

 **XIII CONGRESSO ANUAL**

**SOCIEDADE PORTUGUESA DE ESTATÍSTICA**

# Introdução às Equações Diferenciais Estocásticas e Aplicações

Carlos A. Braumann



**ERICEIRA**

**28 de SETEMBRO a 1 de OUTUBRO de 2005**

**25º Aniversário**



**SOCIEDADE PORTUGUESA  
DE ESTATÍSTICA**



I. N. E.  
BIBLIOTECA

ISBN 972-9890-06-0  
Depósito Legal n°229951/05

**FICHA TÉCNICA:**

**Título:** Uma Introdução às Equações Diferenciais Estocásticas e Aplicações

**Autor:** Carlos A. Braumann

**Editora:** Sociedade Portuguesa de Estatística

**Produção gráfica e Impressão:** Instituto Nacional de Estatística

**Tiragem:** 400 exemplares

**ISBN:** 972-8890-06-0

**Depósito legal:** n.º 229951/05

À Manuela  
À memória dos meus Pais



# Agradecimentos

Agradecemos às seguintes entidades pelo valioso apoio concedido à realização do XIII Congresso Anual da Sociedade Portuguesa de Estatística:

Associação Nacional de Farmácias

Banco Português de Investimento

British Council

Câmara Municipal de Mafra

Centro de Estatística e Aplicações da Universidade de Lisboa

Departamento de Estatística e Investigação Operacional (FCUL)

ELIS - Europe Linge Service

Faculdade de Ciências da Universidade de Lisboa

Fundação Calouste Gulbenkian

Fundação para a Ciência e a Tecnologia

Hotel VilaGalé Ericeira

Instituto Nacional de Estatística

Junta de Freguesia da Ericeira

Livraria Escolar Editora

PSE - Produtos e Serviços de Estatística, Lda.

SAS Portugal, SASINST Software, Lda.

SPRUCE - Statistics in Public Resources, Utilities and Care of the Environment

Tapada Nacional de Mafra

Timberlake Consultores



O presente texto foi preparado para servir de base ao minicurso integrado no XIII Congresso Anual da Sociedade Portuguesa de Estatística. É, no entanto, substancialmente mais extenso que o minicurso, visando proporcionar aos seus frequentadores a possibilidade de posteriormente suplementarem esse curso com desenvolvimentos que nele não cabiam. Pode também ser usado como manual de um primeiro curso regular sobre equações diferenciais estocásticas. Como o título indica, trata-se de apresentar de forma sucinta as equações diferenciais estocásticas, ilustrando os conceitos com algumas aplicações relevantes.

É bem sabido que as equações diferenciais estocásticas têm aplicações em praticamente todos os ramos da Ciência e da Tecnologia para o estudo de fenómenos dinâmicos descritos por equações diferenciais que sejam perturbados por flutuações aleatórias. Demos mais ênfase às aplicações biológicas e financeiras, as primeiras por terem sido objecto de vários trabalhos de investigação do autor e as segundas pela sua crescente importância e por serem um poderoso motor para o desenvolvimento da própria teoria.

Foi nosso objectivo apresentar um curso em língua portuguesa que permitisse divulgar mais facilmente uma área de grande importância teórica e aplicada ainda pouco estudada no nosso País. Procurámos, sem quebra de rigor, ligar a teoria às aplicações precisamente porque se trata de uma poderosa ferramenta de modelação matemática que importava não só compreender como ver em acção.

Há sempre duas opções para um texto desta natureza. Uma é de partir de uma descrição com a máxima generalidade, particularizada volta e meia para efeitos de exemplificação. Outra, que adoptamos e que nos parece mais aconselhável num tema que tem algumas dificuldades intrínsecas, é a de trabalhar no contexto mais simples possível, que permita não obscurecer o essencial com complicações técnicas ou de notação. Assim, o nosso tratamento estudará o caso de equações diferenciais estocásticas em que o processo “perturbador” é um processo de Wiener. Aliás, tratamentos mais gerais em contexto de processos perturbadores tipo martingala contínua podem reduzir-se ao tratamento aqui feito, pelo que, nesse contexto, o presente tratamento não perde generalidade. Também trabalharemos em dimensão um para evitar a complicação de notações matriciais, embora indiquemos como proceder à generalização ao caso multidimensional (sistemas de equações diferenciais estocásticas) uma vez apreendidas as principais ideias no contexto



unidimensional.

Tratando-se de uma introdução com espaço limitado, haverá muitos tópicos que não poderão ser tratados e outros que só serão aflorados. Um caso típico, apesar da sua primordial importância nas aplicações, é o do tratamento estatístico das observações (estimação paramétrica e não-paramétrica, escolha de modelos, testes de hipóteses, previsão). A dificuldade do tema em contextos não triviais não nos deixa outra alternativa, já que este estudo só poderia ser posterior ao da apreensão da teoria básica das equações diferenciais estocásticas. Como sempre, a estatística inferencial pressupõe a teoria probabilística.

Daremos mais ênfase à compreensão, que queremos rigorosa, dos conceitos e resultados e à sua aplicação do que propriamente à sua dedução. Não deixaremos, porém, de apresentar as principais demonstrações, ainda que, nos casos mais complexos ou que exijam mais tempo e espaço, apenas em esboço e remetendo os desenvolvimentos para literatura mais especializada. Procuramos assim minimizar os conhecimentos prévios exigidos ao leitor, sem sacrificar o rigor matemático. Os pressupostos são apenas os de uma boa formação básica em probabilidades e, de preferência, também em estatística, ao nível do que é comum em qualquer bom curso superior de várias áreas científicas que usam a Matemática como instrumento privilegiado de trabalho. O leitor que tenha alguns conhecimentos básicos de processos estocásticos e de teoria da medida e integração estará muito mais à vontade, mas procuraremos suprir alguma falha de formação nestas áreas do leitor proveniente das áreas de aplicação apresentando no texto os conceitos essenciais que sejam necessários a uma compreensão rigorosa do tema aqui em estudo; o Capítulo 2 contém os conceitos principais (outros serão apresentados à medida que sejam necessários) e pode ser dispensado pelo leitor mais preparado ou numa primeira leitura. Em certos casos, as demonstrações de alguns resultados exigem conhecimentos mais avançados, mas o leitor com menor formação de base pode simplesmente saltar por cima delas. O texto contém alguns (poucos por limitações de espaço e tempo) exercícios, incluindo algumas demonstrações relativamente simples que deixamos ao cuidado do leitor.

O último Capítulo faz um resumo do que é mais importante com o objectivo de permitir ao leitor um visão informal de síntese final em que se procura dar mais relevo à intuição do que ao rigor matemático. Mas pode também ser lido em primeiro lugar pelo leitor mais apressado que queira ter apenas uma primeira ideia intuitiva da matéria.

Dada a forte limitação temporal em que decorreu a elaboração deste livro, não foi possível o trabalho de revisão cuidada e de aperfeiçoamento técnico e didáctico que seria desejável, facto de que pedimos desculpa ao leitor.

Este trabalho decorreu no CIMA-UE (Centro de Investigação em Matemática e Aplicações da Universidade de Évora), centro financiado pela FCT (Fundação para a Ciência e a Tecnologia) no âmbito do fundo FEDER. Aqui se regista o agradecimento à entidade financiadora.

Resta-nos finalmente agradecer reconhecidos o convite da Comissão Organizadora do Congresso, que assim nos proporciona a oportunidade de uma maior divulgação desta relevante área do conhecimento. Esperemos que o leitor possa também beneficiar.

Universidade de Évora, Agosto de 2005

Carlos A. Braumann



# Índice

Prefácio	vii
<b>1 Introdução</b>	<b>1</b>
<b>2 Revisão de probabilidades e processos estocásticos</b>	<b>7</b>
2.1 Breve revisão de conceitos probabilísticos . . . . .	7
2.2 Esperanças matemáticas e probabilidades condicionais . . . . .	12
2.3 Breve revisão de processos estocásticos . . . . .	13
2.4 Breve revisão de processos estacionários . . . . .	18
2.5 Filtrações, martingalas e tempos de Markov . . . . .	18
2.6 Processos de Markov . . . . .	21
<b>3 Uma introdução informal às equações diferenciais estocásticas</b>	<b>25</b>
<b>4 O processo de Wiener</b>	<b>31</b>
4.1 Definição . . . . .	31
4.2 Principais propriedades . . . . .	32
4.3 Algumas propriedades analíticas . . . . .	35
4.4 Tempos de primeira passagem . . . . .	37
4.5 Processos de Wiener multidimensionais . . . . .	39
<b>5 Processos de difusão</b>	<b>41</b>
5.1 Definição . . . . .	41
5.2 Equações de Kolmogorov . . . . .	43
5.3 Caso multidimensional . . . . .	48
<b>6 Integrais estocásticos</b>	<b>51</b>
6.1 Definição informal dos integrais de Itô e Stratonovich . . . . .	51
6.2 Construção do integral de Itô . . . . .	55
6.3 Existência de sucessões aproximadoras de funções em escada . . . . .	63

6.4	Estudo do integral como função do limite superior de integração . . . . .	65
6.5	Extensões do integral de Itô . . . . .	68
6.6	Teorema e fórmula de Itô . . . . .	72
6.7	Os cálculos de Itô e Stratonovich . . . . .	76
6.8	O caso multidimensional . . . . .	79
<b>7</b>	<b>Equações diferenciais estocásticas</b>	<b>83</b>
7.1	Teorema de existência e unicidade e principais propriedades da solução de uma equação diferencial estocástica . . . . .	83
7.2	Esboço da demonstração do teorema de existência e unicidade . . . . .	86
7.3	Observações e extensões ao teorema de existência e unicidade . . . . .	94
<b>8</b>	<b>Estudo do modelo de Black-Scholes</b>	<b>99</b>
8.1	Estudo pelo cálculo de Itô . . . . .	99
8.2	Estudo pelo cálculo de Stratonovich . . . . .	105
<b>9</b>	<b>A questão dos cálculos de Itô e de Stratonovich</b>	<b>107</b>
9.1	Controvérsia . . . . .	107
9.2	Resolução da controvérsia . . . . .	108
<b>10</b>	<b>Estudo de alguns funcionais</b>	<b>111</b>
10.1	Fórmula de Dynkin . . . . .	111
10.2	Fórmula de Feynman-Kac . . . . .	114
<b>11</b>	<b>Introdução ao estudo das difusões de Itô unidimensionais</b>	<b>117</b>
11.1	O processo de Ornstein-Uhlenbeck . . . . .	117
11.2	Tempo de saída de um intervalo . . . . .	120
11.3	Comportamento nas fronteiras de difusões de Itô e densidades estacionárias . . . . .	126
11.4	Um exemplo de aplicação em crescimento populacional . . . . .	130
<b>12</b>	<b>Teorema de Girsanov</b>	<b>137</b>
12.1	Introdução através de um exemplo . . . . .	137
12.2	Teorema de Girsanov . . . . .	143
<b>13</b>	<b>Opções e fórmula de Black-Scholes</b>	<b>149</b>
13.1	Introdução . . . . .	149
13.2	Fórmula de Black-Scholes e estratégia hedging . . . . .	153
13.3	Um exemplo numérico . . . . .	158
13.4	Obtenção da fórmula de Black-Scholes via teorema de Girsanov . . . . .	160

13.5 O modelo binomial . . . . .	164
13.6 Opções europeias de venda . . . . .	167
13.7 Outros modelos e opções . . . . .	168
<b>14 Síntese</b>	<b>173</b>
<b>Referências</b>	<b>181</b>
<b>Índice Remissivo</b>	<b>185</b>



# Capítulo 1

## Introdução

As *equações diferenciais estocásticas* (EDE) são basicamente equações diferenciais com um termo estocástico adicional. O termo determinístico, que é comum às equações diferenciais ordinárias, descreve o comportamento dinâmico “médio” do fenómeno em estudo e o termo estocástico descreve o “ruído”, ou seja as perturbações aleatórias que influenciam esse fenómeno.

