos leva a um outro conceito, o de *ordem de convergência fraca*. Um determinado método é dito possuir *ordem de convergência fraca* igual a  $\gamma$  se existir uma constante  $C$  tal que para todas as funções  $p$  pertencentes a alguma classe de funções:

$$|E(p(X_n)) - E(p(X(\tau)))| \leq C \Delta t^\gamma \tag{2.4.5}$$

em qualquer ponto fixo  $\tau = n \Delta t \in [0, T]$ , e  $\Delta t$  suficientemente pequeno. Tipicamente, as funções  $p$  utilizadas em (2.4.5) devem satisfazer condições de suavidade e crescimento polinomial. Aqui iremos utilizar  $p$  igual à função identidade. Para funções  $f$  e  $g$  adequadas pode-se mostrar que o método de Euler-Maruyama possui ordem de convergência fraca  $\gamma = 1$ .

Portanto, definimos:

$$e_{\Delta t}^f = |E(X_L) - E(X(T))| \tag{2.4.6}$$

onde  $L \Delta t = T$ , denota o erro no ponto final no sentido de convergência fraca. Portanto, em (2.4.5) com  $\gamma = 1$  é imediato que:

$$e_{\Delta t}^f \leq C \Delta t \tag{2.4.7}$$

para  $\Delta t$  suficientemente pequeno.

O algoritmo *script4.py* apresentado no final deste trabalho mostra a convergência fraca do método de Euler-Maruyama aplicado em (2.3.5) sobre  $[0, 1]$  para  $\lambda = 2$ ,  $\mu = 0, 1$  e  $X_0 = 1$ . O processo aqui foi um pouco diferente do processo utilizado para convergência forte:

1. Foram simulados 50000 caminhos brownianos discretizados com tamanho de passo  $\Delta t = 2^{p-10}$ , com  $1 \leq p \leq 5$ . Este processo possui um nível de vetorização a mais que o *script3.py*, isto aumenta o desempenho com o custo de se utilizar mais memória;
2. Este processo foi feito de maneira simultânea, sem que fossem guardados na memória

todos os caminhos, para posteriormente verificar qual seria o valor da integral.

3. Para cada valor de  $p$ , tomou-se o valor médio da integral pelo método de Euler-Maruyama.
4. A seguir, só foi plotado o gráfico do erro versus  $\Delta t$  e uma linha tracejada com coeficiente angular igual a 1 ao lado para referência.

Ao fazer isso, novamente com o método dos mínimos quadrados obtivemos desta vez  $\gamma=0,975620295825$  e um resíduo igual a 0,00224565. Aparentemente, (2.4.7) vale aproximadamente com a igualdade. Executando algumas vezes este algoritmo foi observado que  $\gamma$  varia em torno de 1.

