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

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: nº 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

1 Introdução 1

2 Revisão de probabilidades e processos estocásticos 7
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

3 Uma introdução informal às equações diferenciais estocásticas 25

4 O processo de Wiener 31
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

5 Processos de difusão 41
5.1 Definição 41
5.2 Equações de Kolmogorov 43
5.3 Caso multidimensional 48

6 Integrais estocásticos 51
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

7 Equações diferenciais estocásticas 83
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

8 Estudo do modelo de Black-Scholes 99
8.1 Estudo pelo cálculo de Itô 99
8.2 Estudo pelo cálculo de Stratonovich 105

9 A questão dos cálculos de Itô e de Stratonovich 107
9.1 Controvérsia 107
9.2 Resolução da controvérsia 108

10 Estudo de alguns funcionais 111
10.1 Fórmula de Dynkin 111
10.2 Fórmula de Feynman-Kac 114

11 Introdução ao estudo das difusões de Itô unidimensionais 117
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

12 Teorema de Girsanov 137
12.1 Introdução através de um exemplo 137
12.2 Teorema de Girsanov 143

13 Opções e fórmula de Black-Scholes 149
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

14 Síntese **173**

Referências **181**

Índice Remissivo **185**

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 fluído; 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 estocática, 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 homogéneos, 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 Capitulo 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* Ω 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, $\Omega = \{(1,1),(1,2),...,(6,6)\}$ 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-σ* [1] , isto é uma classe não-vazia de subconjuntos de Ω fechada para a complementação (se $A \in \mathcal{F}$, então o complemento $A^c := \Omega - A \in \mathcal{F}$) e para uniões contáveis [2] (se $A_n \in \mathcal{F}$, $n = 1,2,...$, 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-σ (se $A_n \in \mathcal{F}$, $n = 1,2,..$, 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

[1] Há quem prefira dizer "σ-álgebra", imitando a norma inglesa de o adjectivo preceder o substantivo.

[2] 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 Ω 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 comportamente "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 ω que realmente ocorre é um elemento de Ω "escolhido ao acaso" de acordo com a lei de probabilidade P. $\mathcal{F}$ é a álgebra-σ dos subconjuntos de Ω (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-σ 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 Ω 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-σ de Borel $\mathcal{B}$ entende-se a álgebra-σ gerada pelos conjuntos abertos de $\mathbb{R}$, ou seja a menor álgebra-σ que inclui tais conjuntos. Ela é também a álgebra-σ 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 ω e escrevendo simplesmente $Z = 11$). Tomando ω 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. Recorda-se que a função de distribuição de X é definida para $x \in \mathbb{R}$ por $F_X(x) := P[X \in (-\infty, x]] = P[X \leq x]$. [3]

Note-se que a classe $\sigma(X)$ formada pelas imagens inversas por X dos conjuntos de Borel é uma sub-álgebra-σ de $\mathcal{F}$, a que se chama *álgebra-σ 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 $S = \{x_1, x_2, ...\}$ tal que $P[X \in S] = 1$; representaremos por $p_X(k) = P[X = x_k]$ $(k = 1, 2, ...)$ 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. [4] Tem-se $f_X(x) = dF_X(x)/dx$ (se $N \neq \emptyset$, podemos atribuir

[3] 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 $P[Z \in (9.5, 11.2)] = P(\{(4,6),(5,5),(6,4),(5,6),(6,5)\}) = 5/36$.

[4] 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 μ 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). [5] 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 $(k = 1, 2, ...)$, 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* [6] I_A de um conjunto $A \in \mathcal{F}$ é uma

(extensão da medida de comprimento) nula.

[5] Convém lembrar que as v.a. são funções do acaso ω e, portanto, uma propriedade relativa a variáveis aleatórias pode ser verdadeira para certos valores de ω 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 ω para os quais $X(\omega) \neq Y(\omega)$ (claro que o conjunto N de tais valores excepcionais terá probabilidade zero de ocorrer, isto é, $P(N) = 0$).

[6] 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 $\int_\Omega Y dP = \lim_{n\to\infty} \int_\Omega Y_n dP$, 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* $\mathbf{X} = [X_1, X_2, ..., X_n]^T$ ("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 $F_{\mathbf{X}}(\mathbf{x}) = F_{X_1,X_2,...,X_n}(x_1, x_2, ..., x_n) := P[X_1 \leq x_1, X_2 \leq x_2, ..., X_n \leq x_n]$ para $\mathbf{x} = [x_1, x_2, ..., x_n]^T \in \mathbb{R}^n$, também chamada f.d. conjunta das v.a. $X_1, X_2, ..., X_n$. 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 $F_{\mathbf{X}}(\mathbf{x}) = \int_{-\infty}^{x_1}\int_{-\infty}^{x_2}\cdots\int_{-\infty}^{x_n} f_{X_1,X_2,...,X_n}(y_1, y_2, ..., y_n) dy_n ... dy_2 dy_1$, então $\mathbf{X}$ diz-se absolutamente contínuo, caso em que $f_{\mathbf{X}}(\mathbf{x}) = \frac{\partial^n F_{X_1,X_2,...,X_n}(x_1,x_2,...,x_n)}{\partial x_1 \partial x_2 ... \partial x_n}$ (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. $f_{\mathbf{X}}(\mathbf{x}) = (2\pi)^{-n/2}\sqrt{det(\mathbf{C^{-1}})}\exp\left(-\frac{1}{2}(\mathbf{x}-\mu)^T\mathbf{C^{-1}}(\mathbf{x}-\mu)\right)$.

Os conceitos de espaços e normas L^p podem ser generalizados a vectores aleatórios n-dimensionais $\mathbf{X}$ interpretando $|\mathbf{X}|$ como a norma euclideana; 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-σ $\mathcal{H} \subset \mathcal{F}$, 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-σ $\mathcal{H} = \{C, C^c, \emptyset, \Omega\}$, onde $C = \{(2,1),(2,2),...,(2,6),(4,1),(4,2),...,(4,6),(6,1),(6,2),...,(6,6)\}$ 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 ω 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 ω) 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-σ $\mathcal{G}$ 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-σ 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 ω; 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 intervado 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" ω não aparecer explicitamente, não devemos esquecer que o valor do processo estocástico depende dele. Esta função de t e ω não é uma função arbitrária pois está sujeita à restrição de ser, para cada t fixo, uma função mensurável de ω, isto é, uma v.a.

Se fixarmos o "acaso" ω, obtemos uma função apenas do tempo, a que se chama uma *trajectória* [7] do processo estocástico. Um processo estocástico pode, portanto, ser também interpretado como uma colecção de trajectórias, uma para cada estado do acaso ω. 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 ω 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 ω que o "acaso ditou" que fosse o cenário efectivamente ocorrido.

[7] Em inglês usa-se "trajectory" ou "sample path".

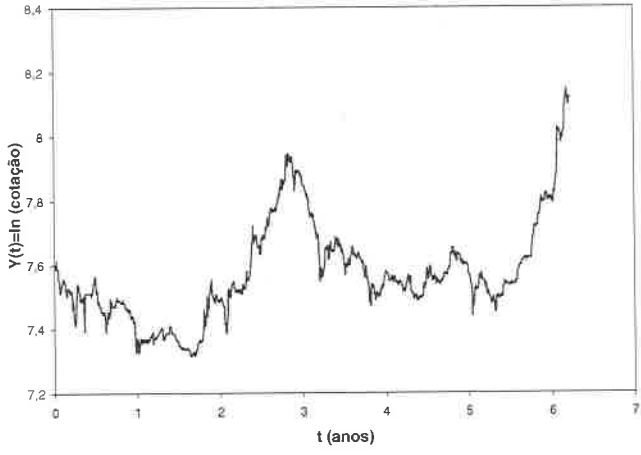


Figura 2.1: Trajectória observada do processo estocástico $Y(t)$, cotação em escala logarítmica das acções do banco BCP entre 8 de Abril de 1991 ($t = 0$) e 30 de Junho de 1997. Corresponde ao ω (cenário do mercado) que efectivamente ocorreu.

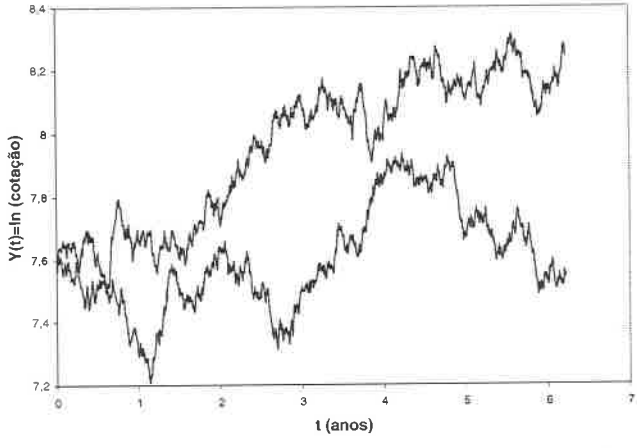


Figura 2.2: Duas trajectórias do processo estocástico "movimento browniano geométrico" com parâmetros $R = r - \sigma^2/2 = 0,084/\text{ano}$ e $\sigma = 0,193/\sqrt{\text{ano}}$ (ver Capítulo 8) simuladas pelo método de Monte Carlo. Este processo foi usado como modelo do processo estocástico $Y(t)$, 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

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

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

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

2. Para $k = 1,...,n-1$, $F_{t_1,t_2,...,t_k,t_{k+1},...,t_n}(x_1,x_2,...,x_k,+\infty,...,+\infty) = F_{t_1,t_2,...,t_k}(x_1,x_2,...,x_k)$ (*propriedade de compatibilidade*).

O *teorema da extensão de Kolmogorov* diz que, dada uma família de funções de distribuição $F_{t_1,t_2,...,t_n}$ (de domínio $\mathbb{R}^n$), definida para todo o $n = 1,2,...$ e todos os $t_1,t_2,...,t_n \in I$ 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. [8] De facto, seria possível construir vários espaços de probabilidade e processos estocásticos neles definidos satisfazendo a propriedade desejada.

[8] 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 $A = \{\omega \in \mathbb{R}^I : (\omega_{t_1},\omega_{t_2},...,\omega_{t_n}) \in B\}$ $(n = 1,2,...)$ cujas bases B sejam intervalos de $\mathbb{R}^n$ (produtos cartesianos de n intervalos reais). A álgebra-σ $\mathcal{B}^I$ gerada por esta classe é conhecida por álgebra-σ 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 $P(B) = \int_B dF_{t_1,t_2,...,t_n}$. 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 difícl 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 ω.

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-σ 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 ω 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 ω 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-σ 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 [9] , 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, α e β 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 ω, é preciso que, como função conjunta das duas variáveis, seja mensurável. Isto é, considerando o espaço mensurável produto [10] $(I \times \Omega, \mathcal{B}_I \times \mathcal{F})$, 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-σ $\mathcal{B}_I \times F$.

Podemos indiferentemente definir um processo estocástico n-dimensional $\{\mathbf{X}_t\}_{t\in I}$ 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.

---

[9] 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

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

e

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

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 $X(t)$.

[10] $\mathcal{B}_I \times \mathcal{F}$ é a álgebra-σ gerada pelos conjuntos da forma $G \times A$ com $G \in \mathcal{B}_I$ (a álgebra-σ 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,

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

para qualquer $n = 1, 2, ...$, quaisquer $x_1, x_2, ..., x_n \in \mathbb{R}$, quaisquer t_1, t_2, $..., t_n \in I$ e qualquer $\tau \in \mathbb{R}$ tal que $t_1 + \tau, t_2 + \tau, ..., t_n + \tau \in I$. Caso $X_t \in L^2$ para todo o $t \in I$ e seja estritamente estacionário, então vem $\mathbb{E}[X_t] = m$ = 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-σ 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-σ 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-$\mathcal{F}_t$. [11] Claro que um processo estocástico $\{X_t\}_{t\in I}$ está sempre adaptado à sua filtração natural ou qualquer outra filtração cujas álgebras-σ contenham, para cada $t \in I$, as correspondentes álgebras-σ $\sigma(X_s; 0 \leq s \leq t)$ 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

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

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 $I = \{0, 1, 2, ..., d\}$ com $0 \leq d \leq +\infty$ (quando $d = +\infty$, interpretamos $I = \mathbb{N}$).

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*:

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

[11] 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 $p > 1$,

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

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 $I = \mathbb{N}_0 = \{0, 1, 2, ...\}$). 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 Ω 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-σ 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 $F_T(+\infty) := \lim_{t\to+\infty} F_T(t) < 1$. 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-σ $\mathcal{F}_T$ que desempenha para T um papel análogo ao papel que $\mathcal{F}_t$ desempenha para t. Para isso considera-se a menor álgebra-σ $\mathcal{F}_{+\infty}$ que contém todas as álgebra-σ $\mathcal{F}_t$ para $t \geq 0$ e define-se $\mathcal{F}_T$ como a álgebra-σ 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 $a \in \mathbb{R}$

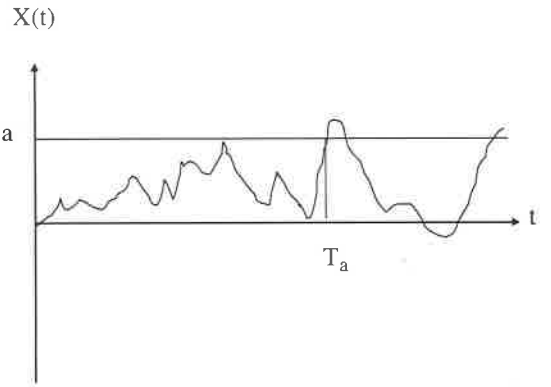


Figura 2.3: Tempo de primeira passagem $T_a = T_a(\omega)$ para uma trajectória ω de um processo estocástico $X(t)$ pelo limiar a.

(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 trajectória ω 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 trajectória 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 $0 \leq d \leq +\infty$ [12] , 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* [13] se, para quaisquer $s, t \in I$, $s \leq t$ e qualquer conjunto de Borel $B \in \mathcal{B}$, tivermos [14]

$$P[X_t \in B | X_u,\ 0 \leq u \leq s] = P[X_t \in B | X_s]. \tag{2.4}$$

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-σ gerada pelas v.a. X_u, $0 \leq u \leq s$.

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

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

para qualquer $n = 1, 2, ...$, quaisquer $t_1 \leq ... \leq t_{n-1} \leq t_n \leq t$, quaisquer $x_1, ..., x_{n-1}, x_n \in \mathbb{R}$ 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:

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

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 $s, t \in I$, $s \leq t$,

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

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

[12] O intervalo pode começar noutro instante t_0 mas, por simplicidade, suporemos que começa em 0.

[13] 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 $I = \{0, 1, 2, ..., d\}$ com $0 \leq d \leq +\infty$.

[14] 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 $0 < t_1 < t_2 < ... < t_{n-1} < t_n$,

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

e

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

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*

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

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

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

Um *processo de Markov homogéneo* (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 homogéneo 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 homogéneo 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-σ 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 homogéneo $\{X_t\}_{t\geq 0}$ 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

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

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

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

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*

$$\frac{dX}{dt} = rX, \tag{3.1}$$

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)

$$X(t) = x_0 \exp(rt). \tag{3.2}$$

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 $\Delta t \to 0$ do modelo em tempo discreto (equação às diferenças)

$X(t+\Delta t) - X(t) = (r\Delta t)X(t)$, onde $r\Delta t$ é aproximadamente, para Δt pequeno, a taxa de crescimento/rendimento/juro no intervalo $(t, t+\Delta t)$.

Contudo, o ambiente tem flutuações aleatórias que afectam a taxa de crescimento da população. Os mercados também flutuam e essas flutuações certamente afectam as taxas de crescimento dos preços dos bens no consumidor e as taxas de rendimento dos produtos financeiros. Assim, se $X(t)$ é, por exemplo, a cotação de uma acção, não esperamos que a sua taxa de rendimento seja constante, antes que flutue aleatoriamente. Seja $B(t)$ (ou B_t como notação alternativa) o efeito acumulado entre os instantes 0 e t das flutuações do ambiente/mercado na taxa de crescimento/rendimento.

Obviamente, $B(t)$ depende do acaso ω (estado da natureza/cenário do mercado), onde ω varia em Ω (conjunto dos estados/cenários possíveis) e temos uma estrutura de espaço de probabilidade $(\Omega, \mathcal{F}, P)$ que nos dá a probabilidade de ocorrência dos diversos conjuntos (mensuráveis) de estados/cenários. Devíamos escrever $B(t,\omega)$ mas adoptamos a convenção usual, que permite simplificar a notação, de não explicitar a referência à dependência de ω. Trata-se, pois, de um processo estocástico. Claro que $B(0) = 0$ e que a taxa de crescimento/rendimento num pequeno intervalo de tempo $(t, t+\Delta t)$ é aproximadamente $r\Delta t + \Delta B(t)$, onde $\Delta B(t) = B(t+\Delta t) - B(t)$ é o *incremento* do processo $B(t)$ no intervalo $(t, t+\Delta t)$. Obtemos, assim, como modelo aproximado, a *equação às diferenças estocástica* $X(t+\Delta t) - X(t) = (r\Delta t + \Delta B(t))X(t)$. No limite quando $\Delta t \to 0$, obtemos $dX = (rdt + dB(t))X$ ou

$$dX = rXdt + XdB(t). \tag{3.3}$$

Como $B(t)$ é um processo estocástico, esta é uma *equação diferencial estocástica*, conhecida como o *modelo de Black-Scholes* para a cotação de uma acção. A sua solução $X(t)$ também depende do acaso ω (estado do ambiente/cenário do mercado) e é, portanto, um processo estocástico $X(t,\omega)$. O termo $rXdt$ descreve a tendência média do comportamento dinâmico desta variável [r é assim uma taxa (instantânea) "média" de crescimento/rendimento] e o termo $XdB(t)$ descreve as perturbações aleatórias em volta desta tendência.

Quais as características que devemos exigir do processo estocástico $B(t)$ para que se obtenha um modelo razoável do seu comportamento? Supondo que há muitas fontes de flutuações aleatórias do ambiente/mercado afectando as taxas de crescimento/rendimento, esperamos, pelo teorema do limite central, que o seu efeito acumulado $B(t)$ sobre tais taxas tenha aproximadamente uma distribuição gaussiana. Podemos então supor que o incremento $\Delta B(t)$ no intervalo $(t, t+\Delta t)$ é uma v.a. gaussiana com média zero (se a média não fosse nula, podíamos absorvê-

la no parâmetro r de modo a que a interpretação deste parâmetro como taxa "média" faça sentido).

Por outro lado, dados dois intervalos não sobrepostos, é razoável supor que as fontes das variações aleatórias num dos intervalos (que afectam o incremento de B nesse intervalo) são aproximadamente independentes das fontes das variações aleatórias no outro intervalo. Assim, podemos supor com razoável aproximação que os incrementos de B em intervalos não-sobrepostos são variáveis aleatórias independentes.

Quanto à variância do incremento $\Delta B(t)$, ela deve ser a soma das variâncias das muitas fontes de flutuações aleatórias que ocorrem no intervalo de tempo $(t, t + \Delta t)$ (supondo que há independência dessas fontes, pelo menos com razoável aproximação). Como o número destas fontes deve ser aproximadamente proporcional à duração Δt do intervalo, o mesmo deve suceder à variância do incremento. Assim, podemos supor que a variância de $\Delta B(t)$ é $\sigma^2 \Delta t$, onde σ é um parâmetro que mede a intensidade das flutuações aleatórias e que é conhecido na literatura financeira por *volatilidade*. Se pusermos $W(t) = B(t)/\sigma$, então $\Delta W(t) \backsim \mathcal{N}(0, \Delta t)$ (distribuição normal com média zero e variância Δt). Os incrementos são estacionários (não dependem do intervalo de tempo específico mas apenas da sua duração Δt).

Um processo estocástico com as propriedades acima mencionadas para $W(t)$ é chamado *processo de Wiener padrão* (ou standard) e será estudado no Capítulo 4. Também é conhecido por *movimento browniano* porque foi o modelo usado por Einstein no seu ano de ouro de 1905 para descrever o movimento browniano de uma partícula suspensa num fluído. Mais exactamente, se o movimento ocorre numa superfície plâna, ele é descrito por $(x_0 + \sigma W_1(t), y_0 + \sigma W_2(t))$, onde (x_0, y_0) é a posição inicial da partícula, $W_1(t)$ e $W_2(t)$ são dois processos de Wiener padrão independentes entre si e σ é um coeficiente (a σ^2 chama-se coeficiente de difusão). A generalização a movimentos tridimensionais é óbvia. Claro que em 1905 ainda o processo $W(t)$ não era conhecido. O movimento browniano foi inicialmente observado pelo botânico Brown em 1827 quando constatou ao microscópio o movimento extremamente irregular de uma partícula de pólen suspensa numa superfície líquida. Einstein atribui o fenómeno aos choques da partícula com as moléculas do fluído que estão em constante movimento. Olhando para a projecção da posição da partícula num dos eixos coordenados, obtemos (à parte um parâmetro de localização inicial da partícula) um processo de Wiener. Isso pode ser justificado pelo facto de, em cada pequeno intervalo de tempo Δt, a partícula sofrer inúmeros choques, cujo efeito acumulado sobre a posição da partícula é a soma de muitas v.a. independentes, esperando-se uma distribuição aproximadamente normal; além disso, a independência dos incrementos e a variância proporcional ao intervalo de tempo explicam-

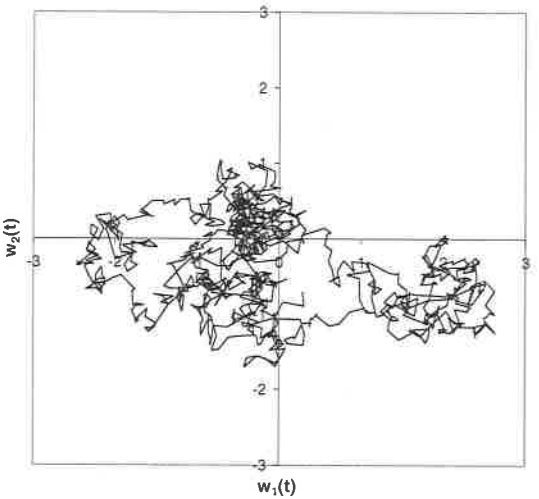


Figura 3.1: Simulação, usando o modelo de Einstein com $x_0 = y_0 = 0$ e $\sigma = 1$, isto é $(W_1(t,\omega), W_2(t,\omega))$, do movimento browniano para t a variar entre 0 e 10 e ω fixo escolhido aleatoriamente. O tempo foi discretizado em passos de 0.01 unidades e os pontos $(W_1(0.01k), W_2(0.01k))$ $(k = 0, 1, 2, ..., 1000)$ unidos por segmentos de recta.

se de forma semelhante à acima descrita para crescimento de populações ou cotação de acções. O coeficiente de difusão depende naturalmente da temperatura (quanto maior ela for, maior será a agitação das moléculas do fluído e o número de choques por unidade de tempo com a partícula suspensa), da massa da partícula, da viscosidade do fluído e também do número de Avogadro. Einstein, através de princípios físicos, determinou a expressão dessa dependência funcional. Para mais pormenores pode ver-se [28]. Numa experiência em que sejam controlados e conhecidos os valores da temperatura, massa da partícula e viscosidade, estimando σ^2 (o que pode ser feito com grande precisão) através da variância amostral dos incrementos da posição da partícula em pequenos intervalos Δt, facilmente se pode determinar o número de Avogadro com grande aproximação. Pessoalmente, não gosto de chamar a $W(t)$ movimento browniano porque, embora o modelo de Einstein (que coincide em cada eixo, à parte constantes de localização e escala, com $W(t)$) descreva razoavelmente este movimento, foram posteriormente desenvolvidos outros modelos mais realistas.

Embora fosse habitual atribuir a "paternidade" do processo de Wiener a Einstein, o desenvolvimento das aplicações financeiras das equações

diferenciais estocásticas chamou a atenção para o trabalho de Bachelier em 1900 ([4]), que usou precisamente o modelo $x_0 + \sigma W(t)$ para descrever a cotação de acções na bolsa de Paris. A não ser talvez a curto prazo, trata-se de um modelo pouco realista, que prevê um valor esperado constante para a cotação (pois a média de $W(t)$ é nula).

Podemos preferir escrever (3.3) na forma

$$dX = rX\,dt + \sigma X\,dW_t \qquad \text{ou} \qquad \frac{dX}{dt} = rX + \sigma X\varepsilon(t), \tag{3.4}$$

com $\varepsilon(t) = \varepsilon(t,\omega) = dW(t)/dt$. Sucede, porém, que as trajectórias de $W(t)$, embora q.c. contínuas, são extremamente irregulares (serviram precisamente para descrever o movimento browniano de partículas) e, como veremos no Capítulo 4, são q.c. não-diferenciáveis. Ou seja, a derivada $dW(t)/dt$ não existe no sentido corrente, mas apenas no sentido das funções generalizadas. Assim, $\varepsilon(t)$, chamado *ruído branco padrão*, não é um processo estocástico mas um processo estocástico generalizado (as suas trajectórias são funções generalizadas do tempo). Como $W(t)$ tem incrementos independentes estacionários, $\varepsilon(t)$ é um processo estacionário cujos valores em dois instantes diferentes (por mais próximos que estejam um do outro) são independentes. Para processos em tempo discreto, a independência em instantes diferentes não causa qualquer dificuldade porque esses instantes não podem estar arbitrariamente próximos e é bem conhecido do estudo das séries temporais (ou sucessões cronológicas) o ruído branco gaussiano em tempo discreto, que é simplesmente uma sucessão de v.a. gaussianas independentes identicamente distribuídas. Porém, em tempo contínuo a independência em instantes diferentes exige o uso de processos estocásticos generalizados. Assim, $\varepsilon(t)$ é um processo estocástico generalizado gaussiano com função de auto-covariância $C(t) = 0$ para todo o $t \neq 0$. Curiosamente, $\Delta W(t)/\Delta t$ tem variância $1/\Delta t \to +\infty$ quando $\Delta t \to 0$ e, portanto, $C(0) = VAR[\varepsilon(t)] = +\infty$. Logo, $C(t) = \delta(t)$ é a *função delta de Dirac*, que é uma função generalizada caracterizada por $\delta(x) = 0$ para $x \neq 0$, $\delta(x) = +\infty$ para $x = 0$ e $\int_{-\infty}^{+\infty} \delta(x)dx = 1$. A sua transformada de Fourier é uma função constante e, portanto, $\varepsilon(t)$ tem função densidade espectral constante (todas as frequências têm igual intensidade, um comportamento que faz lembrar a luz branca). É de esperar que os ruídos naturais $\tilde{\varepsilon}(t)$ sejam coloridos, isto é, tenham alguma auto-correlação para instantes próximos e, portanto, tenham uma densidade espectral não constante e uma função de auto-covariância $C(t)$ que seja uma função propriamente dita com um pico na origem. Se este pico for aguçado, ela poderá ser razoavelmente aproximada por uma função delta de Dirac e o *ruído colorido* correspondente

ser aproximado por um ruído branco. Ao contrário do ruído branco, o integral de um tal ruído colorido não será um processo de Wiener mas um processo $\tilde{W}(t)$ com incrementos "ligeiramente" dependentes e com trajectórias mais suaves. Contudo, o processo de Wiener ou o ruído branco são matematicamente muito melhor tratáveis e constituem frequentemente uma boa aproximação, razões pelas quais os iremos utilizar como modelos.

A equação (3.4) pode ser escrita na forma

$$\frac{dX}{dt} = (r + \sigma X \varepsilon(t))X, \tag{3.5}$$

que tem a interessante interpretação de ser o modelo (3.1) com a taxa (instantânea) de crescimento/rendimento r perturbada por ruído branco induzido pelo ambiente/mercado. Usualmente, os livros de texto evitam referir-se ao ruído branco e preferem usar a notação $dX = rXdt + \sigma X dW$.

Capítulo 4

O processo de Wiener

4.1 Definição

O processo de Wiener desempenha um papel essencial nas equações diferenciais estocásticas que iremos estudar. Ele traduz o efeito acumulado das perturbações aleatórias que afectam a dinâmica do fenómeno em estudo, ou seja, é o integral do ruído perturbador que se supõe ser um ruído branco em tempo contínuo. Como referimos no Capítulo 3, Bachelier usou o processo de Wiener em 1900 para modelar a cotação de uma acção na bolsa e Einstein usou-o em 1905 para modelar o movimento browniano de uma partícula suspensa num fluído. Só, porém, a partir de 1920 é que ele foi rigorosamente estudado por Wiener e por Lévy. Vamos examinar aqui as suas principais propriedades. Antes de mais, notaremos o processo de Wiener padrão indiferentemente por W_t (abreviatura de $W_t(\omega)$) ou $W(t)$ (abreviatura de $W(t,\omega)$). Comecemos pela sua definição formal.

Definição. *Dado um espaço de probabilidade $(\Omega, \mathcal{F}, P)$, um processo estocástico $\{W_t\}_{t\in[0,+\infty)}$ definido nesse espaço diz-se um processo de Wiener padrão (ou movimento browniano) se satisfizer as seguintes propriedades:*

- $W(0) = 0$ *q.c.*
- *Os incrementos $W(t) - W(s)$ $(s < t)$ têm distribuição normal com média 0 e variância $t - s$.*
- *Os incrementos $W(t_i) - W(s_i)$ $(i = 1, ..., n)$ em intervalos de tempo*

$(s_i, t_i]$ $(i = 1, 2, ..., n)$ não-sobrepostos são v.a. independentes [1] *(diz-se simplesmente que tem incrementos independentes).*

Observação. *Um processo da forma $a + \sigma W(t)$, com a e σ constantes, diz-se um processo de Wiener. Por vezes, porém, para abreviar a linguagem, diremos simplesmente "processo de Wiener" querendo-nos referir ao processo de Wiener padrão.*

Observação. *Como a distribuição dos incrementos do processo de Wiener num intervalo de tempo só depende do comprimento do intervalo, concluímos que tem incrementos estacionários. Mas o próprio processo de Wiener não é estacionário; basta reparar que a variância $VAR[W(t)] = VAR[W(t) - W(0)] = t - 0 = t$ não é constante. Também os incrementos do processo de Wiener são independentes, mas os valores do próprio processo em instantes distintos não são independentes (a Propriedade 2 abaixo mostra-o claramente).*

4.2 Principais propriedades

Estudemos agora as principais propriedades do processo de Wiener $W(t)$.

Propriedades:

1. *O processo de Wiener tem uma versão separável contínua (isto é, com trajactórias q.c. contínuas). Suporemos, daqui por diante, que trabalhamos sempre com uma tal versão. Por isso, alguns autores incluem logo a continuidade das trajectórias na definição.*

   Este resultado decorre da aplicação do critério de Kolmogorov tendo em conta que, pelas propriedades da distribuição normal, se tem $\mathbb{E}[|W_t - W_s|^4] = 3|t - s|^2$.

[1] A independência é um importante conceito probabilístico. Num espaço de probabilidade $(\Omega, \mathcal{F}, P)$, os acontecimentos $A_1, A_2, ..., A_n \in \mathcal{F}$ dizem-se *independentes* se $P(A_1 \cap A_2 \cap ... \cap A_n) = P(A_1)P(A_2)...P(A_n)$. Se A e B são independentes, então $P(A|B) = P(A)$. Dadas álgebras-σ $\mathcal{F}_1, \mathcal{F}_2, ..., \mathcal{F}_n$ contidas em $\mathcal{F}$, elas dizem-se *independentes* se os acontecimentos $A_1, A_2, ..., A_n$ forem independentes para todas as escolhas possíveis de n e de $A_1 \in \mathcal{F}_1$, $A_2 \in \mathcal{F}_2$, ..., $A_n \in \mathcal{F}_n$. As v.a. $X_1, X_2, ..., X_n$ dizem-se *independentes* se as álgebras-σ $\sigma(X_1)$, $\sigma(X_2)$, ..., $\sigma(X_n)$ por elas geradas forem independentes; isto equivale à propriedade de a f.d. conjunta das v.a. ser igual ao produto das f.d. individuais. Se $X_1, X_2, ..., X_n$ forem v.a. independentes e forem integráveis (isto é, com esperança matemática), vem $\mathbb{E}[X_1 X_2 ... X_n] = \mathbb{E}[X_1]\mathbb{E}[X_2]...\mathbb{E}[X_n]$ e $\mathbb{E}[X_i|X_j] = \mathbb{E}[X_i]$ para $i \neq j$. Também se pode definir independência para uma colecção infinita de acontecimentos, álgebras-σ ou variáveis aleatórias; considera-se que há independência quando toda a subcolecção finita for independente.

2. *$W(t)$ tem distribuição normal com média 0 e variância t, o que representamos por $W_t \frown \mathcal{N}(0,t)$.*

   Basta notar que $W(t)$ é igual ao incremento $W(t) - W(0)$.

3. *A covariância entre $W(s)$ e $W(t)$ é $COV[W(s), W(t)] = \mathbb{E}[W(s)W(t)] = \min(s,t)$.*

   O caso $s = t$ decorre da propriedade 2. Para $s < t$, vem $\mathbb{E}[W(s)W(t)] = \mathbb{E}[W^2(s)] + \mathbb{E}[(W(t) - W(s))(W(s) - W(0))]$ e basta ver que $\mathbb{E}[W^2(s)] = s$ e que, como os intervalos $(0,s]$ e $(s,t]$ são não-sobrepostos, a independência dos incrementos dá $\mathbb{E}[(W(t) - W(s))(W(s) - W(0))] = \mathbb{E}[W(t) - W(s)]\mathbb{E}[W(s) - W(0)] = 0 \times 0 = 0$. O caso $s > t$ é semelhante.

4. *$W(t)$ é um processo gaussiano e, para $0 < t_1 < ... < t_n$, a f.d.p. conjunta $f_{t_1,...,t_n}(x_1, ..., x_n)$ de $W(t_1), ..., W(t_n)$ é dada por*

$$f_{t_1,...,t_n}(x_1, ..., x_n) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi(t_i - t_{i-1})}} \exp\left(-\frac{(x_i - x_{i-1})^2}{2(t_i - t_{i-1})}\right), \tag{4.1}$$

   onde se pôs $t_0 = 0$ e $x_0 = 0$.

   Os incrementos $W(t_1) - W(0)$, $W(t_2) - W(t_1)$, ..., $W(t_n) - W(t_{n-1})$ são independentes, pelo que a f.d.p. conjunta é o produto das densidades gaussianas de cada um (com média 0 e variância igual ao comprimento do correspondente intervalo de tempo). Essa densidade conjunta dá o produto do lado direito da expressão supra. Basta agora notar que a transformação que passa dos incrementos referidos para as variáveis $W(t_1)$, $W(t_2)$, ..., $W(t_n)$ tem jacobiano igual a um.