Como muitos fenómenos naturais podem ser descritos por equações diferenciais, as EDE têm importantes aplicações em praticamente todos os ramos da Ciência e da Tecnologia para descrever esses fenómenos sempre que haja perturbações aleatórias relevantes que os afectem. A primeira EDE surgiu na literatura, tanto quanto sei, em 1930. Trata-se do modelo de Ornstein-Uhlenbeck ([58]) para o *movimento browniano*, que é o movimento irregular de uma partícula suspensa num fluido; a designação provém de ter sido o botânico Brown, no século XIX, que observou esse fenómeno ao microscópio. Só, porém, em meados do século XX, é que uma teoria matemática rigorosa foi desenvolvida por Itô (ver [36]). As aplicações percorrem áreas como a Física, a Astronomia, a Electrónica, as Telecomunicações, a Química, a Sismologia, a Astronáutica, a Oceanografia, a Física da Atmosfera, a Biologia, a Economia, as Finanças, etc. Permitem estudar fenómenos como, por exemplo, a dispersão de um poluente na água ou no ar, o efeito dos ruídos na transmissão de sinais em telecomunicações, a trajectória de um satélite artificial, a localização de um navio, o ruído térmico num circuito eléctrico, a dinâmica de uma reacção química, a dinâmica de uma ou várias populações de seres vivos quando o ambiente sofre perturbações aleatórias que afectam as suas taxas de crescimento, a determinação da política óptima de pescas de uma população de peixes em ambiente com perturbações



aleatórias, a variação das taxas de juro, a flutuação das taxas de câmbio, a variação da cotação de uma acção na bolsa, o valor de uma opção de compra ou de venda, a imunização de riscos de carteiras de investimento ou de planos de poupança reforma, etc.

Os fenómenos biológicos (e particularmente a dinâmica de populações de seres vivos) e financeiros têm frequentemente tendências vincadas mas também componentes imprevisíveis devidas à complexidade e variabilidade das condições do ambiente ou do mercado. São, por isso, particularmente propícios ao uso de modelos baseados em EDE para o seu estudo. Iremos dar, nos exemplos de aplicação, maior relevo a estas aplicações. No que se refere ao estudo de dinâmica de populações, por ser uma área em que tenho desenvolvido trabalho de investigação. No que se refere às aplicações financeiras, por ser uma das áreas mais activas de investigação com grande desenvolvimento nos últimos 35 anos (vejam-se os trabalhos pioneiros [6], [49], [50]) e em que as necessidades provocadas pelas aplicações mais têm contribuído para o desenvolvimento da própria teoria numa das mais fecundas interacções entre teoria e aplicação. Merton e Scholes receberam o Prémio Nobel da Economia em 1997 pelo seu trabalho em matemática financeira, particularmente no que se refere à valorização das opções, que se baseou no cálculo estocástico que aqui vamos apresentar de forma introdutória.

Dado o carácter introdutório desta apresentação, introduziremos as EDE no contexto mais simples possível, evitando obscurecer as ideias principais com questões técnicas ou notação pesada e encaminhando o leitor para literatura mais especializada quando apropriado. Assim, iremos estudar apenas equações diferenciais estocásticas quando o ruído perturbador é um *ruído branco* em tempo contínuo, cujo integral (o ruído acumulado) é o bem conhecido *processo de Wiener*.

O processo de Wiener, rigorosamente estudado por Wiener e Lévy a partir de 1920, é também conhecido nalguma literatura por movimento browniano pelo facto de ter sido usado por Einstein em 1905 como primeiro modelo do movimento browniano de uma partícula. Nós pessoalmente não gostamos de usar esta designação, já que o movimento browniano de uma partícula pode ser descrito por outros modelos mais realistas, como o modelo de Ornstein-Uhlenbeck acima referido. A invenção do processo de Wiener é frequentemente atribuída a Einstein por se pensar que terá sido o primeiro a tê-lo utilizado (ainda que sem estar então “baptizado” como processo de Wiener). No entanto, Bachelier em 1900 (ver [4]) já o tinha usado como modelo (não muito adequado) para as cotações de acções na Bolsa de Paris. O uso de ruído branco em tempo contínuo pode ser justificado como uma aproximação adequada de ruídos perturbadores mais complexos.

Teremos como principal preocupação o estudo das equações dife-

renciais estocásticas unidimensionais, mas apresentaremos as indicações necessárias para o tratamento do caso multidimensional. Claro que, se queremos estudar simultaneamente várias variáveis (como, por exemplo, a cotação de vários activos financeiros na bolsa ou o tamanho de várias populações interactuantes), precisamos de equações multidimensionais (isto é, sistemas de equações diferenciais estocásticas), mas não queremos complicar a apresentação inicial dos conceitos com a notação matricial que o caso multidimensional envolve.

O Capítulo 2 apresenta uma revisão dos principais conceitos e propriedades relativos a processos estocásticos, particularmente os processos de Markov, aproveitando-se para rever as noções de medida e probabilidade essenciais ao seu estudo. Não serão apresentadas demonstrações.

O Capítulo 3 apresenta um exemplo de equação diferencial estocástica que pode ser utilizada para estudar o crescimento de uma população de seres vivos crescendo num ambiente com recursos abundantes mas sujeito a perturbações aleatórias que afectam a taxa de crescimento da população. O mesmo modelo é conhecido por *modelo de Black-Scholes* e foi utilizado por estes na modelação da cotação de uma acção numa bolsa de valores. Este exemplo serve para apresentar ao leitor o processo de Wiener (que alguns leitores já conhecerão) e as equações diferenciais estocásticas de uma forma informal.

O Capítulo 4 faz um estudo dos aspectos mais relevantes do processo de Wiener. No Capítulo 5 faz-se uma breve revisão dos processos de difusão, que são de certa forma generalizações do processo de Wiener e que vão desempenhar um papel fundamental no estudo das equações diferenciais estocásticas. Mostrar-se-á mais tarde que, sob certas condições de regularidade, há uma equivalência entre processos de difusão e soluções de equações diferenciais estocásticas.

Dada uma condição inicial e uma equação diferencial estocástica, temos um problema de Cauchy que é equivalente a uma equação integral estocástica. De certa forma, quer em ambiente determinístico, quer em ambiente estocástico, um problema de Cauchy não é mais que uma equação integral disfarçada, pois é esta que normalmente é o fulcro do tratamento teórico. No mundo estocástico, é a versão integral da EDE que faz verdadeiramente sentido já que as derivadas aqui, como veremos, não existem no sentido corrente, mas apenas como processos estocásticos generalizados. Assim, para que estas equações integrais estocásticas ganhem sentido, é preciso definir e estudar os integrais estocásticos. Esse é o objecto do Capítulo 6. Infelizmente a clássica definição de integral de Riemann-Stieltjes, agora naturalmente ao longo das trajectórias, é inaplicável porque o processo integrador (o processo de Wiener) é quase certamente de variação ilimitada. Diferentes escolhas de pontos intermédios nas somas de Riemann-Stieltjes conduzem a resultados diferentes.

Há, pois, várias definições possíveis de integrais estocásticos. A definição de Itô é a que tem melhores propriedades probabilísticas e é por isso a mais habitualmente adoptada, como aliás faremos aqui. Não satisfaz, porém, as regras usuais de cálculo diferencial e integral, pelo que apresentaremos o teorema de Itô, que nos dá a regra chave desse novo cálculo estocástico. Mas não deixaremos de falar de definições alternativas de integral, particularmente do integral de Stratonovich, que, não tendo as boas propriedades probabilísticas do integral de Itô, satisfaz, porém, as regras usuais de cálculo. Discutiremos a utilização de um e outro cálculo e apresentaremos uma fórmula de conversão entre eles, que nos será muito útil. A generalização do integral a várias dimensões é também apresentada.

No Capítulo 7 trataremos então do problema de Cauchy para equações diferenciais estocásticas, que é equivalente à correspondente equação integral estocástica. Usaremos o cálculo de Itô. Apresentaremos o teorema mais comum sobre a existência e unicidade de solução, bem como sobre as principais propriedades desta, particularmente a de, sob certas condições de regularidade, ser um processo de difusão. Apresentaremos outros resultados de existência e unicidade exigindo hipóteses mais fracas. A generalização ao caso multidimensional (sistemas de EDE) será então apresentada.

No Capítulo 8 examinaremos o caso particular do modelo de Black-Scholes, obtendo a solução explícita e estudando as suas propriedades. Como as soluções pelos cálculos de Itô e Stratonovich são distintas (mesmo em aspectos qualitativos), discutiremos a controvérsia que tem existido na literatura sobre qual o cálculo, de Itô ou de Stratonovich, mais apropriado nas aplicações. Este exemplo serve também de pretexto, no Capítulo 9, para verificar que afinal a controvérsia sobre qual o cálculo mais adequado a utilizar não tem sentido e se deve a uma confusão semântica. Referem-se ainda resultados do autor que generalizam esta constatação a uma classe bastante geral de EDE.

As equações diferenciais estocásticas autónomas, em que os coeficientes dos termos determinístico e estocástico são funções do estado do processo mas não do tempo, são particularmente importantes nas aplicações e, sob certas condições de regularidade, são processos de difusão homogêneos, também conhecidos por difusões de Itô.

No Capítulo 10 falaremos das fórmulas de Dynkin e de Feynman-Kac e da relação entre soluções de certas equações às derivadas parciais e esperanças matemáticas de certos funcionais de soluções de equações diferenciais estocásticas autónomas.

No Capítulo 11 estudaremos as difusões de Itô unidimensionais (soluções de equações diferenciais estocásticas autónomas unidimensionais) quanto a aspectos como tempos de primeira passagem, classificação

de fronteiras, existência de densidades estacionárias. Apresentaremos também alguns exemplos de aplicação, começando com o processo de Ornstein-Uhlenbeck, solução da primeira EDE surgida na literatura, que será usado como pretexto para o estudo das difusões de Itô, categoria a que pertencem a maioria dos modelos usados nas aplicações.