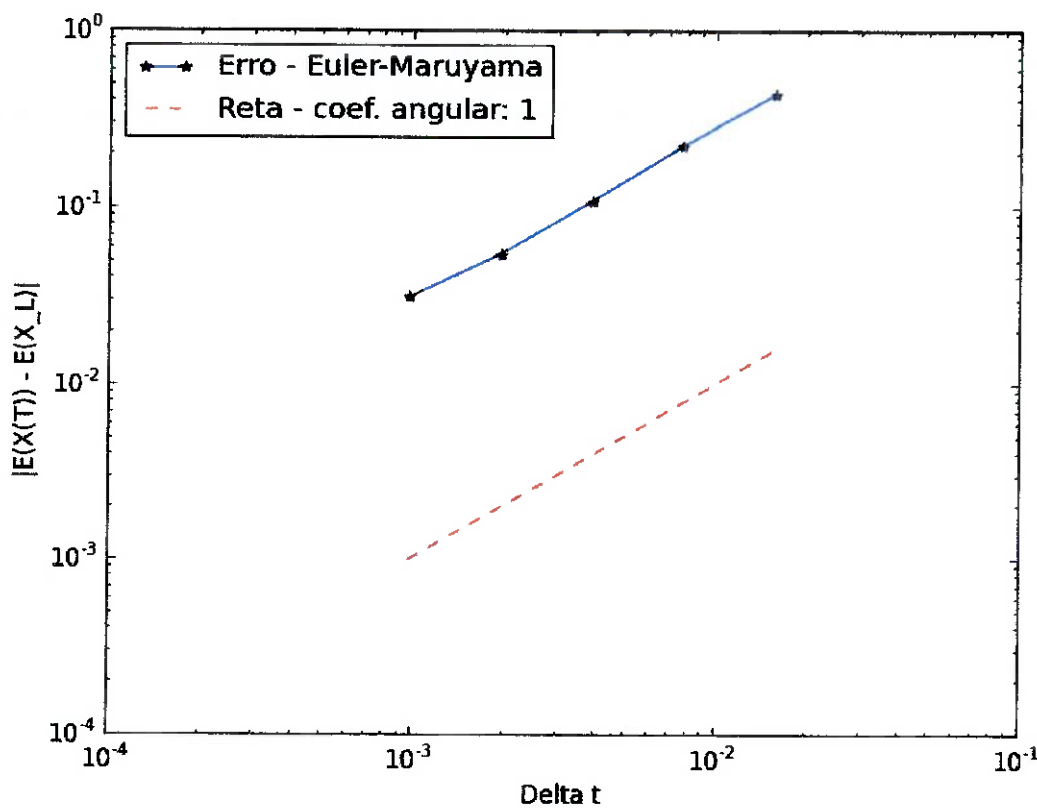


Figura 2.4.1: Convergência fraca. Linha vermelha tracejada como referência, e azul com asteriscos são os erros para cada tamanho de passo.

## 2.5 O método de Milstein

O objetivo inicial deste trabalho era chegar ao método de Euler-Maruyama, e implementá-lo de maneira simples e, além disso, de alguma forma verificar sua convergência com a solução exata de uma equação diferencial estocástica. Mas vale a pena verificar também o método de Milstein, que nada mais é que um melhoramento do método de Euler-Maruyama. No método de Euler-Maruyama o termo estocástico  $dB(\omega)$  é aproximado por

uma soma parecida com a soma de Riemann. Já nesse método de Milstein como veremos logo adiante, este termo é tratado de maneira ligeiramente diferente, levando em consideração o lema de Ito. Fazendo isso, veremos ainda que este método possui uma ordem de convergência melhor que o método de Euler-Maruyama.

Vamos ao método. Aqui consideraremos a mesma equação que a apresentada na seção 2.3. Então, rerepresentando a equação (2.3.1), temos:

$$X(t) = X_0 + \int_0^t f(X(s)) ds + \int_0^t g(X(s)) dB_s(\omega) \quad (2.3.1)$$

Novamente,  $f$  e  $g$  são funções escalares e a condição inicial  $X_0$  é uma variável aleatória. A segunda integral do lado direito de (2.3.1) é calculada com respeito a um movimento browniano, e ainda estamos considerando a versão de Ito.

O método de Milstein é da seguinte forma:

$$X_j = X_{j-1} + f(X_{j-1})\Delta t + g(X_{j-1})(B_{\tau_j}(\omega) - B_{\tau_{j-1}}(\omega)) + \frac{1}{2}g(X_{j-1})g'(X_{j-1})((B_{\tau_j}(\omega) - B_{\tau_{j-1}}(\omega))^2 - \Delta t) \quad (2.5.1)$$

Como já mencionado acima, a diferença deste método para o de Euler-Maruyama está no termo em que está sendo truncada a equação. Na expansão de Taylor feita pelo método de Euler-Maruyama não foi considerado o lema de Ito. Já no método de Milstein é considerado um termo a mais na expansão.

O método de Milstein foi aplicado para o mesmo problema em que aplicamos o Euler-Maruyama na seção 2.3. Segue abaixo o gráfico obtido:

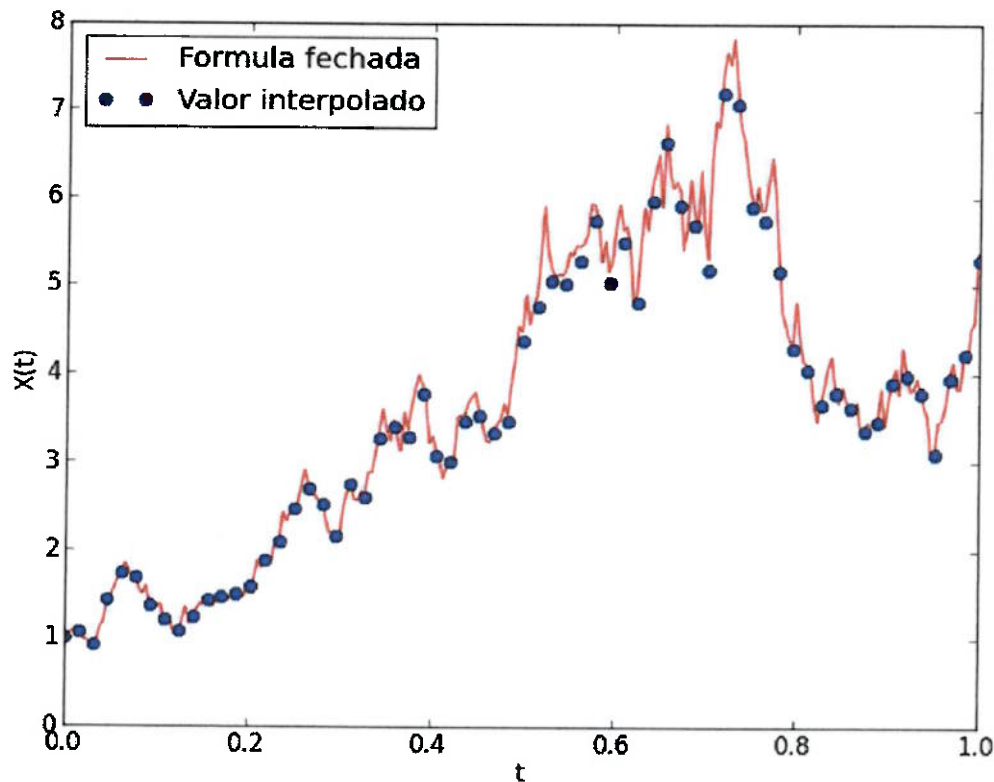


Gráfico 2.5.1: O método de Milstein.

Para a implementação deste método foi aproveitado o *script2.py*, o que foi feito foi apenas uma pequena alteração na forma como foi calculado o valor da variável *Xtemp*.

Olhando para este gráfico acima apenas fica difícil notar uma diferença entre ambos os métodos. Apenas parece ser tão bom quanto. Não apresentaremos aqui neste trabalho, mas se executarmos várias vezes esses métodos, é possível observar graficamente que o método de Euler-Maruyama as vezes não parece aproximar tão bem a solução exata. Uma forma de constatar isso foi executando ambos os métodos 10 vezes, e em cada vez foi calculado o erro absoluto em cada ponto da aproximação, e somamos esse erro, obtendo assim um erro total em cada vez que executamos o algoritmo. Segue abaixo a tabela obtida:

Erro EM	Erro Milstein
10,3950	30,5893
9,6283	6,9105
20,3060	18,0695
97,3952	14,3360
23,1852	16,9689
27,2297	18,1173
10,8095	7,2809
69,8430	14,5309
17,2815	7,6308
9,6973	10,0684

Tabela 2: Comparação – Métodos EM e Milstein.

O algoritmo calcula com até dez casas decimais este erro, mas apresentamos acima somente quatro casas decimais para facilitar a leitura da tabela. Em algumas ocasiões o método de Euler-Maruyama até obteve um valor de erro absoluto menor que o de Milstein, no entanto, na média não é o que acontece. A média dos erros do método de Euler-Maruyama nesta tabela acima foi de 29,577067402566, enquanto que a média dos erros do método de Milstein foi 14,450240631564, ou seja, menos da metade.

A variância dos métodos foi de 890,83049633408 para Euler-Maruyama e de 51,9257153042813 para Milstein. Isso indica que o método de Euler-Maruyama é muito mais instável que o de Milstein.

Por último, nesta tabela acima, o menor valor de erro absoluto foi de 9,62831989111 para Euler-Maruyama e de 6,91048513599 para Milstein.