5. *$W(t)$ é um processo de Markov homogéneo com densidades de transição*

$$p(\tau, y|x) = (2\pi\tau)^{-1/2} \exp\left(-\frac{(y-x)^2}{2\tau}\right) \qquad (\tau > 0). \tag{4.2}$$

   Isto é, a distribuição condicional de $W(s+\tau)$ dado que $W(s) = x$ é normal de média x e variância τ:

$$W(s+\tau)|(W(s) = x) \frown \mathcal{N}(x, \tau). \tag{4.3}$$

Devido à independência dos incrementos, vê-se que a média condicional é $\mathbb{E}[W_{s+\tau}|W_s = x] = \mathbb{E}[W_{s+\tau} - W_s|W_s - W_0 = x] + \mathbb{E}[W_s|W_s = x] = \mathbb{E}[W_{s+\tau} - W_s] + x = 0 + x = x$ e a variância

condicional é $VAR[W_{s+\tau}|W_s = x] = \mathbb{E}[(W_{s+\tau} - x)^2|W_s = x] = \mathbb{E}[(W_{s+\tau} - W_s)^2|W_s - W_0 = x] = \mathbb{E}[(W_{s+\tau} - W_s)^2] = (s+\tau-s) = \tau$. O facto desta distribuição só depender da duração τ do intervalo de tempo garante a homogeneidade. Falta verificar se o processo é de Markov, para o que basta mostrar a propriedade (2.6). Ora, usando a independência dos incrementos, vem

$$
\begin{aligned}
&F_{W_t|W_{t_1}=x_1,\ldots,W_{t_{n-1}}=x_{n-1},W_{t_n}=x_n}(x) = \\
&P[W_t \leq x|W_{t_1} = x_1, W_{t_2} = x_2, \ldots, W_{t_{n-1}} = x_{n-1}, W_{t_n} = x_n] = \\
&P[W_t - W_{t_n} \leq x - x_n|W_{t_1} - W_0 = x_1, W_{t_2} - W_{t_1} = x_2 - x_1, \ldots, \\
&\quad W_{t_{n-1}} - W_{t_{n-2}} = x_{n-1} - x_{n-2}, W_{t_n} - W_{t_{n-1}} = x_n - x_{n-1}] = \\
&P[W_t - W_{t_n} \leq x - x_n] = P[W_t - W_{t_n} \leq x - x_n|W_{t_n} - W_0 = x_n] = \\
&P[W_t \leq x|W_{t_n} - W_0 = x_n] = F_{W_t|W_{t_n}=x_n}(x).
\end{aligned}
$$

6. *$W(t)$ é uma martingala, subentende-se que com respeito à filtração natural $\{\mathcal{M}_s\}_{s\geq 0}$, com $\mathcal{M}_s = \sigma(W(u) : 0 \leq u \leq s)$. Isto é, para $s \leq t$, vem $\mathbb{E}[W(t)|\mathcal{M}_s] := \mathbb{E}[W(t)|W(u), 0 \leq u \leq s] = W(s)$.*

   De (2.4), vem $P[W(t) \in B|\mathcal{M}_s] = P[W(t) \in B|W(s)]$ para qualquer conjunto de Borel B. Logo $\mathbb{E}[W(t)|W(u), 0 \leq u \leq s] = \mathbb{E}[W(t)|W(s)]$. Basta agora atender a que, devido a (4.3), se tem, para qualquer x, $\mathbb{E}[W(t)|W(s) = x] = x$.

Uma questão importante é a da existência de processos de Wiener, isto é, de processos que satisfaçam as propriedades da Definição da Secção 4.1, pois caso contrário estaríamos a estudar um objecto de conteúdo vazio. A Propriedade 4 anterior define as funções de distribuição de dimensão finita (que se obtêm por integração das f.d.p. obtidas nessa propriedade), as quais são efectivamente funções de distribuição (gaussianas) e facilmente se verifica satisfazerem as condições de simetria e compatibilidade. O teorema da extensão de Kolmogorov garante a existência de pelo menos um tal processo num certo espaço de probabilidade e dá-nos um método construtivo de o obter.

Uma outra forma de construir um processo de Wiener é considerar um *passeio aleatório*, isto é um processo de Markov em tempo discreto ($t = 0, \Delta t, 2\Delta t, \ldots$) com espaço de estados discreto $S = \{\ldots, -2\Delta x, -\Delta x, 0, \Delta x, 2\Delta x, \ldots\}$ (com $\Delta t > 0$ e $\Delta x > 0$) tal que, em cada passo, o processo se mova para um dos dois estados vizinhos (à direita ou à esquerda) com igual probabilidade. Podemos pensar no movimento de uma partícula sujeita a colisões. Pondo $\Delta x = (\Delta t)^{1/2}$ e fazendo $\Delta t \to 0$, obtemos o processo de Wiener padrão. Para obter $\sigma W(t)$, basta usar $\Delta x = \sigma(\Delta t)^{1/2}$.

Exercício: Demonstre que:

a) Para $s \geq 0$ fixo, o processo $Y(t) = W(s+t) - W(s)$ $(t \geq 0)$ é também um processo de Wiener padrão.

b) Para $c > 0$ constante, o processo $Z(t) = \frac{1}{c}W(c^2 t)$ $(t \geq 0)$ é também um processo de Wiener padrão.

c) O processo $H(t) = W^2(t) - t$ é uma martingala-$\mathcal{M}_t$.

Pode provar-se (ver, por exemplo, [37]) que, dado um processo estocástico $X(t)$ $(t \geq 0)$ adaptado a uma filtração $\mathcal{F}_t \subset \mathcal{F}$, que tenha trajectórias q.c. contínuas, que verifique $X(0) = 0$ q.c., que seja martingala-$\mathcal{F}_t$ e tal que $X^2(t) - t$ também seja martingala-$\mathcal{F}_t$, então $X(t)$ é um processo de Wiener. Esta propriedade é conhecida por *caracterização de Lévy.*

4.3 Algumas propriedades analíticas

Apesar de as trajectórias do processo de Wiener serem q.c. contínuas, elas são q.c. não-diferenciáveis, como já se referiu no Capítulo 3. Elas são até q.c. de *variação ilimitada* em qualquer intervalo finito $[a, b]$.

Para o reconhecer, decomponhamos o intervalo em 2^n subintervalos de igual duração $\delta_n = (b-a)/2^n$ usando os pontos de decomposição $t_{n,k} = a + k\delta_n$ $(k = 0, 1, ..., 2^n)$. Note-se que, quando $n \to +\infty$, o diâmetro da decomposição $\delta_n \to 0$ e $\sum_{k=1}^{2^n} \delta_n = b - a < +\infty$. Seja $V_n = \sum_{k=1}^{2^n} \left(W(t_{n,k}) - W(t_{n,k-1})\right)^2$.

Exercício: Com $V_{n,k} = \left(W(t_{n,k}) - W(t_{n,k-1})\right)^2 - \delta_n$, mostre que

$$\mathbb{E}[V_n] = \sum_k \delta_n = (b-a)$$

$$VAR[V_n] = \mathbb{E}\left[\left(\sum_k V_{n,k}\right)^2\right] = 2\sum_k \left(t_{n,k} - t_{n,k-1}\right)^2 \leq 2(b-a)\delta_n \to 0.$$

Note que $\left(\sum_k V_{n,k}\right)^2 = \sum_k V_{n,k}^2 + 2\sum_k \sum_{j>k} V_{n,k} V_{n,j}$ e que devido à independência dos incrementos do processo de Wiener, vem, para $k \neq j$, $\mathbb{E}[V_{n,k} V_{n,j}] = \mathbb{E}[V_{n,k}]\mathbb{E}[V_{n,j}]$.

A propriedade referida no exercício anterior mostra que V_n *converge em média quadrática* (isto é, com respeito à norma-L^2) para $b-a$. Como $\sum_n VAR[V_n] < +\infty$, isto implica (pode consultar qualquer bom livro de probabilidades) que V_n *converge quase certamente* (q.c., isto é, com probabilidade um) para $b-a$, ou seja, $P[V_n \to b-a \text{ quando } n \to +\infty] = 1$. Como

$$V_n \leq \max_k |W(t_{n,k}) - W(t_{n,k-1})| \sum_k |W(t_{n,k}) - W(t_{n,k-1})|$$

e como $\max_k |W(t_{n,k}) - W(t_{n,k-1})| \to 0$ q.c. (porque $W(t)$ é q.c. contínuo em $[a,b]$ e, portanto, q.c. uniformemente contínuo nesse intervalo fechado), resulta que $\sum_k |W(t_{n,k}) - W(t_{n,k-1})| \to +\infty$ q.c., assim se concluindo pela variação ilimitada q.c. das trajectórias do processo de Wiener.

Para dar a ideia de como se comportam as trajectórias do processo de Wiener, a Figura 4.1 mostra uma simulação de duas trajectórias de $W(t)$ no intervalo de tempo $[0,10]$. [2] Se amplifiássemos com uma lente um pequeno pedaço de uma trajectória, o seu aspecto pareceria tão irregular como o que observa na figura para toda a trajectória. A trajectória não tem regiões suaves, é toda formada por "bicos". Como a trajectória é q.c. de variação ilimitada, o comprimento de qualquer pedaço da trajectória correspondente a um intervalo de tempo $[a,b]$ será infinito, por menor que seja $[a,b]$.

Embora em qualquer instante t a trajectória tenha valor médio 0, sabemos que a variância está a aumentar com o tempo e é interessante ver qual é o comportamento assintótico de $W(t)$ quando $t \to +\infty$. Verifica-se que $W(t)/t \to 0$ q.c. quando $t \to +\infty$. Isso é uma consequência da *lei forte dos grandes números.* [3] Uma ideia mais precisa do comportamento assintótico é dada pela *lei do logaritmo iterado*

$$\limsup_{t\to+\infty} \frac{W(t)}{\sqrt{2t \ln\ln t}} = 1 \quad \text{q.c.} \tag{4.4}$$

$$\liminf_{t\to+\infty} \frac{W(t)}{\sqrt{2t \ln\ln t}} = -1 \quad \text{q.c.} \tag{4.5}$$

A demonstração é complicada e pode ver-se, por exemplo, em [31]. Da lei do logaritmo iterado decorre que, dado $\varepsilon > 0$ arbitrário, se tem q.c.

$$-(1+\varepsilon)\sqrt{2t \ln\ln t} \leq W(t,\omega) \leq (1+\varepsilon)\sqrt{2t \ln\ln t}$$

[2] De facto, como não podemos simular para todos os valores de t, simulámos os valores de $W(t)$ para $t = 0, 0.01, 0.02, ..., 10$ e unimos os pontos de coordenadas $(t, W(t))$ por segmentos de recta. Em vez de simular directamente os valores dos vários $W(0.01i)$ $(i = 1, 2, ..., 1000)$ usando uma distribuição normal multivariada (cuja estrutura de correlações é complicada), é preferível gerar os incrementos $\Delta W_{(k)} = W(0.01k) - W(0.01(k-1))$ $(k = 1, 2, ..., 1000)$, os quais são independentes entre si e têm distribuição $\mathcal{N}(0, 0.01)$, e adicioná-los sucessivamente para obter os $W(0.01i) = \sum_{k=1}^{i} \Delta W_{(k)}$ $(i = 1, 2, ..., 1000)$.

[3] No caso de $t = n$ $(n = 1, 2, ...)$, a demonstração é simples. Basta notar que as v.a. $W(n) - W(n-1) \frown \mathcal{N}(0,1)$ $(n = 1, 2, ...)$ são independentes entre si e aplicar a lei forte dos grandes números para reconhecer que a sua média $W(n)/n \to 0$ q.c. O caso geral, com valores de $t \to +\infty$ não necessariamente inteiros, pode demonstrar-se recorrendo à lei do logaritmo iterado de que falaremos em seguida e da qual é uma consequência trivial.

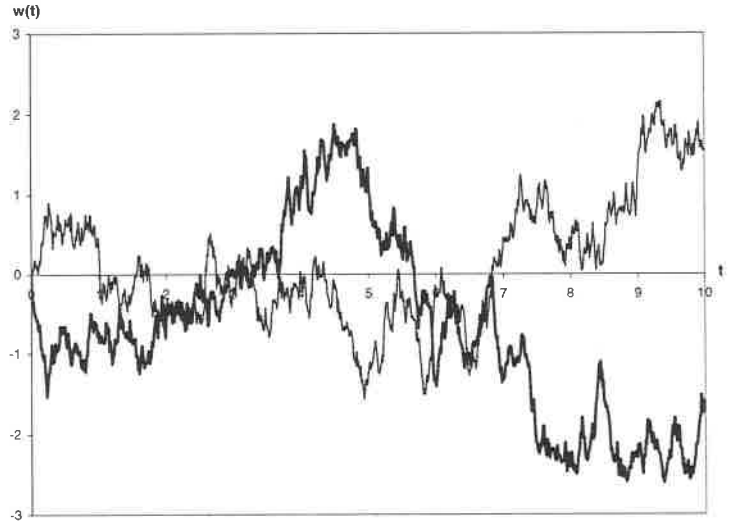


Figura 4.1: Duas trajectórias simuladas (dois valores diferentes de ω) do processo de Wiener padrão no intervalo $[0, 10]$.

para t suficientemente grande (isto é, existe um $t_0(\omega)$ tal que a propriedade vale para $t \geq t_0(\omega)$). Isto dá-nos curvas majorantes e minorantes das trajectórias para t grande.

Exercício: Mostre que o processo

$$X(t) := \begin{cases} tW(1/t) & \text{se } t > 0 \\ 0 & \text{se } t = 0 \end{cases}$$

é um processo de Wiener padrão.

Veremos no Capítulo 5 que um processo de Wiener é um caso particular de um processo de difusão.

4.4 Tempos de primeira passagem

O processo de Wiener, que vimos ser um processo de Matkov homogéneo, é mesmo um *processo de Markov forte*. Assim, dado um tempo de Markov S, como $W(0) = 0$, podemos obter o processo $Y(u) := W(S+u) - W(S)$ $(u \geq 0)$ e, utilizando (2.12), reconhecer que ele tem a mesma distribuição que $W(u)$ e tem a mesma distribuição dos incremen-

tos que o processo de Wiener original. Como é um processo gaussiano, os incrementos, sendo não-correlacionados, são indepedentes. Conclui-se que $Y(u)$ é um processo de Wiener independente de $\mathcal{F}_S$, propriedade que já tinhamos visto na Secção 4.2 mas para s determinístico em vez de aleatório. A demonstração destas propriedades pode ver-se, por exemplo, em [34] ou em [37]. Daqui decorre que

$$\mathbb{E}[h(W(S+u))|\mathcal{F}_S] = \mathbb{E}[h(W(u))] \tag{4.6}$$

para funções h mensuráveis-Borel e limitadas e $u \geq 0$. O facto de $Y(u)$ ser processo de Wiener independente de $\mathcal{F}_S$ pode traduzir-se, usando terminologia de Itô e McKean, na ideia de que o processo de Wiener "recomeça de fresco" nos tempos de Markov isto é, mudando o sistema de eixos tempo/estado do gráfico das trajectórias do processo de modo a que os novos eixos sejam paralelos aos anteriores mas tenham origem no ponto com as antigas coordenadas $(S, W(T_S))$, obtemos as trajectórias de $Y(u)$.

Consideremos o *tempo de primeira passagem* $T_a = \inf\{t \geq 0 : W(t) = a\}$ de um processo de Wiener por um limiar constante $a \neq 0$ (o caso $a = 0$ é trivial). Trata-se de um tempo de Markov. Assim, para saber se o acontecimento $[T \leq t]$ ocorreu ou não, precisamos de conhecer apenas a trajectória do processo de Wiener até ao instante t. Vamos determinar a distribuição de T_a, supondo $a > 0$, usando um argumento de simetria e a propriedade de Markov forte.

Como as trajectórias são contínuas e $W(0) = 0 < a$, se $T_a > t$ a trajectória no instante t tem de estar abaixo de a (caso contrário, o processo teria passado por a antes do instante t e não viria $T_a > t$), pelo que $P[W(t) \geq a | T_a > t] = 0$. Logo

$$\begin{aligned} &P[W(t) \geq a] = \\ &P[W(t) \geq a | T_a \leq t] P[T_a \leq t] + P[W(t) \geq a | T_a > t] P[T_a > t] = \\ &P[W(t) \geq a | T_a \leq t] P[T_a \leq t]. \end{aligned}$$

Sabemos que $Y(u) = W(T_a + u) - W(T_a) = W(T_a + u) - a$ é um processo de Wiener. Logo $P[W(t) \geq a | T_a \leq t] = P[Y(t - T_a) \geq 0]$; esta última probabilidade só pode ser igual a $P[Y(t - T_a) \leq 0]$ devido à simetria da distribuição do processo de Wiener, donde ambas terem de ser iguais a $1/2$ (já que a probabilidade da intersecção $P[Y(t - T_a) = 0] = 0$). Daqui resulta que

$$P[T_a \leq t] = 2P[W(t) \geq a] = 2(2\pi t)^{-1/2} \int_a^{+\infty} \exp\left(-\frac{x^2}{2t}\right) dx$$

e, fazendo uma mudança de variável no integral, obtemos a f.d. de T_a

$$F_{T_a}(t) = 2(2\pi)^{-1/2} \int_{a/\sqrt{t}}^{+\infty} \exp(-y^2/2)dy \qquad (t > 0). \tag{4.7}$$

A f.d.p. de T_a obtém-se por derivação:

$$f_{T_a}(t) = |a|(2\pi t^3)^{-1/2} \exp\left(-\frac{a^2}{2t}\right) \qquad (t > 0). \tag{4.8}$$

Pusemos $|a|$ para que a expressão também seja válida para $a < 0$, ficando esse caso como exercício.

Exercício:

a) Apesar de $T_a < +\infty$ q.c. (isto é, $P[T_a = +\infty] = 0$), o valor médio de T_a é $E[T_a] = +\infty$ (cometemos o pequeno abuso de linguagem de chamar esperança matemática ao integral infinito $\int_\Omega T_a dP$ desta v.a. não-integrável). Mostre que esta afirmação é corrrecta.

b) Como $W(t)$ tem trajectórias q.c. contínuas, existe q.c. o *máximo* $X(t) = \max_{0\le u\le t} W(u)$. Determine a f.d. de $X(t)$.

Sugestão: Pela continuidade das trajectórias, há uma relação entre o acontecimento $[X(t) \ge a]$ e um acontecimento relativo a T_a.

c) Seja $0 < t_1 < t_2$. Mostre que a probabilidade de $W(t)$ não tomar o valor 0 no intervalo de tempo (t_1, t_2) é $(2/\pi) \arcsin \sqrt{t_1/t_2}$ (*lei do arcoseno*).

Sugestão: Considere o acontecimento A que consiste em $W(t)$ tomar o valor 0 pelo menos uma vez no intervalo de tempo (t_1, t_2). Para determinar $P(A)$, condicione relativamente a $W(t_1) = x$ e use o teorema das probabilidades totais. Note que $P(A|W(t_1)=x) = P[T_{|x|} \le t_2 - t_1]$, que já sabe calcular, e substitua. Depois é só uma questão de fazer uma mudança de variáveis apropriada no integral duplo que obtém.

4.5 Processos de Wiener multidimensionais

Podemos definir um *processo de Wiener (padrão) m-dimensional*

$$\mathbf{W}(t) = [W_1(t), W_2(t), ..., W_m(t)]^T.$$

Trata-se apenas de um vector (coluna) de m processos de Wiener padrão (unidimensionais) independentes entre si. Se $\mathbf{a}$ é um vector (coluna)

m-dimensional constante e $\mathbf{C}$ é uma matriz constante $m \times m$ definida positiva, então $\mathbf{a} + \mathbf{C}\mathbf{W}(t)$ também se diz um processo de Wiener m-dimensional, podendo agora haver correlação (e, portanto, dependência) entre as suas coordenadas.

Capítulo 5

Processos de difusão

5.1 Definição

Há várias definições não equivalentes do que se entende por processo de difusão. Vamos usar uma definição um pouco mais forte do que o habitual. Para simplificar, vamos apresentar a notação abreviada $\mathbb{E}_{s,x}[...]$ para as esperanças matemáticas condicionais $\mathbb{E}[...|X_s = x]$, onde "..." representa alguma v.a. e X_t é um processo estocástico. Também ocasionalmente usaremos notação semelhante para abreviar probabilidades condicionais: $P_{s,x}[...]$ em vez de $P[...|X_s = x]$, onde "..." será algum acontecimento. Usaremos indiferentemente as notações X_t e $X(t)$ para o mesmo processo estocástico.

Seja $\{X_t\}_{t\in[0,d]}$ um processo estocástico num espaço de probabilidade $(\Omega, \mathcal{F}, P)$. Dizemos que é um *processo de difusão* se for um processo de Markov com trajectórias q.c. contínuas tal que $X_t \in L^2$ $(t \in [0,d])$ e, para todo o $x \in \mathbb{R}$ e $s \in [0,d)$, vier, com convergências uniformes com respeito a $s \in [0,d)$,

$$\lim_{\Delta\to 0+} \frac{1}{\Delta} P_{s,x}\left[|X_{s+\Delta} - x| > \varepsilon\right] = 0 \text{ para todo o } \varepsilon > 0 \quad (5.1)$$

$$\lim_{\Delta\to 0+} \mathbb{E}_{s,x}\left[\frac{X_{s+\Delta} - x}{\Delta}\right] = a(s,x) \quad (5.2)$$

$$\lim_{\Delta\to 0+} \mathbb{E}_{s,x}\left[\frac{(X_{s+\Delta} - x)^2}{\Delta}\right] = b(s,x). \quad (5.3)$$

Nota. A definição mais vulgar é semelhante mas não exige nem a con-

vergência uniforme nem que $X_t \in L^2$ $(t \in [0,d])$. Daí não haver garantia que existam os momentos utilizados em (5.2) e (5.3), pelo que se substituem pelos momentos truncados (que existem sempre):

$$\lim_{\Delta \to 0+} \mathbb{E}_{s,x} \left[\frac{1}{\Delta} (X_{s+\Delta} - x) I_{|X_{s+\Delta}-x| \leq \varepsilon} \right] = a(s,x),$$

$$\lim_{\Delta \to 0+} \mathbb{E}_{s,x} \left[\frac{1}{\Delta} (X_{s+\Delta} - x)^2 I_{|X_{s+\Delta}-x| \leq \varepsilon} \right] = b(s,x),$$

Aqui I é a função indicatriz e $\varepsilon > 0$ é arbitrário. A propriedade (5.1) não sofre alteração.

O *momento infinitésimal de primeira ordem* $a(s,x)$, chamado *coeficiente de tendência* (em inglês, é também conhecido por "drift"), é a velocidade média de X no instante s quando $X_s = x$. Também se lhe pode chamar *média infinitésimal*. Quanto ao *momento infinitésimal de segunda ordem* $b(s,x)$, chamado *coeficiente de difusão*, ele mede a intensidade das flutuações e é a velocidade da variância do processo X no instante s quando $X_s = x$. Também se lhe pode chamar *variância infinitésimal*. De (5.2) e (5.3) vem, quando $\Delta \to 0+$,

$$\mathbb{E}_{s,x} [X_{s+\Delta} - Xs] = a(s,x)\Delta + o(\Delta)$$

$$VAR_{s,x} [X_{s+\Delta} - Xs] = b(s,x)\Delta + o(\Delta),$$

onde $VAR_{s,x}$ tem o significado óbvio. Logo, $X_{s+\Delta} - X_s \approx a(s,x)\Delta + \sqrt{b(s,x)}Z$, onde Z é uma v.a. com média zero e desvio padrão $\sqrt{\Delta}$, que pode ser aproximada por $W_{s+\Delta} - W_s$. Passando para diferenciais e pondo $g(t,x) = \sqrt{b(t,x)}$, obtém-se $dX_s = a(s,x)dt + g(s,x)dW_s$, que é a forma geral de uma equação diferencial estocástica (EDE); o modelo de Black-Scholes referido no Capítulo 3 é apenas um exemplo em que $a(s,x) = rx$ e $g(s,x) = \sigma x$. Assim, não haverá surpresa quando virmos mais tarde que, sob certas condições, as soluções das equações diferenciais estocásticas são processos de difusão.

Como (5.1) pode ser escrita na forma $P[|X_{s+\Delta} - X_s| \leq \varepsilon | X_s = x] = 1 - o(\Delta)$ quando $\Delta t \to 0+$, o seu significado é o de dizer que são improváveis grandes alterações do valor do processo num pequeno intervalo de tempo Δ.

Se $a(t,x)$ e $b(t,x)$ não dependerem de t, temos um *processo de difusão homogéneo*.

Exercício:

a) Mostre que o processo de Wiener W_t é um processo de difusão homogéneo com coeficiente de tendência nulo e coeficiente de difusão unitário.

b) Mostre que $X_t = x_0 + \sigma W_t$, com x_0 e σ constantes, que é um processo de Wiener (não-padrão), é um processo de difusão homogéneo com coeficiente de tendência nulo e coeficiente de difusão σ^2.

c) Mostre que $Z_t = x_0 + \mu t + \sigma W_t$, com x_0, μ e σ constantes, chamado *movimento browniano com tendência* (em ingês "brownian motion with drift") é um processo de difusão homogéneo com coeficiente de tendência μ e coeficiente de difusão σ^2.

5.2 Equações de Kolmogorov

Dado um processo de difusão $\{X_t\}_{t\in[0,d]}$, podemos definir o *operador de difusão*

$$\mathcal{D} = a(s,x)\frac{\partial}{\partial x} + \frac{1}{2}b(s,x)\frac{\partial^2}{\partial x^2}. \tag{5.4}$$

Suponhamos que $a(s,x)$ e $b(s,x)$ são funções contínuas. Seja $h(x)$ uma função contínua limitada e, para um t fixo com $t > s$, defina-se

$$u^*(s,x) = \mathbb{E}_{s,x}[h(X_t)]. \tag{5.5}$$

Se u^* for limitada e contínua com primeiras e segundas derivadas parciais com respeito a x também limitadas e contínuas, então u^* é diferenciável em s e satisfaz a *equação de Kolmogorov regressiva* (EKR)

$$\frac{\partial u^*}{\partial s} + \mathcal{D}u^* = 0 \tag{5.6}$$

com a condição terminal

$$\lim_{s\uparrow t} u^*(s,x) = h(x). \tag{5.7}$$

Uma demonstração rigorosa pode ver-se em [33]. Para isso, recorre-se a um desenvolvimento de Taylor de u^* em torno de (s,x) [1] . Em teoria, a

[1] Para dar uma ideia da técnica da demonstração, suponhamos para facilitar a notação que existe densidade de transição. Vem $u^*(s,x) = \mathbb{E}_{s,x}[h(X_t)] = \int_{\mathbb{R}} h(y)p(t,y|s,x)dy$. Usando as equações de Chapman-Kolmogorov para decompor a transição entre s e t nas transições entre s e $s+\Delta$ (com $\Delta > 0$) e entre $s+\Delta$ e t, temos $u^*(s,x) = \int h(y)\int p(t,y|s+\Delta,z)p(s+\Delta,z|s,x)dzdy$. Façamos agora um desenvolvimento de Taylor de primeira ordem em s e de segunda ordem em x, $p(t,y|s+\Delta,z) = p(t,y|s,x) + \Delta\frac{\partial p}{\partial s} + (z-x)\frac{\partial p}{\partial x} + \frac{1}{2}(z-x)^2\frac{\partial^2 p}{\partial x^2} + resto$ (as derivadas parciais são calculadas no ponto (s,x)), e substituamos na expressão anterior. Vem $u^*(s,x) = u^*(s,x) + \Delta\int h(y)\frac{\partial p}{\partial s}dy + \mathbb{E}_{s,x}[X_{s+\Delta}-x]\int h(y)\frac{\partial p}{\partial x}dy + \frac{1}{2}\mathbb{E}_{s,x}[(X_{s+\Delta}-x)^2]$ $\int h(y)\frac{\partial^2 p}{\partial x^2}dy + resto$. Dividindo membro a membro por Δ e fazendo $\Delta \to 0$, vem, desde que se prove que o termo correspondente ao resto tende para zero (e essa é a parte delicada), $0 = \frac{\partial u^*}{\partial s} + a(s,x)\frac{\partial u^*}{\partial x} + \frac{1}{2}b(s,x)\frac{\partial^2 u^*}{\partial x^2} = (\frac{\partial}{\partial s} + \mathcal{D})u^*$. O caso

EKR permite determinar as probabilidades de transição pois estas ficam univocamente definidas se conhecermos $u^*(s,x) = \mathbb{E}_{s,x}[h(X_t)]$ (solução da EKR) para todas as funções h num conjunto denso no espaço das funções contínuas limitadas. Ou seja, para caracterizar probabilisticamente um processo de difusão, precisamos apenas dos dois primeiros momentos infinitésimais (únicos que intervêm na EKR). Há um método mais fácil de obter as probabilidades de transição no caso de existir densidade de transição $p(t,y|s,x)$ contínua em s e com primeiras e segundas derivadas parciais relativamente a x também contínuas em s. De facto, nesse caso, a densidade de transição pode obter-se directamente (para t fixo com $t > s$ e y fixo) como solução fundamental da EKR

$$\frac{\partial p}{\partial s} + \mathcal{D}p = 0; \tag{5.8}$$

por *solução fundamental* entendemos a que satisfaz a condição terminal

$$\lim_{s \uparrow t} p(t,y|s,x) = \delta(x-y), \tag{5.9}$$

onde δ é a função delta de Dirac (ver Capítulo 3). [2]

Se tivermos um processo de difusão homogéneo e usarmos a notação $p(\tau,y|x) = p(t,y|t-\tau,x)$ (não depende de t), agora os coeficientes de tendência e difusão não dependem do tempo, $\mathcal{D} = a(x)\frac{\partial}{\partial x} + \frac{1}{2}b(x)\frac{\partial^2}{\partial x^2}$, e, como $s = t - \tau$ implica $\frac{\partial}{\partial s} = -\frac{\partial}{\partial \tau}$, obtemos para EKR

$$\left(-\frac{\partial}{\partial \tau} + \mathcal{D}\right) p(\tau,y|x) = 0 \tag{5.10}$$

e para condição terminal

$$\lim_{\tau \downarrow 0} p(\tau,y|x) = \delta(x-y). \tag{5.11}$$

Neste caso homogéneo, vem também (ver (5.5)) que $u(\tau,x) := u^*(t-\tau,x) = \mathbb{E}_{t-\tau,x}[h(X_t)] = \mathbb{E}_{0,x}[h(X_{t-(t-\tau)})] = \mathbb{E}_{0,x}[h(X_\tau)]$. Logo, de (5.6) e (5.7), resulta que

$$u(\tau,x) := \mathbb{E}_{0,x}[h(X_\tau)] \tag{5.12}$$

$\Delta < 0$ também pode ser tratado com alguns ajustamentos. A condição terminal é consequência das propriedades de h e de se ter $\mathbb{E}_{s,x}[h(X_s)] = h(x)$.

[2] A demonstração é semelhante à feita para (5.6), mas agora parte-se de $p(t,y|s,x) = \int p(t,y|s+\Delta,z)p(s+\Delta,z|s,x)dz$. A condição terminal resulta de a distribuição de transição entre o instante s e ele próprio ter toda a massa de probabilidade concentrada em x, pelo que a densidade de probabilidade de transição entre s e s não é própria (não existe no sentido corrente).

satisfaz para $\tau \geq 0$ a EKR

$$-\frac{\partial u}{\partial \tau} + \mathcal{D}u = 0 \tag{5.13}$$

com a condição terminal

$$\lim_{\tau \downarrow 0} u(\tau, x) = h(x). \tag{5.14}$$

Também existe uma *equação de Kolmogorov progressiva* (EKP) para a densidade de transição $p(t, y|s, x)$ no caso de ela existir e ter derivadas parciais $\frac{\partial p}{\partial t}$, $\frac{\partial (a(t,y)p)}{\partial y}$ e $\frac{\partial^2 (b(t,y)p)}{\partial y^2}$ contínuas. Para s fixo tal que $s < t$ e x fixo, a densidade de transição é a solução fundamental da EKP, também chamada *equação de Fokker-Planck* ou *equação de difusão* [3] :

$$\frac{\partial p}{\partial t} + \frac{\partial}{\partial y}(a(t, y)p) - \frac{1}{2}\frac{\partial^2}{\partial y^2}(b(t, y)p) = 0. \tag{5.15}$$

Por solução fundamental entende-se a que satisfaz a condição inicial

$$\lim_{t \downarrow s} p(t, y|s, x) = \delta(x - y). \tag{5.16}$$

Se existir a densidade de probabilidade não-condicional $p(t, y) = f_{X_t}(y)$ (f.d.p. da distribuição de X_t) e tiver derivadas parciais $\frac{\partial p}{\partial t}$, $\frac{\partial (a(t,y)p)}{\partial y}$ e $\frac{\partial^2 (b(t,y)p)}{\partial y^2}$ contínuas, ela também satisfaz a EKP (5.15) mas com condição inicial

$$\lim_{t \downarrow s} p(t, y) = p(s, y). \tag{5.17}$$

Quando o processo de difusão é homogéneo, se usarmos a notação $p(\tau, y|x) = p(s + \tau, y|s, x)$ (não depende de s), obtemos para EKP

$$\frac{\partial p(\tau, y|x)}{\partial \tau} + \frac{\partial}{\partial y}(a(y)p(\tau, y|x)) - \frac{1}{2}\frac{\partial^2}{\partial y^2}(b(y)p(\tau, y|x)) = 0 \tag{5.18}$$

[3] A técnica de demonstração é análoga à da EKR (5.8) mas partindo de $p(t, y|s, x) = \int p(t, y|t - \Delta, z)p(t - \Delta, z|s, x)dz$, isto é, decompondo a transição entre s e t nas transições entre s e $t - \Delta$ (com $\Delta > 0$) e entre $t - \Delta$ e t (olha-se para o que se passa na vizinhança-Δ do presente t, ao contrário do que se fazia na EKR que olhava para a vizinhança-Δ do passado s). Depois faz-se um desenvolvimento de Taylor de $p(t - \Delta, z|s, x)dz$.

e para condição terminal

$$\lim_{\tau \downarrow 0} p(\tau, y|x) = \delta(x - y). \tag{5.19}$$

É interessante ver se o processo de difusão homogéneo tem uma *densidade invariante* ou *densidade de equilíbrio* $p(y)$ que, como o nome indica, seja invariante em relação ao tempo, isto é, tal que, se a f.d.p. de X_0 for $p(y)$, então será também $p(y)$ a f.d.p. de X_t para todo o $t \in I$. Se existir, ela terá de satisfazer (5.15) e, como $\frac{\partial p(y)}{\partial t} = 0$, ela será solução da equação diferencial ordinária

$$\frac{d}{dy}(a(y)p(y)) - \frac{1}{2}\frac{d^2}{dy^2}(b(y)p(y)) = 0. \tag{5.20}$$

Existindo distribuição invariante $p(y)$, mais interessante é ver o que se passa com a densidade (suposta existente) $p(t, y) = f_{X_t}(y)$ de X_t quando a distribuição inicial de X_0 é diferente da distribuição invariante. É particularmente interessante saber se, quando $t \to +\infty$, a distribuição de X_t converge para a distribuição invariante. Se isso suceder, diremos que $p(y) = f_{X_{+\infty}}(y)$ é a *densidade estacionária* e podemos chamar $X_{+\infty}$ a uma v.a. cuja distribuição tenha essa f.d.p. Nesse caso X_t *converge em distribuição* para $X_{+\infty}$, isto é, $F_{X_t}(y) \to F_{X_{+\infty}}(y)$ (claro que $F_{X_{+\infty}}(y) = \int_{-\infty}^{y} p(z)dz$ é a f.d. de $X_{+\infty}$). Quando existe densidade estacionária, é habitual, dadas condições de regularidade adequadas, que X_t seja um *processo ergódico*, o que significa basicamente que os *momentos amostrais* (médias ao longo da trajectória de certas funções de X_t) convergem (quando o intervalo de tempo envolvido na média tende para $+\infty$) para os correspondentes *momentos de conjunto* da distribuiçao estacionária (esperanças matemáticas dessas funções de $X_{+\infty}$). Isso facilita bastante pois evita estimar os momentos de conjunto através da utilização de muitas trajectórias calculando médias para o conjunto das trajectórias. Ora, é frequente em muitas aplicações não podermos repetir a experiência e, portanto, dispormos apenas de uma trajectória. Por exemplo, as taxas de juro de curto prazo ou a evolução do tamanho de uma população natural num certo período de tempo não podem ser repetidas para outros cenários do mercado ou estados da natureza aleatoriamente escolhidos. Só dispomos de uma única trajectória. Se, porém, houver ergodicidade, os momentos amostrais ao longo dessa única trajectória permitem estimar os momentos de conjunto. Isso permite utilizar métodos estatísticos em muitas aplicações em que o fenómeno dinâmico pode ser modelado por um processo de difusão (como seja o caso de fenómenos modelados por equações diferenciais estocásticas satisfazendo

certas condições de regularidade).