No Capítulo 12 apresentaremos o problema da mudança de medida de probabilidade como forma de alterar a tendência (termo “determinístico”) da EDE, através do teorema de Girsanov.

No Capítulo 13 usaremos duas formas alternativas de obter a fórmula de Black-Scholes, que resolve o problema da determinação do valor de uma opção europeia de compra na hipótese de não arbitragem dos mercados. Uma delas recorre ao teorema de Girsanov. Falaremos também de generalizações deste resultado.

No Capítulo 14 apresentamos um resumo do que é mais importante com o objectivo de permitir ao leitor um visão informal de síntese final em que se procura dar mais relevo à intuição do que ao rigor matemático. Mas pode também ser lido em primeiro lugar pelo leitor mais apressado que queira ter apenas uma primeira ideia intuitiva da matéria.



## Capítulo 2

# Revisão de probabilidades e processos estocásticos

### 2.1 Breve revisão de conceitos probabilísticos

Considere um *espaço de probabilidade*  $(\Omega, \mathcal{F}, P)$ , onde  $(\Omega, \mathcal{F})$  é um *espaço mensurável* e  $P$  uma *probabilidade* nele definida.

Habitualmente, o *universo*, *conjunto universal* ou *espaço amostral*  $\Omega$  representa o conjunto (suposto não-vazio) de todos os possíveis resultados de uma experiência ou fenómeno aleatório. Por exemplo, se lançarmos dois dados, um vermelho e um preto,  $\Omega = \{(1, 1), (1, 2), \dots, (6, 6)\}$  é um conjunto de 36 elementos, cada um deles representando o possível resultado de um lançamento; assim, o elemento  $(3, 4)$  representa o resultado “três pintas no dado vermelho e quatro pintas no dado preto”.  $\mathcal{F}$  é uma *álgebra- $\sigma$* <sup>1</sup>, isto é uma classe não-vazia de subconjuntos de  $\Omega$  fechada para a complementação (se  $A \in \mathcal{F}$ , então o complemento  $A^c := \Omega - A \in \mathcal{F}$ ) e para uniões contáveis<sup>2</sup> (se  $A_n \in \mathcal{F}$ ,  $n = 1, 2, \dots$ , então  $\bigcup_n A_n \in \mathcal{F}$ ). Os conjuntos  $A \in \mathcal{F}$  são chamados *acontecimentos* ou *conjuntos mensuráveis*. A probabilidade  $P$  é uma função de  $\mathcal{F}$  em  $[0, 1]$ , normada ( $P(\Omega) = 1$ ) e aditiva- $\sigma$  (se  $A_n \in \mathcal{F}$ ,  $n = 1, 2, \dots$ , é uma colecção contável de conjuntos disjuntos dois a dois, então  $P(\bigcup_n A_n) = \sum_n P(A_n)$ ). No exemplo dos dois dados, supostos

---

<sup>1</sup>Há quem prefira dizer “ $\sigma$ -álgebra”, imitando a norma inglesa de o adjectivo preceder o substantivo.

<sup>2</sup>Usamos o termo “contável” como sinónimo de finito ou infinito numerável.

honestos, temos  $P(A) = N/36$ , onde  $N$  é o número de elementos de  $A$ ; por exemplo, se  $A = \{(4, 6), (5, 5), (6, 4), (5, 6), (6, 5)\}$  (acontecimento “soma de 10 or 11 pintas nos dois dados”), temos  $P(A) = 5/36$ .

Se estivermos a estudar, por exemplo, a evolução da cotação de uma acção, ela é influenciado pelo “cenário do mercado” que se tiver realizado. Por cenário do mercado podemos entender uma descrição multifactorial que inclui a evolução ao longo do tempo (passado, presente e futuro) de tudo o que possa afectar a cotação da acção e que pode incluir as vendas da empresa cotada, as cotações de outras acções, o comportamento de variáveis económicas a nível nacional e internacional, a situação política, os conflitos armados, etc., etc. Poderá então tomar-se para  $\Omega$  o conjunto de todos os possíveis cenários de mercado. Felizmente, não é habitualmente necessário trabalhar com um espaço tão complexo pois o que de facto interessa é o comportamento da cotação da acção resultante dos possíveis cenários do mercado, e esse comportamento “vive” num espaço mais simples e manejável (e também mais fácil de probabilizar). É assim legítima a visão simplificada de adoptar esse espaço mais simples como o espaço dos cenários do mercado, uma vez que essa simplificação produz exactamente os mesmos resultados. A mesma questão se põe quando, por exemplo, se estuda a evolução do tamanho de uma população de seres vivos, a qual é influenciada pelo “estado da natureza” (o equivalente ao cenário do mercado), que inclui aspectos como a evolução temporal do clima, do habitat, de outras populações que interajam com a população em estudo, etc. Também aqui se pode adoptar uma visão simplificada. O cenário do mercado ou o estado da natureza  $\omega$  que realmente ocorre é um elemento de  $\Omega$  “escolhido ao acaso” de acordo com a lei de probabilidade  $P$ .  $\mathcal{F}$  é a álgebra- $\sigma$  dos subconjuntos de  $\Omega$  (acontecimentos) para os quais a probabilidade  $P$  está definida. A probabilidade  $P$  faz corresponder a cada acontecimento (conjunto de cenários de mercado ou de estados da natureza)  $A \in \mathcal{F}$  a sua probabilidade  $P(A)$ , que é a probabilidade de o cenário do mercado ou estado da natureza que efectivamente ocorre pertencer ao conjunto  $A$ .

Podemos, supor, sem perda de generalidade, que o espaço de probabilidade  $(\Omega, \mathcal{F}, P)$  é *completo*, isto é, dado qualquer conjunto  $N \in \mathcal{F}$  tal que  $P(N) = 0$ , todos os subconjuntos  $Z$  de  $N$  também pertencem a  $\mathcal{F}$ . Com efeito, se o espaço não for completo, podemos sempre completá-lo alargando  $\mathcal{F}$  de forma a incluir todos os conjuntos da forma  $A \cup Z$  com  $A \in \mathcal{F}$  e estendendo a probabilidade  $P$  à álgebra- $\sigma$  alargada pondo  $P(A \cup Z) = P(A)$ .

Recorda-se que uma *variável aleatória* (v.a.) ou *função mensurável*- $\mathcal{F}$  (abreviadamente *função mensurável*)  $X$  definida no espaço mensurável  $(\Omega, \mathcal{F})$  é uma função de  $\Omega$  em  $\mathbb{R}$  tal que, seja qual for o *conjunto de Borel*  $B \in \mathcal{B}$ , a sua imagem inversa  $X^{-1}(B) = [X \in B] := \{\omega \in \Omega : X(\omega) \in B\}$

pertence a  $\mathcal{F}$ . Esta propriedade é essencial para mais tarde podermos definir a distribuição de probabilidade de  $X$  para uma probabilidade  $P$  definida em  $(\Omega, \mathcal{F})$ . Por álgebra- $\sigma$  de Borel  $\mathcal{B}$  entende-se a álgebra- $\sigma$  gerada pelos conjuntos abertos de  $\mathbb{R}$ , ou seja a menor álgebra- $\sigma$  que inclui tais conjuntos. Ela é também a álgebra- $\sigma$  gerada pelos intervalos (ou mesmo pelos intervalos da forma  $] -\infty, x]$  com  $x \in \mathbb{R}$ ). No exemplo dos dois dados, se  $Z$  é a v.a. que representa a soma do número de pintas dos dois dados, vem, para  $\omega = (6, 5)$ ,  $Z(\omega) = 11$  (habitualmente simplifica-se a notação omitindo a dependência de  $\omega$  e escrevendo simplesmente  $Z = 11$ ). Tomando  $\omega$  como representando o “acaso”, podemos dizer que uma v.a. é uma função real do acaso.

A cotação de fecho de uma acção amanhã, a taxa de câmbio do dólar daqui a 90 dias, a altura de uma pessoa escolhida ao acaso, o tamanho de uma população daqui a um ano são exemplos de variáveis aleatórias.

Dado o espaço de probabilidade  $(\Omega, \mathcal{F}, P)$  e uma v.a.  $X$  (no espaço mensurável  $(\Omega, \mathcal{F})$ ), a sua *função de distribuição* (f.d.), que caracteriza completamente a sua distribuição de probabilidade, será aqui representada por  $F_X$ . Recordar-se que a função de distribuição de  $X$  é definida para  $x \in \mathbb{R}$  por  $F_X(x) := P[X \in (-\infty, x]] = P[X \leq x]$ .<sup>3</sup>

Note-se que a classe  $\sigma(X)$  formada pelas imagens inversas por  $X$  dos conjuntos de Borel é uma sub-álgebra- $\sigma$  de  $\mathcal{F}$ , a que se chama *álgebra- $\sigma$  gerada* por  $X$ ; ela contém toda a informação que é pertinente para a determinação do comportamento de  $X$ .

No exemplo dos dois dados,  $Z$  é um exemplo de v.a. *discreta*. Uma v.a.  $X$  é discreta se existir um conjunto contável  $S = \{x_1, x_2, \dots\}$  tal que  $P[X \in S] = 1$ ; representaremos por  $p_X(k) = P[X = x_k]$  ( $k = 1, 2, \dots$ ) a sua *função massa de probabilidade* (f.m.p.), verificando-se que  $F_X(x) = \sum_{x_k \leq x} p_X(k)$ .

Uma v.a.  $X$  diz-se *absolutamente contínua* (vulgarmente, e não muito correctamente, abrevia-se para “v.a. contínua”) se existir uma função não-negativa integrável  $f_X(x)$ , chamada *função densidade de probabilidade* (f.d.p.), tal que  $F_X(x) = \int_{-\infty}^x f_X(y) dy$ . Se  $X$  é absolutamente contínua, então a sua f.d.  $F_X(x)$  é uma função contínua. É até uma função diferenciável em quase toda a parte, isto é, o conjunto excepcional  $N$  de números reais onde a função não é diferenciável é um conjunto negligenciável.<sup>4</sup> Tem-se  $f_X(x) = dF_X(x)/dx$  (se  $N \neq \emptyset$ , podemos atribuir

<sup>3</sup>Note-se que, se  $P$  é uma probabilidade em  $(\Omega, \mathcal{F})$ , a probabilidade  $P_X(B) = P[X \in B] := P(X^{-1}(B))$  está bem definida para todos os conjuntos de Borel  $B$  e temos um novo espaço de probabilidade  $(\mathbb{R}, \mathcal{B}, P_X)$ . No exemplo dos dois dados,  $P_Z(\{10, 11\}) = P[Z \in \{10, 11\}] = P[Z = 10 \text{ ou } Z = 11] = P(Z^{-1}(\{10, 11\})) = P(\{(4, 6), (5, 5), (6, 4), (5, 6), (6, 5)\}) = 5/36$ . Note-se que também  $P[Z \in (9, 5, 11, 2)] = P(\{(4, 6), (5, 5), (6, 4), (5, 6), (6, 5)\}) = 5/36$ .