## 2.6 Convergência forte do método de Milstein

Para encerrar este assunto, vamos falar da convergência forte do método de Milstein. Dissemos na seção anterior que este método possuía uma ordem de convergência maior que o método de Euler-Maruyama. Apenas lembrando o que foi apresentado na equação (2.4.1)

acima, um método possui ordem de convergência forte igual a  $\gamma$  quando:

$$E(|X_n - X(\tau_n)|) \leq C \Delta t^\gamma \quad (2.4.1)$$

Lembrando (2.4.2), temos:

$$e_{\Delta t}^F := E(|X_L - X(T)|), \quad L\Delta t = T.$$

Então, dessas duas equações sai que:

$$e_{\Delta t}^F \leq C \Delta t^\gamma \quad (2.6.1)$$

Aplicando o mesmo método que na seção 2.4 onde linearizamos esta desigualdade aplicando logaritmo em ambos os lados, foi possível obter a imagem abaixo:

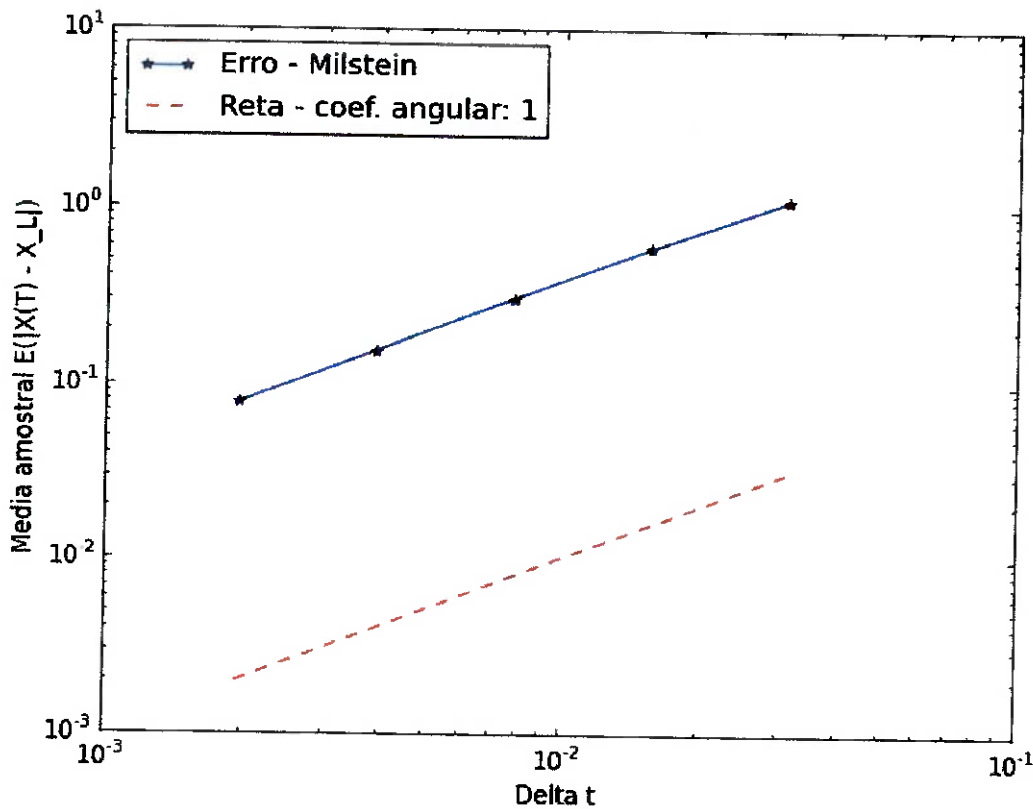


Figura 2.6.1: Convergência forte. Linha vermelha tracejada como referência, e azul com asteriscos são os erros para cada tamanho de passo.

Aqui foi aproveitado o *script3.py*. A única alteração necessária foi a forma de se calcular o valor da variável *Xtemp*, aplicando aqui o método de Milstein.

Nesta execução do algoritmo, foi obtido um  $\gamma = 0,955788223326$ , com resíduo igual a 0,00155068. Nas vezes em que este algoritmo foi executado observamos  $\gamma$  um pouco acima de 0,95. Em nenhuma das ocasiões foi obtido  $\gamma = 1$ , no entanto, na referência [2] utilizada para este tópico menciona que este método possui ordem de convergência igual a 1.

## Conclusão

O objetivo inicial deste trabalho era chegar ao método de Euler-Maruyama, e implementá-lo de maneira simples e, além disso, de alguma forma verificar sua convergência com a solução exata de uma equação diferencial estocástica.

Observamos, com a ajuda do método dos mínimos quadrados, que a ordem de convergência forte do método de Euler-Maruyama é de aproximadamente  $\gamma = 0,5$ , e sua ordem de convergência fraca é aproximadamente  $\gamma = 1$ . Vimos ainda que é possível com um algoritmo simples se obter resultados consistentes com o resultado real.