No caso do processo de Wiener $W(t)$, que é processo de difusão homogéneo com coeficiente de tendência nulo e coeficiente de difusão unitário, a EKR e a EKR para a densidade de transição $p(t, y|s)$ têm o mesmo aspecto

$$\frac{\partial p(t,y|x)}{\partial t} = \frac{1}{2}\frac{\partial^2}{\partial y^2}(p(t,y|x)). \tag{5.21}$$

Utilizando a solução fundamental desta equação às derivadas parciais, conhecida por *equação do calor* (dá a evolução da intensidade do calor quando se propaga ao longo de um eixo), obtemos para densidade de transição de $W(t)$

$$p(t,y|x) = (2\pi t)^{-1/2}\exp\left(-\frac{(y-x)^2}{2t}\right) \tag{5.22}$$

(note-se que a solução é determinada a menos de uma constante multiplicativa, que se obtém atendendo a que se trata de uma densidade e, portanto, tem de vir $\int_{\mathbb{R}} p(t,y|x)dy = 1$). Este resultado mostra que, como já sabíamos por (4.3),

$$W(t+s)|(W(s)=x) \frown \mathcal{N}(x,t). \tag{5.23}$$

As equações de Kolmogorov permitem reduzir o problema probabilístico da determinação das densidades de transição de um processo de difusão ao problema determinístico de resolução de uma equação às derivadas parciais. Como as soluções das equações diferenciais estocásticas são, sob certas condições de regularidade, processos de difusão, a caracterização probabilística das suas soluções (as suas probabilidades de transição) fica assim reduzida a um problema determinístico. Mas também, inversamente, algumas questões relevantes sobre teoria das equações às derivadas parciais de difusão têm avançado graças ao estudo por métodos probabilísticos das propriedades dos processos de difusão (ou soluções de equações diferenciais estocásticas) a elas associados.

Dito isto, geralmente é difícl obter soluções das equações de Kolmogorov, mesmo que por métodos numéricos, sendo por vezes preferível usar métodos de simulação. O leitor pode fazer a experiência de tentar obter através da solução fundamental da EKP a expressão das densidades de transição do movimento browniano com tendência $Z(t) = x_0 + \mu t + \sigma W(t)$ (x_0, μ e σ constantes), que é um processo de difusão homogéneo com coeficiente de tendência constante μ e coeficiente de difusão constante σ^2. Neste caso relativamente feliz é possível, mas dá algum trabalho, encontrar a solução explícita. Porventura, se o leitor soubesse apenas o coeficiente de tendência e de difusão e não soubesse

a relação de $Z(t)$ com o processo de Wiener, seria este o caminho que seria tentado a tomar, pelo menos nesta fase do estudo. Encontrando a solução, verifica que ela é a densidade de uma distribuição normal e que

$$Z(t+s)|(Z(s)=x) \frown \mathcal{N}(x+\mu t, \sigma^2 t). \tag{5.24}$$

Mas certamente que chegará a este resultado muito mais facilmente usando as propriedades do processo de Wiener, o que deixamos como exercício. Claro que, agora que conhece a solução, não será difícil mostrar que ela é a solução fundamental da EKP. Em todo o caso, as equações de Kolmogorov são um importante instrumento para o estudo teórico dos processos de difusão e, portanto, das equações diferenciais estocásticas.

5.3 Caso multidimensional

O estudo pode ser estendido a *processos de difusão n-dimensionais* $\mathbf{X}(t) = [X_1(t), ..., X_n(t)]^T$. Agora o coeficiente de tendência é um vector n-dimensional $\mathbf{a}(s, \mathbf{x})$ cujas coordenadas são os coeficientes de tendência dos processos unidimensionais constituintes. O coeficiente de difusão é uma matriz $\mathbf{b}(s, \mathbf{x})$ de elementos

$$b_{i,j}(s,\mathbf{x}) = \lim_{\Delta \to 0+} E_{s,\mathbf{x}} \left[\frac{(X_i(s+\Delta) - x_i)\,(X_j(s+\Delta) - x_j)}{\Delta} \right].$$

O operador de difusão é

$$\mathcal{D} = \sum_{i=1}^{n} a_i(s,\mathbf{x}) \frac{\partial}{\partial x_i} + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} b_{ij}(s,\mathbf{x}) \frac{\partial^2}{\partial x_i \partial x_j}.$$

A EKP toma a forma

$$\frac{\partial p}{\partial t} + \sum_{i=1}^{n} \frac{\partial}{\partial y_i}(a_i(t,\mathbf{y})p) - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial^2}{\partial y_i \partial y_j}(b_{ij}(t,\mathbf{y})p) = 0.$$

Suponhamos que são largadas na atmosfera partículas de um poluente. Devido aos choques com as moléculas do ar, elas difundem-se, mas, se houver vento, há, além disso, um movimento tendencial na direcção do vento. Seja $b_{ij}(s,\mathbf{x})$ a velocidade de alteração da covariância (da variância se $i=j$) entre os movimentos devidos à difusão nas direcções de x_i e x_j no local $\mathbf{x}$ e no instante s. Seja $\mathbf{a}(s,\mathbf{x})$ o vector velocidade do vento no local $\mathbf{x}$ no instante s. Então a posição $\mathbf{X}(t)$ da partícula no instante t será um processo de difusão a $n=3$ dimensões com coeficiente de tendência

$\mathbf{a}(s, \mathbf{x})$ e coeficiente de difusão $\mathbf{b}(s, \mathbf{x})$. Frequentemente, é razoável supor que a matriz $\mathbf{b}(s, \mathbf{x})$ é o produto de uma função escalar $\sigma(s, \mathbf{x})$ pela matriz identidade (isotropia da difusão da partícula nas várias direcções do espaço); admite-se que $\sigma(s, \mathbf{x})$ possa depender do instante s e do local $\mathbf{x}$ pois as condições de temperatura e outras que possam afectar a difusão podem variar com o local e o tempo. Resolvendo uma das equações de Kolmogorov podemos ober a f.d.p. $p(t, \mathbf{y})$ de $\mathbf{X}(t)$, que será assim a f.d.p. no instante t para uma partícula ser encontrada no local $\mathbf{y}$. Se for largado um número muito grande N de partículas, então, no instante t, o número de partículas localizadas num pequeno volume ΔV à volta do local $\mathbf{y}$ será aproximadamente $p(t, \mathbf{y}) N \Delta V$. Problema análogo ocorre com a forma como se espalha petróleo derramado no oceano.

Capítulo 6

Integrais estocásticos

6.1 Definição informal dos integrais de Itô e Stratonovich

O modelo de Black-Scholes (3.4) é um caso particular de uma equação diferencial estocástica. Mas podemos pensar em situações mas gerais. Neste livro vamos estudar *equações diferenciais estocásticas* (EDE) num intervalo de tempo $[0, d]$ $(d > 0)$ da forma

$$dX(t) = f(t, X(t))dt + g(t, X(t))dW(t) \qquad X(0) = X_0, \qquad (6.1)$$

onde $f(s, x)$ e $g(s, x)$ são funções reais com domínio $[0, d] \times \mathbb{R}$ e X_0 é uma v.a. (condição inicial) independente do processo de Wiener $W(t)$. A condição inicial pode, em particular, ser uma v.a. degenerada que assuma um valor real constante x_0 com probabilidade um.

Equações deste tipo aparecem nos mais variados domínios da Ciência e da Tecnologia, sempre que queremos modelar um fenómeno com dinâmica descrita por uma equação diferencial que seja perturbado por flutuações aleatórias (desde que estas possam, ainda que aproximadamente, considerar-se contínuas no tempo e com incrementos independentes).

Como sucede também com as equações diferenciais ordinárias (EDO), o *problema de Cauchy* (6.1) não é mais que uma equação integral "disfarçada":

$$X(t) = X_0 + \int_0^t f(s, X(s))ds + \int_0^t g(s, X(s))dW(s). \qquad (6.2)$$

Entende-se por *solução* $X(t) = X(t, \omega)$ de (6.1) um processo estocástico

que satisfaça a *equação integral estocástica* (6.2), que não é mais que a forma integral da EDE. No caso das EDO, a equação integral obtém-se por integração membro a membro da EDO. Aqui, a definição de solução da EDE como sendo a solução da sua forma integral é a mais natural, já que as derivadas do processo de Wiener $W(t)$ e, portanto, de $X(t)$, não existem no sentido corrente do termo, apenas existem como processos estocásticos generalizados. Esta definição de solução exige, contudo, que os integrais em (6.2) estejam bem definidos.

Quanto ao primeiro integral $\int_0^t f(s, X(s,\omega))ds$, podemos fixar o acaso (cenário do mercado, estado da natureza, ...) $\omega = \omega_0$, caso em que se torna (supondo que f é uma função suficientemente bem comportada) um integral de Riemann $\int_0^t F(s,\omega_0)ds$ com $F(s,\omega) = f(s, X(s,\omega))$; o seu valor dependerá do valor de $\omega = \omega_0$ e, portanto, o integral é uma v.a. (função mensurável de ω).

Será que este truque resulta para o integral $\int_0^t g(s, X(s,\omega))dW(s,\omega)$, usando agora "integrais de Riemann-Stieltjes" $\int_0^t G(s,\omega_0)dW(s,\omega_0)$ com $G(s,\omega) = g(s, X(s,\omega))$? A reposta é negativa, pois estes alegados *integrais de Riemann-Stieltjes* (RS) não existem como tal. De facto, a definição usual de integral RS como limite comum (quando o diâmetro das decomposições de $[0,t]$ tende para zero) de todas as *somas de Riemann-Stieltjes* (somas RS) não funciona porque escolhas diferentes dos pontos intermédios da função integranda produzem limites diferentes. A razão reside no facto de a função integradora, o processo de Wiener $W(t)$, ser q.c. de variação ilimitada. Assim, faz falta uma definição diferente do segundo integral de (6.2).

Neste Capítulo iremos definir integrais da forma $\int_0^t G(s,\omega)dW(s,\omega)$, abreviadamente

$$\int_0^t G(s)dW(s),$$

para funções bastante gerais G.

Mas primeiro vamos mostrar ao leitor, com um exemplo, que a definição de RS do integral não funciona aqui. Consideremos para isso o caso particular de $G(s) = W(s)$ e tentemos determinar o integral $\int_0^t W(s)dW(s)$.

Se se aplicassem as regras usuais de cálculo, o leitor, notando que $W(0) = 0$, diria imediatamente que o integral seria igual a

$$\frac{1}{2}W^2(t).$$

Será? Consideremos decomposições

$$0 = t_{n,0} < t_{n,1} < ... < t_{n,n-1} < t_{n,n} = t \quad (n = 1, 2, ...) \tag{6.3}$$

do intervalo de integração $[0, t]$ e suponhamos que os seus diâmetros $\delta_n = \max_{k=1,...,n}(t_{n,k} - t_{n,k-1}) \to 0$ quando $n \to +\infty$. Para cada sub-intervalo $[t_{n,k}, t_{n,k-1}]$ da decomposição, seja $\tau_{n,k}$ um ponto situado nesse intervalo (chamemos-lhe "ponto intermédio").

As somas RS tomam a forma

$$\sum_{k=1}^{n} W(\tau_{n,k}) \left(W(t_{n,k}) - W(t_{n,k-1})\right). \tag{6.4}$$

Vamos ver quais os limites destas somas RS quando $n \to +\infty$. As somas são v.a. e escolheremos limites em média quadrática (limites m.q., que são limites na norma L^2), representando tais limites por "l.i.m.".

Se escolhermos para pontos intermédios os pontos iniciais de cada sub-intervalo, $\tau_{n,k} = t_{n,k-1}$, obtemos as somas RS

$$S_n = \sum_{k=1}^{n} W(t_{n,k-1}) \left(W(t_{n,k}) - W(t_{n,k-1})\right) \tag{6.5}$$

e, no limite m.q., o *integral não-antecipativo*, chamado *integral de Itô*:

$$I = \int_0^t W(s) dW(s) = \text{ l.i.m. } S_n. \tag{6.6}$$

Note que em (6.5) os valores da função integranda (que, quando trabalharmos com EDE, medem o impacto das flutuações aleatórias no fenómeno por elas descrito) são independentes dos futuros incrementos $W(t_{n,k}) - W(t_{n,k-1})$ do processo integrador de Wiener, os quais descrevem as perturbações aleatórias (por exemplo, as perturbações dos mercados financeiros) no sub-intervalo de tempo $(t_{n,k} - t_{n,k-1}]$. Sendo independentes, não os antecipam, não os "adivinham". Usar o integral de Itô é uma forma de dizer que o comportamento presente do fenómeno não depende das futuras perturbaçoes aleatórias (ausência de capacidades "adivinhatórias"). O integral de Itô, contudo, não segue as regras usuais de cálculo.

Com efeito, vem

$$I = \int_0^t W(s) dW(s) = \frac{1}{2} \left(W^2(t) - t\right). \tag{6.7}$$

Para mostrar que isso é verdade, precisamos de mostrar que

$$E\left[\left(S_n - \frac{1}{2}\left(W^2(t) - t\right)\right)^2\right] \to 0$$

quando $n \to +\infty$. Note que $W^2(t) = \sum_{k=1}^{n}\left(W^2(t_{n,k}) - W^2(t_{n,k-1})\right)$ e, portanto, $S_n - \frac{1}{2}\left(W^2(t) - t\right) = -\frac{1}{2}\left(\sum_{k=1}^{n}\left(W(t_{n,k}) - W(t_{n,k-1})\right)^2 - t\right)$ $= -\frac{1}{2}\sum_{k=1}^{n} h_{n,k}$, com $h_{n,k} = \left(W(t_{n,k}) - W(t_{n,k-1})\right)^2 - (t_{n,k} - t_{n,k-1})$. Como os $h_{n,k}$ $(k = 1, 2, ..., n)$ são independentes e $E[h_{n,k}] = 0$, obtemos $4E\left[\left(S_n - \frac{1}{2}\left(W^2(t) - t\right)\right)^2\right] = E\left[\left(\sum_k h_{n,k}\right)^2\right] = VAR\left[\sum_k h_{n,k}\right] = \sum_k VAR[h_{n,k}] = 2\sum_k (t_{n,k} - t_{n,k-1})^2 \leq 2\delta_n t \to 0$, conforme pretendido.

Se, porém, escolhermos como pontos intermédios de cada sub-intervalo $[t_{n,k-1}, t_{n,k}]$ os pontos terminais $\tau_{n,k} = t_{n,k}$, obtemos as somas RS

$$S_n^+ = \sum_{k=1}^{n} W(t_{n,k})\left(W(t_{n,k}) - W(t_{n,k-1})\right) \tag{6.8}$$

e, no limite m.q., o integral (que é antecipativo)

$$I^+ = \text{ l.i.m. } S_n^+ = \frac{1}{2}\left(W^2(t) + t\right). \tag{6.9}$$

Este integral também não segue as regras usuais de cálculo.

Exercício: Demonstre (6.9). Obtenha também os integrais $I^{(\alpha)}$ correspondentes aos limites m.q. das somas RS quando se escolhem como pontos intermédios $\tau_{n,k} = (1 - \alpha)t_{n,k-1} + \alpha t_{n,k}$ para $0 \leq \alpha \leq 1$. Note que $I = I^{(0)}$ e que $I^+ = I^{(1)}$.

O ponto importante é que diferentes somas RS (correspondentes a diferentes escolhas dos pontos intermédios) produzem diferentes limites m.q. das somas RS e, portanto, não existe integral de Riemann-Stieltjes. Será que a situação se mantém se trabalharmos com outro tipo de convergência (como a convergência em probabilidade ou a convergência quase certa, por exemplo)? A resposta é negativa pois, como se sabe, se esses limites existissem teriam de coincidir com os limites m.q. e, portanto, também variariam com a escolha dos pontos intermédios. Há, assim, uma infinidade de variedades de integrais estocásticos, das quais o integral de Itô, apesar de não seguir as regras usuais de cálculo, é usualmente a preferida na literatura por causa da sua ausência de capacidades "adivinhatórias" (o que parece mais adequado na mode-

lação de fenómenos naturais) e pelas boas propriedades probabilísticas que iremos estudar.

Outra variedade de integral estocástico que iremos considerar por parecer ser a mais adequada em certas circunstâncias é o *integral de Stratonovich*

$$(S)\int_0^t W(s)dW(s) = \int_0^t W(s)\circ dW(s) = \text{ l.i.m. } \frac{S_n + S_n^+}{2} = \frac{1}{2}W^2(t), \tag{6.10}$$

o qual é antecipativo e não tem as boas propriedades probabilísticas do integral de Itô, mas segue as regras usuais de cálculo. Aplica-se também a uma classe mais restrita de funções integrandas. Deixaremos o integral de Stratonovich para mais tarde, pois só o vamos considerar para funções integrandas que sejam soluções de equações diferenciais estocásticas, e vamos ocupar-nos do estudo do integral de Itô.

Obtivémos acima o integral de Itô para uma função integranda especial $W(t)$. O mesmo tipo de definição é aplicável para funções integrandas arbitrárias $G(s,\omega)$ desde que sejam *não-antecipativas* (isto é, em cada instante t, a função seja independente dos incrementos futuros do processo de Wiener) e sejam *contínuas em média quadrática* (contínuas m.q.). Para este caso, podemos definir o *integral de Itô* como

$$I(G) = \int_0^t G(s)dW(s) = \text{ l.i.m. } \sum_{k=1}^{n} G(t_{n,k-1})\left(W(t_{n,k}) - W(t_{n,k-1})\right). \tag{6.11}$$

No entanto, há necessidade de construir o integral de Itô para uma classe mais ampla de funções que podem não ser contínuas m.q., generalizando a definição anterior. É o que faremos na Secção 6.2. O leitor menos interessado nas questões técnicas poderá contentar-se com a definição de integral para funções contínuas m.q. e, na leitura da Secção 6.2, limitar a atenção às principais conclusões sobre as propriedades do integral.

6.2 Construção do integral de Itô

Consideremos um processo de Wiener padrão $W(t) = W(t,\omega)$ $(t \geq 0)$ num espaço de probabilidade $(\Omega, \mathcal{F}, P)$ e seja $\mathcal{M}_s = \sigma(W(u), 0 \leq u \leq s)$ $(s \geq 0)$ a sua filtração natural. O objectivo é o de definir o integral de Itô num intervalo de tempo $[0,t]$ $(t \geq 0)$. Dizemos que $\{\mathcal{A}_s\}_{s\in[0,t]}$ é uma *filtração não-antecipativa* se, para $0 \leq s \leq t$, $\mathcal{A}_s \supset \mathcal{M}_s$ e $\mathcal{A}_s$ for independente dos incrementos futuros $W(u) - W(s)$ $(u \geq s)$ do processo de Wiener (o mesmo é dizer, independente da álgebra-σ $\sigma(W(u) - W(s), u \geq s)$). Usualmente escolhe-se para $\{\mathcal{A}_s\}_{s\in[0,t]}$ a

própria filtração natural $\{\mathcal{M}_s\}_{s\in[0,t]}$, mas há por vezes necessidade de incluir informação adicional à do processo de Wiener (por exemplo, quando considerarmos a informação da condição inicial duma EDE) e daí a razão de se poder escolher uma filtração maior, desde que, tal como sucede na filtração natural, não antecipemos os incrementos futuros do processo de Wiener. Seja λ a medida de Lebesgue no intervalo $[0,t]$. [1]

Vamos agora definir a classe de funções integrandas com que vamos trabalhar.

Seja $G(s,\omega)$ (usualmente abrevia-se a notação para $G(s)$) um função de domínio $[0,t]\times\Omega$ com valores em $\mathbb{R}$ conjuntamente mensurável com respeito a ambas as variáveis (s,ω) [2] . Obviamente, $\{G(s)\}_{s\in[0,t]}$ é um processo estocástico. É costume cometer o abuso de linguagem de identificar duas funções conjuntamente mensuráveis que sejam quase iguais com respeito à medida produto $\lambda\times P$ (elas são quase iguais quando o conjunto dos pontos (s,ω) para os quais as duas funções diferem tem medida $\lambda\times P$ nula). Esta identficação significa que, na realidade, vamos trabalhar com as classes de equivalência destas funções com respeito à relação de quase igualdade, mas que, para simplificar a linguagem, em vez de falarmos de uma determinada classe de equivalência, falamos de uma qualquer das funções representantes da classe (isto é, pertencentes à classe).

Assim, podemos definir o espaço L^{2*} de funções (rigorosamente devíamos falar de classes de equivalência de funções) conjuntamente mensuráveis G tais que $\int_{[0,t]\times\Omega}|G(s,\omega)|^2 d(\lambda\times P)<+\infty$; note-se que, pelo *teorema de Fubini*, $\int_{[0,t]\times\Omega}|G(s,\omega)|^2 d(\lambda\times P)=\int_0^t\left(\int_\Omega|G(s,\omega)|^2 dP(\omega)\right)ds=\int_0^t\mathbb{E}[|G(s)|^2]ds=\mathbb{E}\left[\int_0^t|G(s)|^2 ds\right]$. Este é um espaço L^2 com respeito à medida produto $\lambda\times P$, e, portanto, é um espaço de Hilbert com *norma-*

[1] Uma *medida* μ definida num espaço mensurável $(\Omega,\mathcal{F})$ é uma função de $\mathcal{F}$ em $[0,+\infty]$ aditiva-σ tal que $\mu(\emptyset)=0$. $(\Omega,\mathcal{A},\mu)$ diz-se um *espaço de medida*. Note-se que uma probabilidade é um caso particular de medida, que satisfaz a propriedade suplementar de ser normada. A medida de Lebesgue λ é uma extensão do conceito de comprimento. A medida de Lebesgue λ em $[0,t]$ pode ser definida para os conjuntos de Borel $B\in\mathcal{B}_{[0,t]}$ de $[0,t]$ (pode provar-se que existe uma e uma só medida nestes conjuntos de Borel tal que a medida de qualquer intervalo contido em $[0,t]$ é o seu comprimento). Temos então o espaço de medida $([0,t],\mathcal{B}_{[0,t]},\lambda)$, que pode ser estendido por completação a um espaço de medida $([0,t],\mathcal{M}_{[0,t]},\lambda)$. A completação consiste em tomar para $\mathcal{M}_{[0,t]}$ a classe dos conjuntos da forma $B\cup N$ (com $B\in\mathcal{B}_{[0,t]}$ e N qualquer subconjunto de um conjunto de Borel de medida nula) e estender a medida a esses conjuntos pondo $\lambda(B\cup N)=\lambda(B)$.

[2] Isto significa que a imagem inversa por G de qualquer conjunto de Borel pertence à álgebta-σ produto $\mathcal{B}_{[0,t]}\times\mathcal{F}$.

L^{2*}

$$\|G\|_{2*} = \left(\int_0^t \mathbb{E}[|G(s)|^2] ds \right)^{1/2} \tag{6.12}$$

e *produto interno* $\langle G_1, G_2 \rangle_* = \left(\int_0^t \mathbb{E}[G_1(s) G_2(s)] ds \right)$.

Dizemos que uma função conjuntamente mensurável G é uma *função não-antecipativa-$\mathcal{A}_s$* se estiver adaptada a $\mathcal{A}_s$, isto é, se, para todo o $s \in [0,t]$ fixo, $G(s,.)$ for, como função de ω, mensurável-$\mathcal{A}_s$. No caso de ser claro qual é a filtração $\mathcal{A}_s$ ou no caso de ser a filtração natural do processo de Wiener, é costume abeviar para "função não-antecipativa". Da definição resulta que as funções não-antecipativas são, em cada instante s, independentes dos incrementos futuros $W(u) - W(s)$ $(u \geq s)$ do processo de Wiener. Recordamos que, quando na Secção 6.1 definíamos o integral de Itô $\int_0^t W(t) dW(t)$ como limite m.q. das somas RS (6.5), o carácter não-antecipativo do integral, que era essencial, resultava de o valor $W(t_{n,k-1})$ da função integranda ser independente dos incrementos (futuros) $W(t_{n,k}) - W(t_{n,k-1})$ do processo integrador. Se G for não-antecipativa, esta propriedade mantém-se quando substituirmos, em (6.5), $W(t_{n,k-1})$ por $G(t_{n,k-1})$ e está aberto o caminho à definição do integral de Itô $\int_0^t G(s) dW(s)$.

A família de funções integrandas que iremos adoptar é precisamente a família das funções G (rigorosamente devíamos falar de classes de equivalência de funções) conjuntamente mensuráveis do espaço L^{2*} que sejam não-antecipativas-$\mathcal{A}_s$. Vamos representar esta família, que é um espaço de Hilbert, por $H^2[0,t]$.

Vamos seguir um caminho tradicional na construção de um integral, que é o de o definir para funções mais simples e depois generalizar por continuidade.

Começaremos por definir o integral de Itô para *funções em escada* (alguns também lhe chamam funções simples), que são funções $G \in H^2[0,t]$ constantes (relativamente à variável tempo s mas não relativamente à variável acaso ω) nos sub-intervalos de alguma decomposição $0 = t_0 < t_1 < ... < t_n = t$ do intervalo $[0,t]$, isto é $G(s,\omega) = G(t_{k-1},\omega)$ para todo o $s \in [t_{k-1}, t_k)$ $(k = 1, ..., n)$. Representemos por $H_E^2[0,t]$ o espaço destas funções em escada, que é um subespaço vectorial de $H^2[0,t]$. A definição do integral de Itô para tais funções é a óbvia:

$$I(G) = \int_0^t G(s) dW(s) := \sum_{k=1}^{n} G(t_{k-1}) \left(W(t_k) - W(t_{k-1}) \right). \tag{6.13}$$

Note-se que, como estas funções em escada são constantes em cada sub-

intervalo, seria indiferente usar como ponto intermédio para o cálculo da função integranda o ponto inicial ou qualquer outro ponto do sub-intervalo, pelo que para estas funções não há distinção entre o integral de Itô e os outros tipos de integral estocástico.

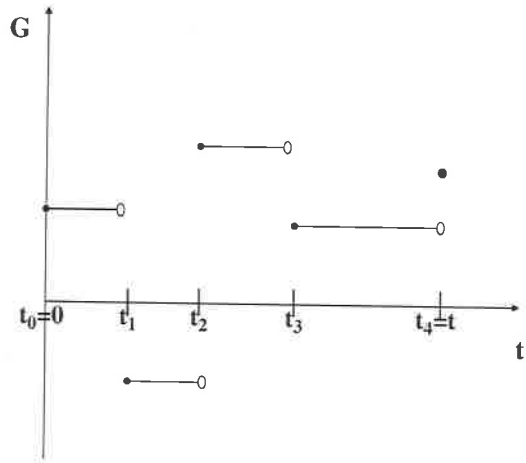


Figura 6.1: Exemplo de gráfico de uma função em escada $G(s,\omega)$ com $n = 4$ para ω fixo. Se mudássemos para outro valor de ω, a altura dos degraus mudaria mas os pontos t_0, t_1, t_2, t_3, t_4 manter-se-iam.

Nota: Note-se também que o integral depende de ω, isto é, trata-se de uma v.a. Como o integral está definido não propriamente para a função G (esse é o abuso de linguagem) mas para a classe de equivalência a que ela pertence, quando substitui a função G, que está a representar a classe a que pertence, por outra função representante da mesma classe (isto é, por uma função quase igual a G), vai obter como integral outra v.a. Mas essa outra v.a. é q.c. igual à primeira, isto é, difere da primeira num conjunto de valores de ω com probabilidade nula. Por isso, consideramos o integral definido a menos de uma quase igualdade. Identificando, como é habitual, v.a. quase iguais, isto é, trabalhando com classes de equivalência de v.a., o integral fica univocamente definido. Neste sentido, rigorosamente, o integral é uma aplicação I que transforma uma classe de equivalência de funções G numa classe de equivalência de v.a. $I(G)$ e, como dissemos, nesse sentido é uma aplicação univocamente definida. Na prática, porém, nós abusamos da linguagem e falamos do integral de uma função G (que representa a sua classe de equivalência) como sendo

a v.a $I(G)$ definida em (6.13) (v.a que representa a sua classe de equivalência). Mas, ao abusar da linguagem, temos de ter sempre presente que não devemos fazer distinção entre funções G quase iguais nem entre v.a. quase iguais.

Exercício: Mostre que a definição anterior é consistente no sentido de, se aplicar (6.13) usando duas diferentes decomposições de $[0,t]$ onde assente a mesma função em escada, obtém o mesmo valor para o seu integral. Para isso convém notar que a mesma função em escada pode também assentar na decomposição obtida por fusão das duas decomposições (a que tem como pontos de decomposição os de ambas as decomposições iniciais).

Exercício: Mostre que o integral de funções em escada é uma *aplicação linear*, isto é, dadas constantes reais α_1 e α_2, vem

$$I(\alpha_1 G_1 + \alpha_2 G_2) = \alpha_1 I(G_1) + \alpha_2 I(G_2). \tag{6.14}$$

Note que, fundindo as decomposições onde assentam G_1 e G_2, obtém uma decomposição onde ambas assentam e onde assenta a sua combinação linear $\alpha_1 G_1 + \alpha_2 G_2$.

Dada uma função em escada $G \in H_E^2[0,t]$, vem

$$\mathbb{E}[I(G)] = 0. \tag{6.15}$$

De facto, basta olhar para (6.13) e atender a que a esperança matemática de cada parcela da soma é, devido à independência entre G e os incrementos futuros do processo de Wiener, o produto das esperanças matemáticas dos dois factores, uma das quais é obviamente nula. Só há que ter um cuidado, que é o de garantir que $\mathbb{E}[G(t_{k-1})]$ existe e é finita, isto é, $\mathbb{E}[|G(t_{k-1})|] < +\infty$. De facto, por definição, para qualquer função $G \in H^2[0,T]$, seja ou não função em escada, tem-se $\int_0^t \mathbb{E}[|G|^2(s)]ds < +\infty$, o que implica que $\mathbb{E}[|G(s)|^2]$ (e, portanto, também $\mathbb{E}[|G(s)|]$) é finito para quase todos os valores de $s \in [0,t]$ (ou seja, o conjunto dos valores de s para os quais isso pode falhar tem medida de Lebesgue nula).

No espaço de probabilidade $(\Omega, \mathcal{F}, P)$, recordamos que designámos por L^2 o espaço de Hilbert das v.a. X (rigorosamente, o espaço das suas classes de equivalência para a relação de quase igualdade) com norma-L^2 $\|X\|_2 = (\mathbb{E}[|X|^2])^{1/2}$ finita. Para uma função em escada $G \in H_E^2[0,t]$, o

integral $I(G)$ pertence a L^2. De facto, tem-se que

$$(\|I(G)\|_2)^2 = \mathbb{E}[(I(G))^2] = VAR[I(G)] = \int_0^t \mathbb{E}[|G(s)|^2]ds = (\|G\|_{2*})^2 \tag{6.16}$$

é finita. Isto mostra que o integral $I(.)$ é uma aplicação linear de $H_E^2[0,t]$ em L^2 que preserva a norma (e, portanto, é contínua).

Incidentemente, (6.16) diz-nos que a variância do integral estocástico é um vulgar integral (determinístico) de Riemann do momento de segunda ordem da função integranda. Usando o facto de G ser independente dos incrementos futuros do processo de Wiener e de este ter incrementos independentes, facilmente se reconhece que (6.16) é verdadeira; com efeito,

$$\begin{array}{l}\mathbb{E}[(I(G))^2] = \sum_k \mathbb{E}\left[G^2(t_{k-1})\right] \mathbb{E}\left[(W(t_k) - W(t_{k-1}))^2\right] + \\ 2\sum_k \sum_{i<k} \mathbb{E}\left[G(t_{i-1})\right] \cdot \\ \cdot \mathbb{E}\left[G(t_{k-1})\ (W(t_i) - W(t_{i-1}))\ (W(t_k) - W(t_{k-1}))\right] = \\ \sum_k \mathbb{E}\left[G^2(t_{k-1})\right](t_k - t_{k-1}) + 0 = \int_0^t \mathbb{E}[G^2(s)]ds.\end{array}$$

A parte crítica da construção do integral de Itô para funções $G \in H^2[0,t]$ consiste em mostrar que existe uma *sucessão aproximadora* de funções em escada $G_n \in H_E^2[0,t]$ $(n = 1, 2, ...)$ convergente em norma L^{2*} para G, isto é, tal que

$$\int_0^t \mathbb{E}\left[|G_n(s) - G(s)|^2\right] ds \to 0 \text{ quando } n \to +\infty. \tag{6.17}$$

Deixamos a demonstração para a Secção 6.3. O leitor menos interessado nos detalhes técnicos pode sem prejuízo de maior saltar a Secção 6.3.

Com este resultado, conclui-se que $H_E^2[0,t]$ é denso em $H^2[0,t]$. A sucessão dos integrais $I(G_n) = \int_0^t G_n(s)dW(s)$ das funções em escada aproximadoras é uma sucessão de Cauchy no espaço L^2. Com efeito, G_n é convergente e, portanto, uma sucessão de Cauchy em $H^2[0,t]$. Como a aplicação $I(.)$ preserva a norma, $I(G_n)$ é uma sucessão de Cauchy em L^2. Como L^2 é completo, a sucessão $I(G_n)$ converge na norma de L^2, isto é, converge em média quadrática. Por definição, o *integral de Itô* de G,

$$I(G) = \int_0^t G(s)dW(s) := \text{l.i.m. } I(G_n), \tag{6.18}$$

é precisamente esse limite m.q.

Exercício: Mostre que esta definição do integral de Itô de funções $G \in$

$H^2[0,t]$ é consistente em dois sentidos:

a) Não depende da sucessão aproximadora de funções em escada, isto é, dadas duas sucessões aproximadoras G_n e G_n^* de funções em escada, vem l.i.m. $I(G_n)$ = l.i.m. $I(G_n^*)$ q.c. *Sugestão:* Combine as duas sucessões numa sucessão única com termos alternadamente de uma e de outra e use a unicidade do limite m.q.

b) Não se altera (a menos de uma quase igualdade) se substituirmos G por outra função quase igual $G^* \in H^2[0,t]$. Isto é, $I(G) = I(G^*)$ q.c. Vemos, pois, que, rigorosamente, o integral transforma classes de equivalência de funções G em classes de equivalência de v.a., como já tínhamos referido no caso das funções em escada.

Exercício: Utilizando as propriedades análogas já demonstrada para funções em escada, mostre que o integral de Itô para funções integrandas em $H^2[0,t]$ satisfaz as seguintes propriedades:

a) É uma aplicação linear, isto é, dadas funções $G_1, G_2 \in H^2[0,t]$ e constantes reais α_1 e α_2, vem

$$I(\alpha_1 G_1 + \alpha_2 G_2) = \alpha_1 I(G_1) + \alpha_2 I(G_2). \tag{6.19}$$

b) Esperança matemática nula:

$$\mathbb{E}[I(G)] = 0. \tag{6.20}$$

c) Preservação da norma:

$$(\|I(G)\|_2)^2 = \mathbb{E}[(I(G))^2] = VAR[I(G)] = \int_0^t \mathbb{E}[|G(s)|^2]ds = (\|G\|_{2*})^2 . \tag{6.21}$$

Vimos acima que, para funções em escada, o integral $I(.)$ era uma aplicação linear de $H_E^2[0,t]$ em L^2 que preservava a norma e, portanto, era contínua. Estendemos por continuidade a definição do integral a funções de $H^2[0,t]$. Vimos que o integral de Itô estendido a $H^2[0,t]$ continua a ser uma aplicação linear de $H^2[0,t]$ em L^2 e que preserva a norma, pelo que o integral estendido permanece uma *aplicação contínua.* Logo, dada uma sucessão G_n $(n = 1, 2, ...)$ de funções de $H^2[0,t]$ convergente em norma-L^{2*} para $G \in H^2[0,t]$, verifica-se que $I(G_n)$ converge na norma L^2 (isto é, converge m.q.) para $I(G)$.