<sup>4</sup>Por conjunto negligenciável entende-se um conjunto com medida de Lebesgue



arbitrariamente os valores da derivada nos pontos  $x \in N$  de forma a que a f.d.p. fique definida, embora não univocamente, para todo o  $x \in \mathbb{R}$ . Uma v.a. *normal* ou *gaussiana*  $X$  com média  $\mu$  e variância  $\sigma^2 > 0$  é uma v.a. absolutamente contínua com f.d.p.  $f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$ ; tem-se  $\mathbb{E}[X] = \mu$  e  $VAR[X] := \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] = \sigma^2$ . Note-se que há v.a. que não são nem discretas nem absolutamente contínuas.

Duas v.a.  $X$  e  $Y$  no mesmo espaço de probabilidade dizem-se *equivalentes* ou *quase iguais* ou *iguais com probabilidade um* se  $P[X = Y] = 1$ . Como elas só diferem num conjunto (irrelevante para os nossos fins) de probabilidade nula e têm, portanto, a mesma f.d. e as mesmas propriedades probabilísticas, adoptaremos frequentemente a prática habitual e, abusando da linguagem, escreveremos simplesmente  $X = Y$ , querendo contudo significar  $X = Y$  com probabilidade 1. Esta forma habitual de identificar v.a. equivalentes é uma maneira informal de dizer que vamos trabalhar com as *classes de equivalência* de variáveis aleatórias em vez das próprias v.a. e que identificamos uma classe de equivalência por qualquer das suas representantes (por representante entendemos uma v.a. pertencente à classe).<sup>5</sup> Se fizermos isso, podemos definir, para  $p \geq 1$ , o espaço  $L^p(\Omega, \mathcal{F}, P)$ , abreviadamente  $L^p$ , das variáveis aleatórias  $X$  (na realidade, é o espaço das suas classes de equivalência) tais que  $\mathbb{E}[|X|^p] = \int_{\Omega} |X|^p dP < +\infty$ .

Recorda-se que, se  $\int_{\Omega} |Y(\omega)| dP < \infty$ , dizemos que  $\mathbb{E}[Y] = \int_{\Omega} Y(\omega) dP = \int_{-\infty}^{+\infty} y dP_Y = \int_{-\infty}^{+\infty} y dF_Y(y)$  é a *esperança matemática*, *valor esperado*, *valor médio* ou *média* da v.a.  $Y$ . Quando  $Y$  é absolutamente contínua com f.d.p.  $f_Y$ , resulta  $\mathbb{E}[Y] = \int_{-\infty}^{+\infty} y f_Y(y) dy$ . Quando  $Y$  é v.a. discreta com átomos  $y_k$  ( $k = 1, 2, \dots$ ), resulta  $\mathbb{E}[Y] = \sum_k y_k P[Y = y_k]$ . Estas são consequências das propriedades do integral (no sentido de Lebesgue). O integral define-se:

- Para funções simples  $Y(\omega) = \sum_{i=1}^n c_i I_{A_i}(\omega)$  (onde os conjuntos  $A_i$  são disjuntos dois a dois com  $\bigcup_{i=1}^n A_i = \Omega$ ,  $I_{A_i}$  são as suas funções indicatrizes e  $c_i$  são números reais) por  $\int_{\Omega} Y dP = \sum_{i=1}^n c_i P(A_i)$ .

Note-se que a *função indicatriz*<sup>6</sup>  $I_A$  de um conjunto  $A \in \mathcal{F}$  é uma

(extensão da medida de comprimento) nula.

<sup>5</sup>Convém lembrar que as v.a. são funções do acaso  $\omega$  e, portanto, uma propriedade relativa a variáveis aleatórias pode ser verdadeira para certos valores de  $\omega$  e falsa para outros. Quando o conjunto de valores de  $\omega \in \Omega$  para os quais a propriedade é verdadeira tem probabilidade um, dizemos que a propriedade se verifica com probabilidade um ou *quase certamente* (q.c.). Portanto,  $X = Y$  com probabilidade 1 (ou  $X = Y$  q.c.) significa que  $P[X = Y] = 1$  e não exclui que possam existir valores excepcionais de  $\omega$  para os quais  $X(\omega) \neq Y(\omega)$  (claro que o conjunto  $N$  de tais valores excepcionais terá probabilidade zero de ocorrer, isto é,  $P(N) = 0$ ).

<sup>6</sup>A função indicatriz (também chamada *função característica*)  $I_A$  de um conjunto

função simples e facilmente se reconhece que  $\mathbb{E}[I_A] = P(A)$ .

- Para v.a. não-negativas  $Y$  por  $\int_{\Omega} Y dP = \lim_{n \rightarrow \infty} \int_{\Omega} Y_n dP$ , onde  $Y_n$  é qualquer sucessão não-decrescente de funções simples não-negativas convergente para  $Y$  com probabilidade um.
- Para v.a.  $Y$  arbitrárias por  $\int_{\Omega} Y dP = \int_{\Omega} Y^+ dP - \int_{\Omega} Y^- dP$ , onde  $Y^+(\omega) = Y(\omega)I_{[Y \geq 0]}(\omega)$  e  $Y^-(\omega) = -Y(\omega)I_{[Y < 0]}(\omega)$ .

Um espaço  $L^p$  é um espaço de Banach para a norma- $L^p$  definida por  $\|X\|_p = (\mathbb{E}[|X|^p])^{1/p}$ . Para  $p = 2$  é mesmo um espaço de Hilbert real com produto interno  $\langle X, Y \rangle = \mathbb{E}[XY]$ .

O conceito de v.a. pode ser generalizado a várias dimensões. Uma v.a.  $n$ -dimensional ou *vector aleatório*  $\mathbf{X} = [X_1, X_2, \dots, X_n]^T$  (“ $T$ ” significa “transposto”, isto é, usualmente consideramos vectores coluna) é simplesmente um vector de  $n$  variáveis aleatórias definidas no mesmo espaço mensurável. Podemos definir a sua função de distribuição  $F_{\mathbf{X}}(\mathbf{x}) = F_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) := P[X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n]$  para  $\mathbf{x} = [x_1, x_2, \dots, x_n]^T \in \mathbb{R}^n$ , também chamada f.d. conjunta das v.a.  $X_1, X_2, \dots, X_n$ . Podemos definir a esperança matemática de um vector aleatório (ou mesmo de uma matriz aleatória) como sendo o vector (matriz) das esperanças matemáticas das suas coordenadas. Se existir um conjunto contável  $S \in \mathbb{R}^n$  tal que  $P[\mathbf{X} \in S] = 1$ , dizemos que o vector aleatório  $\mathbf{X}$  é discreto. Se existir uma f.d.p.  $f_{\mathbf{X}}(\mathbf{x})$  tal que  $F_{\mathbf{X}}(\mathbf{x}) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \dots \int_{-\infty}^{x_n} f_{X_1, X_2, \dots, X_n}(y_1, y_2, \dots, y_n) dy_1 \dots dy_2 dy_1$ , então  $\mathbf{X}$  diz-se absolutamente contínuo, caso em que  $f_{\mathbf{X}}(\mathbf{x}) = \frac{\partial^n F_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n)}{\partial x_1 \partial x_2 \dots \partial x_n}$

(definida com possível excepção de pontos de um conjunto negligenciável, isto é, cuja medida de Lebesgue  $n$ -dimensional é nula). Por exemplo, um vector aleatório *normal* ou *gaussiano*  $\mathbf{X}$  com vector médio  $\boldsymbol{\mu}$  e matriz de variância-covariância  $\mathbf{C}$  (matriz cujos elementos  $c_{ij} = \mathbb{E}[(X_i - \mathbb{E}[X_i])(X_j - \mathbb{E}[X_j])]$  são as covariâncias dos pares de variáveis ou, no caso dos elementos diagonais, as variâncias) é, supondo  $\mathbf{C}$  matriz definida positiva, um vector aleatório absolutamente contínuo com f.d.p.  $f_{\mathbf{X}}(\mathbf{x}) = (2\pi)^{-n/2} \sqrt{\det(\mathbf{C}^{-1})} \exp(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \mathbf{C}^{-1}(\mathbf{x} - \boldsymbol{\mu}))$ .

Os conceitos de espaços e normas  $L^p$  podem ser generalizados a vectores aleatórios  $n$ -dimensionais  $\mathbf{X}$  interpretando  $|\mathbf{X}|$  como a norma euclídeana; no caso  $p = 2$ , o conceito de produto interno pode ser generalizado usando  $\langle \mathbf{X}, \mathbf{Y} \rangle = \mathbb{E}[\mathbf{X}^T \mathbf{Y}]$ .

---

A define-se por  $I_A(\omega) = 1$  se  $\omega \in A$  e  $I_A(\omega) = 0$  se  $\omega \notin A$ .

## 2.2 Esperanças matemáticas e probabilidades condicionais

Esta secção pode ser dispensada pelo leitor menos preocupado com o rigor matemático desde que tenha uma ideia informal das esperanças matemáticas e probabilidades condicionais.

Dada uma v.a.  $X \in L^1$  e uma sub-álgebra- $\sigma$   $\mathcal{H} \subset \mathcal{F}$ , existe uma v.a.  $Y = \mathbb{E}[X|\mathcal{H}]$ , chamada a *esperança matemática condicional* de  $X$  dada  $\mathcal{H}$ , a qual é mensurável- $\mathcal{H}$  e tal que  $\int_H X dP = \int_H Y dP$  para todo o  $H \in \mathcal{H}$ . Isto é,  $Y$  é uma v.a. mensurável- $\mathcal{H}$  que tem as mesmas médias que  $X$  sobre os conjuntos de  $\mathcal{H}$ . O teorema de Radon-Nikodym assegura a existência e a unicidade quase certa (q.c.), isto é, podem existir várias v.a. satisfazendo as condições referidas mas, dadas duas quaisquer dessas v.a., elas são iguais com probabilidade um.

No exemplo dos dois dados, o valor esperado de  $X$  = “soma do número de pintas dos dois dados” é um número real  $\mathbb{E}[X] = \sum_{\omega \in \Omega} X(\omega) \frac{1}{36} = \sum_{x=2}^{12} xP[X=x] = 7$ . Suponhamos que só tínhamos informação sobre se o número de pintas no dado vermelho era par ou ímpar; esta informação é dada pela sub-álgebra- $\sigma$   $\mathcal{H} = \{C, C^c, \emptyset, \Omega\}$ , onde  $C = \{(2, 1), (2, 2), \dots, (2, 6), (4, 1), (4, 2), \dots, (4, 6), (6, 1), (6, 2), \dots, (6, 6)\}$  é o acontecimento “número par de pintas no dado vermelho”. Então tem-se  $\mathbb{E}[X|\mathcal{H}](\omega) = \frac{1}{P(C)} \sum_{\omega \in C} X(\omega) \frac{1}{36} = 145/18$  (valor médio de  $X(\omega)$  no conjunto  $C$ ) para qualquer  $\omega \in C$  (não conseguimos distinguir entre os diferentes  $\omega$  de  $C$ ) e  $\mathbb{E}[X|\mathcal{H}](\omega) = \frac{1}{P(C^c)} \sum_{\omega \in C^c} X(\omega) \frac{1}{36} = 107/18$  (valor médio de  $X(\omega)$  no conjunto  $C^c$ ) para qualquer  $\omega \in C^c$ .