Apesar desta consistência, é possível concluir também que com uma pequena alteração no método de Euler-Maruyama, o método de Milstein consegue obter melhoras significativas na convergência do método. A convergência forte passa a ter sua ordem de convergência aproximadamente dobrada.

É interessante notar que os métodos numéricos aqui apresentados não são somente aplicáveis à EDE mencionada neste trabalho, podem ser aplicados a muitos casos diferentes, como por exemplo, no problema da dinâmica populacional. A equação que descreve a dinâmica de uma determinada população [4] é a seguinte:

$$dX(t) = rX(t)(K - X(t))dt + \beta X(t)dB_t(\omega), X(0) = X_0.$$

Aqui  $r$ ,  $K$  e  $\beta$  são constantes. Na referência [4] há a solução analítica desta equação. No entanto, note que esta equação é da forma da equação (2.3.2), e portanto, é possível também aplicar os métodos de Milstein e de Euler-Maruyama nesta equação.

Modelos estocásticos se aplicam a muitos fenômenos da natureza, mas também são bastante aplicados em finanças, e como muitas vezes não possuem uma solução fechada, estes métodos numéricos são bem vindos.

# Algoritmos

Segue abaixo os algoritmos criados e utilizados durante este trabalho:

- **Script1.py:**

```
import numpy as np
from numpy import arange
from pylab import *

def stochasticIntegral():
    T = 1.0
    N = 1000
    dt = T/N
    dW = sqrt(dt)*np.random.normal(0, 1, N) # Incrementos
    W = np.cumsum(dW); # Soma cumulativa
    dW = np.array(dW)
    W = np.array(W)

    dW_strat = np.array(0.5*sqrt(dt)*np.random.normal(0, 1, N)) # Incrementos para
    integral de Stratonovich
    W_strat = 0.5*( concatenate( [[0], W[0:-1]] ) + W ) + dW_strat
    ito = np.dot( concatenate( [[0], W[0:-1]] ), dW )
    itoerr = abs( ito - 0.5*(W[-1]**2 - T) )
    strat = np.dot( W_strat, dW )
    straterr = abs( strat - 0.5*W[-1]**2 )
    print ito, "\t ", itoerr, "\t ", strat, "\t ", straterr, "\t ", strat - ito
##### Funcao Principal #####
def main():
    print "ito \t itoerr \t strat \t straterr \t diferenca"
    for i in range(5):
        stochasticIntegral()
##### Chamada da funcao #####
main()
```

- **Script2.py:**

```

import numpy as np
from numpy import arange
from pylab import *
import matplotlib.pyplot as plt

lbd = 2
mu = 1
def f(x): return lbd*x
def g(x): return mu*x
def EulerMaruyama(Xzero, T, N):
    dt = T/N
    dW = np.array(sqrt(dt)*np.random.normal(0, 1, N))
    W = np.cumsum(dW);           # Soma cumulativa

    # Solucao exata para futura comparacao
    Xtrue = Xzero*exp((lbd-0.5*mu**2)*np.array(arange(dt, T+dt, dt))+mu*W)

    # Inicio da implementacao do metodo de EM
    R = 2**2
    Dt = R*dt
    L = N/R
    Xem = zeros( L )    # Variavel para armazenar dados da simulacao via EM
    Xtemp = Xzero
    for j in range(L):
        Winc = dW[R*j:R*(j+1)].sum() # Armazena variacao do movimento
        Browniano desde Rj ate R(j+1)
        Xtemp = Xtemp + f(Xtemp)*Dt + g(Xtemp)*Winc # Calcula o proximo ponto
        via EM
        Xem[j] = Xtemp           # Armazena valor calculado em Xem

    emerr = abs((Xem[-1]-Xtrue[-1])/Xtrue[-1])           # Erro - diferenca entre
    solucoes analitica e interpolada
    print "emerr = ", emerr

```



```
##### Plotando #####
plot(arange(0, T+dt, dt), concatenate([[Xzero], Xtrue]), 'r-', label="X x t")
plot(arange(0, T+Dt, Dt), concatenate([[Xzero], Xem]), 'o', label="X x t")
show()
```

```
##### Funcao main #####
```

```
def main():
```

```
    Xzero = 1
```

```
    T = 1.
```

```
    N = 2**8
```

```
    EulerMaruyama(Xzero, T, N)
```

```
##### Chamada da funcao main #####
```

```
main()
```