Também se reconhece que o integral de Itô $I(G)$ é uma *função mensurável-$\mathcal{A}_t$*, pois isso é evidente para os integrais $I(G_n)$ de funções em escada G_n convergentes para G e $I(G)$ é o limite m.q. (que preserva a mensurabilidade) desses integrais.

Há dois casos particulares especialmente interessantes.

Um deles é o caso de uma *função $G \in H^2[0,t]$ contínua m.q.* [3] Neste caso, pode escolher-se qualquer sucessão de decomposições $0 = t_{n,0} < t_{n,1} < ... < t_{n,n-1} < t_{n,n} = t$ $(n = 1, 2, ...)$ de $[0,t]$ com diâmetros $\delta_n = \max_{k=1,...,n}(t_{n,k} - t_{n,k-1}) \to 0$ e "escadizar" G, isto é, escolher como sucessão de funções em escada aproximadoras a definida por $G_n(s) = G(t_{n,k-1})$ para todo o $s \in [t_{n,k-1}, t_{n,k})$. Veremos na Secção 6.3 que esta sucessão converge para G na norma L^{2*}

Usando a sucessão aproximadora acima definida, a definição (6.18) toma a forma

$$I(G) = \int_0^t G(s)dW(s) = \text{l.i.m.} \sum_{k=1}^{n} G(t_{n,k-1})\left(W(t_{n,k}) - W(t_{n,k-1})\right). \tag{6.22}$$

Esta é precisamente a definição (6.11) dada no final da Secção 6.1 para funções não-antecipativas contínuas m.q., que assim fica justificada e que nos dá uma expressão explícita para o cálculo dos integrais destas funções.

O outro caso particular interessante é o caso de uma *função G determinística* (isto é, não dependente de ω), caso em que podemos escolher para funções G_n funções determinísticas. É óbvio, de (6.13), que $I(G_n)$ tem distribuição normal com média nula e variância $\int_0^t G_n^2(s)ds$ (ver (6.16)). Logo o seu limite m.q. $I(G)$ tem também distribuição normal com média nula e variância $\lim \int_0^t G_n^2(s)ds = \int_0^t G^2(s)ds$, isto é

$$\int_0^t G(s)dW(s) \frown \mathcal{N}\left(0, \int_0^t G^2(s)ds\right) \text{ se G determinística.} \tag{6.23}$$

[3] Significa isto que, para qualquer $s \in [0,t]$, l.i.m. $G(s_n) = G(s)$ sempre que $s_n \to s$ quando $n \to +\infty$.

6.3 Existência de sucessões aproximadoras de funções em escada

Na Secção 6.2 procedeu-se à construção do integral de Itô para funções integrandas $G \in H^2[0,t]$. Começou-se por definir o integral para funções em escada $G \in H_E^2[0,t]$ e depois estendeu-se por continuidade a funções $G \in H^2[0,t]$. Note-se que $H_E^2[0,t]$ é um subespaço do espaço de Hilbert $H^2[0,t]$ (com a norma L^{2*} definida em (6.12)). Na extensão do integral foi crucial o facto de $H_E^2[0,t]$ ser *denso* em $H^2[0,t]$, isto é, de, dada uma função arbitrária $G \in H^2[0,t]$, existir pelo menos uma *sucessão aproximadora* de funções em escada $G_n \in H_E^2[0,t]$ $(n = 1,2,...)$ convergente na norma L^{2*} para G. A prova desse facto ficou pendente. Apresentamos aqui a sua demonstração (pode ver [3] ou [51], por exemplo, para demonstrações semelhantes).

Consideremos então uma função $G \in H^2[0,t]$.

Demonstração para funções G contínuas m.q.

Já vimos na Secção 6.2 como construir uma sucessão aproximadora G_n. Como vimos, basta escolher qualquer sucessão de decomposições $0 = t_{n,0} < t_{n,1} < ... < t_{n,n-1} < t_{n,n} = t$ $(n = 1,2,...)$ de $[0,t]$ com diâmetros $\delta_n = \max_{k=1,...,n}(t_{n,k} - t_{n,k-1}) \to 0$ e "escadizar" G, isto é, escolher como sucessão de funções em escada aproximadoras a definida por $G_n(s) = G(t_{n,k-1})$ para todo o $s \in [t_{n,k-1}, t_{n,k})$.

Dado $\varepsilon > 0$ arbitrário, como G é uniformemente contínua m.q. no intervalo fechado $[0,t]$, vem $\mathbb{E}[(G(t) - G(s))^2] < \varepsilon/t$ quando $|t-s| < \delta$, onde $\delta > 0$ é suficientemente pequeno e não dependente de s. Podemos escolher um n suficientemente grande e não dependente de s de modo a que $\delta_n < \delta$. Determinemos k tal que $t_{n,k-1} \leq s < t_{n,k}$ e escolhamos $t = t_{n,k-1}$. Vem $|t-s| < \delta$ e, portanto, $\mathbb{E}[(G_n(s) - G(s))^2] = \mathbb{E}[(G(t) - G(s))^2] < \varepsilon/t$, donde $\int_0^t \mathbb{E}[(G_n(s) - G(s))^2]ds < \varepsilon$. Isso mostra que a sucessão G_n acima referida converge para G na norma L^{2*}.

Demonstração para funções G limitadas

Seja $|G| \leq M$, onde M é uma constante finita. Para facilitar a notação, vamos trabalhar com a extensão de $G(s)$ ao intervalo de tempo $(-\infty, +\infty)$, que é igual a $G(s)$ para $s \in [0,t]$ e é igual a zero para $s \notin [0,t]$, extensão que continuamos a designar pela letra G e que é igualmente limitada por M.

Defina-se $H_n(s) = \int_0^{+\infty} e^{-\tau} G(s - \tau/n)d\tau$ (note que não é uma função em escada). Claro que é conjuntamente mensurável e é não-antecipativa (o seu valor no instante s só depende dos valores de G no instante s

ou instantes anteriores). Vem $H_n(s) \leq M \int_0^{+\infty} e^{-\tau} d\tau = M$, pelo que $\mathbb{E}\left[\int_0^t H_n^2(s)ds\right] \leq Mt < +\infty$ e, portanto, $H_n \in H^2[0,t]$.

Para $s \in [0,t]$, seja $h \to 0$. Vem $H_n(s+h) = \int_0^{+\infty} e^{-\tau} G(s+h-\tau/n)d\tau = e^{-nh} \int_{-nh}^{+\infty} e^{-\theta} G(s-\theta/n)d\theta$, pelo que, atendendo à conhecida desigualdade $(a+b)^2 \leq 2(a^2+b^2)$, vem, quando $h \to 0$,

$$
\begin{aligned}
&(H_n(s+h) - H_n(s))^2 = \\
&\left((e^{-nh}-1)\int_{-nh}^{+\infty} e^{-\tau} G(s-\tau/n)d\tau + \int_{-nh}^{0} e^{-\tau} G(s-\tau/n)d\tau\right)^2 \leq \\
&2\left((e^{-nh}-1)\int_{-nh}^{+\infty} e^{-\tau} G(s-\tau/n)d\tau\right)^2 + 2\left(\int_{-nh}^{0} e^{-\tau} G(s-\tau/n)d\tau\right)^2 \leq \\
&2M^2\left(\left((e^{-nh}-1)\int_{-nh}^{+\infty} e^{-\tau} d\tau\right)^2 + \left(\int_{-nh}^{0} e^{-\tau} d\tau\right)^2\right) = \\
&2M^2\left(e^{2nh}(e^{-nh}-1)^2 + (e^{nh}-1)^2\right) \to 0.
\end{aligned}
$$

Isto prova que *as funções H_n são contínuas em $[0,t]$*. Pelo teorema da convergência dominada, vem $\mathbb{E}\left[(H_n(s+h) - H_n(s))^2\right] \to 0$, pelo que *as funções H_n são contínuas m.q. em $[0,t]$*.

Também vem $H_n(s) \to G(s)$ para quase todos os $s \in [0,t]$ [4]. Logo, aplicando o teorema da convergência dominada (para a medida de Lebesgue de $[0,t]$), vem $\int_0^t (H_n(s) - G(s))^2 ds \to 0$ q.c. Aplicando de novo o teorema da convergência dominada (para a probabilidade P) vem $\mathbb{E}\left[\int_0^t (H_n(s) - G(s))^2 ds\right] \to 0$. *Isto mostra que existe uma sucessão $H_n \in H^2[0,t]$ de funções contínuas m.q. (e também contínuas) que converge para G na norma L^{2*}.*

Como cada H_n é contínua m.q., pode ser aproximada por uma sucessão de funções em escada $G_{n,m}$ $(m = 1, 2, ...)$. Dado n, seja $H_{k(n)}$ o primeiro termo da sucessão H_k $(k = 1, 2, ...)$ que esteja na vizinhança-$1/2^{n+1}$ de G (para a norma L^{2*}), tendo-se então $\|H_{k(n)} - G\|_{2*} < 1/2^{n+1}$. Seja $G_n := G_{k(n),m(k(n))}$ o primeiro termo da sucesão $G_{k(n),m}$ $(m = 1, 2, ...)$ que esteja na vizinhança-$1/2^{n+1}$ de $H_{k(n)}$, tendo-se então

[4] Como consequência do *teorema de Lusin* (ver [52], p. 56-57), como estamos a trabalhar com funções mensuráveis limitadas que são nulas fora do intervalo $[0,t]$ (que tem medida de Lebesgue finita), dado um número natural k, existe uma função contínua $J_k(s)$ coincidente com $G(s)$ excepto possivelmente num conjunto N_k com medida de Lebesgue inferior a $1/2^k$. Como $\sum_k 1/2^k < +\infty$, usando o lema de Borel-Cantelli (adaptado para medidas finitas), concluímos que é nula a medida de Lebesgue do conjunto dos s que estão em infinitos conjuntos N_k. Logo, para quase todo o s, $J_k(s) = G(s)$ a partir de certa ordem. Se a função G fosse contínua, seria óbvio que $H_n(s)$ convergiria para $G(s)$ para todo o s. Assim, só podemos garantir a convergência para quase todo o s.

$\|G_{k(n),m(k(n))} - H_{k(n)}\|_{2*} < 1/2^{n+1}$. Então, como $\|G_n - G\|_{2*} < 1/2^n$, *a sucessão* G_n $(n = 1, 2, ...)$ *é uma sucessão de funções em escada que converge para* G *na norma* L^{2*}, o que prova o pretendido.

Do facto de se ter $\mathbb{E}\left[\int_0^t (G_n(s) - G(s))^2 ds\right] < 1/2^n$ e de $\sum_n 1/2^n < +\infty$, concluímos até que $\int_0^t (G_n(s) - G(s))^2 ds \to 0$ q.c.

Demonstração para funções $G \in H^2[0,t]$ *arbitrárias*

Para cada $n = 1, 2, ...$ seja

$$F_n(s,\omega) = \begin{cases} -n & \text{se } G(s,\omega) < -n \\ G(s,\omega) & \text{se } |G(s,\omega)| \leq n \\ n & \text{se } G(s,\omega) > n. \end{cases}$$

São funções limitadas de $H^2[0,t]$ e é óbvio que $|F_n(s) - G(s)|^2 \to 0$ e que $|F_n(s) - G(s)|^2 \leq 2G^2(s)$ (onde G^2 é integrável $\lambda \times P$). Aplicando o teorema da convergência dominada duas vezes (com respeito à medida de Lebesgue λ e com respeito a P), obtemos $\mathbb{E}\left[\int_0^t (F_n(s) - G(s))^2 ds\right] \to 0$, o que mostra que F_n converge na norma L^{2*} para G. Cada F_n pode ser aproximada por uma sucessão de funções em escada. Então, por raciocínio semelhante ao usado na parte final do caso anterior, existe uma sucessão de funções em escada G_n que converge na norma L^{2*} para G, o que conclui a demonstração de que $H_E^2[0,t]$ é denso em $H^2[0,t]$. Vem até que $\int_0^t (G_n(s) - G(s))^2 ds \to 0$ q.c.

6.4 Estudo do integral como função do limite superior de integração

Seja $G \in H^2[0,d]$ e seja $[a,b] \subset [0,d]$. Como $G(s)I_{[a,b]}(s) \in H^2[0,d]$, pode-se definir o integral de Itô em $[a,b]$ por

$$\int_a^b G(s)dW(s) := \int_0^d G(s)I_{[a,b]}(s)dW(s). \tag{6.24}$$

Facilmente se reconhece que, para $0 \leq a \leq b \leq c \leq d$, vem

$$\int_a^c G(s)dW(s) = \int_a^b G(s)dW(s) + \int_b^c G(s)dW(s). \tag{6.25}$$

Para estudar o *integral de Itô como função do limite superior de*

integração t, vamos supor t a variar no intervalo $[0,d]$. Claro que, sendo $G \in H^2[0,d]$, também se tem $G \in H^2[0,t]$ para $t \in [0,d]$ (usando a mesma letra G para designar a restrição de $G = G(s,\omega)$ a $[0,t] \times \Omega$). Seja

$$Z(t) = \int_0^t G(s)dW(s) = \int_0^d G(s)I_{[0,t]}(s)dW(s). \qquad (6.26)$$

Propriedade: *Como $Z(t)$ é mensurável-$\mathcal{A}_t$, o processo estocástico $\{Z(t)\}_{[0,d]}$ está adaptado à filtração $\{\mathcal{A}_t\}_{t\in[0,d]}$. Além disso, é uma martingala-$\mathcal{A}_t$.*

Para o demonstrar, consideremos uma sucessão $G_n \in H_E^2[0,d]$ $(n = 1,2,...)$ de funções em escada aproximadora de G. Sejam $0 = t_0^{(n)} < t_1^{(n)} < ... < t_{m_n}^{(n)} = d$ os pontos da decomposição em cujos subintervalos $[t_{k-1}^{(n)}, t_k^{(n)})$ $(k = 1,...,m_n)$ a função $G_n(s)$ é constante (relativamente a s). Vamos ver que cada $Z_n(t) = \int_0^t G_n(s)dW(s)$ é uma martingala (abreviatura de "martingala-$\mathcal{A}_t$"). De facto, dado $s < t$ e sendo $N_n = \{k \in \{1,2,...,m_n\} : s \le t_{k-1}^{(n)}, t_k^{(n)} \le t\}$, vem

$$\begin{aligned}
&\mathbb{E}[Z_n(t)\,|\mathcal{A}_s] = \mathbb{E}\left[\textstyle\int_0^s G_n(s)dW(s) + \int_s^t G_n(s)dW(s)\,|\mathcal{A}_s\right] = \\
&\textstyle\int_0^s G_n(s)dW(s) + \sum_{k\in N_n} \mathbb{E}\left[G_n(t_{k-1}^{(n)})(W(t_k^{(n)}) - W(t_{k-1}^{(n)}))\,|\mathcal{A}_s\right] = \\
&\textstyle\int_0^s G_n(s)dW(s) + \\
&\quad \textstyle\sum_{k\in N_n} \mathbb{E}\left[\mathbb{E}\left[G_n(t_{k-1}^{(n)})(W(t_k^{(n)}) - W(t_{k-1}^{(n)}))\,\middle|\mathcal{A}_{t_{k-1}^{(n)}}\right]|\mathcal{A}_s\right] = \\
&\textstyle\int_0^s G_n(s)dW(s) + \\
&\quad \textstyle\sum_{k\in N_n} \mathbb{E}\left[G_n(t_{k-1}^{(n)})\mathbb{E}\left[W(t_k^{(n)}) - W(t_{k-1}^{(n)})\,\middle|\mathcal{A}_{t_{k-1}^{(n)}}\right]|\mathcal{A}_s\right] = \\
&\textstyle\int_0^s G_n(s)dW(s) + \sum_{k\in N_n} \mathbb{E}\,[0\,|\mathcal{A}_s] = Z_n(s).
\end{aligned}$$

Daqui resulta, pela *desigualdade de Schwarz* [5] e por $Z_n(s)$ convergir m.q. para $Z(s)$, que

$$\begin{aligned}
&\mathbb{E}\left[(\mathbb{E}[Z_n(t)|\mathcal{A}_s] - \mathbb{E}[Z(t)|\mathcal{A}_s])^2\right] = \mathbb{E}\left[(\mathbb{E}[Z_n(t) - Z(t)|\mathcal{A}_s])^2\right] \le \\
&\mathbb{E}\left[\mathbb{E}[1^2|\mathcal{A}_s]\mathbb{E}[(Z_n(t) - Z(t))^2|\mathcal{A}_s]\right] = \mathbb{E}\left[(Z_n(t) - Z(t))^2\right] \to 0.
\end{aligned}$$

Logo $\mathbb{E}[Z_n(t)|\mathcal{A}_s] = Z_n(s)$ converge m.q. para $\mathbb{E}[Z(t)|\mathcal{A}_s]$. Como o limite

[5] A *desigualdade de Schwarz* diz que, para uma medida μ e funções f, g de quadrado integrável, $(\int fg d\mu)^2 \le (\int f d\mu)^2 (\int g d\mu)^2$. Em particular, se μ é uma probabilidade P, vem $(\mathbb{E}[XY])^2 = \mathbb{E}[X^2]\mathbb{E}[Y^2]$.

m.q. é q.c. único, vem $\mathbb{E}[Z(t)|\mathcal{A}_s] = Z(s)$ q.c., provando-se assim que $Z(t)$ é uma martingala-$\mathcal{A}_t$.

Então, dados $a, b \in [0, d]$ com $a \leq b$, $Z(t) - Z(a)$ é, para $t > a$, uma martingala-$\mathcal{A}_t$. Aplicando as desigualdades maximais das martingalas (2.2) e (2.3), vem

$$P\left[\sup_{a\leq t\leq b} |Z(t) - Z(a)| \geq c\right] \leq \frac{1}{c^2}\mathbb{E}[|Z(b) - Z(a)|^2] = \frac{1}{c^2}\int_a^b \mathbb{E}[G^2(s)]ds \tag{6.27}$$

e

$$\mathbb{E}\left[\sup_{a\leq t\leq b} |Z(t) - Z(a)|^2\right] \leq 4\mathbb{E}[|Z(b) - Z(a)|^2] = 4\int_a^b \mathbb{E}[G^2(s)]ds. \tag{6.28}$$

Propriedade: O *processo $Z(t)$ tem uma versão contínua (trajectórias q.c. contínuas) e supomos que trabalhamos sempre com uma tal versão.*

Vamos demonstrá-lo. É fácil ver de (6.13) que $Z_n(t) = \int_0^t G_n(s)dW(s)$ é função q.c. contínua de t em $[0, d]$. Como $Z_n(t) - Z_m(t)$ é martingala, de (2.2) vem

$$P\left[\sup_{0\leq t\leq d} |Z_n(t) - Z_m(t)| > \varepsilon\right] \leq \frac{1}{\varepsilon^2}\mathbb{E}\left[|Z_n(d) - Z_m(d)|^2\right].$$

Quando $n, m \to +\infty$, o lado direito tende para zero. Escolha-se $\varepsilon = 1/2^k$. Escolha-se uma subsucessão $n_k \to +\infty$ tal que

$$P\left[\sup_{0\leq t\leq d} |Z_{n_{k+1}}(t) - Z_{n_k}(t)| > 1/2^k\right] < 1/2^k.$$

Como $\sum_k 1/2^k < +\infty$, o *lema de Borel-Cantelli* [6] implica

$$P\left[\sup_{0\leq t\leq d} |Z_{n_{k+1}}(t) - Z_{n_k}(t)| > 1/2^k \text{ para infinitos valores de } k\right] = 0.$$

Portanto, com probabilidade um, vem $\sup_{0\leq t\leq d} |Z_{n_{k+1}}(t) - Z_{n_k}(t)| \leq 1/2^k$ a partir de certa ordem. Isso implica que $Z_{n_k}(t)$ é q.c. uniforme-

[6] O *lema de Borel-Cantelli* (ver qualquer bom livro de probabilidades) aplica-se a uma sucessão A_n $(n = 1, 2, ...)$ de acontecimentos e ao acontecimento $A = \{\omega : \omega \in A_n$ para infinitos valores de $n\}$. Se $\sum_n P(A_n) < +\infty$, então $P(A) = 0$. Se $\sum_n P(A_n) = +\infty$ e se os acontecimentos A_n forem independentes, então $P(A) = 1$.

mente convergente no intervalo $[0,d]$. Designemos por $J(t)$ o seu limite q.c. quando $k \to +\infty$, que, sendo limite uniforme de funções q.c. contínuas, é uma função q.c. contínua. Como $Z_{n_k}(t)$ converge m.q. para $Z(t)$, vem $Z(t) = J(t)$ q.c., o que prova o pretendido.

Exercício: Mostre que $Z(t)$ tem *incrementos não-correlacionados* (não confundir com independentes), isto é, para intervalos não-sobrepostos $[s,t]$ e $[u,v]$ com $0 \leq s \leq t \leq u \leq v \leq d$,

$$COV[Z(t) - Z(s), Z(v) - Z(u)] = E[(Z(t) - Z(s))(Z(v) - Z(u))] = E\left[\int_s^t G(z)dW(z) \int_u^v G(z)dW(z)\right] = 0. \tag{6.29}$$

Basta demonstrar para funções em escada e transferir a propriedade para o caso geral por passagem ao limite m.q.

6.5 Extensões do integral de Itô

Esta secção é dispensável numa primeira leitura.

O integral de Itô pode ser generalizado à classe $M^2[0,t]$ das funções conjuntamente mensuráveis não-antecipativas-$\mathcal{A}_s$ tais que

$$\int_0^t |G(s)|^2 ds < +\infty \text{ q.c.} \tag{6.30}$$

Como esta condição é mais fraca que a condição $\int_0^t E\left[|G(s)|^2\right] ds < +\infty$, vem $M^2[0,t] \supset H^2[0,t]$.

Para funções em escada de $M^2[0,t]$, a definição do integral de Itô é igualmente dada por (6.13), mas agora nada garante que existam as esperanças matemáticas de G, pelo que pode não estar definida nem a esperança matemática nem a variância do integral, situação que permanece quando se definir o integral para funções arbitrárias de $M^2[0,t]$.

Para definir o integral para uma função arbitrária $G \in M^2[0,t]$, usa-se a mesma técnica de mostrar que ela pode ser aproximada por uma *sucessão aproximadora* de funções $G_n \in M^2[0,t]$ $(n = 1,2,...)$ em escada. Agora a convergência não é obviamente a definida por L^{2*}, mas antes uma convergência mais fraca definida por

$$\int_0^t (G_n(s) - G(s))^2 ds \to 0 \text{ q.c. quando } n \to +\infty. \tag{6.31}$$

A técnica para mostrar a existência desta sucessão aproximadora é semelhante à usada na Secção 6.3. Começa-se por mostar que é ver-

dade para funções $G \in M^2[0,t]$ que sejam contínuas, usando o método da "escadização" já usado na Secção 6.3. É um exercício simples, que deixamos ao cuidado do leitor adaptar a demonstração aí feita (substituindo a continuidade m.q. pela continuidade e retirando as esperanças matemáticas). Estende-se em seguida a funções $G \in M^2[0,t]$ limitadas, mostrando-se que uma tal função pode ser aproximada por uma sucessão de funções $H_n \in M^2[0,t]$ contínuas. A demonstração é semelhante à usada na Secção 6.3, embora mais simples, ficando ao cuidado do leitor fazer a adaptação daquela demonstração. A extensão a funções $G \in M^2[0,t]$ arbitrárias usando uma sucessão aproximadora de funções limitadas usa também a mesma técnica da Secção 6.3 (embora só seja preciso usar o teorema da convergência dominada uma vez e não duas).

Como a aproximação a $G \in M^2[0,t]$ por uma sucessão aproximadora G_n de funções em escada usa um modo de convergência mais fraco, já não é possível mostrar que a sucessão dos integrais $I(G_n) = \int_0^t G_n(s)dW(s)$ converge em média quadrática, mas vamos provar que *converge em probabilidade.* [7]

Primeiro vamos demonstrar, seguindo [3], que, para uma função $J \in M^2[0,t]$ em escada, se tem, para todo o $N > 0$ e todo o $\delta > 0$,

$$P\left[\left|\int_0^t J(s)dW(s)\right| > \delta\right] \leq \frac{N}{\delta^2} + P\left[\int_0^t J^2(s)dW(s) > N\right] \qquad (6.32)$$

Seja $0 = t_0 < t_1 < ... < t_n = t$ a decomposição do intervalo $[0,t]$ onde J assenta, isto é $J(s,\omega) = J(t_{k-1},\omega)$ para todo o $s \in [t_{k-1},t_k)$ $(k = 1,...,n)$. Para $s \in [t_{i-1},t_i)$, defina

$$J_N(s) := \begin{cases} J(s) & \text{se } \int_0^{t_i} J^2(s)ds \leq N \\ 0 & \text{se } \int_0^{t_i} J^2(s)ds > N, \end{cases}$$

que é uma função em escada de $M^2[0,t]$ tal que $\int_0^{t_i} J_N^2(s)ds \leq N$. Logo

$$\|J_N\|_{2*}^2 = \mathbb{E}\left[\int_0^{t_i} J_N^2(s)ds\right] \leq N$$

[7] Uma sucessão de v.a. X_n $(n = 1,2,...)$ *converge em probabilidade* para a v.a. X se, para $\varepsilon > 0$ arbitrário, $P[|X_n - X| > \varepsilon] \to 0$ quando $n \to +\infty$. Esta convergência é mais fraca que a convergência m.q., no sentido de que, sempre que se verifica a convergência m.q., também se verifica a convergência em probabilidade mas o inverso pode falhar. Também é mais fraca que a convergência com probabilidade um ou quase certa.

e, portanto, $J_N \in H_E^2[0,t]$. Logo, pela propriedade de preservação da norma, $\mathbb{E}[(I(J_N))^2] = \|I(J_N)\|_2^2 \leq N$. Como $J \equiv J_N$ sse $\int_0^t J^2(s)ds > N$, vem

$$P\left[\int_0^t J^2(s)ds > N\right] = P\left[\sup_{0\leq s\leq t}|J(s) - J_N(s)| > 0\right] \geq$$
$$P\left[\left|\int_0^t (J(s) - J_N(s))dW(s)\right| > 0\right].$$

Então, usando a *desigualdade de Tchebyshev*,

$$P[|I(J)| > \delta] \leq P[|I(J_N)| > \delta] + P[|I(J) - I(J_N)| > 0] \leq$$
$$\mathbb{E}[(I(J_N))^2]/\delta^2 + P\left[\int_0^t J^2(s)ds > N\right] \leq N/\delta^2 + P\left[\int_0^t J^2(s)ds > N\right],$$

o que mostra que (6.32) é verdadeira.

Demonstração de que $I(G_n)$ converge em probabilidade quando $n \to +\infty$

Como $\int_0^t (G_n(s) - G(s))^2 ds \to 0$ q.c., vem $\int_0^t (G_n(s) - G(s))^2 ds \to 0$ em probabilidade, donde $\int_0^t (G_n(s) - G_m(s))^2 ds \leq 2\int_0^t (G_n(s) - G(s))^2 ds + 2\int_0^t (G_m(s) - G(s))^2 ds \to 0$ em probabilidade quando $n, m \to +\infty$. Seja $\delta > 0$ arbitrário. Aplicando (6.32) a $G_n - G_m$, vem

$$\limsup_{n,m\to+\infty} P[|I(G_n) - I(G_m)| > \varepsilon] \leq$$
$$\delta/\varepsilon^2 + \limsup_{n,m\to+\infty} P\left[\int_0^t |G_n(s) - G_m(s)|^2 > \delta\right] \leq \delta/\varepsilon^2 + 0.$$

Como δ é arbitrário, conclui-se que $\lim_{n,m\to+\infty} P[|I(G_n) - I(G_m)| > \varepsilon] = 0$ e, portanto, $I(G_n)$ é uma sucessão de Cauchy em probabilidade, pelo que converge em probabilidade.

Tem-se então uma sucessão de integrais $I(G_n)$ de funções em escada que converge em probabilidade. O seu limite em probabilidade diz-se o integral de Itô

$$I(G) = \int_0^t G(s)dW(s) = \int_0^t G(s,\omega)dW(s,\omega)$$

da função G. Este limite é insensível à escolha da sucessão aproximante, o que deixamos como exercício para o leitor.

Fica também como exercício mostrar a linearidade do integral. O leitor pode facilmente provar que (6.32) permanece válida para qualquer $J \in M^2[0,t]$, mesmo que não seja uma função em escada. A partir daí, pode também provar que, se $G_n \in M^2[0,t]$ $(n = 1, 2, ...)$ é uma

sucessão (mesmo que estas funções não sejam em escada) convergente para $G \in M^2[0,t]$ no sentido de $\int_0^t (G_n(s) - G(s))^2 ds \to 0$ q.c., então $\int_0^t G_n(s)dW(s)$ converge em probabilidade para $\int_0^t G(s)dW(s)$.

As propriedades (6.20) e (6.21) não podem agora ser garantidas por poderem não existir os respectivos momentos. Pela mesma razão, também não podemos garantir que o integral, como função do limite superior de integração, seja uma martingala. Claro que $Z(t) = \int_0^t G(s)dW(s)$ ($t \in [0,d]$), com $G \in M^2[0,t]$, está adaptado à filtração não-antecipativa $\mathcal{A}_t$. Pode mesmo provar-se que as trajectórias são q.c. contínuas.

Com efeito, para $G \in M^2[0,d]$, considerem-se funções truncadas

$$G_N(t) := \begin{cases} G(t) & \text{se } \int_0^t G^2(u)du \leq N \\ 0 & \text{se } \int_0^t G^2(u)du > N \end{cases},$$

que estão em $H^2[0,d]$, e seja $Z_N(t) = \int_0^t G_N(t)dW(t)$. Logo, as funções $Z_N(t)$ têm, como sabemos, trajectórias q.c. contínuas. Ora $Z(t,\omega) = Z_N(t,\omega)$ para todo o $t \in [0,d]$ e para $\omega \in A_N := \{\omega : \int_0^t G^2(s,\omega)ds) \leq N\}$. Escolhendo N suficientemente grande podemos tornar $P(A_N)$ tão próximo de 1 quanto se queira, pelo que o conjunto dos pontos ω para os quais $Z(t,\omega)$ é contínua tem probabilidade 1.

Uma outra extensão óbvia são os integrais estocásticos, também chamados *processos de Itô*, definidos para $t \in [0,d]$ por

$$X(t,\omega) = X_0(\omega) + \int_0^t F(s,\omega)ds + \int_0^t G(s,\omega)dW(s,\omega), \qquad (6.33)$$

onde:

- X_0 é uma v.a. mensurável-$\mathcal{A}_0$ e, portanto, independente do processo de Wiener. Pode, em particular, ser uma constante determinística.
- F é um função conjuntamente mensurável adaptada à filtração $\mathcal{A}_s$ e tal que $\int_0^d |F(s)|ds < +\infty$ q.c.
- $G \in M^2[0,d]$.

O integral estocástico pode ser escrito na forma diferencial (usa-se a habitual abreviatura de não indicar a dependência de ω)

$$dX(t) = F(t)dt + G(t)dW(t). \qquad (6.34)$$

6.6 Teorema e fórmula de Itô

Vimos na Secção 6.1, através de um exemplo, que os integrais de Itô, e consequentemente os processos de Itô (que podem ser escritos em forma diferencial ou integral), não satisfazem as regras usuais de cálculo. Por isso, necessitamos de um novo cálculo diferencial e integral, o *cálculo de Itô*. Basta para isso darmos a *regra de diferenciação de uma função composta* ou *regra da cadeia*. Trata-se da conhecida *fórmula de Itô*, veiculada pelo *teorema de Itô*.

Teorema de Itô. *Seja $X(t) = X(t,\omega)$ ($t \in [0,d]$) um processo de Itô dado por (6.33) ou (6.34), abreviadamente $dX(t) = F(t)dt + G(t)dW(t)$. Seja $Y(t) = h(t, X(t))$, onde $h(t,x)$ é uma função contínua com derivada parcial contínua com respeito a t e com segunda derivada parcial contínua com respeito a x. Então $Y(t) = Y(t,\omega)$ é também um processo de Itô com condição inicial $Y_0 = h(0, X_0)$ dado na forma diferencial pela fórmula de Itô*

$$dY(t) = \left(\frac{\partial h(t,X(t))}{\partial t} + \frac{\partial h(t,X(t))}{\partial x}F(t) + \frac{1}{2}\frac{\partial^2 h(t,X(t))}{\partial x^2}G^2(t)\right)dt + \frac{\partial h(t,X(t))}{\partial x}G(t)dW(t). \tag{6.35}$$

A forma integral será então

$$Y(t) = Y_0 + \int_0^t \left(\frac{\partial h(s,X(s))}{\partial s} + \frac{\partial h(s,X(s))}{\partial x}F(s) + \frac{1}{2}\frac{\partial^2 h(t,X(s))}{\partial x^2}G^2(s)\right)ds + \int_0^t \frac{\partial h(s,X(s)}{\partial x}G(s)dW(s). \tag{6.36}$$

Uma boa mnemónica para obter (6.36) é usar um desenvolvimento de Taylor até à primeira ordem em t e até à segunda ordem em x

$$dY(t) = \frac{\partial h(t,X(t))}{\partial t}dt + \frac{\partial h(t,X(t))}{\partial x}dX(t) + \frac{1}{2}\frac{\partial^2 h(t,X(t))}{\partial x^2}(dX(t))^2$$

seguido da substituição de $dX(t)$ por $F(t)dt + G(t)dW(t)$. Nesta substituição, ao calcular $(dX(t))^2$, devemos usar as regras de multiplicação

$\times$	dt	dW
dt	0	0
dW	0	dt

ou seja, $(dX(t))^2 = G^2(t)dt$.

Vamos agora apresentar um esboço da demosnstração do teorema de Itô, a qual o leitor pode perfeitamente dispensar numa primeira leitura.

Esboço da demonstração do teorema de Itô

Mais detalhes podem ser vistos, por exemplo, em [3] (que seguimos de perto) ou [51]. Sejam F e G satisfazendo as condições para que o processo $X(t)$ seja um processo de Itô, designadamente serem conjuntamente mensuráveis e adaptadas à filtração, $\int_0^d |F(s,\omega)|ds < +\infty$ q.c. e $\int_0^d |G(s,\omega)|^2 ds < +\infty$ q.c. Em relação a G, sabemos, pela Secção 6.5, que existe uma sucessão de funções em escada G_n $(n = 1, 2, ...)$ com as mesmas propriedades que converge para G no sentido $\int_0^t (G(s) - G_n(s))^2 ds \to 0$ q.c. e tal que $\int_0^t G_n(s)dW(s)$ converge em probabilidade para $\int_0^t G(s)dW(s)$. Também não seria difícil adaptar o raciocínio de modo a provar que existe uma sucessão de funções em escada F_n $(n = 1, 2, ...)$ (com as mesmas propriedades de F) que converge para F no sentido $\int_0^t |F(s) - F_n(s)|ds \to 0$ q.c. Daqui resulta imediatamente que $\int_0^t F_n(s)ds$ converge para $\int_0^t F(s)ds$ q.c. (e logo em probabilidade). Então, se conseguirmos provar a fórmula de Itô para funções em escada, ela vale em geral, bastando fazer a passagem ao limite (em probabilidade) ao longo das sucessões aproximadoras de funções em escada. Como os limites em probabilidade são q.c. únicos, a fórmula de Itô verifica-se q.c.