Note-se que  $\mathbb{E}[X|\mathcal{H}]$  é uma v.a. (depende de  $\omega$ ) e, portanto, podemos calcular a sua esperança matemática. Da definição de esperança condicional é óbvio que  $\mathbb{E}[\mathbb{E}[X|\mathcal{H}]] = \mathbb{E}[X]$ .

Citamos mais algumas propriedades importantes das esperanças matemáticas condicionais, onde se supõe que as v.a. envolvidas estão em  $L^1$  e que as álgebras- $\sigma$   $\mathcal{G}$  e  $\mathcal{H}$  estão contidas em  $\mathcal{F}$ :

$$\begin{aligned} X \text{ mensurável-}\mathcal{H} &\Rightarrow \mathbb{E}[X|\mathcal{H}] = X \\ X \text{ mensurável-}\mathcal{H} &\Rightarrow \mathbb{E}[XY|\mathcal{H}] = X\mathbb{E}[Y|\mathcal{H}] \\ \mathcal{G} \subset \mathcal{H} &\Rightarrow \mathbb{E}[\mathbb{E}[X|\mathcal{H}]|\mathcal{G}] = \mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}] = \mathbb{E}[X|\mathcal{G}]. \end{aligned}$$

Para  $A \in \mathcal{F}$ , podemos definir a *probabilidade condicional*  $P(A|\mathcal{H}) := \mathbb{E}[I_A|\mathcal{H}]$ . Este conceito é uma extensão do conceito clássico  $P(A|C) = P(A \cap C)/P(C)$  definido para  $A \in \mathcal{F}$  e  $C \in \mathcal{F}$  tal que  $P(C) > 0$ . De facto, pondo  $\mathcal{H} = \{C, C^c, \emptyset, \Omega\}$ ,  $P(A|C)$  não é mais que o valor comum de  $\mathbb{E}[I_A|\mathcal{H}](\omega)$  para qualquer dos  $\omega \in C$ .

Quando  $\mathcal{H} = \sigma(Y)$  é a álgebra- $\sigma$  gerada por uma v.a.  $Y$ , definimos

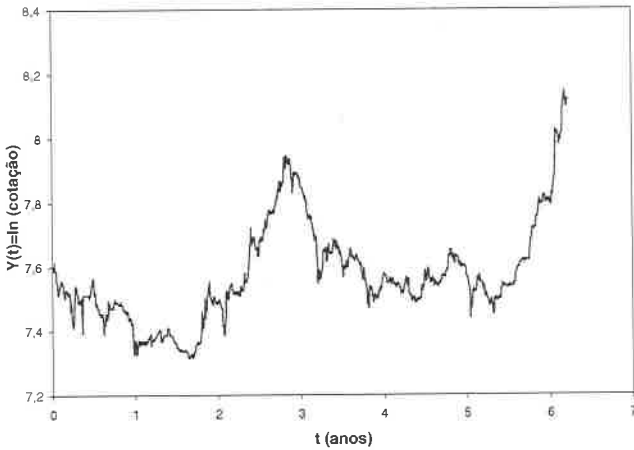
$\mathbb{E}[X|Y] := \mathbb{E}[X|\mathcal{H}]$  e  $P[X \in B|Y] := P(X^{-1}(B)|\mathcal{H})$  para conjuntos de Borel  $B$ . Estas quantidades são v.a., isto é, dependem do acaso  $\omega$ ; de facto, elas dependem essencialmente do valor de  $Y(\omega)$ . Portanto, para  $y \in \mathbb{R}$ , podemos definir  $\mathbb{E}[X|Y = y]$  como o valor (q.c. único) de  $\mathbb{E}[X|Y]$  quando  $Y = y$ . Semelhantemente, pode definir-se  $P[X \in B|Y = y]$ .

## 2.3 Breve revisão de processos estocásticos

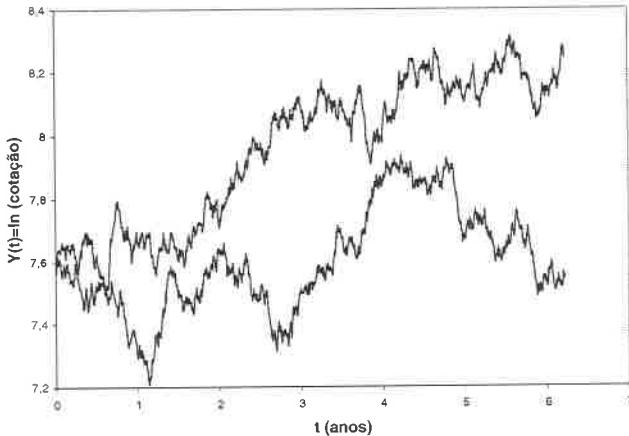
Um *processo estocástico* no espaço de probabilidade  $(\Omega, \mathcal{F}, P)$  é simplesmente uma colecção indexada  $\{X_t\}_{t \in I}$  de variáveis aleatórias. No nosso caso,  $t$  será interpretado como tempo e o *conjunto de índices*  $I$  será usualmente um intervalo de tempo da forma  $[0, +\infty)$ ,  $(-\infty, +\infty)$  ou  $[a, b]$  (processos estocásticos em tempo contínuo). Noutras situações,  $I$  pode ser o conjunto dos inteiros ou dos inteiros não-negativos (processos estocásticos em tempo discreto), um intervalo de  $\mathbb{R}^d$  (processos espaciais) ou qualquer conjunto conveniente. Como cada variável aleatória  $X_t = X_t(\omega)$  é função do “acaso”  $\omega \in \Omega$ , um processo estocástico pode ser considerado uma função de duas variáveis,  $t \in I$  e  $\omega \in \Omega$ , isto é uma função do tempo e do acaso; como no caso das variáveis aleatórias, é hábito abreviar a notação e escrever simplesmente  $X_t$  em vez de  $X_t(\omega)$ , mas, apesar de o “acaso”  $\omega$  não aparecer explicitamente, não devemos esquecer que o valor do processo estocástico depende dele. Esta função de  $t$  e  $\omega$  não é uma função arbitrária pois está sujeita à restrição de ser, para cada  $t$  fixo, uma função mensurável de  $\omega$ , isto é, uma v.a..

Se fixarmos o “acaso”  $\omega$ , obtemos uma função apenas do tempo, a que se chama uma *trajectória*<sup>7</sup> do processo estocástico. Um processo estocástico pode, portanto, ser também interpretado como uma colecção de trajectórias, uma para cada estado do acaso  $\omega$ . A cotação  $X_t$  (abreviatura de  $X_t(\omega)$ ) de uma acção no instante  $t$  para  $t \in I = [0, +\infty)$  é um exemplo de processo estocástico. Naturalmente, diferentes cenários do mercado resultarão em cotações diferentes. Para um  $t \in I$  fixo,  $X_t$  é uma v.a. e, portanto, uma função de  $\omega$  que associa a cada cenário do mercado  $\omega \in \Omega$  a correspondente cotação  $X_t(\omega)$  da acção no instante  $t$ . Para um cenário do mercado  $\omega \in \Omega$  fixo, a trajectória  $X_t(\omega)$  é uma função do tempo  $t$  que associa a cada instante  $t \in I$  a correspondente cotação da acção debaixo desse cenário. Quando observamos a variação da cotação da acção ao longo do tempo e desenhamos o respectivo gráfico, estamos de facto a desenhar uma trajectória, a correspondente ao cenário do mercado  $\omega$  que o “acaso ditou” que fosse o cenário efectivamente ocorrido.

<sup>7</sup>Em inglês usa-se “trajectory” ou “sample path”.



**Figura 2.1:** Trajectória observada do processo estocástico  $Y(t)$ , cotação em escala logarítmica das acções do banco BCP entre 8 de Abril de 1991 ( $t = 0$ ) e 30 de Junho de 1997. Corresponde ao  $\omega$  (cenário do mercado) que efectivamente ocorreu.



**Figura 2.2:** Duas trajectórias do processo estocástico “movimento browniano geométrico” com parâmetros  $R = r - \sigma^2/2 = 0,084/\text{ano}$  e  $\sigma = 0,193/\sqrt{\text{ano}}$  (ver Capítulo 8) simuladas pelo método de Monte Carlo. Este processo foi usado como modelo do processo estocástico  $Y(t)$ , cotação em escala logarítmica das acções do BCP no período 8 de Abril de 1991 ( $t = 0$ ) e 30 de Junho de 1997. Se o modelo for correcto, estas trajectórias correspondem a outros dois cenários do mercado escolhidos aleatoriamente.

Por vezes é mais conveniente usar a notação alternativa  $X(t, \omega)$  (abreviadamente  $X(t)$ ) em vez de  $X_t(\omega)$  (abreviadamente  $X_t$ ). Salvo menção em contrário, não faremos distinção entre as duas notações. Usualmente chama-se *espaço de estados* ao contradomínio da função  $X(t, \omega)$  se bem que, por vezes e por razões de comodidade, se chame espaço de estados a um conjunto que contenha esse contradomínio e que seja mais conveniente para trabalhar.

As *distribuições de dimensão finita* de um processo estocástico  $\{X_t\}_{t \in I}$  são as funções de distribuição conjuntas

$$F_{t_1, t_2, \dots, t_n}(x_1, x_2, \dots, x_n) := P[X_{t_1} \leq x_1, X_{t_2} \leq x_2, \dots, X_{t_n} \leq x_n].$$

A família de todas as distribuições de dimensão finita (isto é, definidas para todo o  $n = 1, 2, \dots$  e todos os  $t_1, t_2, \dots, t_n \in I$ ) determina as propriedades probabilísticas do process estocástico (mas não necessariamente todas as suas propriedades). Esta família obviamente satisfaz, para todo o  $n = 1, 2, \dots$ , todos os  $t_1, t_2, \dots, t_n \in I$  e todos os  $x_1, x_2, \dots, x_n \in \mathbb{R}$ , as seguintes propriedades:

1.  $F_{t_{\sigma(1)}, t_{\sigma(2)}, \dots, t_{\sigma(n)}}(x_{\sigma(1)}, x_{\sigma(2)}, \dots, x_{\sigma(n)}) = F_{t_1, t_2, \dots, t_n}(x_1, x_2, \dots, x_n)$  para todas as permutações  $\sigma$  em  $\{1, 2, \dots, n\}$  (*propriedade de simetria*, isto é, não é importante a ordem das variáveis aleatórias).
2. Para  $k = 1, \dots, n-1$ ,  $F_{t_1, t_2, \dots, t_k, t_{k+1}, \dots, t_n}(x_1, x_2, \dots, x_k, +\infty, \dots, +\infty) = F_{t_1, t_2, \dots, t_k}(x_1, x_2, \dots, x_k)$  (*propriedade de compatibilidade*).