- **Script3.py:**

```

import numpy as np
from numpy import arange
from pylab import *
import matplotlib.pyplot as plt

#####          Condiçoes iniciais          #####
lbd = 2
mu = 1
Xzero = 1
#####
#####          Para discretizacao          #####
T = 1.
N = 2**9
dt = T/N
M=1000
#####

Xerr = zeros( ( M, 5 ) )

for s in range(M):
    dW = np.array(sqrt(dt)*np.random.normal(0, 1, N))
    W = np.cumsum(dW);          # Soma cumulativa
    Xtrue = Xzero*exp( (lbd-0.5*mu**2) + mu*W[-1] )
    for p in range(5):
        R = 2**(p+1)
        Dt = R*dt
        L = N/R
        Xtemp = Xzero
        for j in range(L):
            Winc = dW[R*j:R*(j+1)].sum()
            Xtemp = Xtemp + Dt*lbd*Xtemp + mu*Xtemp*Winc

    Xerr[s][p] = abs(Xtemp - Xtrue)

```

```
Dtvals = dt * ( 2.**np.array(arange(0,5,1) ) )
print np.array(Xerr.mean(axis=0))
A = array([ log(Dtvals), ones(5)])
z = log(Xerr.mean(axis=0))
w = linalg.lstsq(A.T,z) # Obtendo os parametros por regressao

print w[0][0]
print w[1]

plt.loglog(Dtvals, exp(z), 'b*- ', Dtvals, Dtvals**.5, 'r--', label="Delta t")
plt.xlabel('Delta t')
plt.ylabel('Media amostral |X(T) - X_L|')

show()
```

- **Script4.py:**

```

import numpy as np
from numpy import arange
from pylab import *
import matplotlib.pyplot as plt
#####          Condiçoes iniciais          #####
lbd = 2
mu = 0.1
Xzero = 1
#####
#####          Para discretizacao          #####
T = 1.
M=50000
#####
Xem = zeros( 5 )
for p in range(5):
    Dt = 2**(p+1-10)
    L = T/Dt
    Xtemp = Xzero*ones(M)
    for j in range(int(L)):
        Winc = np.array(sqrt(Dt)*np.random.normal(0, 1, M))
        Xtemp = Xtemp + Dt*lbd*Xtemp + mu*Xtemp*Winc
    Xem[p] = Xtemp.mean(axis=0)
Xerr = abs(Xem - exp(lbd))
Dtvals = 2.**np.array(arange(0,5,1) - 10)
A = array([ log(Dtvals), ones(5)])
z = log(Xerr)
w = linalg.lstsq(A, T,z) # Obtendo os parametros por regressao
print w[0][0]
print w[1]
plt.loglog(Dtvals, Xerr, 'b*- ', Dtvals , Dtvals, 'r--', label="Delta t")
plt.xlabel('Delta t')
plt.ylabel('Media amostral |X(T) - X_L|')
show()

```

## Bibliografia

- [1] MIKOSCH, Thomas. **Elementary stochastic calculus: with finance in view**. World Scientific, 1998.
- [2] HIGHAM, Desmond J. An algorithmic introduction to numerical simulation of stochastic differential equations. **SIAM review**, v. 43, n. 3, p. 525-546, 2001.
- [3] HIGHAM, Desmond J. Stochastic ordinary differential equations in applied and computational mathematics. **IMA journal of applied mathematics**, v. 76, n. 3, p. 449-474, 2011.
- [4] ØKSENDAL, Bernt. **Stochastic differential equations**. Springer Berlin Heidelberg, 2003.
- [5] GARD, Thomas C. **Introduction to stochastic differential equations**. New York: M. Dekker, 1988.
- [6] KLOEDEN, Peter E.; PEARSON, R. A. The numerical solution of stochastic differential equations. **The Journal of the Australian Mathematical Society. Series B. Applied Mathematics**, v. 20, n. 01, p. 8-12, 1977.
- [7] [http://wiki.scipy.org/Tentative\\_NumPy\\_Tutorial](http://wiki.scipy.org/Tentative_NumPy_Tutorial)