Para funções em escada, podemos partir o intervalo de integração de acordo com a decomposição em que assentam F e G (pode-se fazer uma comum a ambas as funções) e provar a fórmula para cada um dos sub-intervalos, onde as funções são constantes (com respeito ao tempo, não com respeito a ω). Então, basta demonstrar a fórmula de Itô para funções constantes num intervalo $[a, b] \subset [0, d]$.

Sejam então $F(s,\omega) \equiv F(\omega)$ (uma v.a. que abreviamos para F) e $G(s,\omega) \equiv G(\omega)$ (uma v.a. que abreviamos para G) funções constantes (relativamente ao tempo) em $[a, b]$. Nesse caso, para $t \in [a, b]$, temos

$$X(t) = X(a) + (t - a)F + (W(t) - W(a))G,$$

com $X(a)$, F e G mensuráveis-$\mathcal{A}_a$, logo independentes dos incrementos $W(t) - W(a)$.

Vem

$$Y(t) = h(t, X(t)) = h(t, X(a) + (t - a)F + (W(t) - W(a))G)$$

com condição inicial $Y_a = h(a, X(a))$. Para cada n natural, vamos considerar uma decomposição $a = t_{n,0} < t_{n,1} < ... < t_{n,n} = b$ do intervalo $[a, b]$ com diâmetro $\delta_n = \max_{1 \le k \le n}(t_{n,k} - t_{n,k-1})$ tal que $\delta_n \to 0$ quando $n \to +\infty$. Vamos abreviar a notação e escrever $t_k = t_{n,k}$, $X_k = X(t_{n,k})$, $\Delta t_k = t_{n,k} - t_{n,k-1}$, $\Delta X_k = X(t_{n,k}) - X(t_{n,k-1})$ e

$\Delta W_k = W(t_{n,k}) - W(t_{n,k-1})$. Vem

$$Y(t) - Y(a) = \sum_{k=1}^{n} (h(t_k, X_k) - h(t_{k-1}, X_{k-1}))$$

e, usando a fórmula de Taylor , vem

$$h(t_k, X_k) - h(t_{k-1}, X_{k-1}) = h_t(t_{k-1} + \theta_{n,k}\Delta t_k, X_{k-1})\Delta t_k + h_x(t_{k-1}, X_{k-1})\Delta X_k + \tfrac{1}{2} h_{xx}(t_{k-1}, X_{k-1} + \bar{\theta}_{n,k}\Delta X_k)(\Delta X_k)^2,$$

onde $h_t = \partial h/\partial t$, $h_x = \partial/\partial x$, $h_{xx} = \partial^2/\partial x^2$ e onde $0 < \theta_{n,k}, \bar{\theta}_{n,k} < 1$. Isto deve-se à continuidade das derivadas parciais utilizadas. Como o intervalo é fechado, a continuidade é uniforme (no tempo) pelo que há majorantes

$$\max_{1\le k\le n} |h_t(t_{k-1} + \theta_{n,k}\Delta t_k, X_{k-1}) - h_t(t_{k-1}, X_{k-1})| \le \alpha_n$$

$$\max_{1\le k\le n} \left|h_{xx}(t_{k-1}, X_{k-1} + \bar{\theta}_{n,k}\Delta X_k) - h_{xx}(t_{k-1}, X_{k-1})\right| \le \beta_n,$$

com $\alpha_n, \beta_n \to 0$ q.c., pelo que os erros de substituir $\theta_{n,k}$ e $\bar{\theta}_{n,k}$ por 0 são controlados por esses majorantes. Vem então

$$Y(t) - Y(a) = \sum_{k=1}^{n} (h(t_k, X_k) - h(t_{k-1}, X_{k-1})) = \sum_{k=1}^{n} h_t(t_{k-1}, X_{k-1})\Delta t_k + \sum_{k=1}^{n} h_x(t_{k-1}, X_{k-1})\Delta X_k + \frac{1}{2}\sum_{k=1}^{n} h_{xx}(t_{k-1}, X_{k-1})(\Delta X_k)^2 + erro,$$

com $|erro| \le \alpha_n(b-a) + \frac{1}{2}\beta_n \sum_{k=1}^{n}(\Delta X_k)^2$ convergente para zero em probabilidade pois $\sum_{k=1}^{n}(\Delta X_k)^2$ converge em probabilidade para $(b-a)G^2$ (que é q.c. finito).

Falta então apenas provar que

$$\sum_{k=1}^{n}(\Delta X_k)^2 \to (b-a)G^2 \text{ em probabilidade} \tag{6.37}$$

e que

$$\begin{array}{l} \sum_{k=1}^{n} h_t(t_{k-1}, X_{k-1})\Delta t_k + \\ \sum_{k=1}^{n} h_x(t_{k-1}, X_{k-1})(\Delta X_k) + \\ \frac{1}{2}\sum_{k=1}^{n} h_{xx}(t_{k-1}, X_{k-1})(\Delta X_k)^2 \end{array} \tag{6.38}$$

converge em probabildade para

$$\begin{array}{l}\int_a^b h_t(s,X(s))ds+\\ \int_a^b h_x(s,X(s))Fds+\int_a^b h_x(s,X(s))GdW(s)+\\ \frac{1}{2}\int_a^b h_{xx}(s,X(s))G^2ds,\end{array} \tag{6.39}$$

havendo convergência de cada linha de (6.38) para a linha correspondente de (6.39). O resultado para a primeira linha é evidente. A segunda linha também é evidente se notarmos que $\Delta X_k = F\Delta t_k + G\Delta W_k$. A terceira linha vai dar mais trabalho.

Note-se que

$$\begin{array}{l}\sum_{k=1}^n h_{xx}(t_{k-1},X_{k-1})(\Delta X_k)^2 =\\ F^2\sum_{k=1}^n h_{xx}(t_{k-1},X_{k-1})(\Delta t_k)^2+\\ 2FG\sum_{k=1}^n h_{xx}(t_{k-1},X_{k-1})\Delta t_k\Delta W_k+\\ G^2\sum_{k=1}^n h_{xx}(t_{k-1},X_{k-1})(\Delta W_k)^2;\end{array}$$

os dois primeiros termos do lado direito da igualdade convergem em probabilidade para zero devido à continuidade de h_{xx} e de W. Falta só provar (repare que aqui G é constante) que

$$\sum_{k=1}^n h_{xx}(t_{k-1},X_{k-1})(\Delta W_k)^2 \to \int_a^b h_{xx}(s,X(s))ds \quad \text{em probabilidade.} \tag{6.40}$$

De facto, a outra demonstração que ainda faltava, a da validade de (6.37), aparece como caso particular desta quando $h(s,x)\equiv 1$. Seja

$$H_{nk} := (W_k - W_{k-1})^2 - (t_k - t_{k-1}) = (\Delta W_k)^2 - \Delta t_k.$$

Como $\sum_{k=1}^n h_{xx}(t_{k-1},X_{k-1})\Delta t_k \to \int_a^b h_{xx}(s,X(s))ds$ em probabilidade, basta mostrar que

$$S_n := \sum_{k=1}^n h_{xx}(t_{k-1},X_{k-1})H_{nk} \to 0 \quad \text{em probabilidade.} \tag{6.41}$$

Pode suceder que S_n não tenha momentos mas já terá momentos truncados para a truncatura

$$S_n^{(N)} := \sum_{k=1}^n h_{xx}(t_{k-1},X_{k-1})H_{nk}I_{N,k}(\omega)$$

com $N > 0$ e

$$I_{N,k}(\omega) := \begin{cases} 1 & \text{se } |X(t_{n,i},\omega)| \leq N \text{ para todos os } i < k \\ 0 & \text{caso contrário.} \end{cases}$$

Como $\mathbb{E}[H_{nk}] = 0$, $\mathbb{E}[H_{nk}^2] = 2(\Delta t_k)^2$ e os H_{nk} $(k = 1, 2, ..., n)$ são independentes, vem $\mathbb{E}[S_n^{(N)}] = 0$ e

$$\begin{aligned} &\mathbb{E}[(S_n^{(N)})^2] = \textstyle\sum_{k=1}^n \mathbb{E}\left[(h_{xx}(t_{k-1}, X_{k-1})I_{N,k})^2\right] \mathbb{E}\left[H_{nk}^2\right] \\ &\leq 2 \max_{a\leq t\leq b, |x|\leq N} |h_{xx}(t,x)| (\textstyle\sum_{k=1}^n (\Delta t_k)^2 \to 0 \end{aligned}$$

quando $\delta_n \to 0$. Para todo o $N > 0$ fixo, vem $S_n^{(N)}$ a convergir m.q. e, portanto, em probabilidade, para zero. O erro de truncatura é $P[S_m \neq S_n^{(N)}] = P[\max_{a\leq t\leq b} |X(t)| > N]$. Ora $\max_{a\leq t\leq b} |X(t)| = \max_{a\leq t\leq b} |X(a) + (t-a)F + (W(t) - W(a))G| \leq |X(a)| + (b-a)|F| + |G| \max_{a\leq t\leq b} |W(t) - W(a)|$ é uma v.a q.c. finita, pelo que $P[S_n \neq S_n^{(N)}]$ pode tornar-se tão pequeno quanto se queira para N suficientemente grande. Como $P[|S_n| > \varepsilon] = P[|S_n^{(N)}| > \varepsilon] + P[S_m \neq S_n^{(N)}]$, reconhecemos que $P[|S_n| > \varepsilon] \to 0$ e, portanto, S_n converge em probabilidade para zero como pretendido.

Exercício: Determine $d(tW(t))$ e use o resultado para mostrar que $\int_0^t s dW(s) = tW(t) - \int_0^t W(s)ds$.

Exercício: Mostre que a equação $dY(t) = Y(t)dW(t)$ com $Y(0) = 1$ tem solução $Y(t) = \exp\left(W(t) - \frac{t}{2}\right)$ para $t \geq 0$.

6.7 Os cálculos de Itô e Stratonovich

Vamos dar um exemplo de aplicação do teorema de Itô para efeitos de ilustração das regras especiais de cálculo deste integral.

Seja $X(t) = W(t)$ e $h(t,x) = x^2$. Como $X(0) = 0$ e $X(t) = W(t) = \int_0^t 0ds + \int_0^t 1dW(s)$, temos $F(t,\omega) \equiv 0$ e $G(t,\omega) \equiv 1$ e que $X(t)$ é um processo de Itô. Usámos o facto de a função integranda do segundo integral ser uma função em escada com um único degrau (onde toma o valor 1 para todo o ω) correspondente à decomposição $0 = t_0 < t_1 = t$ do intervalo $[0,t]$, pelo que $\int_0^t 1dW(s) = 1(W(t_1) - W(t_0)) = W(t)$. Na forma diferencial vem $dX(t) = 0dt + 1dW(t) = dW(t)$. Vem $Y(t) = h(t, X(t)) = W^2(t)$. Pela fórmula de Itô, obtemos (pode usar a mnemónica) $dY(t) = 0dt + 2X(t)dX(t) + \frac{1}{2}2(dX(t))^2 = 2W(t)dW(t) + dt$. Na form integral, atendendo a que $Y(0) = 0$, obtemos

$Y(t) = 2\int_0^t W(s)dW(s) + \int_0^t dt = 2\int_0^t W(s)dW(s) + t$. Daqui vem o resultado já familiar (ver (6.7)) $\int_0^t W(s)dW(s) = \frac{1}{2}\left(W^2(t) - t\right)$.

A diferença entre a fórmula de Itô e a correspondente fórmula do cálculo usual é que nesta o termo do desenvolvimento de Taylor de segunda ordem em x é desprezável. Na fórmula de Itô, porém, este termo aparece. De facto, consideremos um pequeno intervalo de tempo $(t, t+\Delta t)$ com $\Delta t \to 0$ e seja $\Delta W = W(t+\Delta t) - W(t)$. Vem $E\left[(\Delta W(t))^2\right] = \Delta t$, que não é desprezável (não é de ordem $o(\Delta t)$) e sugere a regra $(dW(t))^2 = dt$. Isto deve-se à irregulatidade das trajectórias do processo de Wiener. Com efeito, se $W(t)$ fosse aproximado por um processo diferenciável $\tilde{W}(t)$, teríamos $(\Delta\tilde{W}(t))^2 = \left(\frac{d\tilde{W}(t)}{dt}\Delta t + o(\Delta t)\right)^2 = o(\Delta t)$, que seria desprezável, aplicando-se as regrais usuais de cálculo.

Esta é a razão porque o *cálculo de Stratonovich*, que verifica as regras usuais de cálculo, é recomendado como o cálculo estocástico adequado quando temos uma equação diferencial aleatória "realista" com ruído colorido (cujo integral, que aparece na forma integral da equação, seja um processo diferenciável) e a aproximamos, por razões de conveniência matemática, por uma EDE "idealizada" com ruído branco (cujo integral, que aparece na forma integral da equação, é o processo de Wiener). De facto, sob condições apropriadas (ver [3], [24], [35], [40], [60]), se considerarmos aproximações suaves (ou mesmo poligonais) $\tilde{W}_n(t)$ de $W(t)$ convergindo para $W(t)$ quando $n \to +\infty$, as soluções das equações diferenciais aleatórias "realistas" (com $\tilde{W}_n(t)$), que não são processos de Markov, convergem para a solução pelo cálculo de Stratonovich da EDE "idealizada" (com o processo de Wiener $W(t)$).

O cálculo de Stratonovich não tem as boas propriedades probabilísticas do integral de Itô (como mostra o exemplo da Secção 6.1, em que falha a propriedade da martingala e a do integral ter média nula) e só é aplicável a uma classe muito mais restrita de funções integrandas. Mas, como vimos, parece ser o mais apropriado quando se usa uma EDE (com ruído branco em tempo contínuo) como aproximação conveniente do comportamento dinâmico de um fenómeno ocorrendo em tempo contínuo e perturbado por um ruído em tempo contínuo ligeiramente colorido. Já o cálculo de Itô parece ser apropriado quando se usa uma EDE (com ruído branco em tempo contínuo $\xi(t)$) como aproximação conveniente do comportamento dinâmico de um fenómeno ocorrendo em tempo discreto e perturbado por um ruído branco discreto $\bar{\xi}(t)$. De facto, como decorre da própria forma de construção do integral, as soluções das equações às diferenças estocásticas $X(t_k + \Delta t) - X(t_k) = f(t_k, X(t_k))\Delta t + g(t_k, X(t_k))\bar{\xi}(t_k)\Delta t$ ($t_k = k\Delta t$ com $\Delta t > 0$, $k = 0, 1, 2, ...$) convergem, debaixo de condições adequadas, para as soluções

com $N > 0$ e

$$I_{N,k}(\omega) := \begin{cases} 1 & \text{se } |X(t_{n,i},\omega)| \leq N \text{ para todos os } i < k \\ 0 & \text{caso contrário.} \end{cases}$$

Como $\mathbb{E}[H_{nk}] = 0$, $\mathbb{E}[H_{nk}^2] = 2(\Delta t_k)^2$ e os H_{nk} $(k = 1, 2, ..., n)$ são independentes, vem $\mathbb{E}[S_n^{(N)}] = 0$ e

$$\mathbb{E}[(S_n^{(N)})^2] = \sum_{k=1}^n \mathbb{E}\left[(h_{xx}(t_{k-1}, X_{k-1})I_{N,k})^2\right] \mathbb{E}\left[H_{nk}^2\right]$$
$$\leq 2 \max_{a\leq t\leq b, |x|\leq N} |h_{xx}(t,x)| (\sum_{k=1}^n (\Delta t_k)^2 \to 0$$

quando $\delta_n \to 0$. Para todo o $N > 0$ fixo, vem $S_n^{(N)}$ a convergir m.q. e, portanto, em probabilidade, para zero. O erro de truncatura é $P[S_m \neq S_n^{(N)}] = P[\max_{a\leq t\leq b} |X(t)| > N]$. Ora $\max_{a\leq t\leq b} |X(t)| = \max_{a\leq t\leq b} |X(a) + (t-a)F + (W(t) - W(a))G| \leq |X(a)| + (b-a)|F| + |G| \max_{a\leq t\leq b} |W(t) - W(a)|$ é uma v.a q.c. finita, pelo que $P[S_n \neq S_n^{(N)}]$ pode tornar-se tão pequeno quanto se queira para N suficientemente grande. Como $P[|S_n| > \varepsilon] = P[|S_n^{(N)}| > \varepsilon] + P[S_m \neq S_n^{(N)}]$, reconhecemos que $P[|S_n| > \varepsilon] \to 0$ e, portanto, S_n converge em probabilidade para zero como pretendido.

Exercício: Determine $d(tW(t))$ e use o resultado para mostrar que $\int_0^t s dW(s) = tW(t) - \int_0^t W(s) ds$.

Exercício: Mostre que a equação $dY(t) = Y(t)dW(t)$ com $Y(0) = 1$ tem solução $Y(t) = \exp\left(W(t) - \frac{t}{2}\right)$ para $t \geq 0$.

6.7 Os cálculos de Itô e Stratonovich

Vamos dar um exemplo de aplicação do teorema de Itô para efeitos de ilustração das regras especiais de cálculo deste integral.

Seja $X(t) = W(t)$ e $h(t,x) = x^2$. Como $X(0) = 0$ e $X(t) = W(t) = \int_0^t 0 ds + \int_0^t 1 dW(s)$, temos $F(t,\omega) \equiv 0$ e $G(t,\omega) \equiv 1$ e que $X(t)$ é um processo de Itô. Usámos o facto de a função integranda do segundo integral ser uma função em escada com um único degrau (onde toma o valor 1 para todo o ω) correspondente à decomposição $0 = t_0 < t_1 = t$ do intervalo $[0,t]$, pelo que $\int_0^t 1 dW(s) = 1(W(t_1) - W(t_0)) = W(t)$. Na forma diferencial vem $dX(t) = 0dt + 1dW(t) = dW(t)$. Vem $Y(t) = h(t, X(t)) = W^2(t)$. Pela fórmula de Itô, obtemos (pode usar a mnemónica) $dY(t) = 0dt + 2X(t)dX(t) + \frac{1}{2}2(dX(t))^2 = 2W(t)dW(t) + dt$. Na form integral, atendendo a que $Y(0) = 0$, obtemos

$Y(t) = 2\int_0^t W(s)dW(s) + \int_0^t dt = 2\int_0^t W(s)dW(s) + t$. Daqui vem o resultado já familiar (ver (6.7)) $\int_0^t W(s)dW(s) = \frac{1}{2}\left(W^2(t) - t\right)$.

A diferença entre a fórmula de Itô e a correspondente fórmula do cálculo usual é que nesta o termo do desenvolvimento de Taylor de segunda ordem em x é desprezável. Na fórmula de Itô, porém, este termo aparece. De facto, consideremos um pequeno intervalo de tempo $(t, t+\Delta t)$ com $\Delta t \to 0$ e seja $\Delta W = W(t+\Delta t) - W(t)$. Vem $E\left[(\Delta W(t))^2\right] = \Delta t$, que não é desprezável (não é de ordem $o(\Delta t)$) e sugere a regra $(dW(t))^2 = dt$. Isto deve-se à irregulatidade das trajectórias do processo de Wiener. Com efeito, se $W(t)$ fosse aproximado por um processo diferenciável $\tilde{W}(t)$, teríamos $(\Delta\tilde{W}(t))^2 = \left(\frac{d\tilde{W}(t)}{dt}\Delta t + o(\Delta t)\right)^2 = o(\Delta t)$, que seria desprezável, aplicando-se as regrais usuais de cálculo.

Esta é a razão porque o *cálculo de Stratonovich*, que verifica as regras usuais de cálculo, é recomendado como o cálculo estocástico adequado quando temos uma equação diferencial aleatória "realista" com ruído colorido (cujo integral, que aparece na forma integral da equação, seja um processo diferenciável) e a aproximamos, por razões de conveniência matemática, por uma EDE "idealizada" com ruído branco (cujo integral, que aparece na forma integral da equação, é o processo de Wiener). De facto, sob condições apropriadas (ver [3], [24], [35], [40], [60]), se considerarmos aproximações suaves (ou mesmo poligonais) $\tilde{W}_n(t)$ de $W(t)$ convergindo para $W(t)$ quando $n \to +\infty$, as soluções das equações diferenciais aleatórias "realistas" (com $\tilde{W}_n(t)$), que não são processos de Markov, convergem para a solução pelo cálculo de Stratonovich da EDE "idealizada" (com o processo de Wiener $W(t)$).

O cálculo de Stratonovich não tem as boas propriedades probabilísticas do integral de Itô (como mostra o exemplo da Secção 6.1, em que falha a propriedade da martingala e a do integral ter média nula) e só é aplicável a uma classe muito mais restrita de funções integrandas. Mas, como vimos, parece ser o mais apropriado quando se usa uma EDE (com ruído branco em tempo contínuo) como aproximação conveniente do comportamento dinâmico de um fenómeno ocorrendo em tempo contínuo e perturbado por um ruído em tempo contínuo ligeiramente colorido. Já o cálculo de Itô parece ser apropriado quando se usa uma EDE (com ruído branco em tempo contínuo $\varepsilon(t)$) como aproximação conveniente do comportamento dinâmico de um fenómeno ocorrendo em tempo discreto e perturbado por um ruído branco discreto $\bar{\varepsilon}(t)$. De facto, como decorre da própria forma de construção do integral, as soluções das equações às diferenças estocásticas $X(t_k + \Delta t) - X(t_k) = f(t_k, X(t_k))\Delta t + g(t_k, X(t_k))\bar{\varepsilon}(t_k)\Delta t$ ($t_k = k\Delta t$ com $\Delta t > 0$, $k = 0, 1, 2, ...$) convergem, debaixo de condições adequadas, para as soluções

pelo cálculo de Itô da EDE $dX(t) = f(t, X(t))dt + g(t, X(t))dW(t)$ quando $\Delta t \to 0$.

Como os cálculos de Itô e Stratonovich dão resultados diferentes, parece ser importante nas aplicações, mesmo admitindo que as condições são apropriadas ao uso das aproximações acima referidas, destrinçar se o fenómeno dinâmico ocorre intrinsecamente em tempo discreto ou em tempo contínuo, o que nem sempre é fácil. Iremos ver mais tarde, para uma ampla classe de modelos, que também não é necessário.

Vamos agora falar do *cálculo de Stratonovich* num contexto mais geral que o do exemplo da Secção 6.1. Vamos definir *integrais estocásticos de Stratonovich* apenas para funções $G(s, \omega)$ em que a dependência do acaso não seja arbitrária mas feita através da solução de uma EDE. Para isso vamos trabalhar informalmente com EDE apesar de ainda não as termos estudado matematicamente.

Por solução segundo o cálculo de Stratonovich de uma EDE

$$(S) \quad dX_t = f(t, X(t))dt + g(t, X(t))dW(t), \qquad X(0) = X_0, \tag{6.42}$$

queremos dizer uma solução da correspondente equação integral

$$X(t) = X_0 + \int_0^t f(s, X(s))ds + (S)\int_0^t g(s, X(s))dW(s) \tag{6.43}$$

quando se usa o integral de Stratonovich $(S)\int_0^t g(s, X(s))dW(s)$. Também dizemos que a EDE (6.42) é uma *equação diferencial estocástica de Stratonovich*. Usa-se o "(S)" para não confundir com o integral de Itô e o cálculo de Itô. Também se usa em alternativa escrever $\int_0^t g(s, X(s)) \circ dW(s)$ e $dX = fdt + g \circ dW(t)$. Para definir o *integral de Stratonovich* $(S)\int_0^t g(s, X(s))dW(s)$, vamos exigir que $G(s, \omega) = g(s, X(s, \omega))$ (onde $X(s, \omega)$ é a solução de (6.42)) seja contínua m.q. para adoptar uma definição semelhante à usada no exemplo da Secção 6.1 e, claro, que a EDE (6.42) tenha solução única (questões a estudar mais tarde). Definimos então

$$(S)\int_0^t g(s, X(s))dW(s) = \int_0^t g(s, X(s)) \circ dW(s) = \\ \text{l.i.m.} \sum_{k=1}^n g\left(t_{n,k-1}, \frac{X(t_{n,k-1}) + X(t_{n,k})}{2}\right)(W(t_{n,k}) - W(t_{n,k-1})), \tag{6.44}$$

onde $0 = t_{n,0} < t_{n,1} < \ldots < t_{n,n-1} < t_{n,n} = t$ $(n = 1, 2, \ldots)$ são decomposições do intervalo de integração $[0, t]$ cujos diâmetros $\delta_n = \max_{k=1,\ldots,n}(t_{n,k} - t_{n,k-1}) \to 0$ quando $n \to +\infty$.

Note-se a diferença para o que resulta da definição do integral de Itô

em condições semelhantes

$$\int_0^t g(s, X(s))dW(s) = \\ \text{l.i.m.} \sum_{k=1}^n g\left(t_{n,k-1}, X(t_{n,k-1})\right)\left(W(t_{n,k}) - W(t_{n,k-1})\right),$$

integral que aparece na forma integral da EDE de Itô

$$dX_t = f(t, X(t))dt + g(t, X(t))dW(t), \qquad X(0) = X_0. \tag{6.45}$$

Comparando a expressão do integral de Itô com a expressão (6.44) do integral de Stratonovich e usando um desenvolvimento de Taylor de primeira ordem em torno de $x = X(t_{n,k-1})$ e técnicas semelhantes às usadas na demonstração do teorema de Itô (embora sem algumas das complicações técnicas para a demonstração das convergências visto trabalharmos com convergências m.q. em vez de convergências em probabilidade), vem, para funções $g(t,x)$ contínuas em t e com derivada parcial contínua em x, que a EDE de Stratonovich (6.42) é equivalente (tem a mesma solução) que a EDE de Itô

$$dX(t) = \left(f(t, X(t)) + \tfrac{1}{2}\tfrac{\partial g(t,X(t))}{\partial x} g(t, X(t))\right) dt + g(t, X(t))dW(t) \\ X(0) = X_0. \tag{6.46}$$

O leitor mais motivado poderá tentar a demonstração, devida a [61] (ver também [3]). Um trabalho relevante de Stratonovich é [54].

Esta relação entre os cálculos de Itô e Stratonovich (quando aplicada a EDE) pode ser usada em sentido oposto, mostrando que a solução da EDE de Itô (6.45) é equivalente à EDE de Stratonovich

$$(S)dX(t) = \left(f(t, X(t)) - \tfrac{1}{2}\tfrac{\partial g(t,X(t))}{\partial x} g(t, X(t))\right) dt + g(t, X(t))dW(t) \\ X(0) = X_0. \tag{6.47}$$

Esta relação é muito útil porque podemos usar as regras usuais de cálculo, a que estamos mais habituados do que as regras do cálculo ce Itô, para resolver a EDE de Stratonovich (6.47), já que a solução assim obtida é a solução da EDE de Itô (6.45).

Das fórmulas de conversão entre os dois cálculos, vê-se que eles coincidem quando $g(t,x)$ não depende de x.

6.8 O caso multidimensional

O integral de Itô pode ser facilmente generalizado ao caso multidimensional. Nesse caso, trabalha-se com processos de Wiener padrão m-

dimensionais $\mathbf{W}(s,\omega) = [W_1(s,\omega), W_2(s,\omega), ..., W_m(s,\omega)]^T$ como processos integradores e com *funções integrandas* $\mathbf{G}(s,\omega) = [G_{ij}(s,\omega)]_{i=1,...,n;j=1,...,m}$ com *valores matriciais-$n \times m$* (isto é, com valores em $\mathbb{R}^{n\times m}$). Note-se que as coordenadas do processo de Wiener serem v.a. independentes não é restrição porque qualquer estrutura de correlação se obteria pré-multiplicando $\mathbf{W}(s)$ por uma matriz, a qual pode ser absorvida em $\mathbf{G}(s)$. Vem para *integral de Itô multidimensional*

$$\int_0^t \mathbf{G}(s,\omega)d\mathbf{W}(s,\omega) = \left[\sum_{j=1}^m \int_0^t G_{1j}dW_j(s), ..., \sum_{j=1}^m \int_0^t G_{nj}dW_j(s)\right]^T,$$

que, como se vê, se obtém à custa dos integrais de Itô unidimenisonais já aqui definidos. Os resultados são também semelhantes. As famílias $H^2[0,t]$ e $M^2[0,t]$ para as funções integrandas definem-se de forma semelhante ao caso unidimensional com a *norma-L^{2*}* a ser definida da mesma forma que (6.12) mas substituindo nela $|G|^2(s)$ (que era igual a $G^2(s)$) por $|\mathbf{G}|^2 = \left(\sum_{i=1}^n \sum_{j=1}^m G_{ij}^2\right) = \text{traço}(\mathbf{G}\mathbf{G}^T)$. Quanto à expressão do *produto interno* em $H^2[0,t]$, deve usar-se

$$\langle \mathbf{G}, \mathbf{H}\rangle_* = \left(\int_0^t \mathbb{E}[\text{traço}(\mathbf{G}(s)\mathbf{H}^T(s))]ds\right).$$

Estamos agora habilitados a tratar *sistemas de equações diferenciais estocásticas* ou *equações diferenciais estocásticas multidimensionais*. Sejam dados um processo estocástico n-dimensional $\mathbf{X}(s,\omega) = [X_1(s,\omega), X_2(s,\omega), ..., X_n(s,\omega)]^T$, uma função com valores n-dimensionais $\mathbf{f}(s,\mathbf{x}) = [f_1(s,\mathbf{x}), f_2(s,\mathbf{x}), ..., f_n(s,\mathbf{x})]^T$ definida para $(s,\mathbf{x}) \in [0,t] \times \mathbb{R}^n$ e uma função com valores matriciais-$n \times m$ $\mathbf{g}(s,\mathbf{x}) = [g_{ij}(s,\mathbf{x})]_{i=1,...,n;j=1,...,m}$ definida para $(s,\mathbf{x}) \in [0,t]\times\mathbb{R}^n$. Podemos agora considerar um *sistema de equações diferenciais estocásticas* (omitimos, como é habitual, o ω para aliviar a notação)

$$d\mathbf{X}(t) = \mathbf{f}(t,\mathbf{X}(t))dt + \mathbf{g}(t,\mathbf{X}(t))d\mathbf{W}(t), \qquad \mathbf{X}(0) = \mathbf{X}_0,$$

onde $\mathbf{X}_0$ é um vector aleatório mensurável-$\mathcal{A}_0$ (portanto, independente do processo de Wiener $\mathbf{W}(t)$). O sistema de EDE é equivalente ao *sistema de equações integrais*

$$\mathbf{X}(t) = \mathbf{X}_0 + \int_0^t \mathbf{f}(s,\mathbf{X}(s))ds + \int_0^t \mathbf{g}(s,\mathbf{X}(s))d\mathbf{W}(s).$$

Quanto aos *processos de Itô n-dimensionais* $\mathbf{X(t)}$ $(t \in [0,d])$, eles tomam a forma diferencial

$$d\mathbf{X}(t,\omega) = \mathbf{F}(t,\omega)dt + \mathbf{G}(t,\omega)d\mathbf{W}(t,\omega)$$

ou a forma integral

$$\mathbf{X}(t,\omega) = \mathbf{X}_0(\omega) + \int_0^t \mathbf{F}(s,\omega)ds + \int_0^t \mathbf{G}(s,\omega)d\mathbf{W}(s,\omega),$$

onde

- $\mathbf{X_0}$ é uma vector aleatório n-dimensional mensurável-$\mathcal{A}_0$ e, portanto, independente do processo de Wiener. Pode, em particular, ser uma constante n-dimensional determinística.
- $\mathbf{F}$ é um função com valores n-dimensionais conjuntamente mensurável adaptada à filtração $\mathcal{A}_s$ e tal que $\int_0^d |\mathbf{F}(s)|ds < +\infty$ q.c., onde $|\mathbf{F}(s)|^2 = \sum_{i=1}^n (F_i(s))^2$
- $\mathbf{G} \in M^2[0,d]$ é uma função com valores matriciais-$n \times m$.

Na versão multidimensional do teorema e fórmula de Itô, agora $\mathbf{h}(t,\mathbf{x}) = [h_1(t,\mathbf{x}),...,h_k(t,\mathbf{x})]^T$ é um vector coluna de dimensão k de funções reais contínuas definidas em $[0,d] \times \mathbb{R}^n$ com derivadas parciais contínuas de primeira ordem em t, de primeira ordem em x_i $(i = 1,...,n)$ e de segunda ordem em x_i e x_j $(i,j = 1,...,n)$. Seja $\mathbf{Y}(t) = \mathbf{h}(t,\mathbf{X}(t))$, onde $\mathbf{X}(t)$ é o processo de Itô n-dimensional acabado de referir. Representamos por $\frac{\partial \mathbf{h}}{\partial t}$ o vector coluna de dimensão k das derivadas parciais $\frac{\partial h_r}{\partial t}$, por $\frac{\partial \mathbf{h}}{\partial \mathbf{x}}$ a matriz $k \times n$ das derivadas parciais $\frac{\partial h_r}{\partial x_i}$, por $\frac{\partial^2 \mathbf{h}}{\partial x_i \partial x_j}$ o vector columa de dimensão k das derivadas parciais $\frac{\partial^2 h_r}{\partial x_i \partial x_j}$ e por $\frac{\partial^2 \mathbf{h}}{\partial \mathbf{x} \partial \mathbf{x}}$ a matriz $n \times n$ cujas coordenadas são os vectores $\frac{\partial^2 \mathbf{h}}{\partial x_i \partial x_j}$. A fórmula de Itô toma o aspecto

$$d\mathbf{Y}(t) = \frac{\partial \mathbf{h}}{\partial t}dt + \frac{\partial \mathbf{h}}{\partial \mathbf{x}}d\mathbf{X}(t) + \frac{1}{2}\text{traço}\left(\mathbf{G}\mathbf{G}^T \frac{\partial^2 \mathbf{h}}{\partial \mathbf{x} \partial \mathbf{x}}\right)dt,$$

sendo as derivadas calculadas no ponto $(t, \mathbf{X}(t))$.

Exercício: Mostre que, se $X_1(t)$ e $X_2(t)$ são processos de Itô, então $Y(t) = X_1(t)X_2(t)$ é um processo de Itô e $d(Y(t)) = X_1(t)dX_2(t) + X_2(t)dX_1(t) + dX_1(t)dX_2(t)$, usando mo cálculo de $dX_1(t)dX_2(t)$ a tabela de multiplicação $(dt)^2 = 0$, $dtdW_i = 0$, $(dW_i)^2 = dt$, $dW_i dW_j = 0$ $(i \neq j)$.

Capítulo 7

Equações diferenciais estocásticas

7.1 Teorema de existência e unicidade e principais propriedades da solução de uma equação diferencial estocástica

Seja dado um espaço de probabilidade $(\Omega, \mathcal{F}, P)$, um processo de Wiener padrão $W(t)$ nele definido e uma filtração não-antecipativa $(\mathcal{A}_s)_{s\in[0,d]}$. Seja X_0 uma v.a. mensurável-$\mathcal{A}_0$ (logo independente do processo de Wiener); pode, em particular, ser uma constante determinística. Tendo definido os integrais estocásticos de Itô no Capítulo 6, a equação integral estocástica

$$X(t,\omega) = X_0(\omega) + \int_0^t f(s, X(s,\omega))ds + \int_0^t g(s, X(s,\omega))dW(s,\omega) \quad (7.1)$$

tem agora um significado para $t \in [0,d]$, desde que as funções $F(s,\omega) = f(s, X(s,\omega))$ e $G(s,\omega) = g(s, X(s,\omega))$ sejam conjuntamento mensuráveis não-antecipativas-$\mathcal{A}_s$, $\int_0^t |F(s,\omega)|ds < +\infty$ q.c. e $\int_0^t |G(s,\omega)|^2 dW(s,\omega) < +\infty$ q.c. (isto é, $G \in M^2[0,d]$). São estas as propriedades de um processo de Itô que utilizámos na Secção 6.5. Claro que f e g precisam de satisfazer certas condições para que se possa garantir que F e G satisfazem essas propriedades. Atribuído agora um significado à forma integral (7.1), a correspondente *equação diferencial estocástica* (de Itô)

$$dX(t,\omega) = f(t, X(t,\omega))dt + g(t, X(t,\omega))dW(t,\omega) \qquad X(0,\omega) = X_0(\omega) \quad (7.2)$$

tem agora sentido para $t \in [0, d]$.