O *teorema da extensão de Kolmogorov* diz que, dada uma família de funções de distribuição  $F_{t_1, t_2, \dots, t_n}$  (de domínio  $\mathbb{R}^n$ ), definida para todo o  $n = 1, 2, \dots$  e todos os  $t_1, t_2, \dots, t_n \in I$  e satisfazendo as propriedades de simetria e compatibilidade, existe pelo menos um espaço de probabilidade e um processo estocástico nesse espaço para o qual a família das distribuições finitas é a família dada.<sup>8</sup> De facto, seria possível construir vários espaços de probabilidade e processos estocásticos neles definidos satisfazendo a propriedade desejada.

<sup>8</sup>A demonstração é construtiva. O espaço amostral usado foi o conjunto  $\mathbb{R}^I$  das funções reais  $\omega = \omega(\cdot)$  definidas em  $I$ . Consideremos nesse espaço a classe dos conjuntos cilíndricos da forma  $A = \{\omega \in \mathbb{R}^I : (\omega_{t_1}, \omega_{t_2}, \dots, \omega_{t_n}) \in B\}$  ( $n = 1, 2, \dots$ ) cujas bases  $B$  sejam intervalos de  $\mathbb{R}^n$  (produtos cartesianos de  $n$  intervalos reais). A álgebra- $\sigma$   $\mathcal{B}^I$  gerada por esta classe é conhecida por álgebra- $\sigma$  de Borel em  $\mathbb{R}^I$ . Uma probabilidade no espaço mensurável  $(\mathbb{R}^I, \mathcal{B}^I)$  fica univocamente caracterizada se conhecermos as probabilidades dos conjuntos cilíndricos acima referidos. Kolmogorov utilizou a probabilidade definida por  $P(B) = \int_B dF_{t_1, t_2, \dots, t_n}$ . Nesse espaço de probabilidade  $(\mathbb{R}^I, \mathcal{B}^I, P)$  considerou então o processo estocástico  $X_t(\omega) = \omega(t)$  e mostrou (como não é difícil de verificar) que a sua família de distribuições finita é a família dada. Repare-se que o que se fez foi utilizar as trajectórias do processo estocástico como sendo os acontecimentos elementares  $\omega$ .

Dois processos estocásticos  $\{X_t\}_{t \in I}$  e  $\{Y_t\}_{t \in I}$  no mesmo espaço de probabilidade dizem-se *equivalentes* (também se diz que cada um deles é uma *versão* do outro) se, para cada  $t \in I$ , se tiver  $X_t = Y_t$  com probabilidade 1. Processos equivalentes têm as mesmas funções de distribuição finitas e, portanto, as mesmas propriedades probabilísticas, mas podem ter diferentes propriedades analíticas, como veremos a seguir.

No que se segue desta secção, iremos supor que  $I$  é um intervalo da forma  $[a, b]$  ou  $(-\infty, b]$  ou  $[a, +\infty)$  ou  $(-\infty, +\infty)$ . Pode suceder que, dados dois processos estocásticos equivalentes, uma versão tenha todas as trajectórias contínuas e a outra tenha todas as trajectórias descontínuas. É o que acontece se, para  $\Omega = [0, 1]$ ,  $I = [0, 1]$ ,  $\mathcal{F} = \mathcal{B}_{[0,1]}$  (onde  $\mathcal{B}_{[0,1]}$  é a álgebra- $\sigma$  de Borel do intervalo  $[0, 1]$ , que é gerada pelos conjuntos abertos de  $[0, 1]$ ) e  $P$  a distribuição uniforme em  $[0, 1]$  (a probabilidade de cada intervalo é o seu comprimento), considerarmos os processos estocásticos

$$X_t(\omega) = \begin{cases} 0 & \text{se } \omega \neq t \\ 1 & \text{se } \omega = t, \end{cases}$$

(cujas trajectórias são obviamente todas descontínuas) e  $Y_t(\omega) \equiv 0$  (cujas trajectórias são obviamente todas contínuas). Não é difícil constatar que os dois processos são equivalentes pois, para qualquer  $t \in [0, 1]$ , se tem  $P[X_t = Y_t] = P\{\omega \in \Omega : X_t(\omega) = Y_t(\omega)\} = P([0, 1] - \{t\}) = P([0, t[ + ]t, 1]) = t + (1 - t) = 1$ . A razão porque não são ambos processos contínuos reside no facto de os conjuntos excepcionais  $N_t$  dos valores de  $\omega$  para os quais  $X_t(\omega) \neq Y_t(\omega)$  (conjuntos que têm probabilidade zero) poderem variar com  $t$ . Com efeito, o conjunto  $N = \bigcup_{t \in I} N_t$  dos valores de  $\omega$  para os quais as trajectórias dos dois processos diferem (em pelo menos um instante), sendo união não-numerável de conjuntos com probabilidade zero, poderá já não ter probabilidade zero (a propriedade aditiva- $\sigma$  das probabilidades só garante probabilidade zero para uniões numeráveis de conjuntos com probabilidade zero); se isso suceder, embora se tenha  $P[X_t = Y_t] = 1$  para cada  $t$ , vem  $P[X_t = Y_t \text{ para todos os } t \in I] < 1$ . Um conceito mais forte que o de processos equivalentes é o de processos com *trajectórias idênticas* (com probabilidade um), ou seja  $P[X_t = Y_t \text{ para todo o } t \in I] = 1$ , caso em que  $P(N) = 0$ . Quando os dois processos são ambos contínuos, os dois conceitos coincidem. Por *processo contínuo* entende-se um processo cujas trajectórias são contínuas com probabilidade um (dito de outra forma, o conjunto das trajectórias descontínuas tem probabilidade zero). Para evitar o problema surgido no exemplo acima, há vantagem em trabalhar apenas com *processos separáveis*, o que não se traduz em qualquer perda de generalidade do ponto de vista probabilístico visto todo o processo

estocástico admitir uma versão separável. Basicamente, um processo separável é um processo que fica, para efeitos de continuidade, caracterizado pelos seus valores num conjunto numerável denso em  $I$ <sup>9</sup>, o que elimina o problema acima referido, já que podemos trabalhar com uniões numeráveis de conjuntos excepcionais. Assim, se um de dois processos separáveis equivalentes é contínuo, o outro também o será.

O *critério de Kolmogorov* diz que, se existem constantes positivas  $C$ ,  $\alpha$  e  $\beta$  tais que  $\mathbb{E}[|X_t - X_s|^\alpha] \leq C|t - s|^{1+\beta}$  para todo o  $s, t \in I$ , então existe uma versão separável de  $X_t$  que é contínua.

Isso não significa que o próprio processo  $X_t$  seja contínuo, como mostra o exemplo acima em que o processo  $X_t$  satisfaz o critério de Kolmogorov e não é contínuo. Tem, porém, uma versão separável contínua  $Y_t$ .

Daqui por diante, convencionamos que trabalharemos sempre com versões separáveis dos processos estocásticos. Com esta convenção, um processo que satisfaça o critério de Kolmogorov é contínuo.

Um processo contínuo é também um processo mensurável. Para um processo  $X(t, \omega)$  ser *mensurável* não basta que seja mensurável com respeito a cada uma das variáveis  $t$  e  $\omega$ , é preciso que, como função conjunta das duas variáveis, seja mensurável. Isto é, considerando o espaço mensurável produto<sup>10</sup>  $(I \times \Omega, \mathcal{B}_I \times \mathcal{F})$ , a imagem inversa (conjunta)  $\{(t, \omega) \in I \times \Omega : X(t, \omega) \in B\}$  de qualquer conjunto de Borel  $B \in \mathcal{B}$  pertence à álgebra- $\sigma$   $\mathcal{B}_I \times \mathcal{F}$ .

Podemos indiferentemente definir um processo estocástico  $n$ -dimensional  $\{X_t\}_{t \in I}$  como uma colecção indexada de vectores aleatórios  $n$ -dimensionais ou como um vector  $n$ -dimensional cujos elementos são processos estocásticos com o mesmo conjunto de índices definidos no mesmo espaço de probabilidade.

---

<sup>9</sup>Mais precisamente, diz-se que um processo  $\{X_t\}_{t \in I}$  é separável se existir um conjunto numerável  $J$  denso em  $I$  tal que, para todo o subintervalo aberto  $(a, b) \subset I$  e todo o subconjunto fechado  $A \subset \mathbb{R}$ , se verifique que a diferença entre os conjuntos

$$\{\omega : X_t(\omega) \in A \text{ para todo } t \in (a, b) \cap J\} \text{ (que pertence a } \mathcal{F})$$

e

$$\{\omega : X_t(\omega) \in A \text{ para todo } t \in (a, b)\} \text{ (que pode não pertencer a } \mathcal{F})$$

está contida num conjunto de probabilidade nula. Se convencionarmos trabalhar num espaço de probabilidade completo e o processo for separável, então ambos os conjuntos são mensuráveis (pertencem a  $\mathcal{F}$ ) e a sua diferença é mensurável e tem mesmo probabilidade nula. No exemplo acima,  $X_t$  não é separável mas  $Y_t$  é uma versão separável de  $X(t)$ .

<sup>10</sup> $\mathcal{B}_I \times \mathcal{F}$  é a álgebra- $\sigma$  gerada pelos conjuntos da forma  $G \times A$  com  $G \in \mathcal{B}_I$  (a álgebra- $\sigma$  de Borel no intervalo  $I$ ) e  $A \in \mathcal{F}$ .



## 2.4 Breve revisão de processos estacionários

Um processo estocástico diz-se *estritamente estacionário* se as suas distribuições de dimensão finita forem invariantes para translações no tempo, isto é,

$$F_{t_1, t_2, \dots, t_n}(x_1, x_2, \dots, x_n) = F_{t_1 + \tau, t_2 + \tau, \dots, t_n + \tau}(x_1, x_2, \dots, x_n)$$

para qualquer  $n = 1, 2, \dots$ , quaisquer  $x_1, x_2, \dots, x_n \in \mathbb{R}$ , quaisquer  $t_1, t_2, \dots, t_n \in I$  e qualquer  $\tau \in \mathbb{R}$  tal que  $t_1 + \tau, t_2 + \tau, \dots, t_n + \tau \in I$ . Caso  $X_t \in L^2$  para todo o  $t \in I$  e seja estritamente estacionário, então vem  $\mathbb{E}[X_t] = m = \text{constante}$  e  $COV[X_s, X_t] = C(t - s)$  ( $C$  é chamada *função de auto-covariância*), o que constitui a definição de *processo estacionário em sentido lato*, também chamado *estacionário de segunda ordem* ou simplesmente *estacionário*. Contudo, um processo estocástico estacionário em sentido lato pode não ser estritamente estacionário. Nos *processos gaussianos*, ou seja os processos com distribuições de dimensão finita gaussianas, pelo facto de estas serem completamente caracterizadas pelos momentos de primeira e segunda ordem, verifica-se a coincidência dos dois conceitos de estacionaridade.