Isso não significa que a EDE (7.2) tenha necessariamente solução ou que, tendo solução, a solução seja única. Para isso, precisamos de exigir que as funções f e g satisfaçam certas condições. Como no caso das equações diferenciais ordinárias, faz falta uma condição de restrição ao crescimento destas funções para evitar a explosão da solução (solução a divergir para ∞ em tempo finito) e uma condição de Lipschitz para garantir a unicidade.

Designemos por $\mathcal{E}$ o conjunto das funções reais conjuntamente mensuráveis-Borel [1] $h(t, x)$ de domínio $[0, d] \times \mathbb{R}$ satisfazendo, para algum $K > 0$ e para todo o $t \in [0, d]$ e $x, y \in \mathbb{R}$, as propriedades seguintes:

- $|h(t, x) - h(t, y)| \leq K|x - y|$ (*condição de Lipschitz*)
- $|h(t, x)| \leq K\left(1 + |x|^2\right)^{1/2}$ (*restrição ao crescimento*).

No que se segue vamos usar a convenção habitual de não explicitar a dependência do acaso ω nas v.a e processos estocásticos que utilizarmos, mas convém termos sempre presentes que ela está implícita e, por isso, quer o processo de Wiener, quer a solução das EDE, dependem de ω.

Teorema (Existência e unicidade e algumas propriedades da solução de uma EDE de Itô)

Seja $X_0 \in L^2$ (isto é, X_0 tem variância finita) uma v.a. mensurável-$\mathcal{A}_0$ (logo independente do processo de Wiener) e considere-se no intervalo de tempo $[0, d]$ $(d > 0)$ uma equação diferencial estocástica (subentende-se de Itô)

$$dX_t = f(t, X(t))dt + g(t, X(t))dW(t), \qquad X(0) = X_0, \tag{7.3}$$

ou a correspondente equação integral estocástica

$$X(t) = X_0 + \int_0^t f(s, X(s))ds + \int_0^t g(s, X(s))dW(s) \tag{7.4}$$

(com o segundo integral interpretado como integral de Itô). Suponhamos que $f, g \in \mathcal{E}$.

Então:

a) *Existe um processo estocástico $X(t) = X(t, \omega)$ $(t \in [0, d])$ q.c. contínuo que é solução de (7.3), isto é, com probabilidade um, $X(t)$ satisfaz (7.4) para todo o $t \in [0, d]$.*

[1] Por funções conjuntamente mensuráveis-Borel em s e x, queremos dizer que são funções mensuráveis com respeito à álgebra-σ produto das álgebras-σ de Borel de $[0, d]$ e de $\mathbb{R}$.

b) *A solução é q.c. única (diremos simplesmente "única") no sentido de que, dadas duas soluções q.c. contínuas $X(t)$ e $X^*(t)$, se tem*

$$\sup_{0\le t\le d} |X(t) - X^*(t)| = 0$$

com probabilidade um.

c) *Tem-se*

$$\mathbb{E}\left[(X(t))^2\right] \le \left(1 + \mathbb{E}\left[(X_0)^2\right]\right) \exp(K(K+2)t) - 1$$
$$\mathbb{E}\left[(X(t) - X_0)^2\right] \le 2K^2(d+1)\left(1 + \mathbb{E}\left[(X_0)^2\right]\right) t \exp(K(K+2)t). \tag{7.5}$$

d) *A solução $X(t)$ está em $H^2[0,d]$ e é contínua em média quadrática.*

e) *A solução $X(t)$ é um processo de Markov com distribuição inicial igual à distribuição de X_0 e com probabilidades de transição $P(t,B|s,x) = P[X_{s,x}(t) \in B]$ $(s \le t)$. Denotámos por $X_{s,x}(t)$ a solução q.c. contínua única de $dX(t) = f(t,X(t))dt + g(t,X(t))dW(t)$ com condição inicial $X(s) = x$, isto é, a solução de*

$$X(t) = x + \int_s^t f(u,X(u))du + \int_s^t g(u,X(u))dW(u).$$

f) *Se f e g forem também contínuas em t, então a solução $X(t)$ é um processo de difusão com coeficiente de tendência $a(s,x) = f(s,x)$ e coeficiente de difusão $b(s,x) = |g(s,x)|^2$.*

Nota: Admitindo condições de regularidade suficientes para f e g, se considerássemos uma EDE de Stratonovich

$$(S) \quad dX_t = f(t,X(t))dt + g(t,X(t))dW(t),$$

como ela é equivalente à EDE de Itô

$$dX_t = \left(f(t,X(t)) + \frac{1}{2}\frac{\partial g(t,X(t))}{\partial x} g(t,X(t))\right) dt + g(t,X(t))dW(t),$$

a sua solução será um processo de difusão com coeficiente de tendência $a(s,x) = f(s,x) + \frac{1}{2}\frac{\partial g(s,x)}{\partial x} g(s,x) = f(s,x) + \frac{1}{4}\frac{\partial b(s,x)}{\partial x}$ e coeficiente de difusão $b(s,x) = |g(s,x)|^2$.

7.2 Esboço da demonstração do teorema de existência e unicidade

Apresentamos o esboço da demonstração do teorema da Secção 7.1, o qual pode ser dispensado numa primeira leitura. Para mais pormenores sobre sugerimos a consulta de [3], [51], [32], [53] ou [59] (obras também úteis sobre outros aspectos das EDE).

Demonstração da unicidade

Sejam $X(t)$ e $X^*(t)$ duas soluções q.c. contínuas. Note-se que $f(t,X(t))$, $g(t,X(t))$, $f(t,X^*(t))$ e $g(t,X^*(t))$, sendo funções mensuráveis-Borel de funções não-antecipativas, também são não-antecipativas. Poderíamos trabalhar com $\mathbb{E}[|X(t)-X^*(t)|^2]$, mas, neste momento, ainda não provámos que este momento existe, pelo que o iremos substituir pelo momento truncado $\mathbb{E}[|X(t)-X^*(t)|^2 I_N(t)]$ com $I_N(t)=1$ se $|X(u)|\leq N$ e $|X^*(u)|\leq N$ para $0\leq u\leq t$ e $I_N(t)=0$ caso contrário. Note-se que $I_N(t)$ é não-antecipativa e $I_N(t)=I_N(t)I_N(s)$ para $s\leq t$. Como $X(t)$ e $X^*(t)$ satisfazem (7.4), vem, atendendo a que $(a+b)^2\leq 2a^2+2b^2$,

$$\begin{aligned}
&\mathbb{E}[|X(t)-X^*(t)|^2 I_N(t)] = \\
&\mathbb{E}\left[I_N(t)\left(\int_0^t I_N(s)(f(s,X(s))-f(s,X^*(s)))ds + \right.\right. \\
&\quad \left.\left.\int_0^t I_N(s)(g(s,X(s))-g(s,X^*(s)))dW(s)\right)^2\right] \leq \\
&2\mathbb{E}\left[\left(\int_0^t I_N(s)(f(s,X(s))-f(s,X^*(s)))ds\right)^2\right] + \\
&2\mathbb{E}\left[\left(\int_0^t I_N(s)(g(s,X(s))-g(s,X^*(s)))dW(s)\right)^2\right].
\end{aligned}$$

Usando a desigualdade de Schwarz (que, em particular, diz que $(\int_0^t \varphi(s)ds)^2 \leq \int_0^t 1^2 ds \int_0^t \varphi^2(s)ds$), o facto de $t\leq d$ e a propriedade de preservação da norma dos integrais estocásticos (quando as funções integrandas estão, como é o caso aqui em virtude da truncatura, em $H^2[0,d]$), vem

$$\begin{aligned}
&\mathbb{E}[|X(t)-X^*(t)|^2 I_N(t)] \leq \\
&\quad 2d\int_0^t \mathbb{E}\left[I_N(s)(f(s,X(s))-f(s,X^*(s)))^2\right]ds + \\
&\quad 2\int_0^t \mathbb{E}\left[I_N(s)(g(s,X(s))-g(s,X^*(s)))^2\right]ds.
\end{aligned}$$

Logo, com $L = 2(d+1)K^2$, vem, pela condição de Lipschitz,

$$\mathbb{E}[|X(t) - X^*(t)|^2 I_N(t)] \leq L \int_0^t \mathbb{E}\left[|X(s) - X^*(s)|^2 I_N(s)\right] ds. \quad (7.6)$$

Aplicando o *lema de Bellman-Gronwall* [2] , também conhecido por *desigualdade de Gronwall*, obtemos $\mathbb{E}[|X(t) - X^*(t)|^2 I_N(t)] = 0$, o que implica que $I_N(t)X(t) = I_N(t)X^*(t)$ q.c. Ora

$$P\left[I_N(t) \not\equiv 1 \text{ em } [0,d]\right] \leq P\left[\sup_{0\leq t\leq d} |X(t)| > N\right] + P\left[\sup_{0\leq t\leq d} |X^*(t)| > N\right]$$

e, como $X(t)$ e $X^*(t)$ são (devido à continuidade q.c.) limitadas q.c. em $[0,d]$, ambas as probabilidades do lado direito podem tornar-se arbitrariamente pequenas para N suficientemente grande. Logo $X(t) = Y(t)$ q.c. para cada t e, portanto também para todo o $t \in \mathbb{Q} \cap [0,d]$ ($\mathbb{Q}$ é o conjunto dos racionais). Devido à continuidade, vem q.c. $X(t) = Y(t)$ para todo o $t \in [0,d]$ e, portanto, $\sup_{0\leq t\leq d} |X(t) - X^*(t)| = 0$ q.c.

Demonstração da existência

A demonstração da existência baseia-se no mesmo método de Picard que é usado para as equações diferenciais ordinárias. Trata-se de um método recursivo de aproximações sucessivas que começa com

$$X_0(t) \equiv X_0 \quad . \quad (7.7)$$

e usa a recursão

$$X_{n+1}(t) = X_0 + \int_0^t f(s, X_n(s))ds + \int_0^t g(s, X_n(s))dW(s). \quad (7.8)$$

Vamos provar que $X_n(t)$ $(n = 0, 1, 2, ...)$ são funções q.c. contínuas e que a sucessão por elas formada converge uniformemente q.c. O seu limite, que resulta então uma função q.c. contínua, é, como mostraremos, solução da EDE.

Note que, por ser $X_0 \in L^2$, vem $\sup_{0\leq t\leq d} \mathbb{E}[(X_0(t))^2] < +\infty$ e vamos ver que $X_n(t) \in L^2$ e $\sup_{0\leq t\leq d} \mathbb{E}[(X_n(t))^2] < +\infty$ para qualquer n, para o que se usa o método de indução. Com efeito, supondo a propriedade

[2] O *lema de Bellman-Gronwall* diz que, dadas funções $\varphi(t) \geq 0$ e $h(t)$ integráveis em $[a,b]$, se $\varphi(t) \leq L\int_0^t \varphi(s)ds + h(t)$ para todo o $t \in [a,b]$ com L constante positiva, então $\varphi(t) \leq L\int_0^t \exp(L(t-s))h(s)ds + h(t)$ para todo o $t \in [a,b]$. Uma demonstração pode ser vista em [33]. Aqui aplicou-se o lema às funções $\varphi(t) = \mathbb{E}[|X(t) - X^*(t)|^2 I_N(t)]$ e $h(t) \equiv 0$.

válida para $n-1$, vemos que ela é válida para n porque

$$\begin{aligned}\mathbb{E}\left[(X_n(t))^2\right] \leq \; & 3\mathbb{E}\left[(X_0)^2\right] + 3K^2 d \int_0^t \left(1 + \mathbb{E}\left[(X_{n-1}(s))^2\right]\right) ds + \\ & 3K^2 \int_0^t \left(1 + \mathbb{E}\left[(X_{n-1}(s))^2\right]\right) ds = \\ & 3\mathbb{E}\left[(X_0)^2\right] + B \int_0^t \left(1 + \mathbb{E}\left[(X_{n-1}(s))^2\right]\right) ds \leq \\ & 3\mathbb{E}\left[(X_0)^2\right] + Bd\left(1 + \sup_{0\leq t\leq d} \mathbb{E}\left[(X_{n-1}(t))^2\right]\right),\end{aligned} \tag{7.9}$$

com $B = 3K^2(d+1)$, onde usámos a restrição ao crescimento e a propriedade de preservação da norma do integral, bem como $(a+b+c)^2 \leq 3a^2 + 3b^2 + 3c^2$ e a desigualdade de Schwarz.

Então, podemos usar um raciocínio análogo ao da demonstração de (7.6), mas sem necessidade agora de recorrer a truncaturas (porque os momentos não truncados existem), para obter

$$\mathbb{E}[|X_{n+1}(t) - X_n(t)|^2] \leq L \int_0^t \mathbb{E}\left[|X_n(s) - X_{n-1}(s)|^2\right] ds. \tag{7.10}$$

Iterando (7.10) obtemos (pode demonstrar por indução)

$$\mathbb{E}[|X_{n+1}(t) - X_n(t)|^2] \leq L^n \int_0^t \frac{(t-s)^{n-1}}{(n-1)!} \mathbb{E}\left[|X_1(s) - X_0(s)|^2\right] ds.$$

Pela restrição ao crescimento, vem

$$\begin{aligned}&\mathbb{E}\left[|X_1(s) - X_0(s)|^2\right] = \mathbb{E}\left[\left|\int_0^t f(s, X_0) ds + \int_0^t g(s, X_0) dW(s)\right|^2\right] \leq \\ &2K^2(d+1) \int_0^t (1 + \mathbb{E}[(X_0)^2]) ds \leq 2K^2(d+1)d(1 + \mathbb{E}[(X_0)^2]) =: M,\end{aligned}$$

pelo que

$$\sup_{0\leq t\leq d} \mathbb{E}[|X_{n+1}(t) - X_n(t)|^2] \leq M(Ld)^n/n!.$$

Vem

$$\begin{aligned}d_n := \; & \sup_{0\leq t\leq d} |X_{n+1}(t) - X_n(t)| \leq \\ & \int_0^d |f(s, X_n(s)) - f(s, X_{n-1}(s))| \, ds + \\ & \sup_{0\leq t\leq d} \left|\int_0^t \left(g(s, X_n(s)) - g(s, X_{n-1}(s))\right) dW(s)\right|.\end{aligned}$$

Utilizando (6.28) e a condição de Lipschitz, vem

$$\begin{aligned}&\mathbb{E}\left[(d_n)^2\right] \leq 2K^2 d \int_0^d \mathbb{E}\left[|X_n(s) - X_{n-1}(s)|^2\right] ds + \\ &\qquad 8K^2 \int_0^d \mathbb{E}\left[|X_n(s) - X_{n-1}(s)|^2\right] ds \leq \\ &(2K^2 d + 8K^2) d M (Ld)^{n-1}/(n-1)! = C(Ld)^{n-1}/(n-1)!.\end{aligned}$$

Pela desigualdade de Tchebyshev, vem

$$\sum_{n=1}^{+\infty} P[d_n > 1/2^{n-1}] \leq \sum_n 2^{2(n-1)}\mathbb{E}[(d_n)^2] \leq C\sum_n (4Ld)^{n-1}/(n-1)! < +\infty.$$

Pelo lema de Borel-Cantelli, vem

$$P[d_n > 1/2^{n-1} \text{ para infinitos valores de } n] = 0,$$

pelo que, a partir de certa ordem, vem $d_n \leq 1/2^{n-1}$ q.c. Então, como a série $\sum_{k=0}^{+\infty} d_k$ é convergente, a série cuja sucessão das somas iniciais é

$$X_n(t) = X_0(t) + \sum_{k=0}^{n-1} (X_{k+1}(t) - X_k(t)))$$

(os termos são majorados pelos d_k) converge uniformemente q.c. em $[0, d]$.

Isto mostra a convergência q.c. uniforme em $[0, d]$ de $X_n(t)$ quando $n \to +\infty$. Designemos o seu limite por $X(t)$. Como as funções $X_n(t)$ são obviamente não-antecipativas, o mesmo sucede a $X(t)$. A continuidade q.c. das $X_n(t)$ e a convergência uniforme implicam a continuidade q.c. de $X(t)$. Obviamente, pela restrição ao crescimento e continuidade q.c. de $X(t)$, vem $\int_0^t |f(s, X(s))|ds \leq \int_0^t (1 + |X(s)|^2)^{1/2} ds < +\infty$ q.c. e $\int_0^t |g(s, X(s))|^2 ds \leq \int_0^t (1 + |X(s)|^2) ds < +\infty$ q.c., pelo que $X(t)$ é um processo de Itô e os integrais que aparecem em (7.4) fazem sentido.

Só falta ver que $X(t)$ satisfaz (7.4) para $t \in [0, d]$. Apliquemos limites em probabilidade a ambos os membros de (7.8). O lado esquerdo $X_n(t)$ converge em probabilidade para $X(t)$ (a convergência é até q.c. e uniforme). Do lado direito, temos:

- Os integrais $\int_0^t f(s, X_n(s))ds$ convergem q.c. (e, portanto, em probabilidade) para $\int_0^t f(s, X(s))ds$. Com efeito, a condição de Lipschitz implica $\left|\int_0^t f(s, X_n(s))ds - \int_0^t f(s, X(s))ds\right| \leq K \int_0^t |X_n(s) - X(s)|ds \to 0$ q.c.

- Os integrais $\int_0^t g(s, X_n(s))dW(s)$ convergem em probabilidade para $\int_0^t g(s, X(s))dW(s)$. Com efeito, a condição de Lipschitz implica $\int_0^t |g(s, X_n(s)) - g(s, X(s))|^2 ds \leq K^2 \int_0^t |X_n(s) - X(s)|^2 ds \to 0$ q.c., e daí (ver Secção 6.5) resulta o pretendido.

- Logo, o lado direito $X_0 + \int_0^t f(s, X_n(s))ds + \int_0^t g(s, X_n(s))dW(s)$

converge em probabilidade para

$$X_0 + \int_0^t f(s, X(s))ds + \int_0^t g(s, X(s))dW(s).$$

Como os limites em probabilidade são q.c. iguais, vem

$$X(t) = X_0 + \int_0^t f(s, X(s))ds + \int_0^t g(s, X(s))dW(s) \text{ q.c.},$$

ou seja $X(t)$ satisfaz (7.4).

Demonstração de que $X(t)$ está em $H^2[0,d]$ e $\sup_{0\le t\le d}\mathbb{E}\left[(X(t))^2\right] < +\infty$

De (7.9), fazendo $h_n(t) = 1 + \mathbb{E}[(X_n(t))^2]$, tira-se $h_n(t) \le 3h_0 + B\int_0^t h_{n-1}(s)ds$. Iterando sucessivamente, vem

$$\begin{aligned} h_n(t) \le \quad & 3h_0 + 3h_0Bt + 3h_0\tfrac{(Bt)^2}{2!} + ... + 3h_0\tfrac{(Bt)^{n-1}}{(n-1)!} + \\ & B^n\int_0^t \tfrac{(t-s)^{n-1}}{(n-1)!}h_0(s)ds \le \\ & 3h_0\left(1 + Bt + \tfrac{(Bt)^2}{2!} + ... + \tfrac{(Bt)^n}{n!}\right) \le 3h_0e^{Bt}, \end{aligned}$$

donde $\mathbb{E}[(X_n(t))^2] \le 3\left(1 + \mathbb{E}\left[(X_0)^2\right]\right)e^{Bt}$ e, fazendo $n \to \infty$, vem, pelo teorema da convergência dominada,

$$\mathbb{E}\left[(X(t))^2\right] \le 3\left(1 + \mathbb{E}\left[(X_0)^2\right]\right)e^{Bt}. \tag{7.11}$$

Claro que daqui resulta que $X(t) \in H^2[0,d]$ e $\sup_{0\le t\le d}\mathbb{E}\left[(X(t))^2\right] < +\infty$.

Demonstração de que $F(s,\omega) = f(s, X(s,\omega))$ e $G(s,\omega) = g(s, X(s,\omega))$ estão em $H^2[0,d]$

Do resultado anterior resulta que $\int_0^t \mathbb{E}[f^2(s, X(s))]ds \le K^2\int_0^t\left(1 + \mathbb{E}\left[X^2(s)\right]\right)ds < +\infty$ e que $\int_0^t \mathbb{E}[g^2(s, X(s))]ds \le K^2\int_0^t\left(1 + \mathbb{E}\left[X^2(s)\right]\right)ds < +\infty$, pelo que F e G estão em $H^2[0,t]$.

Daqui resulta, em particular, que o integral estocástico $\int_0^t g(s, X(s))dW(s)$ é uma martingala.

Dedução de (7.5)

Pela fórmula de Itô (ver (6.36)), vem

$$(X(t))^2 = (X_0)^2 + \int_0^t \left(2X(s)f(s,X(s)) + g^2(s,X(s))\right) ds + \int_0^t 2X(s)g(s,X(s))dW(s). \tag{7.12}$$

Aplicando esperanças matemáticas vem

$$\mathbb{E}\left[(X(t))^2\right] = \mathbb{E}\left[(X_0)^2\right] + \int_0^t \mathbb{E}\left[2X(s)f(s,X(s)) + g^2(s,X(s))\right] ds.$$

Admitiu-se que o integral estocástico tem esperança nula apesar de não termos mostrado que a função integranda esteja em $H^2[0,d]$. Rigorosamente devíamos ter truncado $X(t)$ em módulo por N, o que garante a nulidade do integral estocástico, e passado ao limite quando $N \to +\infty$.

Usando a restrição ao crescimento e o facto de $|x| \leq (1+x^2)^{1/2}$, vem

$$\begin{aligned}
&\mathbb{E}\left[(X(t))^2\right] \leq \\
&\mathbb{E}\left[(X_0)^2\right] + \int_0^t \mathbb{E}\left[2K|X(s)|(1+|X(s)|^2)^{1/2} + K^2(1+|X(s)|^2)\right] ds \leq \\
&\mathbb{E}\left[(X_0)^2\right] + K(K+2)\int_0^t \mathbb{E}\left[\left(1+\mathbb{E}[|X(s)|^2]\right)\right] ds.
\end{aligned}$$

Ponha-se $\varphi(s) = 1 + \mathbb{E}[|X(s)|^2]$ e $h(t) \equiv h = 1 + \mathbb{E}[|X_0|^2]$, note-se que $\varphi(t) = h + K(K+2)\int_0^t \varphi(s)ds$ e aplique-se a desigualdade de Gronwall para obter $1 + \mathbb{E}[|X(s)|^2] \leq h + K(K+2)e^{K(K+2)t}\int_0^t e^{-K(K+2)s}h ds$, de onde resulta a primeira igualdade de (7.5).

De (7.4), elevando ao quadrado e aplicando esperanças matemáticas, vem

$$\begin{aligned}
\mathbb{E}\left[(X(t)-X_0)^2\right] &\leq 2\mathbb{E}\left[\left|\int_0^t f(s,X(s))ds\right|^2\right] + \\
&\quad 2\mathbb{E}\left[\left|\int_0^t g(s,X(s))dW(s)\right|^2\right] \leq \\
&2t\int_0^t \mathbb{E}\left[|f(s,X(s))|^2\right] ds + 2\int_0^t \mathbb{E}\left[|g(s,X(s))|^2\right] ds \leq \\
&2(d+1)K^2\int_0^t \left(1+\mathbb{E}\left[|X(s)|^2\right]\right) ds.
\end{aligned}$$

Aplicando a primeira expressão (7.5) e majorando e^{Ls} no integral por e^{Lt}, obtemos a segunda expressão.

A solução é contínua m.q.

Para $t \geq s$, $X(t)$ é também solução de $X(t) = X(s) + \int_s^t f(u,X(u))du$

$+\int_s^t g(u,X(u)dW(u)$. Como a condição inicial é agora $X(s)$ no instante s e o comprimento do intervalo é $t-s$, vem da segunda relação de (7.5), que

$$\begin{aligned}&\mathbb{E}\left[(X(t)-X(s))^2\right]\leq\\&2K^2(d+1)\left(1+\mathbb{E}\left[(X(s))^2\right]\right)(t-s)\exp(K(K+2)(t-s));\end{aligned}\tag{7.13}$$

atendendo à primeira relação de (7.5), podemos majorar $1+\mathbb{E}\left[(X(s))^2\right]$ por $\left(1+\mathbb{E}\left[(X(0))^2\right]\right)e^{K(K+2)d}$. Concluímos que $\mathbb{E}\left[(X(t)-X(s))^2\right]\to 0$ quando $t-s\to 0$, o que prova o pretendido.

A solução é um processo de Markov

Seja $0\leq s\leq t\leq d$. A justificação intuitiva de que $X(t)$ é um processo de Markov vem do facto de $X(t)=X_{s,X(s)}(t)$, isto é, de $X(t)$ poder ser obtido no intervalo $[s,t]$ como solução de $X(t)=X(s)+\int_s^t f(u,X(u))du+\int_s^t g(u,X(u)dW(u)$. Logo $X(t)$ fica definido à custa (como função) de X_s e de $W(u)-W(s)$ $(u\in[s,d])$, pelo que é mensurável para $\sigma(X(s),W(u)-W(s):s\leq u\leq t)$. Como $X(s)$ é mensurável-$\mathcal{A}_s$ e, portanto, independente de $\sigma(W(u)-W(s):s\leq u\leq t)$, concluímos que, dado $X(s)$, $X(t)$ só depende de $\sigma(W(u)-W(s):s\leq u\leq t)$. Como $\sigma(X(u):0\leq u\leq s)\subset\mathcal{A}_s$ é independente de $\sigma(W(u)-W(s):s\leq u\leq t)$, também $X(t)$, dado $X(s)$, é independente de $\sigma(X(u):0\leq u\leq s)$ e, portanto, é um processo de Markov.

Uma demonstração formal pode ver-se, por exemplo, em [34], [3] e [32].

A solução é, para f e g contínuas em t, um processo de difusão

Seja $0\leq s\leq t\leq d$ e $\Delta=t-s$. Condicionando em $X(s)=x$, obtemos de (7.13)

$$\mathbb{E}_{s,x}\left[(X(t)-X(s))^2\right]\leq 2K^2(d+1)(1+x^2)(t-s)\exp(K(K+2)(t-s));\tag{7.14}$$

expressões semelhantes podem ser obtidas para momentos de ordem par

$$\mathbb{E}_{s,x}\left[(X(t)-X(s))^{2n}\right]\leq C(1+x^{2n})(t-s)^n\exp(D(t-s)),\tag{7.15}$$

com C e D constantes positivas adequadas. Fica como exercício para o leitor obter expressões do género da segunda expressão de (7.5) para outros momentos de ordem par $2n$ no caso de $X_0\in L^{2n}$ e, daí, deduzir (7.15) notando que a condição inicial x é agora determinística e, portanto, tem momentos de todas as ordens. Dado $X(s)=x$, vem

$X(t) = X_{s,x}(t)$, pelo que

$$\mathbb{E}_{s,x}[(X(t) - X(s))^k] := \mathbb{E}\left[(X(t) - x)^k | X(s) = x\right] = \mathbb{E}\left[(X_{s,x}(t) - x)^k\right].$$

Devido a (7.15), estes momentos existem para todo o $k = 1, 2, 3, 4, ...$ Como já sabemos que $X(t)$ é de Markov e tem trajectórias q.c. contínuas, basta mostrar (5.1), (5.2) e (5.3).

Claro que f e g são funções contínuas na variável x também, devido à condição de Lipschitz.

De (7.15) com $n = 2$ vem que $\mathbb{E}_{s,x}\left[(X(s+\Delta) - x)^4\right] \leq A\Delta^2$, com $A = C(1+x^{2n})\exp(Dd)$ constante, pelo que $\frac{1}{\Delta}\mathbb{E}_{s,x}\left[(X(s+\Delta) - x)^4\right] \to 0$ quando $\Delta \to 0$. Logo

$$\frac{1}{\Delta} P_{s,x}\left[|X(s+\Delta) - x| \geq \varepsilon\right] \leq \frac{1}{\Delta}\frac{1}{\varepsilon^4}\mathbb{E}_{s,x}\left[(X(s+\Delta) - x)^4\right] \to 0$$

e resulta (5.1).

Partindo de (7.4), como o integral estocástico tem esperança nula, obtemos

$$\begin{array}{l} \mathbb{E}_{s,x}\left[X(s+\Delta) - x\right] = \int_s^{s+\Delta} \mathbb{E}_{s,x}\left[f(u, X_{s,x}(u))\right] du = \\ \int_s^{s+\Delta} f(u,x)du + \int_s^{s+\Delta} \mathbb{E}_{s,x}\left[(f(u, X_{s,x}(u)) - f(u,x))\right] du. \end{array} \tag{7.16}$$

Vem também, usando a condição de Lipschitz, a desigualdade de Schwarz e (7.14),

$$\begin{array}{l} \left|\int_s^{s+\Delta} \mathbb{E}_{s,x}\left[(f(u, X_{s,x}(u)) - f(u,x))\right] du\right| \leq \\ \Delta^{1/2}\left(\int_s^{s+\Delta} \mathbb{E}_{s,x}\left[\left|f(u, X_{s,x}(u)) - f(u,x)\right|^2\right] du\right)^{1/2} \leq \\ K\Delta^{1/2}\left(\int_s^{s+\Delta} \mathbb{E}_{s,x}\left[\left|X_{s,x}(u) - x\right|^2\right] du\right)^{1/2} \leq \\ K\Delta^{1/2}\left(2K^2(d+1)(1+x^2)\Delta\exp(K(K+2)d)\right)^{1/2}\Delta^{1/2} = B\Delta^{3/2} \end{array} \tag{7.17}$$

com B constante não-dependente de s. A continuidade de f em t no intervalo fechado $[0, d]$ é uniforme, pelo que, dado $\varepsilon > 0$ arbitrário, existe um $\Delta > 0$ independente de s tal que

$$\begin{array}{l} \left|\int_s^{s+\Delta} f(u,x)du - f(s,x)\Delta\right| = \\ \left|\int_s^{s+\Delta}(f(u,x) - f(s,x))du\right| \leq \int_0^t \varepsilon ds = \varepsilon\Delta. \end{array} \tag{7.18}$$

De (7.16), (7.17) e (7.18), obtém-se (5.2) com $a(s,x) = f(s,x)$.

Para obter (5.3) com $b(s,x) = g^2(s,x)$, parte-se de (7.12) em vez de

(7.4), usando também a condição inicial $X(s) = x$ em vez de $X(0) = X_0$, e usam-se técnicas semelhantes.

Está concluída a demonstração do teorema.

7.3 Observações e extensões ao teorema de existência e unicidade

O teorema usa uma técnica recursiva ((7.7) e (7.8)) para construir a solução da EDE. Esta técnica permitiria mesmo obter uma aproximação tão boa quanto desejado da solução, mas normalmente a convergência para a solução é lenta e o método pouco prático. Para métodos mais expeditos de resolução numérica e de simulação de EDE, remetemos o leitor para [41] e [42].

A condição $X_0 \in L^2$ (X_0 ter variância finita) não é necessária para a existência se solução e, embora essa condição tenha sido usada na demonstração, seria dispensável usando uma técnica de truncatura de X_0 a um intervalo $[-N, N]$ (a v.a. truncada já estaria em L^2) seguida da passagem ao limite quando $N \to +\infty$. A condição também não faz falta para provar a unicidade e continuidade da solução ou para o facto de ela ser um processo de Markov ou de difusão, mas faz falta para as propriedades que envolvam existência de primeiros e segundos momentos da solução. Aliás, se $X_0 \in L^{2n}$, as soluções da EDE têm momentos de ordem $2n$.

Para a existência e unicidade da solução também não é essencial a condição de Lipschitz, basta que f e g satisfaçam uma condição de Lipschitz local. Diz-se que uma função $h(s,x)$ de domínio $[0,d] \times \mathbb{R}$ satisfaz uma *condição de Lipschitz local* se, para todo o N, existir um $K_N > 0$ tal que, para todo o $t \in [0,d]$ e $|x| \leq N$, $|y| \leq N$, vier

$$|h(t,x) - h(t,y)| \leq K_N |x - y|. \tag{7.19}$$

A demonstração da existência de solução usa um técnica de truncatura de $X(s)$ ao intervalo $[-N, N]$ seguida da passagem ao limite quando $N \to +\infty$.

A restrição ao crescimento e a condição de Lipschiz de $f(t,x)$ e $g(t,x)$ nem sempre precisam de verificar-se para todos os pontos $x \in \mathbb{R}$, ou $x \in \mathbb{R}$ e $y \in \mathbb{R}$. Basta que se verifiquem para um conjunto $D \subset \mathbb{R}$ caso a solução $X(t)$ da equação tenha contradomónio contido em D com probabilidade um. Com efeito, se isso suceder, podemos substituir f e g por outras funções a elas idênticas em D e nulas fora de D, que nada se altera.

Dados f e g, podemos dizer que a EDE $dX(t) = f dt + g dW(t)$, $X(0) = X_0$, ou a equação integral estocástica correspondente $X(t) = X_0 + \int_0^t f(s, X(s))ds + \int_0^t g(s, X(s))dW(s)$, é uma correspondência que transforma uma v.a. X_0 e um dado processo de Wiener $\{W(t)\}_{t\in[0,d]}$ (num dado espaço de probabilidade com uma certa filtração não antecipativa $\mathcal{A}_s$) na solução $\{X(t)\}_{t\in[0,d]}$ da EDE, a qual está adaptada à filtração. Se mudarmos o processo de Wiener escolhido, altera-se a solução. Este tipo de solução, que foi a que nós estudámos (e que se subentende quando nada se diz), também se chama *solução forte*. Claro que, uma vez escolhido o processo de Wiener, a solução é quase certamente única e, a cada ω corresponde o valor de $X_0(\omega)$ e uma trajectória $\{W(t,\omega)\}_{t\in[0,d]}$ do processo de Wiener, que é transformada pela EDE (ou equação integral correspondente) numa trajectória $\{X(t,\omega)\}_{t\in[0,d]}$ da solução da EDE.

Podem também definir-se *soluções fracas*, que correspondem à situação em que são apenas dados f, g e X_0 e se procura um espaço de probabilidade e um par de processos $\{\bar{X}(t)\}_{t\in[0,d]}$ e $\{\bar{W}(t)\}_{t\in[0,d]}$ de tal forma que se tenha $\bar{X}(t) = X_0 + \int_0^t f(s, \bar{X}(s))ds + \int_0^t g(s, \bar{X}(s))d\bar{W}(s)$. A diferença é que, enquanto na solução forte, o processo de Wiener era dado *a priori* e escolhido arbitrariamente, ficando a solução (forte) dependente do processo escolhido, agora o processo de Wiener é obtido *a posteriori* e faz parte da solução do problema. Há que ter cuidado para que se considere uma filtração $\bar{\mathcal{A}}_s$ à qual $X(t)$ esteja adaptado e para a qual o processo de Wiener $\bar{W}(t)$ seja uma martingala-$\bar{\mathcal{A}}_t$. É possível estender a definição do integral estocástico $\int_0^t g(s, \bar{X}(s))d\bar{W}(s)$ a esta situação, mesmo quando $\bar{X}(t)$ não está adaptado à filtração gerada pelo processo de Wiener e pela condição inicial.

Claro que uma solução forte é também uma solução fraca, mas o recíproco é falso. Um contra-exemplo pode ser visto em [51], em que se mostra que uma dada EDE não tem soluções fortes mas tem soluções fracas.

A unicidade que considerámos no teorema da Secção 7.1 é a *unicidade forte*, isto é, dadas duas soluções, as suas trajectórias coincidem completamente com probabilidade um. Também se pode definir *unicidade fraca*, que significa que, dadas duas soluções (fortes ou fracas), as suas trajectórias têm as mesmas funções de distribuição de dimensão finita. Claro que a unicidade forte implica a unicidade fraca, mas o recíproco é falso. Nas condições do teorema da Secção 7.1, verifica-se que quaisquer soluções, fortes ou fracas, são fracamente únicas. De facto, dadas duas soluções, sejam fortes ou fracas, elas são também soluções fracas. Sejam elas $(\bar{X}(t), \bar{W}(t))$ e $(\bar{X}^*(t), \bar{W}^*(t))$. Então, como existem soluções fortes, sejam $X(t)$ e $Y(t)$ as soluções fortes correspondentes a

$\bar{W}(t)$ e $\bar{W}^*(t)$). Pelo método de Picard das aproximações sucessivas dado por (7.7)-(7.8), as sucessões aproximadoras têm as mesmas distribuições de dimensão finita, pelo que o mesmo sucede aos seus limites q.c., que são $X(t)$ e $Y(t)$.

Repare-se que, satisfeitas as condições do teorema de existência e unicidade, do ponto de vista probabilístico, isto é, das funções de distribuição de dimensão finita, não há diferença entre soluções, sejam fortes ou fracas, pelo que se pode trabalhar com soluções fracas para determinar as propriedades probabilísticas da solução forte.

Por vezes não são válidas as condições de Lipschitz ou de restrição ao crescimento e não há garantia de existência de solução forte. Importa nesses casos ver se existem soluções fracas. Condições de existência de soluções fracas podem ver-se em [55] e [37]. Um outro resultado interessante (ver [37]) é que, se existir solução fraca e se houver unicidade forte, então há solução forte.

Vimos que, quando f e g, além de satisfazerem as restantes propriedades do teorema da Secção 7.1, eram contínuas no tempo, a solução da EDE era um processo de difusão com coeficiente de tendência $f(t,x)$ e coeficiente de difusão $g^2(t,x)$.

O problema recíproco é também relevante. Dado um processo de difusão $X(t)$ ($t \in [0,d]$) num espaço de probabilidade $(\Omega, \mathcal{F}, P)$ com coeficiente de tendência $a(s,x)$ e coeficiente de difusão $b(s,x)$, será que existe uma EDE da qual o processo seja solução fraca? Sob condições de regularidade adequadas, a resposta é positiva. Para alguns resultados ver [55], [37] e [34]. Assim há de certa forma uma correspondência entre soluções de EDE e processos de difusão.

No caso particular de f e g não dependerem do tempo (e satisfazerem as condições do teorema da Secção 7.1), serão automaticamente funções contínuas do tempo e a solução da EDE será um processo de difusão homogéneo com coeficiente de tendência $f(x)$ e coeficiente de difusão $g^2(x)$. Neste caso, a EDE diz-se uma *equação diferencial estocástica autónoma* e à respectiva solução também se diz uma *difusão de Itô*. Note-se que neste caso não é necessário verificar a restrição ao crescimento porque ela é uma consequência da condição de Lipschitz (a única que é então preciso verificar). No caso autónomo podemos trabalhar com $t \in [0, +\infty)$ já que f e g não dependem do tempo.

No caso autónomo, nas condições do teorema da Secção 7.1, pode mesmo concluir-se que a solução da EDE é um *processo de Markov forte*. Veja-se [34] ou [51].

No caso autónomo, desde que estejamos a trabalhar em dimensão um (isto é, não podemos generalizar a sistemas de EDE), podemos obter resultados mesmo quando falha a condição de Lipschitz mas $f(x)$ e $g(x)$

têm derivada contínua (note que a existência de derivada contínua limitada implica a condição de Lipschitz mas, se a derivada não for limitada, a condição de Lipschitz pode falhar). Neste caso, prova-se que existe uma solução (forte) única até a um instante de explosão $T_\infty(\omega)$; a condição de Lipschitz é suficiente mas não necessária para evitar explosões (isto é, para ter $T_\infty = +\infty$ q.c.). Há um teste que permite saber se haverá ou não explosão, o *teste de Feller* (ver [48]).

Podemos ter *equações diferenciais estocásticas multidimensionais*. O enquadramento é semelhante ao dos processos de Itô multidimensionais da Secção 6.8, só que agora usa-se $\mathbf{G}(s,\omega) = \mathbf{g}(s, \mathbf{X}(s,\omega))$ e $\mathbf{F}(s,\omega) = \mathbf{f}(s, \mathbf{X}(s,\omega))$. A condição de Lipschitz e a restrição ao crescimento de $\mathbf{f}$ e $\mathbf{g}$ no teorema de existência e unicidade são idênticas com $|\mathbf{x} - \mathbf{y}|$ a significar a distância euclideana. A validade dos resultados do teorema da Secção 7.1 permanece. Quando a solução é um processo de difusão, o coeficiente de tendência é o vector $\mathbf{f}$ e o coeficiente de difusão é a matriz $\mathbf{g}\mathbf{g}^T$.

Note-se que, neste caso, dado um processo de difusão com coeficiente de tendência $\mathbf{a}$ e coeficiente de difusão $\mathbf{b}$, há várias matrizes $\mathbf{g}$ para as quais $\mathbf{g}\mathbf{g}^T = \mathbf{b}$, mas todas elas dão as mesmas propriedades probabilísticas para as soluções da EDE, pelo que, do ponto de vista das soluções fracas, é indiferente qual delas se escolhe.

O tratamento deste Capítulo permite incluir EDE do tipo

$$dY(t) = f(t, Y(t), W(t))dt + g(t, Y(t), W(t))dW(t),$$

acrescentando a equação $dW(t) = dW(t)$ e trabalhando a duas dimensões com o vector $\mathbf{X}(t) = [Y(t), W(t)]^T$ e a EDE

$$d\mathbf{X}(t) = \begin{pmatrix} f \\ 0 \end{pmatrix} dt + \begin{pmatrix} g \\ 1 \end{pmatrix} dW(t).$$

A condição inicial é da forma $\mathbf{X}(t) = [Y_0, 0]^T$. Note que $f(t, Y(t), W(t)) = f(t, \mathbf{X}(t))$ e $g(t, Y(t), W(t)) = g(t, \mathbf{X}(t))$.

Também se admitem funções f e g que tenham uma dependência do acaso ω mais geral do que só depender dele através de $X(s)$, mas é preciso que se mantenha o catácter não-antecipativo.

Finalmente, podemos considerar *equações diferenciais estocásticas de ordem superior* da forma (aqui o expoente "(n)" representa a derivada de ordem n)

$$Y^{(n)}(t) = f(t, Y(t), ..., Y^{(n-1)}(t)) + g(t, Y(t), ..., Y^{(n-1)}(t))\varepsilon(t)$$

com condições iniciais $Y^{(i)}(0) = a_i$ $(i = 0, 1, ..., n-1)$. Usa-se a mesma técnica que nas EDO trabalhando com o vector

$$\mathbf{X}(t) = [Y(t), Y^{(1)}(t), ..., Y^{(n-1)}(t)]^T$$

e a EDE

$$d\mathbf{X}(t) = \begin{pmatrix} Y^{(1)}(t) \\ Y^{(2)}(t) \\ ... \\ f(t, \mathbf{X}(t) \end{pmatrix} dt + \begin{pmatrix} 0 \\ 0 \\ 0 \\ g(t, \mathbf{X}(t) \end{pmatrix} dW(t).$$

Capítulo 8

Estudo do modelo de Black-Scholes

8.1 Estudo pelo cálculo de Itô

O *modelo de Black-Scholes* foi referido no Capítulo 3 e pode, entre outras coisas, ser usado para modelar a cotação de uma acção na bolsa e também o crescimento de uma população de seres vivos em ambiente sujeito a perturbações aleatórias que afectem a taxa de crescimento *per capita* da população. Trata-se pois de uma *aleatoriedade ambiental*. O modelo não serve para tratar a *aleatoriedade demográfica*, isto é, as perturbações demográficas do número de nascimentos e mortes devidas ao acaso, que ocorrem mesmo quando as taxas de natalidade e mortalidade são constantes (isto é, mesmo na ausência de perturbações ambientais que afectem tais taxas). O modelo é descrito pela EDE

$$dX(t) = rX(t)dt + \sigma X(t)dW(t), \qquad X(0) = x_0 \tag{8.1}$$

ou pela correspondente forma integral

$$X(t) = x_0 + \int_0^t rX(s)ds + \int_0^t \sigma X(s)dW(s). \tag{8.2}$$

Aqui r representa a taxa "média" de rendimento ou crescimento. Quanto a σ, mede a intensidade dos efeitos das flutuações aleatórias do mercado/ambiente sobre a taxa de rendimento/crescimento. Dá uma ideia do desvio-padrão das flutuações em torno do valor médio das taxas e é conhecido na literatura financeira por *volatilidade*. Estamos a supor que o valor inicial x_0 é conhecido, isto é, determinístico e vamos supor

ainda que $x_0 > 0$. Não esquecer que $W(t) = W(t,\omega)$ e $X(t) = X(t,\omega)$ dependem do acaso ω (cenário do mercado ou estado da natureza nos exemplos dados).

Nesta Secção vamos adoptar o cálculo de Itô, isto é, o segundo integral em (8.2) vai ser interpretado como *integral de Itô*.

Vamos agora determinar a solução da EDE (8.1), ou seja, a solução da forma integral (8.2).

Como $f(t,x) \equiv rx$ e $g(t,x) \equiv \sigma x$ no teorema da Secção 7.1, a EDE é *autónoma* e satisfaz a condição de Lipschitz; pelo facto de ser autónoma, a condição de restrição ao crescimento é consequência da condição de Lipschitz. Sabemos assim que existe uma e uma só solução q.c. contínua, a qual é um processo de Markov e, até, um processo de difusão homogéneo (é uma difusão de Itô) com coeficiente de tendência $a(x) = rx$ e coeficiente de difusão $b(x) = \sigma^2 x^2$.

Como o fenómeno é por natureza multiplicativo, pode ser útil trabalhar em escala logarítmica, usando a nova variável

$$Z(t) = \ln(X(t)/x_0). \tag{8.3}$$

Vamos obter a solução da EDE por três vias diferentes a fim de explorar vários conceitos e técnicas.

Solução via uso da fórmula de Itô

Seja $h(t,x) = \ln(x/x_0)$, pelo que $Z(t) = h(t, X(t))$. Aplicando a *fórmula de Itô* (ver (6.35) ou a mnemónica que se segue), obtemos

$$dZ(t) = Rdt + \sigma dW(t), \quad Z(0) = 0, \tag{8.4}$$

com $R = r - \sigma^2/2$. Na forma integral, vem $Z(t) = \int_0^t R ds + \int_0^t \sigma dW(s)$. Como, por (6.13), $\int_0^t 1 dW(t) = 1(W(t) - W(0)) = W(t)$ (a função integranda é função em escada com um único degrau), a solução desejada é

$$Z(t) = Rt + \sigma W(t). \tag{8.5}$$

Se se preferir escrever a solução em termos da escala original, teremos

$$X(t) = x_0 \exp(Rt + \sigma W(t)),$$

a que se chama *movimento browniano geométrico*.

Breve estudo da solução

Vem

$$Z(t) \frown \mathcal{N}(Rt, \sigma^2 t). \tag{8.6}$$

É de notar que o valor médio (esperança matemática) Rt de $Z(t)$ é diferente da solução ($Z(t) = rt$) do modelo determinístico $dX = rXdt$.

Resulta que $X(t)$ tem distribuição lognormal com valor esperado

$$\mathbb{E}[X(t)] = x_0 \exp\left(Rt + \sigma^2 t/2\right) = x_0 \exp(rt), \tag{8.7}$$

o mesmo valor que a solução determinística $X(t) = x_0 \exp(rt)$.

Como $W(t)/t \to 0$ q.c. quando $t \to +\infty$, obtemos $Z(t) = t(R + W(t)/t) \sim Rt$ q.c.. Portanto, quando $t \to +\infty$, obtemos

$$X(t) \begin{cases} \to +\infty \text{ q.c.} & \text{se } R > 0 \text{ ou seja, se } r > \sigma^2/2 \\ \to 0 \text{ q.c.} & \text{se } R < 0 \text{ ou seja, se } r < \sigma^2/2. \end{cases} \tag{8.8}$$

Este comportamento difere aparentemente do do modelo determinístico, em que $r > 0$ ou $r < 0$ determina limite $+\infty$ ou limite nulo, respectivamente. Este diferente comportamento qualitativo permite, por exemplo no caso do crescimento populacional, que a população se extinga com probabilidade um mesmo para valores positivos (desde que inferiores a $\sigma^2/2$) da taxa "média" de crescimento r.

A partir da distribuição normal de $Z(t)$ facilmente se obém a f.d.p. $p(t, y)$ de $X(t)$:

$$p(t, y) = \frac{1}{y\sqrt{2\pi\sigma^2 t}} \exp\left(-\frac{(\ln(y/x_0) - Rt)^2}{2\sigma^2 t}\right). \tag{8.9}$$

Para $s > t$, podemos também obter $X(t)$ condicionado ao valor de $X(s)$ visto que $X(t) = X(s) + \int_s^t rX(u)du + \int_s^t \sigma X(u)dW(u)$ ou, em termos de $Z(t)$, $Z(t) = Z(s) + \int_s^t Rdu + \int_s^t \sigma dW(u)$. Vem

$$Z(t)\,|(Z(s) = c) \frown \mathcal{N}(c + R(t-s), \sigma^2(t-s)). \tag{8.10}$$

Daqui resulta, pondo $c = \ln(x/x_0)$, que a f.d.p. de transição para $X(t)$ é

$$p(t-s, y|x) = \frac{1}{y\sqrt{2\pi\sigma^2(t-s)}} \exp\left(-\frac{(\ln(y/x) - R(t-s))^2}{2\sigma^2(t-s)}\right), \tag{8.11}$$

que nos dá a f.d.p. de $X(t)$ condicionado a se ter $X(s) = x$.

Solução via equações de Kolmogorov

A expressão (8.11) pode também obter-se resolvendo directamente a

equação de Kolmogorov progressiva

$$\frac{\partial p(\tau, y|x)}{\partial \tau} + \frac{\partial}{\partial y}(a(y)p(\tau, y|x)) - \frac{1}{2}\frac{\partial^2}{\partial y^2}(b(y)p(\tau, y|x)) = 0$$

com a condição terminal $\lim_{\tau \downarrow 0} p(\tau, y|x) = \delta(y - x)$ e com $a(y) = ry$ e $b(y) = \sigma^2 y$. Para resolver esta equação, ajuda fazer a mudança de variável $z = \ln(y/x_0)$.

Quanto à expressão (8.9), pode obter-se resolvendo directamente a equação de Kolmogorov progressiva

$$\frac{\partial p(t, y)}{\partial t} + \frac{\partial}{\partial y}(a(y)p(t, y)) - \frac{1}{2}\frac{\partial^2}{\partial y^2}(b(y)p(t, y)) = 0$$

com a condição terminal $\lim_{\tau \downarrow 0} p(\tau, y) = \delta(y - x_0)$ e com $a(y) = ry$ e $b(y) = \sigma^2 y$. Claro que agora isso é desnecessário pois já obtivemos (8.11) e, como a condição inicial é determinística ($X_0 = x_0$), basta atender a que $p(t, y) = p(t, y|x_0)$.

Note-se que as equações de Kolmogorov nos permitem obter as densidades de transição donde, conhecendo a distribuição inicial (neste caso, em que a condição inicial é determinística, conhecendo o valor de x_0), podemos determinar qualquer distribuição de dimensão finita da solução. Mas é um método mais limitado que o anterior pois não nos dá uma expressão para a solução $X(t)$ como função da condição inicial e da trajectória concreta do processo de Wiener $W(t)$. Na maioria dos casos, a trajectória de $W(t)$ não é conhecida e queremos mesmo é calcular probabilidades de certos acontecimentos relativos ao processo $X(t)$ e, neste caso, o conhecimento das distribuições de dimensão finita é suficiente.

Solução via conversão a uma equação de Stratonovich

Também podemos converter a EDE de Itô (8.1) numa EDE de Stratonovich equivalente (com a mesma solução) usando a equivalência (6.45)-(6.47). Esta pode ser resolvida usando as regras usuais de cálculo, a que estamos habituados, o que se torna nais simples. A EDE de Itô (8.1) é equivalente à EDE de Stratonovich

$$\begin{aligned}(S)\ dX(t) &= \left(rX(t) - \tfrac{1}{2}\sigma\sigma X(t)\right)dt + \sigma X(t)dW(t) \\ &= RX(t)dt + \sigma X(t)dW(t), \quad X(0) = x_0.\end{aligned} \tag{8.12}$$

Utilizando nesta regras usuais de diferenciação de funções compostas vem

$$(S)\ dZ(t) = \frac{1}{X(t)}dX(t) = Rdt + \sigma dW(t),$$

que, na forma integral, dá $Z(t) = \int_0^t R dt + (S)\int_0^t \sigma dW(t) = Rt + \sigma W(t)$, o mesmo resultado que já tínhamos obtido antes. Note-se que $(S)\int_0^t 1 dW(t) = 1(W(t) - W(0)) = W(t)$ é dado por (6.44), neste caso com $g \equiv 1$, e vai coincidir, porque se trata de função $g(t,x)$ não-dependente de x, com o integral de Itô.

Note-se que usámos o cálculo de Stratonovich apenas como auxiliar, já que o objectivo é resolver a equação original (8.1) segundo o cálculo de Itô. Só que, para o fazer, tirámos partido de isso ser equivalente a resolver a equação modificada (8.12) segundo o cálculo de Stratonovich.

Convidamos o leitor a tornar a olhar para as Figuras 2.1 e 2.2 do Capítulo 2. A Figura 2.1 mostra em escala logarítmica a trajectória real (correspondente ao cenário do mercado ω que efectivamente ocorreu) das cotações diárias de fecho de acções do BCP no período de 8 de Abril de 1991 a 30 de Junho de 1997. As estimativas dos parâmetros para o modelo (8.1) feitas a partir destas observações foram $\hat{R} = 0.084/\text{ano}$ ($R = r - \sigma^2/2$) e $\hat{\sigma}^2 = 0.0373/\text{ano}$ ($\hat{\sigma} = 0.193/\sqrt{\text{ano}}$).

Neste caso em que temos uma solução explícita, a estimação dos parâmetros pode fazer-se pelo método da máxima verosimilhança, atendendo à propriedade de Markov da solução. Suponhamos que, além do valor inicial x_0 (suposto determinístico), observámos as cotações $X(t_k) = x_k$ nos instantes t_k ($k = 1, 2, ..., n$), instantes que supomos ordenados crescentemente. É mais cómodo trabalhar com $Z(t_k) = \ln(X(t_k)/x_0) = \ln(x_k/x_0) = z_k$; note-se que, fazendo $t_0 = 0$, vem $Z(t_0) = z_0 = 0$. Claro que, pela propriedade de Markov da solução, a f.d.p. conjunta $f_{Z(t_1),...,Z(t_n)}(z_1, ..., z_n)$ das v.a. $Z(t_1), ..., Z(t_n)$ é igual ao produto $f_{Z(t_1)|Z(t_0)=z_0}(z_1) f_{Z(t_2)|Z(t_1)=z_1}(z_2) ... f_{Z_n|Z_{n-1}=z_{n-1}}(z_n)$ das sucessivas densidades de transição. Estas são fáceis de obter já que $Z(t_i)|(Z(t_{i-1}) = z_{i-1}) \frown \mathcal{N}(z_{i-1} + R(t_i - t_{i-1}) + \sigma(t_i - t_{i-1}))$. Deixamos ao leitor, a título de exercício, o trabalho de explicitar os cálculos. A expressão da f.d.p. conjunta das observações, considerada como função dos parâmetros R e σ, é a função de verosimilhança das observações e os *estimadores de máxima verosimilhança* dos parâmetros são os valores dos parâmetros que maximizam a função de verosimilhança. Foi assim que obtivémos as estimativas acima indicadas.

Para mais informações sobre métodos de estimação, teste e predição neste tipo de equações de Black-Scholes (permitindo até comparações entre várias taxas de rendimento), veja [11] e [15]. Para modelos de dinâmica de populações, podem ver-se alguns destes problemas estatísticos tratados em [9], [10], [11], [13] e [14].

Muitas vezes, porém, não se dispõe de solução de explícita nem das densidades de transição e não é possível obter a função de verosimilhança.

No entanto, mesmo nesses casos, é muito importante para as aplicações a estimação paramétrica ou não-paramétrica dos coeficientes de tendência e de difusão a partir das observações de uma ou mais trajectórias (normalmente só se dispõe de uma trajectória), sejam as observações feitas em tempo contínuo ou (como sucedeu no exemplo) em tempo discreto. Problemas estatísticos relacionados são os da escolha do modelo, os testes de hipóteses sobre parâmetros e a previsão de valores não observados do processo. A leitura sobre estas matérias está ainda muito dispersa. Pode consultar [5], [26] e [43].

A Figura 2.2 mostra na mesma escala logarítmica duas trajectórias simuladas pelo método de Monte Carlo de um movimento browniano geométrico usando como parâmetros as estimativas anteriores, isto é, com $R = r - \sigma^2/2 = 0.084/\text{ano}$, $\sigma = 0.193/\sqrt{\text{ano}}$. Correspondem a dois valores de ω simulados. Não sabendo, poderá ser difícil dizer qual a trajectória real e quais são simuladas, o que é uma indicação da razoabilidade do modelo de Black-Scholes como um primeiro modelo para o estudo de cotações de acções. Os estudos mais aprofundados indicam que o comportamento das acções parece diferir, em vários aspectos mais finos, do comportamento do movimento browniamo geométrico. Por exemplo, a distribuição dos retornos $\ln(X(t+\Delta)/X(t))$, que no modelo são v.a. independentes identicamente distribuídas normais, parece ter alguma assimetria e ter caudas mais pesadas do que a normal.

Na Figura 8.1 vemos em simultâneo duas trajectórias simuladas com o mesmo R mas com valores de σ diferentes. A mais "saltitante" usa $\sigma = 0.193/\sqrt{\text{ano}}$, a outra usa um valor 10 vezes menor, $\sigma = 0.0193/\sqrt{\text{ano}}$. De facto, usámos em ambas a mesma trajectória do ruído perturbador (donde a mesma trajectória do processo de Wiener), pelo que as diferenças entre as duas curvas são devidas unicamente à diferente escolha do valor de σ. Assim, o leitor poderá apreciar o papel desempenhado por σ.

Já agora, indicamos ao leitor como foram feitas as simulações. Neste modelo, a tarefa está muito facilitada pois basta simular o processo de Wiener (pelo método descrito na Secção 4.3) uma vez que temos uma solução explícta directamente em termos do processo de Wiener $W(t)$ (ver(8.5)). Chamamos só a atenção que não apresentamos os gráficos de $Z(t)$ mas de $Y(t) = \ln X(t) = Z(t) + x_0$.

Normalmente não se dispõe de uma solução explícita da EDE e, nesse caso, a simulação pelo método de Monte Carlo é ainda mais importante porque permite obter valores aproximados da solução e estudar propriedades da mesma que, de outra forma, não estariam ao nosso alcance. Nestes casos, as técnicas de simulação terão de ser diferentes, inspirando-se nas técnicas de resolução numérica de equações diferenci-

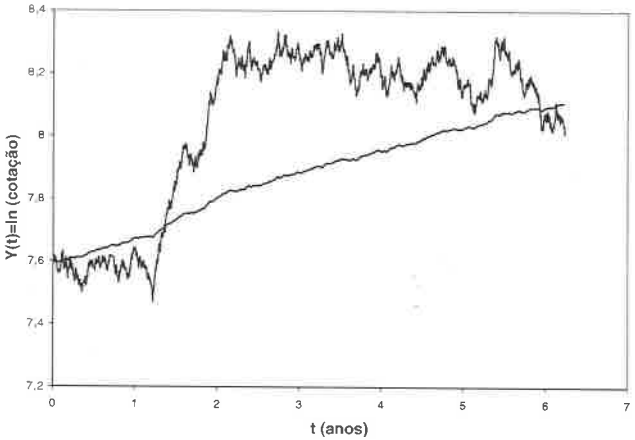


Figura 8.1: Trajectórias simuladas em escala logarítmica do movimento browniano geométrico com $R = r - \sigma^2/2 = 0.084/\text{ano}$ e $\sigma = 0.193/\sqrt{\text{ano}}$ (curva mais "saltitante") ou $\sigma = 0.0193/\sqrt{\text{year}}$ (curva menos "saltitante" usando a mesma trajectória do ruído perturbador).

ais ordinárias (EDO) conjugadas com a simulação das trajectórias do processo de Wiener. A situação, porém, é bastante mais complexa do que nas EDO. O leitor interessado pode consultar [41], [42] e [7].

8.2 Estudo pelo cálculo de Stratonovich

Aqui vamos resolver a EDE com o mesmo aspecto que (8.1) mas usando o cálculo de Stratonovich. O nosso modelo agora é a EDE de Stratonovich

$$(S)\ dX(t) = r_s X(t)dt + \sigma X(t)dW(t), \quad X(0) = x_0. \tag{8.13}$$

Esta equação *não é equivalente* à EDE de Itô (8.1). Por isso usámos r_s em vez de r para facilitar a distinção entre este modelo e o modelo (8.1) do cálculo de Itô (quanto a σ não há necessidade de distinguir).

A solução de (8.13) obtém-se facilmente usando regras usuais de cálculo. Fazendo a mudança de variável (8.3), vem

$$(S)\ dZ(t) = \frac{1}{X(t)} dX(t) = r_s dt + \sigma dW(t), \tag{8.14}$$

ou, na forma integral, $Z(t) = \int_0^t r_s dt + (S)\int_0^t \sigma dW(t)$, donde a solução

almejada

$$Z(t) = r_s t + \sigma W(t) \frown \mathcal{N}(r_s t, \sigma^2 t). \tag{8.15}$$

O aspecto é, de facto, diferente do resultado (8.5)-(8.6) obtido pelo cálculo de Itô. Agora o valor esperado de $Z(t)$ coincide com a solução do modelo determinístico. Mas agora o valor esperado de $X(t)$, $\mathbb{E}[X(t)] = x_0 \exp\left(r_s t + \sigma^2 t/2\right)$, já não coincide com a solução do modelo determinístico.

Agora vem $Z(t) = t\,(r_s + W(t)/t) \sim r_s t$ q.c. Logo, quando $t \to +\infty$, vem

$$X(t) \begin{cases} \to +\infty \text{ q.c.} & \text{se } r_s > 0 \\ \to 0 \text{ q.c.} & \text{se } r_s < 0. \end{cases} \tag{8.16}$$

Este comportamento concorda com o do modelo determinístico, mas discorda do comportamento do modelo (8.1) do cálculo de Itô usado na Secção 8.1. Agora, no caso da aplicação ao crescimento de uma população, sempre que a taxa "média" r_s de crescimento for positiva, a população cresce de forma ilimitada e, sempre que aquela taxa for negativa, a população extingue-se com probabilidade um.

Esta diferença de comportamentos quando se usam cálculos diferentes põe um problema delicado nas aplicações, particularmente quando tem incidência sobre questões tão importantes como a extinção. Qual dos cálculos se deve usar?

Vamos examinar estas questões no Capítulo 9.

Capítulo 9

A questão dos cálculos de Itô e de Stratonovich

9.1 Controvérsia

Verificámos no Capítulo 8, com o modelo de Black-Scholes, que o uso do cálculo de Itô ou do cálculo de Stratonovich na resolução de uma equação diferencial estocástica podia dar resultados diferentes e que os respectivos modelos tinham comportamentos qualitativos diferentes, por exemplo, quando aplicados ao crescimento de populações em ambientes aleatórios, no que dizia respeito à extinção da população.

Estas diferenças são frequentes na literatura aplicada.

Noutros modelos mais realistas de crescimento populacional, como o modelo logístico, há situações em que o cálculo de Stratonovich prevê a não-extinção e a existência de uma densidade estacionária (uma espécie de equilíbrio estocástico) e o modelo de Itô prevê extinção com probabilidade um. Esse facto provocou controvérsias na literatura sobre qual o cálculo mais apropriado nas aplicações. Houve mesmo uma teoria sobre a limitação à sobreposição de nichos ecológicos para uma comunidade de espécies em competição, avançada em 1972 por May e MacArthur (ver [47], [45] e [46]), que (além de outros problemas) falharia se, em vez do cálculo de Itô, se usasse o cálculo de Stratonovich (ver [29] e [57]).

Vimos na Secção 6.7 que o cálculo de Itô parece ser a melhor aproximação quando o fenómeno ocorre intrinsecamente em tempo discreto (com perturbações induzidas por um ruído branco em tempo discreto). Já quando o fenómeno ocorre intrinsecamente em tempo contínuo (com perturbações induzidas por um ruído colorido em tempo contínuo), o cálculo de Stratonovich parece ser a melhor aproximação. Esta discussão

sobre a utilização nas aplicações é tratada em vários livros de texto. Mas, na vida real, não é fácil descortinar qual das duas situações é aplicável.

Com efeito, imaginem que desejam modelar o crescimento de uma população. Se a população tiver gerações sobrepostas com reprodução ocorrendo continuamente no tempo, a indicação será para usar o cálculo de Stratonovich. Se, pelo contrário, existirem épocas reprodutivas periódicas, curtas e bem definidas, ou se as gerações não se sobrepuserem, a indicação será para usar um modelo em tempo discreto. Porém, estas indicações são em boa parte inúteis e pode-se discutir, e discute-se, qual delas é a que mais se aproxima de uma determinada situação real. Devemos lembrar-nos que os nascimentos e mortes são acontecimentos discretos mas que podem ocorrer em qualquer instante num intervalo de tempo contínuo e, mesmo os nascimentos que ocorrem numa curta estação reprodutiva, se espalham de facto ao longo da estação.

Estas diferenças entre os cálculos têm sido um obstáculo à sua aplicação e à credibilidade destes modelos em certas áreas.

Voltemos ao modelo de Black-Scholes (8.1) ou (8.13), conforme se use o cálculo de Itô ou o de Stratonovich. Será que o parâmetro r (que designámos por r_s no cálculo de Stratonovich para ajudar a distinguir), interpretado como taxa "média" de crescimento/rendimento, representa a mesma taxa "média"? Ou será que representa "médias" de tipos diferentes? Veremos que é este o caso e que toda a controvérsia que existiu sobre a questão do cálculo a usar nas aplicações se baseou num equívoco. É que implicitamente a literatura presume que a letra r que usa para ambos os cálculos significa a mesma coisa nos dois cálculos, a chamada taxa "média" de crescimento/rendimento. Também implicitamente se tem presumido que o termo "'média", sem especificação de qual o tipo de média, é inequívoco e que, portanto, a mesma "média" é usada nos dois cálculos. Isso não é verdade e era a fonte de toda a controvérsia.

9.2 Resolução da controvérsia

Vejamos então o que significa taxa de rendimento (ou de crescimento) e taxa "média" de rendimento. Podemos pensar no contexto da cotação de uma acção, ou equivalentemente no valor do capital investido numa acção, para tornar mais vívida a compreensão dos conceitos.

No modelo determinístico $dX(t)/dt = rX(t)$, a taxa de rendimento quando o capital (cotação da acção) é x no instante t é a taxa de crescimento por unidade de capital. No caso do crescimento de uma população, significa a taxa de crescimento *per capita*. Seja qual for a aplicação, ela

é dada por

$$\frac{1}{X(t)}\frac{dX(t)}{dt} = \frac{1}{x}\lim_{\Delta t \downarrow 0}\frac{1}{\Delta t}\left(X(t+\Delta t) - x\right),$$

que é simplesmente r.

No caso estocástico, $X(t+\Delta t)$ é uma v.a. e temos de utilizar uma média. Mas há vários tipos de médias. Se tirarmos a média arutmética, obtemos a *taxa média aritmética de rendimento* quando o capital no instante t é x: [1]

$$r_a = \frac{1}{x}\lim_{\Delta t \downarrow 0}\frac{1}{\Delta t}\left(\mathbb{E}_{t,x}\left[X(t+\Delta t)\right] - x\right). \tag{9.1}$$

Esta média é igual a r se se usar o cálculo de Itô (modelo (8.1)) e é igual a $r_s + \sigma^2/2$ se se usar o cálculo de Stratonovich (modelo (8.13)).

Exercício: Use os resultados obtidos no Capítulo 8 para demonstrar esta afirmação.

Podemos também usar a *taxa média geométrica de rendimento* quando o capital no instante t é x:

$$\begin{aligned} r_g &= \tfrac{1}{x}\lim_{\Delta t \downarrow 0}\tfrac{1}{\Delta t}\left(\exp\left(\mathbb{E}_{t,x}\left[\ln X(t+\Delta t)\right]\right) - x\right) \\ &= \lim_{\Delta t \downarrow 0}\tfrac{1}{\Delta t}\mathbb{E}_{t,x}\left[\ln\tfrac{X(t+\Delta t)}{x}\right]. \end{aligned} \tag{9.2}$$

Vem igual a $R = r - \sigma^2/2$ se se usar o cálculo de Itô e r_s quando se usa o cálculo de Stratonovich.

Exercício: Use os resultados obtidos no Capítulo 8 para demonstrar esta afirmação.

Assim, quando usavam a mesma letra r em ambos os cálculos para se referir à taxa "média" de rendimento, os autores pensavam que estavam a falar da mesma média mas não estavam. *Como vimos, no cálculo de Itô, r significa a taxa média aritmética de rendimento (a taxa média geométrica é $R = r - \sigma^2/2$) e, no cálculo de Stratonovich, significa a taxa média geométrica de rendimento.*

Portanto, se especificarmos qual a média que está em uso, ambos os cálculos produzem exactamente os mesmos resultados qualitativos e

[1] Como estamos a lidar com equações autónomas, esta taxa não depende de t. Neste modelo de Black-Scholes, que é muito simples, também não depende de x, mas isso não é assim noutros modelos autónomos (em que a taxa de rendimento no modelo determinístico também dependa de x).

quantitativos. Nomeadamente, em termos de taxa média geométrica de rendimento (o leitor pode ver o que se passa em termos da taxa média aritmética), ambos os cálculos dão

$$\begin{aligned} Z(t) &= r_g t + \sigma W(t) \\ X(t) &= x_0 \exp(r_g t + \sigma W(t)) \end{aligned} \tag{9.3}$$

e também

$$X(t) \begin{cases} \to +\infty \text{ q.c.} & \text{se } r_g > 0 \\ \to 0 \text{ q.c.} & \text{se } r_g < 0. \end{cases} \tag{9.4}$$

Isto é, no caso do crescimento populacional, há extinção ($X(t) \to 0$ quando $t \to +\infty$) ou crescimento ilimitado ($X(t) \to +\infty$) conforme a taxa média geométrica de crescimento r_g seja negativa ou positiva.

Sendo o fenómeno multiplicativo, é natural que seja o uso da média geométrica que dá resultados qualitativamente análogos ao modelo determinístico e, o que queríamos destacar, resultados idênticos para ambos os cálculos.

A aparente diferença entre os dois cálculos, que gerou muita controvérsia na literatura sobre qual o cálculo mais apropriado em diversas aplicações, era apenas uma confusão semântica. Eliminada a confusão, podemos utilizar indiferentemente qualquer dos dois cálculos, que produzem exactamente os mesmos resultados. Só temos de ter cuidado e usar para r o valor da média apropriada ao cálculo que estamos a usar (a taxa média aritmética de rendimento se usarmos o cálculo de Itô, a taxa média geométrica de rendimento se usarmos o cálculo de Stratonovich). Para mais pormenores, ver [20].