Um processo estacionário em sentido lato que seja *contínuo em média quadrática* (isto é, contínuo com respeito à norma  $L^2$ ) tem *função de distribuição espectral*  $F(\lambda)$  (a qual indica como estão distribuídas as frequências das oscilações harmónicas de  $X_t$ ) e a *função de auto-covariância* é dada por  $C(t) = \int_{-\infty}^{+\infty} \exp(it\lambda) dF(\lambda)$ . Se  $F$  tiver uma *densidade*  $f$ , a *densidade espectral*, e  $C$  for integrável, então a *densidade* é a transformada de Fourier da auto-covariância,  $f(\lambda) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \exp(-it\lambda) C(t) dt$ .

## 2.5 Filtrações, martingalas e tempos de Markov

Consideremos um intervalo de tempo  $I = [0, d]$  com  $0 \leq d \leq +\infty$  (quando  $d = +\infty$ , interpretamos  $I = [0, +\infty)$ ). A suposição de que o intervalo começa em 0 foi feita apenas por comodidade mas não é obrigatória. Considere um espaço de probabilidade  $(\Omega, \mathcal{F}, P)$  e um processo estocástico  $\{X_t\}_{t \in I}$  nesse espaço. Seja  $\{\mathcal{F}_t\}_{t \in I}$  uma família de sub-álgebras- $\sigma$  de  $\mathcal{F}$  tais que  $\mathcal{F}_s \subset \mathcal{F}_t$  quando  $s \leq t$ ; uma tal família chama-se uma *filtração*. É frequente escolher a *filtração natural*  $\mathcal{F}_t = \sigma(X_s; 0 \leq s \leq t)$ , constituída pelas álgebras- $\sigma$  geradas pelo processo até ao instante  $t$  (contendo a informação do passado e do presente do processo  $X_t$ ). Por vezes, porém, podem ser necessárias outras escolhas caso necessitemos que a filtração contenha informação adicional, como, por exemplo, a informação contida na condição inicial de uma equação diferencial estocástica.

Dizemos que o processo estocástico  $\{X_t\}_{t \in I}$  está *adaptado à filtração*  $\{\mathcal{F}_t\}_{t \in I}$  se, para cada  $t \in I$ ,  $X_t$  for mensurável- $\mathcal{F}_t$ .<sup>11</sup> Claro que um processo estocástico  $\{X_t\}_{t \in I}$  está sempre adaptado à sua filtração natural ou qualquer outra filtração cujas álgebras- $\sigma$  contenham, para cada  $t \in I$ , as correspondentes álgebras- $\sigma$   $\sigma(X_s; 0 \leq s \leq t)$  da filtração natural.

Dado um espaço de probabilidade  $(\Omega, \mathcal{F}, P)$ , um processo estocástico  $\{X_t\}_{t \in I}$  e uma filtração  $\{\mathcal{F}_t\}_{t \in I}$  nesse espaço, dizemos que o processo estocástico é uma *martingala- $\mathcal{F}_t$*  se o processo estiver adaptado à filtração, se  $X_t$  for integrável ( $\mathbb{E}[|X_t|] < +\infty$ , ou seja  $X_t \in L^1$ ) para todo o  $t \in I$  e se

$$\mathbb{E}[X_t | \mathcal{F}_s] = X_s \text{ q.c. para qualquer } s \leq t. \quad (2.1)$$

No caso da filtração natural, em vez de  $\mathbb{E}[X_t | \mathcal{F}_s]$  pode também escrever-se  $\mathbb{E}[X_t | X_u, 0 \leq u \leq s]$ . Quando for claro qual é a filtração com que se trabalha, é costume abreviar e falar de *martingala* em vez de *martingala- $\mathcal{F}_t$* . Também é costume dizer simplesmente *martingala* quando não se indica qualquer filtração, caso em que se subentende que se trabalha com a filtração natural.

O conceito de martingala está ligado ao conceito de jogo equilibrado. De facto, se  $X_s$  são os seus ganhos acumulados no instante presente  $s$  e se o jogo é equilibrado, então, dada a informação disponível até ao instante presente  $s$ , espera manter num instante futuro  $t$  os seus ganhos presentes.

Se substituirmos “=” por “ $\leq$ ” (respectivamente por “ $\geq$ ”) em (2.1), temos uma *supermartingala* (respectivamente uma *submartingala*).

Aqui só nos interessam processos em tempo contínuo. Porém, todos estes conceitos (filtração, filtração natural, martingala, supermartingala e submartingala) para processos em tempo discreto têm definições em tudo análogas, com excepção do facto de o conjunto de índices ser da forma  $I = \{0, 1, 2, \dots, d\}$  com  $0 \leq d < +\infty$  (quando  $d = +\infty$ , interpreta-mos  $I = \mathbb{N}$ ).

Se  $X_t$  é uma martingala [supermartingala, submartingala], então  $\mathbb{E}[X_t]$  é uma função constante [não-crescente, não-decrescente] de  $t$ . Se  $X_t \in L^p$  para  $p \geq 1$  é uma martingala, então, para intervalos finitos  $[a, b] \subset I$  e para  $c > 0$ , resultam as *desigualdades maximais das martingalas*:

$$P \left[ \sup_{t \in [a, b]} |X_t| \geq c \right] \leq \frac{|X_b|^p}{c^p} \quad (2.2)$$

<sup>11</sup> Isso garante que as imagens inversas de conjuntos de Borel pelas v.a.  $X_t$  estão em  $\mathcal{F}_t$  e não apenas em  $\mathcal{F}$ .

e, se  $p > 1$ ,

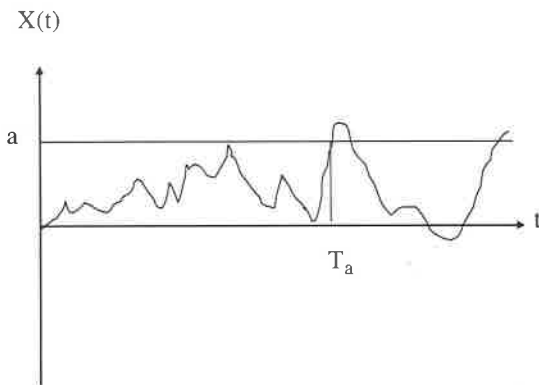
$$\mathbb{E} \left[ \sup_{t \in [a, b]} |X_t|^p \right] \leq \left( \frac{p}{p-1} \right)^p \mathbb{E} [|X_b|^p]. \quad (2.3)$$

Vamos agora definir o conceito de tempo de Markov ou tempo de paragem em contexto de tempo contínuo (trabalhando com o conjunto de índices  $I = [0, +\infty)$ ), fazendo no entanto notar que igual conceito se pode definir de forma análoga no contexto de tempo discreto (usando o conjunto de índices  $I = \mathbb{N}_0 = \{0, 1, 2, \dots\}$ ). Consideremos um espaço de probabilidade  $(\Omega, \mathcal{F}, P)$ , seja  $I = [0, +\infty)$  e consideremos uma filtração  $\{\mathcal{F}_t\}_{t \in I}$  definida naquele espaço. Seja  $T$  uma *v.a. estendida* (também dita “imprópria”), isto é uma aplicação de  $\Omega$  em  $\bar{\mathbb{R}} := \mathbb{R} + \{+\infty\}$  cujas imagens inversas de borelianos de  $\bar{\mathbb{R}}$  sejam conjuntos de  $\mathcal{F}$ . Os referidos borelianos formam a álgebra- $\sigma$  de Borel  $\bar{\mathcal{B}}$  gerada pelos intervalos de  $\bar{\mathbb{R}}$  (que são os intervalos de  $\mathbb{R}$  mas permitindo agora intervalos fechados em  $+\infty$ ). Suporemos que  $T$  toma valores em  $[0, +\infty]$ , isto é, o seu contradomínio está contido nesse intervalo. Dizemos que  $T$  é um *tempo de Markov- $\mathcal{F}_t$*  ou *tempo de paragem- $\mathcal{F}_t$*  se, dado qualquer instante  $t$  fixo, o acontecimento  $[T \leq t] = \{\omega \in \Omega : T(\omega) \leq t\} \in \mathcal{F}_t$ . Frequentemente, usa-se como filtração a filtração natural de um processo estocástico  $\{X_t\}_{t \in I}$  definido no espaço de probabilidade referido. Nesse caso, a definição de tempo de Markov  $T$  garante que, para determinar a sua função de distribuição  $F_T(t) = P[T \leq t]$  no instante  $t$ , só precisamos da informação contida no processo estocástico até esse instante (isto é, não precisamos de “adivinhar” informação futura). Por vezes há necessidade de incorporar informação adicional não contida no processo estocástico, caso em que se usam filtrações não-naturais mas em relação às quais o processo esteja adaptado. Note-se que  $F_T$  pode ser uma *função de distribuição imprópria*, isto é, pode suceder que  $F_T(+\infty) := \lim_{t \rightarrow +\infty} F_T(t) < 1$ . Isso sucede se e só se  $P[T = +\infty] > 0$  e vem  $F_T(+\infty) + P[T = +\infty] = 1$ .

Quando for claro qual a filtração com que se trabalha, é costume abreviar e falar simplesmente em “tempo de Markov” sem indicar qual a filtração. É também o que se faz quando se trabalha com um processo estocástico e a sua filtração natural.

Se  $T$  é um tempo de Markov- $\mathcal{F}_t$ , pode ser útil definir uma álgebra- $\sigma$   $\mathcal{F}_T$  que desempenha para  $T$  um papel análogo ao papel que  $\mathcal{F}_t$  desempenha para  $t$ . Para isso considera-se a menor álgebra- $\sigma$   $\mathcal{F}_{+\infty}$  que contém todas as álgebra- $\sigma$   $\mathcal{F}_t$  para  $t \geq 0$  e define-se  $\mathcal{F}_T$  como a álgebra- $\sigma$  formada por todos os conjuntos  $A \in \mathcal{F}_{+\infty}$  tais que  $A \cap [T \leq t] \in \mathcal{F}_t$ .

O exemplo mais típico dos tempos de Markov são os *tempos de primeira passagem*  $T_a$  de um processo estocástico  $X_t$  por um limiar  $a \in \mathbb{R}$



**Figura 2.3:** Tempo de primeira passagem  $T_a = T_a(\omega)$  para uma trajetória  $\omega$  de um processo estocástico  $X(t)$  pelo limiar  $a$ .

(ver uma ilustração na Figura 2.3), que é por definição

$$T_a = \inf\{t \geq 0 : X_t = a\}.$$

Recorda-se que, por convenção, esse ínfimo é igual a  $+\infty$  quando o conjunto a que se aplica é vazio. Logo  $T_a(\omega) = +\infty$  quando não existe nenhum  $t \geq 0$  tal que  $X_t(\omega) = a$ , ou seja quando a trajetória  $\omega$  do processo  $X_t$  nunca passa por  $a$ . Reconhece-se que  $T_a$  é um tempo de Markov (para a filtração natural do processo estocástico) pois para saber se o acontecimento  $[T_a \leq t]$  ocorreu ou não ocorreu só precisamos de conhecer a trajetória de  $X_t$  até ao instante  $t$ . O mesmo já não é verdade se estivermos a falar do tempo de última passagem por  $a$ , que não é claramente um tempo de Markov.

## 2.6 Processos de Markov

Em palavras correntes, um *processo de Markov* é um processo estocástico em que, conhecido o seu valor presente, os valores futuros são independentes dos valores passados. Por outras palavras, quando alguém conhece *exactamente* o valor presente do processo, conhecer ou não como é que o processo evoluiu no passado para chegar a esse valor presente é irrelevante para o cálculo de probabilidades de acontecimentos futuros.

A palavra “exactamente” é essencial, já que o conhecimento impreciso ou aproximado sobre o valor presente não garante a independência referida.

Como na Secção 2.5, consideramos um intervalo de tempo  $I = [0, d]$  com  $0 \leq d \leq +\infty$ <sup>12</sup>, um espaço de probabilidade  $(\Omega, \mathcal{F}, P)$  e um processo estocástico  $\{X_t\}_{t \in I}$  definido nesse espaço. O processo é um *processo de Markov em tempo contínuo*<sup>13</sup> se, para quaisquer  $s, t \in I$ ,  $s \leq t$  e qualquer conjunto de Borel  $B \in \mathcal{B}$ , tivermos<sup>14</sup>

$$P[X_t \in B | X_u, 0 \leq u \leq s] = P[X_t \in B | X_s]. \quad (2.4)$$

Note-se que  $P[X_t \in B | X_u, 0 \leq u \leq s] = P[X_t \in B | \mathcal{F}_s]$ , onde  $\mathcal{F}_s$  é a álgebra- $\sigma$  gerada pelas v.a.  $X_u$ ,  $0 \leq u \leq s$ .

A propriedade referida, também conhecida por *propriedade de Markov*, é equivalente a

$$P[X_t \in B | X_{t_1} = x_1, \dots, X_{t_{n-1}} = x_{n-1}, X_{t_n} = x_n] = P[X_t \in B | X_{t_n} = x_n] \quad (2.5)$$

para qualquer  $n = 1, 2, \dots$ , quaisquer  $t_1 \leq \dots \leq t_{n-1} \leq t_n \leq t$ , quaisquer  $x_1, \dots, x_{n-1}, x_n \in \mathbb{R}$  e qualquer conjunto de Borel  $B$ . De facto, nem sequer precisamos de verificar a propriedade para todos os conjuntos de Borel, mas apenas para os conjuntos de uma classe geradora de  $\mathcal{B}$ , como por exemplo a classe dos intervalos  $(-\infty, x]$ . Assim, a propriedade de Markov é equivalente à seguinte igualdade entre funções de distribuição condicionais:

$$F_{X_t | X_{t_1} = x_1, \dots, X_{t_{n-1}} = x_{n-1}, X_{t_n} = x_n}(x) = F_{X_t | X_{t_n} = x_n}(x) \quad (2.6)$$

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

Um outra propriedade equivalente à propriedade de Markov é a de se ter, para qualquer v.a.  $Y$  mensurável- $\mathcal{F}_d$  e integrável e para  $s, t \in I$ ,  $s \leq t$ ,

$$\mathbb{E}[Y | X_u, 0 \leq u \leq s] = \mathbb{E}[Y | X_s]. \quad (2.7)$$

Consideremos conjuntos de Borel  $B, B_1, \dots, B_n$  e seja  $x \in \mathbb{R}$ . Podemos definir as probabilidades (incondicionais)  $P_t(B) := P[X_t \in B]$  (distribuição no instante  $t$ ) e  $P_{t_1, \dots, t_n}(B_1, \dots, B_n) := P[X_{t_1} \in B_1, \dots, X_{t_n} \in B_n]$  (distribuição conjunta nos instantes  $t_1, \dots, t_n$ ). Podemos também definir

<sup>12</sup>O intervalo pode começar noutra instante  $t_0$  mas, por simplicidade, suporemos que começa em 0.

<sup>13</sup>Aqui só nos interessam processos de Markov em tempo contínuo. Porém, a definição de processo de Markov em tempo discreto é em tudo análoga, com excepção do facto de o conjunto de índices ser da forma  $I = \{0, 1, 2, \dots, d\}$  com  $0 \leq d \leq +\infty$ .

<sup>14</sup>Embora não explicitamente referido, a igualdade de probabilidades condicionais é apenas quase certa, isto é, verifica-se com probabilidade um. Continuaremos daqui por diante a omitir essa referência explícita.

as probabilidades de transição  $P(t, B|s, x) := P[X_t \in B | X_s = x]$  para  $s \leq t$ . Claro que  $P(s, B|s, x) = I_B(x)$  ( $= 1$  se  $x \in B$ ,  $= 0$  se  $x \notin B$ ). Usando o teorema das probabilidades totais e uma das formas da propriedade de Markov, obtemos, para  $s \leq u \leq t$  e  $0 < t_1 < t_2 < \dots < t_{n-1} < t_n$ ,

$$P_t(B) = \int_{\mathbb{R}} P(t, B|s, z) P_s(dz) \quad (2.8)$$

e

$$P_{t_1, \dots, t_n}(B_1, \dots, B_n) = \int_{\mathbb{R}} \int_{B_1} \dots \int_{B_{n-1}} P(t_n, B_n | t_{n-1}, x_{n-1}) P(t_{n-1}, dx_{n-1} | t_{n-2}, x_{n-2}) \dots P(t_1, dx_1 | 0, x_0) P_0(dx_0). \quad (2.9)$$

Assim, conhecidas a distribuição de probabilidade  $P_0$  de  $X_0$  e as probabilidades de transição, podemos obter as distribuições de dimensão finita do processo, que completamente o caracterizam do ponto de vista probabilístico. Também podemos facilmente obter as equações de Chapman-Kolmogorov

$$P(t, B|s, x) = \int_{\mathbb{R}} P(t, B|u, z) P(u, dz|s, x) \quad (s < u < t). \quad (2.10)$$

No caso de existir uma densidade de transição  $p(t, y|s, x) := f_{X_t | X_s = x}(y) = \frac{\partial}{\partial y} F_{X_t | X_s = x}(y)$ , as equações de Chapman-Kolmogorov tomam a forma

$$p(t, y|s, x) = \int_{\mathbb{R}} p(t, y|u, z) p(u, z|s, x) dz \quad (s < u < t). \quad (2.11)$$

Um processo de Markov homogêneo (no tempo) é um processo de Markov cujas probabilidades de transição são estacionárias, isto é,

$$P(t + \tau, B|s + \tau, x) = P(t, B|s, x),$$

caso em que são apenas função de  $x$ ,  $B$  e  $t - s$ , podendo escrever-se  $P(t - s, B|x) := P(t, B|s, x)$ . Se existir densidade de transição, podemos escrever  $p(t - s, y|x) := p(t, y|s, x)$ . É importante não confundir um processo de Markov homogêneo com um processo de Markov estacionário; o último tem funções de distribuição finitas estacionárias, o primeiro apenas tem probabilidades de transição estacionárias.

Note-se que, para um processo de Markov homogêneo com  $I = [0, d]$ , a propriedade de Markov (2.4) pode escrever-se na forma

$$P[X_{s+t} \in B | \mathcal{F}_s] = P[X_t \in B | X_0]$$

para quaisquer  $t \geq 0$  com  $s, s + t \in I$  e conjunto de Borel  $B$  (pois

$P[X_{s+t} \in B | X_s] = P[X_t \in B | X_0]$ . Aqui  $\mathcal{F}_s$  é a álgebra- $\sigma$  gerada pelas v.a.  $X(u)$  ( $0 \leq u \leq s$ ). A propriedade equivale a  $\mathbb{E}[h(X_{s+t}) | \mathcal{F}_s] = \mathbb{E}[h(X_t) | X_0]$  para funções  $h$  mensuráveis-Borel limitadas.

Um processo de Markov homogêneo  $\{X_t\}_{t \geq 0}$  diz-se um *processo de Markov forte* (ou que satisfaz a *propriedade de Markov forte*) se, para quaisquer  $t \geq 0$  e conjunto de Borel  $B$ , se tiver

$$P[X_{S+t} \in B | \mathcal{F}_S] = P[X_t \in B | X_0] \quad (2.12)$$

para todos os tempos de Markov  $S$  (com respeito à filtração natural do processo). Esta propriedade é equivalente a ter-se

$$\mathbb{E}[h(X_{S+t}) | \mathcal{F}_S] = \mathbb{E}[h(X_t) | X_0] \quad (2.13)$$

para funções  $h$  mensuráveis-Borel limitadas.

Os conceitos desta secção podem facilmente generalizar-se a processos estocásticos  $n$ -dimensionais. Basta substituir  $\mathbb{R}$  por  $\mathbb{R}^n$  e considerar agora conjuntos de Borel em  $\mathbb{R}^n$ .

## Capítulo 3

# Uma introdução informal às equações diferenciais estocásticas

Seja  $X = X(t)$  o tamanho de uma população de seres vivos no instante  $t \geq 0$  e seja  $X(0) = x_0$  o seu tamanho inicial. Suponhamos que não há limitações (alimentares ou territoriais) ao crescimento. A dinâmica da população pode ser descrita pelo *modelo malthusiano*

$$\frac{dX}{dt} = rX, \quad (3.1)$$

que diz simplesmente que a taxa instantânea de crescimento da população é proporcional ao tamanho da população. A taxa (instantânea) de crescimento *per capita*  $\frac{1}{X} \frac{dX}{dt}$  é a constante de proporcionalidade  $r$ . A solução para este modelo de tipo multiplicativo é a lei malthusiana (crescimento exponencial)

$$X(t) = x_0 \exp(rt). \quad (3.2)$$

O mesmo modelo pode aplicar-se se  $X(t)$  representar o valor de uma obrigação com taxa (instantânea) de rendimento  $r$  fixa, ou o capital de um depósito bancário com taxa (instantânea) de juro  $r$  fixa ou o valor de um bem ou recurso quando a taxa (instantânea) de inflação  $r$  é constante.

A equação diferencial ordinária (EDO) (3.1), que podemos também escrever na forma  $dX = (rdt)X$ , pode obter-se como o limite quando  $\Delta t \rightarrow 0$  do modelo em tempo discreto (equação às diferenças)