Mas será esta conclusão válida apemas para o modelo de Black-Scholes? Em [21] este resultado de total harmonia entre os dois cálculos é alargado a modelos autónomos bastante gerais da forma

$$dX(t) = f(X(t))X(t)dt + \sigma X(t)dW(t)$$

(com algumas restrições sobre f). De novo, ambos os cálculos dão os mesmos resultados se usarmos para $f(x)$ a expressão da taxa média aritmética ou geométrica de rendimento/crescimento quando o capital é x, isso conforme usemos o cálculo de Itô ou de Stratonovich, respectivamente. Uma extensão posterior (em fase de publicação) a modelos da forma

$$dX(t) = f(X(t))X(t) + h(X(t))X(t)dW(t)$$

(com algumas restrições sobre f e h) requer uma modificação da taxa média geométrica.

Capítulo 10

Estudo de alguns funcionais

10.1 Fórmula de Dynkin

Consideremos uma *equação diferencial estocástica autónoma*

$$dX(t) = f(X(t))dt + g(X(t))dW(t), \quad t \geq s \quad X(s) = x \tag{10.1}$$

com $f(x)$ e $g(x)$ satisfazendo uma condição de Lipschitz, pelo que se verificam as condições para aplicação do teorema de existência e unicidade (ver Secção 7.1), já que a restrição ao crescimento decorre, no caso autónomo, da condição de Lipschitz. Designemos a solução por $X_{s,x}(t)$. A solução é também um processo de difusão homogéneo com coeficiente de tendência

$$a(x) = f(x)$$

e coeficiente de difusão

$$b(x) = g^2(x),$$

a que se chama *difusão de Itô*.

A homogeneidade implica que $X_{s,x}(s+\tau)$ e $X_{0,x}(\tau)$ têm, como processos em $\tau \geq 0$, as mesmas distribuições iniciais e de transição e, portanto, as mesmas distribuições finitas. É assim indiferente usar um ou outro no cálculo de valores esperados. Se trabalharmos com valores esperados condicionais ao valor da solução num determinado instante (tipo $\mathbb{E}_{s,x}$), só interessam as probabilidades de transição e podemos mesmo esquecer qual é a condição inicial, designando então simplesmente por $X(t)$ uma solução genérica da EDE $dX(t) = f(X(t))dt + g(X(t))dW(t)$ sem

especificação *a priori* de qual é a condição inicial no instante 0. Como sabemos, fixada essa condição inicial, a solução é única q.c.

Como só vamos trabalhar com as distribuições de probabilidade, tudo o que se segue neste Capítulo pode ser aplicado a qualquer processo de difusão homogéneo com coeficiente de tendência $a(x) = f(x)$ e coeficiente de difusão $b(x) = g^2(x)$ contínuos, mesmo que f e g não satisfaçam condições de Lipschitz, desde que haja solução fraca e unicidade fraca da correspondente EDE. Sabemos que, se f e g tiverem derivada contínua, existe mesmo uma única solução forte até um instante de explosão, bastando, nesse caso, verificar que o instante de explosão é q.c. infinito.

Seja

$$u(t,x) := \mathbb{E}_x[h(X_t)], \tag{10.2}$$

onde abreviámos $\mathbb{E}[...|X(0) = x] = \mathbb{E}_{0,x}[...]$ para $\mathbb{E}_x[...]$ e onde h é uma função mensurável-Borel contínua limitada.

Se u for contínua e limitada com primeiras e segundas derivadas parciais em x também contínuas e limitadas, então, de (5.12), (5.13) e (5.14), ela satisfaz a *equação de Kolmogorov regressiva* (EKR)

$$\frac{\partial u}{\partial t} = \mathcal{D}u \tag{10.3}$$

com condição terminal

$$\lim_{t \downarrow 0} u(t,x) = h(x),$$

onde

$$\mathcal{D} = a(x)\frac{\partial}{\partial x} + \frac{1}{2}b(x)\frac{\partial^2}{\partial x^2}$$

é o *operador de difusão*.

Vamos definir o *operador infinitésimal* da difusão de Itô $X(t)$:

$$\mathcal{A}h(x) := \lim_{t \downarrow 0} \frac{\mathbb{E}_x[h(X(t)] - h(x)}{t}. \tag{10.4}$$

O conjunto de funções reais $h(x)$ para as quais o limite exista para todo o x será chamado o *domínio* de $\mathcal{A}$. Note-se que, se $h(x)$ é de classe C^2 e tem suporte compacto, então h está no domínio de $\mathcal{A}$ e $\mathcal{A} = \mathcal{D}$. De facto, podemos aplicar o teorema de Itô a $Y(t) = h(X(t))$ e obter

$$\begin{aligned} dY(t) &= \tfrac{\partial h(X(t))}{\partial x} dX(t) + \tfrac{1}{2}\tfrac{\partial^2 h(X(t))}{\partial x^2}(dX(t))^2 \\ &= \left(f(X(t))\tfrac{\partial h(X(t))}{\partial x} + \tfrac{1}{2}g^2(X(t))\tfrac{\partial^2 h(X(t))}{\partial x^2}\right) dt + \\ &\quad \left(g(X(t))\tfrac{\partial h(X(t))}{\partial x}\right) dw(t). \end{aligned}$$

poder aplicar o operador $\mathcal{A} = \mathcal{D}$) com condição terminal

$$u(0^+, x) = h(x).$$

Para demonstrar (10.9), note-se primeiro, atendendo a (10.8) (que resulta da fórmula de Dynkin), que u é diferenciável em t. Basta agora notar que

$$\begin{aligned}\tfrac{1}{r}\left(\mathbb{E}_x[u(t, X(r))] - u(t,x)\right) &= \tfrac{1}{r}\mathbb{E}_x\left[\mathbb{E}_{X(r)}[h(X(t))] - \mathbb{E}_x[h(X(t))]\right] \\ &= \tfrac{1}{r}\mathbb{E}_x\left[\mathbb{E}_x[h(X(t+r))|\mathcal{A}_r] - \mathbb{E}_x[h(X(t))|\mathcal{A}_r]\right] \\ &= \tfrac{1}{r}\mathbb{E}_x[h(X(t+r)) - h(X(t))] = \tfrac{u(t+r,x)-u(t,x)}{r}.\end{aligned}$$

e fazer $r \downarrow 0$. Obtemos à esquerda $\mathcal{A}u$ e à direita $\frac{\partial u}{\partial t}$, o que prova o pretendido. Como u é, como função de x, de classe C^2, vem $\mathcal{A} = \mathcal{D}$.

Note-se que u é limitada de classe $C^{1,2}$. Em [51] mostra-se mesmo que a solução é única no sentido de que, se $u^*(t,x)$ é uma solução limitada de classe $C^{1,2}$ da EKR (10.9) satisfazendo a condição terminal $u^*(0^+, x) = h(x)$, então $u^*(t,x) = u(t,x)$.

Os resultados desta secção, que nos serão particularmente úteis no Capítulo 11, são facilmente generalizáveis ao caso multidimensional.

10.2 Fórmula de Feynman-Kac

A EKR pode ser generalizada, obtendo-se a *fórmula de Feynman-Kac* que, para $h(x)$ de classe C^2 com suporte compacto e $q(x)$ contínua e limitada inferiormente, diz que

$$v(t,x) = \mathbb{E}_x\left[\exp\left(-\int_0^t q(X(s))ds\right) h(X(t))\right] \tag{10.10}$$

satisfaz

$$\frac{\partial v}{\partial t} = \mathcal{D}v - qv \tag{10.11}$$

com a condição terminal

$$v(0^+, x) = h(x).$$

Vamos esboçar a demonstração, seguindo de perto [51]. Seja $Y(t) = h(X(t))$ e $Z(t) = \exp\left(-\int_0^t q(X(s))ds\right)$. Já aplicámos o teorema de Itô a $Y(t)$, obtendo (10.5). Para $Z(t)$ obtém-se $dZ(t) = -Z(t)q(X(t))dt$. Neste caso, também se obtém $d(Y(t)Z(t)) = Y(t)dZ(t) + Z(t)dY(t) + dY(t)dZ(t) = Y(t)dZ(t) + Z(t)dY(t) + 0$ e, como $Y(t)Z(t)$ é um processo de Itô, então $v(t,x) = E_x[Y(t)Z(t)]$ é diferenciável em t (mesmo raciocínio

que antes usando a fórmula de Dynkin). Então

$$\begin{aligned}
&\tfrac{1}{r}\left(\mathbb{E}_x[v(t,X(r))]-v(t,x)\right)= \\
&\tfrac{1}{r}\mathbb{E}_x\left[\mathbb{E}_{X(r)}[Z(t)h(X(t))]-\mathbb{E}_x[Z(t)h(X(t))]\right]= \\
&\tfrac{1}{r}\mathbb{E}_x\left[\mathbb{E}_x\left[h(X(t+r))e^{-\int_0^t q(X(s+r))ds}|\mathcal{A}_r\right]-E_x[Z(t)h(X(t))|\mathcal{A}_r]\right] \\
&=\tfrac{1}{r}\mathbb{E}_x[Z(t+r)\exp\left(\textstyle\int_0^r q(X(\theta))d\theta\right)h(X(t+r))-Z(t)h(X(t))] \\
&=\tfrac{1}{r}\mathbb{E}_x[h(X(t+r)Z(t+r)-h(X(t))Z(t)] \\
&\quad+\tfrac{1}{r}\mathbb{E}_x\left[h(X(t+r))Z(t+r)\left(\exp\left(\textstyle\int_0^r q(X(s))ds\right)-1\right)\right].
\end{aligned}$$

Fazendo $r\downarrow 0$ e notando que $\frac{1}{r}h(X(t+r))Z(t+r)\left(\exp\left(\int_0^r q(X(s))ds\right)-1\right)$ é limitada e converge pontualmente para $h(X(t))Z(t)q(X(0))$, obtém-se $\mathcal{A}v(t,x)=\frac{\partial}{\partial t}v(t,x)+q(x)v(t,x)$. Mais uma vez, temos $\mathcal{A}=\mathcal{D}$.

Note-se que v é de classe $C^{1,2}$ e é limitada em conjuntos da forma $K\times\mathbb{R}$, com K conjunto real compacto. Em [51] mostra-se mesmo que uma solução com estas propriedades é única no sentido de que, se $v^*(t,x)$ for uma outra solução de (10.11) com tais propriedades satisfazendo a condição terminal $v^*(0^+,x)=h(x)$, então $v^*(t,x)=v(t,x)$.

Os resultados desta secção, que nos serão úteis a propósito da fórmula de Black-Scholes, são facilmente generalizáveis ao caso multidimensional.

Capítulo 11

Introdução ao estudo das difusões de Itô unidimensionais

11.1 O processo de Ornstein-Uhlenbeck

No Capítulo 3, vimos que a projecção num eixo do *movimento browniano* de uma partícula suspensa num fluído foi descrita por Einstein através de um processo de Wiener da forma $X(t) = x_0 + \sigma W(t)$, onde x_0 é a coordenada da posição inicial da partícula e σ é o coeficiente de difusão, que nos dá a velocidade de alteração da variância da posição da partícula. Este modelo não tem em conta que a fricção é finita pois a partícula, após um choque com uma molécula do fluído, não pára instantaneamente depois de mudar de posição, antes se move continuamente com velocidade decrescente. No modelo de Einstein, a partícula não tem mesmo velocidade definida (pois não existe derivada, no sentido corrente, do processo de Wiener).

No sentido de melhorar o modelo de Einstein, apareceu em 1930 o *modelo de Ornstein-Uhlenbeck* (ver [58]), que é a primeira EDE que conheço. Neste modelo, considera-se que as forças que actuam sobre a partícula são a força de fricção e a força devida às colisões aleatórias com as moléculas do fluído. Sendo $V(t)$ a velocidade da partícula na direcção de um eixo coordenado e m a massa da partícula, a força $mdV(t)/dt$ que sobre ela actua será dada pela EDE

$$m\frac{dV(t)}{dt} = -fV(t) + \beta\varepsilon(t), \tag{11.1}$$

onde f é o coeficiente de fricção. Pondo $\alpha = f/m$ e $\sigma = \beta/m$, vem

$$dV(t) = -\alpha V(t)dt + \sigma dW(t). \tag{11.2}$$

Dada a velocidade inicial, podemos resolver a equação para obter a velocidade $V(t)$ da partícula (que agora está definida) e, conhecida a posição incial, podemos integrar a velocidade para obter a posição da particula. A solução é o *processo de Ornstein-Uhlenbeck.* [1]

Uma variante deste modelo é usada para modelar a dinâmica das taxas de juro e é também conhecida por *modelo de Vasicek.* O modelo considera um valor de referência $R > 0$ e propõe para a *taxa de juro* $X(t)$ no instante t a EDE

$$dX(t) = -\alpha(X(t) - R)dt + \sigma dW(t), \quad X(0) = x_0, \tag{11.3}$$

com $\alpha > 0$, $\sigma > 0$ e $x_0 > 0$ (determinístico). Note-se que, na ausência de flutuações do mercado ($\sigma = 0$), há uma tendência para que a taxa reverta para o valor de referência R (chama-se-lhe *reversão para a média*) e, de facto, converge para R quando $t \to +\infty$. Quando há flutuações aleatórias do mercado, esta reversão é perturbada.

Esta é uma EDE autónoma com $f(x) = -\alpha(x - R)$ e $g(x) \equiv \sigma$ satisfazendo as condições do teorema de existência e unicidade do Capítulo 7 (basta verificar a condição de Lipschitz, que é óbvia face a estas funções terem derivada contínua limitada). A solução, que sabemos existir e ser única, é uma *difusão de Itô.* Sabemos que a solução é um processo de difusão homogéneo com coeficiente de tendência $a(x) = -\alpha(x - R)$ e coeficiente de difusão $b(x) \equiv \sigma^2$. Como $g(x)$ não depende de x, os cálculos de Itô e Stratonovich coincidem. Assim, podemos usar regras usuais de cálculo.

Seja $V(t) = X(t) - R$. Vem $dV(t) = dX(t) = -\alpha V(t)dt + \sigma dW(t)$ (modelo de Ornstein-Uhlenbeck), donde $e^{\alpha t}dV(t) + \alpha e^{\alpha t}V(t)dt = \sigma e^{\alpha t}dW(t)$. Consequentemente, $d\left(e^{\alpha t}V(t)\right) = \sigma e^{\alpha t}dW(t)$.

Ponhamos $Z(t) = e^{\alpha t}V(t)$ (vem $Z(0) = x_0 - R$) para obter $Z(t) = (x_0 - R) + \int_0^t \sigma e^{\alpha s}dW(s)$. Logo,

$$\begin{aligned} X(t) = V(t) + R = \quad & e^{-\alpha t}Z(t) + R = \\ & e^{-\alpha t}\left((x_0 - R) + \int_0^t \sigma e^{\alpha s}dW(s)\right) + R = \\ & R + e^{-\alpha t}(x_0 - R) + \sigma e^{-\alpha t}\int_0^t e^{\alpha s}dW(s). \end{aligned}$$

[1] Por vezes, reserva-se a designação de processo de Ornstein-Uhlenbeck para uma solução de (11.2) que seja processo estacionário. Para obter essa solução, há que escolher convenientemente a v.a. $V(0)$.

Como a função integranda é determinística, $\int_0^t e^{\alpha s} dW(s)$ tem distribuição

$$\mathcal{N}\left(0, \int_0^t (e^{\alpha s})^2 \, ds\right) = \mathcal{N}\left(0, \frac{1}{2\alpha}\left(e^{2\alpha t} - 1\right)\right).$$

Concluímos que

$$X(t) \frown \mathcal{N}\left(R + e^{-\alpha t}(x_0 - R), \frac{\sigma^2}{2\alpha}\left(1 - e^{-2\alpha t}\right)\right).$$

Como esta *distribuição transiente* converge quando $t \to +\infty$ para a *distribuição estacionária*

$$\mathcal{N}\left(R, \frac{\sigma^2}{2\alpha}\right),$$

chamando $X_{+\infty}$ a uma v.a com esta distribuição, podemos dizer que $X(t)$ converge em distribuição para $X_{+\infty}$. A *densidade estacionária* é a f.d.p. desta distribuição estacionária.

Assim, ao contrário do que sucede no modelo determinístico $dX(t) = -\alpha(X(t) - R)dt$, em que $X(t)$ converge para o ponto de equilíbrio R, no modelo estocástico (11.3) a taxa de juro $X(t)$ flutua aleatoriamente mas a sua distribuição de probabilidade converge para uma distribuição de equilíbrio, a distribuição estacionária acima referida. Podemos falar de um equilíbrio estocástico. O processo $X(t)$ é *ergódico* e, como se referiu na Secção 5.2, podemos estimar a média assintótica R e a variância assintótica $\frac{\sigma^2}{2\alpha}$ usando médias e variâncias temporais ao longo da única trajectória do processo usualmente disponível (a trajectória observada, correspondente ao cenário do mercado ω que efectivamente ocorreu).

A aplicação do modelo (11.3) a taxas de juro tem o problema de $X(t)$ poder ter valores negativos, embora (para os valores típicos dos parâmetros) com probabilidade insignificante. Como curiosidade, referimos que um modelo alternativo em que esse problema é ultrapassado (se R não for demasiado elevado) é o *modelo de Cox-Ingersoll-Ross* (CIR)

$$dX(t) = -\alpha(X(t) - R)dt + \sigma\sqrt{X(t)}dW(t), \quad X(0) = x_0. \qquad (11.4)$$

No modelo (11.3) conseguimos determinar a solução explícita da EDE e a sua f.d.p.. Passando ao limite quando $t \to +\infty$, verificou-se a convergência em distribuição e pudemos determinar a densidade estacionária.

Geralmente, porém, não é possível obter a expressão explícita da solução $X(t)$ de uma difusão de Itô, nem sequer se consegue resolver a equação de Kolmogorov progressiva (que é uma equação às derivadas

parciais) para determinar a f.d.p. de $X(t)$. Pode, no entanto, ser mesmo assim possível determinar a densidade estacionária caso exista. Ela tem de ser densidade invariante e, portanto, é solução de (5.20), que é uma equação diferencial ordinária muito mais fácil de resolver. Isso não significa que uma solução de (5.20) seja densidade estacionária, no sentido de $X(t)$ convergir em distribuição para uma distribuição com essa densidade. Isso terá de ser deduzido por outra via, nomeadamente estudando o comportamento da solução da EDE nas fronteiras do espaço de estados.

Uma introdução ao estudo das difusões de Itô a uma dimensão será o objecto das próximas Secções. O estudo, infelizmente, não é extensível, pelo menos nos moldes em que aqui é apresentado, a várias dimensões.

11.2 Tempo de saída de um intervalo

Consideremos um processo de difusão homogéneo $X(t)$ $(t \in [0, +\infty))$ com coeficiente de tendência

$$a(x) = f(x)$$

e coeficiente de difusão

$$b(x) = g^2(x),$$

ambos funções contínuas de x, cujo espaço de estados seja um intervalo (aberto, fechado ou semiaberto) de extremo inferior r_1 e extremo superior r_2, com $-\infty \leq r_1 < r_2 \leq +\infty$. Qualquer das fronteiras r_1 e r_2 pode pertencer ou não ao espaço de estados. Vamos supor que $b(x) > 0$ para $x \in (r_1, r_2)$ e que o processo é solução fraca, fracamente única, de uma EDE $dX(t) = f(X(t))dt + g(X(t))dW(t)$ com condição inicial $X(0)$. Vamos supor ainda que o processo é *regular*, isto é, que

$$P_x[T_y < +\infty] := P[T_y < +\infty | X(0) = x] > 0 \quad (r_1 < x, y < r_2),$$

onde

$$T_y = \inf\{t \geq 0 : X(t) = y\} \tag{11.5}$$

é o tempo de primeira passagem por y (quando o processo não passa por y, vem $T_y = +\infty$); basicamente, isso significa que todos os estados do interior do espaço de estados comunicam entre si (pode-se transitar com probabilidade positiva entre quaisquer dois pontos do interior do espaço de estados).

Vamos considerar dois pontos a e b do espaço de estados tal que $r_1 < a < x < b < r_2$ e designemos por

$$T_{a,b} = min\{T_a, T_b\} \tag{11.6}$$

o instante em que o processo atinge a ou b pela primeira vez, ou seja, sai do intervalo (a, b).

Teorema. *Seja h uma função contínua limitada e*

$$v_h(x) := \mathbb{E}_x \left[\int_0^{T_{a,b}} h(X(s))ds \right]. \tag{11.7}$$

Seja $v(x)$ $(x \in [a, b])$ uma solução da equação diferencial

$$\mathcal{D}v(x) = -h(x) \tag{11.8}$$

tal que $v(a) = v(b) = 0$. Então $v_h(x) < +\infty$ e $v_h(x) = v(x)$.

Em particular, fazendo $h(x) \equiv 1$, concluímos que $\mathbb{E}_x [T_{a,b}] < +\infty$ (o que implica que $T_{a,b}$ é q.c. finito) e é dada pela solução da equação diferencial $\mathcal{D}v(x) = -1$ tal que $v(a) = v(b) = 0$.

Demonstração

Seja $\tau = min\{T, T_{a,b}\}$ com $T > 0$ fixo. Vem obviamente $\mathbb{E}_x[\tau] < +\infty$ e podemos aplicar a fórmula de Dynkin (10.7) a v, obtendo

$$\mathbb{E}_x [v(X(\tau))] = v(x) + \mathbb{E}_x \left[\int_0^{\tau} \mathcal{D}v(X(s))ds \right].$$

Para $s \leq \tau$, vem $X(s) \in [a, b]$ e, portanto, $\mathcal{D}v(X(s)) = -h(X(s))$, pelo que

$$\mathbb{E}_x [v(X(\tau))] = v(x) - \mathbb{E}_x \left[\int_0^{\tau} h(X(s))ds \right]. \tag{11.9}$$

Quando $T \uparrow +\infty$, vem $\tau \uparrow T_{a,b}$. Como $\mathbb{E}_x \left[\int_0^{\tau} h(X(s))ds \right] = v(x) - \mathbb{E}_x [v(X(\tau))]$ é uniformemente limitada em T, vem $\mathbb{E}_x \left[\int_0^{\tau} h(X(s))ds \right] \to \mathbb{E}_x \left[\int_0^{T_{a,b}} h(X(s))ds \right] < +\infty$. Também vem $X(\tau) \to X(T_{a,b}) = a$ ou b e, portanto, $v(X(\tau)) \to 0$. Logo, de (11.9), vem $\mathbb{E}_x \left[\int_0^{T_{a,b}} h(X(s))ds \right] = v(x)$, como pretendido. ■

Teorema. *Seja $u(x)$ $(x \in [a, b])$ uma solução da equação diferencial*

$$\mathcal{D}u(x) = 0 \tag{11.10}$$

tal que $u(a) = 0$ e $u(b) = 1$. Então

$$P_x [T_b < T_a] := P [T_b < T_a | X(0) = x] = u(x). \tag{11.11}$$

Demonstração

Façamos τ como na demonstração do teorema anterior e apliquemos a fórmula de Dynkin (10.7) a u. Vem

$$\mathbb{E}_x\left[u(X(\tau))\right] = u(x) + \mathbb{E}_x\left[\int_0^\tau \mathcal{D}u(X(s))ds\right]. \tag{11.12}$$

Para $s \leq \tau$, vem $X(s) \in [a,b]$ e, portanto, $\mathcal{D}u(X(s)) = 0$, pelo que $\mathbb{E}_x\left[u(X(\tau))\right] = u(x)$. Fazendo $T \uparrow +\infty$, vem $\mathbb{E}_x\left[u(X(T_{a,b}))\right] = u(x)$. Como $X(T_{a,b})$ só pode tomar os valores a (quando $T_a < T_b$) ou b (quando $T_b < T_a$), vem $\mathbb{E}_x\left[u(X(T_{a,b}))\right] = u(a)(1 - P_x[T_b < T_a]) + u(b)P_x[T_b < T_a] = P_x[T_b < T_a] = u(x)$ como pretendido. ■

Para obter expressões explícitas, temos que resolver as equações diferenciais que surgem nos dois teoremas anteriores.

Para isso vamos seguir de perto [39] e utilizar duas medidas absolutamente contínuas com respeito à medida de Lebesgue, a *medida de escala* S e a *medida de velocidade* M. As suas densidades estão definidas para $\xi \in (r_1, r_2)$ a menos de uma constante multiplicativa (pois $\xi_0 \in (r_1, r_2)$ é arbitrariamente escolhido) por

$$s(\xi) = \exp\left(-\int_{\xi_0}^{\xi} \frac{2a(\eta)}{b(\eta)}d\eta\right) \tag{11.13}$$

(*densidade de escala*) e

$$m(\xi) = \frac{1}{b(\xi)s(\xi)} \tag{11.14}$$

(*densidade de velocidade*). Podemos então definir, a menos de constantes multiplicativa e aditiva (pois $x_0 \in (r_1, r_2)$ é arbitrariamente escolhido), as "funções de distribuição"

$$S(x) = \int_{x_0}^{x} s(\xi)d\xi \tag{11.15}$$

e

$$M(x) = \int_{x_0}^{x} m(\xi)d\xi, \tag{11.16}$$

a que chamamos, respectivamente, *função de escala* e *função de velocidade*. As medidas de densidade e escala podem então ser definidas (a menos de uma constante multiplicativa) para intervalos semiabertos por

$$S(a,b] = S(b) - S(a) \tag{11.17}$$

e

$$M(a,b] = M(b) - M(a) \tag{11.18}$$

e, para os outros borelianos do espaço de estados, por extensão.

Para uma função $h(x)$ de classe C^2, temos

$$\mathcal{D}h(x) = \frac{1}{2}\left(\frac{1}{m(x)}\right)\frac{d}{dx}\left(\frac{1}{s(x)}\frac{dh(x)}{dx}\right),$$

donde

$$\mathcal{D}h(x) = \frac{1}{2}\frac{d}{dM(x)}\left(\frac{dh(x)}{dS(x)}\right). \tag{11.19}$$

Para obter $u(x)$, podemos escrever (11.10) na forma $\frac{d}{dM(x)}\left(\frac{du(x)}{dS(x)}\right) = 0$. Integrando com respeito a M e depois com respeito a S, obtemos $u(x) = B + AS(x)$ com A e B constantes. Usando $u(a) = 0$ e $u(b) = 1$, vem

$$u(x) := P_x[T_b < T_a] = \frac{S(x) - S(a)}{S(b) - S(a)}. \tag{11.20}$$

Exercício: Escolha uma das funções de escala $S(x)$ e faça $Y(t) = S(X(t))$. Use o teorema de Itô e determine os coeficientes de tendência e difusão de $Y(t)$. Note que o coeficiente de tendência é nulo (no Capítulo 12 veremos outro método de anular o coeficiente de tendência) e, portanto, $S_Y(y) = y$ é uma função de escala de $Y(t)$. Logo, supondo $Y(0) = y$, a probabilidade de $Y(t)$ passar por b antes de passar por a é dada por $\frac{y-a}{b-a}$ (a proporção das distâncias). Diz-se que $Y(t)$ está em *escala natural*. A razão da designação "função de escala" provém de S transformar uma difusão de Itô numa difusão em escala natural.

Para obter $v_h(x)$, que é solução de (11.8), podemos escrever esta equação na forma $\frac{d}{dM(x)}\left(\frac{dv_h(x)}{dS(x)}\right) = -2h(x)$. Integrando com respeito a M, obtemos $\frac{dv_h(\eta)}{dS(\eta)} = -2\int_a^\eta h(\xi)dM(\xi) + A = -2\int_a^\eta h(\xi)m(\xi)d\xi + A$ (A constante). Integrando agora com respeito a S, obtemos

$$v_h(x) = -2\int_a^x \int_a^\eta h(\xi)m(\xi)d\xi dS(\eta) + A(S(x) - S(a)) + B$$

com A e B constantes. Usando $v_h(a) = 0$ e $v_h(b) = 0$, vem $B = 0$ e

$A = \frac{2}{S(b)-S(a)} \int_a^b \int_a^\eta h(\xi)m(\xi)d\xi dS(\eta)$. Logo

$$v_h(x) = 2\left(u(x)\int_a^b \int_a^\eta h(\xi)m(\xi)d\xi dS(\eta) - \int_a^x \int_a^\eta h(\xi)m(\xi)d\xi dS(\eta)\right).$$

Uma mudança de ordem de integração e algumas contas dão:

$$\begin{aligned} v_h(x) := \quad & \mathbb{E}_x\left[\int_0^{T_{a,b}} h(X(s))ds\right] = \\ & 2u(x)\int_x^b (S(b)-S(\xi))m(\xi)h(\xi)d\xi + \\ & 2(1-u(x))\int_a^x (S(\xi)-S(a))m(\xi)h(\xi)d\xi. \end{aligned} \tag{11.21}$$

Como já referido, fazendo $h(x) \equiv 1$, obtemos $\mathbb{E}_x[T_{a,b}]$.

Para interpretar a medida de velocidade, consideremos um processo em escala natural (S(x)=x), seja $\varepsilon > 0$ e façamos $a = x - \varepsilon$, $b = x + \varepsilon$. De (11.20) vem $u(x) = P_x[T_{x-\varepsilon,x+\varepsilon}] = 1/2$ e, de (11.21) com $h(x) \equiv 1$, vem

$$\mathbb{E}_x[T_{x-\varepsilon,x+\varepsilon}] = \int_x^{x+\varepsilon} (x+\varepsilon-\xi)m(\xi)d\xi + \int_{x-\varepsilon}^x (\xi-x+\varepsilon)m(\xi)d\xi,$$

pelo que

$\lim_{\varepsilon\downarrow 0} \frac{1}{\varepsilon^2}\mathbb{E}_x[T_{x-\varepsilon,x+\varepsilon}] =$
$\lim_{\varepsilon\downarrow 0} \frac{1}{\varepsilon^2}\int_x^{x+\varepsilon}(x+\varepsilon-\xi)m(\xi)d\xi + \frac{1}{\varepsilon^2}\int_{x-\varepsilon}^x(\xi-x+\varepsilon)m(\xi)d\xi = m(x).$

Isto mostra que, para ε pequeno, se o processo estiver em escala natural e estiver no ponto x, vai em média estar no intervalo $(x-\varepsilon, x+\varepsilon)$ um tempo $m(x)\varepsilon^2$. Teria talvez sido mais razoável chamar a M "medida de ocupação" mas não foi essa a designação consagrada.

Podemos generalizar o primeiro teorema desta secção de forma a conseguirmos determinar momentos de ordem superior de $T_{a,b}$.

Teorema. *Seja h uma função contínua limitada, q uma função de classe C^2 e*

$$U_{q,h}(x) := \mathbb{E}_x\left[q\left(\int_0^{T_{a,b}} h(X(s))ds\right)\right]. \tag{11.22}$$

Seja $U(x)$ $(x \in [a,b])$ uma solução da equação diferencial

$$\mathcal{D}U(x) = -h(x)V(x) \tag{11.23}$$

tal que $U(a) = U(b) = q(0)$, com $V(x) := \mathbb{E}_x\left[q'\left(\int_0^{T_{a,b}} h(X(s))ds\right)\right]$

(onde q' representa a derivada de q). Então $U_{q,h}(x) < +\infty$ e vem $U_{q,h}(x) = U(x)$.

Em particular, fazendo $h(x) \equiv 1$, $q(x) = x^n$ e $U_n(x) := \mathbb{E}_x\left[(T_{a,b})^n\right]$ $(n = 1, 2, ...)$, concluímos que $\mathbb{E}_x\left[(T_{a,b})^n\right] < +\infty$ é dada pela solução da equação diferencial $\mathcal{D}U_n(x) = -nU_{n-1}(x)$ tal que $U_n(a) = U_n(b) = 0$.

Demonstração

A técnica é semelhante à dos teoremas anteriores, pelo que vamos saltar a passagem de começar com $\tau = min\{T, T_{a,b}\}$ e fazer depois $T \uparrow +\infty$, bem como a demonstração de que a esperança é finita. Aplicando a fórmula de Dynkin (10.7) a U, vem

$$\mathbb{E}_x\left[U(X(T_{a,b})\right] = U(x) + \mathbb{E}_x\left[\int_0^{T_{a,b}} \mathcal{D}U(X(s))ds\right]. \qquad (11.24)$$

Atendendo a (11.23) e designado por $T^y_{a,b}$ o tempo de primeira passagem por a ou b quando o processo se inicia no instante 0 com o valor y (omite-se y quando $y = x$), vem

$$\mathbb{E}_x\left[\int_0^{T_{a,b}} \mathcal{D}U(X(s))ds\right] = -\mathbb{E}_x\left[\int_0^{T_{a,b}} h(X(s))V(X(s))ds\right] =$$
$$-\mathbb{E}_x\left[\int_0^{T_{a,b}} h(X(s))\mathbb{E}_{X(s)}\left[q'\left(\int_0^{T^{X(s)}_{a,b}} h(X(u))du\right)\right]ds\right] =$$
$$-\mathbb{E}_x\left[\int_0^{T_{a,b}} h(X(s))q'\left(\int_s^{T^{X(s)}_{a,b}+s} h(X(u))du\right)ds\right] =$$
$$\mathbb{E}_x\left[-\int_0^{T_{a,b}} h(X(s))q'\left(\int_s^{T_{a,b}} h(X(u))du\right)ds\right] =$$
$$\mathbb{E}_x\left[\int_0^{T_{a,b}} \frac{dq\left(\int_s^{T_{a,b}} h(X(u))du\right)}{ds}ds\right] = q(0) - \mathbb{E}_x\left[q\left(\int_0^{T_{a,b}} h(X(u))du\right)\right]$$

Substituindo em (11.24) e atendendo a que $\mathbb{E}_x\left[U(X(T_{a,b})\right] = q(0)$ (já que $U(a) = U(b) = q(0)$), obtemos o resultado desejado

$$\mathbb{E}_x\left[q\left(\int_0^{T_{a,b}} h(X(u))du\right)\right] = U(x).\blacksquare$$

Exercício: Determine a expressão de $\mathbb{E}_x\left[(T_{a,b})^2\right]$ em termos das medidas de densidade e escala.

Exercício: A partir do teorema anterior, obtenha uma equação diferencial para $R(\lambda, x) = \mathbb{E}_x\left[\exp\left(-\lambda T_{a,b}\right)\right]$ (transformada de Laplace da den-

sidade do tempo de primeira passagem). Um método mais directo pode ser visto em [25]. Note que $\frac{\partial^n R(0,x)}{\partial \lambda^n} = (-1)^n \mathbb{E}_x \left[(T_{a,b})^n\right]$ $(n = 1, 2, ...)$, o que permite obter os momentos de $T_{a,b}$. Também a f.d.p. de $T_{a,b}$ pode ser obtida por inversão da transformada de Laplace, embora sejam raros os casos em que se conseguem expressões explícitas.

11.3 Comportamento nas fronteiras de difusões de Itô e densidades estacionárias

Aqui seguimos de perto [39]. Consideremos o enquadramento previsto no início da Secção 11.2. Em particular, supomos $X(0) = x \in (r_1, r_2)$.

Seja $T_{r_1^+} = \lim_{a \downarrow r_1} T_a$ e $T_{r_2^-} = \lim_{b \uparrow r_2} T_b$; escolhendo $a < x$ e $b > x$, como $T_a \leq T_{r_1}$ e $T_b \leq T_{r_2}$, claro que $T_{r_1^+} \leq T_{r_1}$ e $T_{r_2^-} \leq T_{r_2}$. Se $T_{r_1^+} = +\infty$, então $T_{r_1} = +\infty = T_{r_1^+}$. Se $T_{r_1^+} < +\infty$, dada a continuidade das trajectórias, vem $X(T_{r_1^+}) = \lim_{a \downarrow r_1} X(T_a) = \lim_{a \downarrow r_1} a = r_1$ donde $T_{r_1} \leq T_{r_1^+}$ e, portanto, $T_{r_1} = T_{r_1^+}$. Logo, em qualquer caso, não há necessidade de distinguir $T_{r_1^+}$ de T_{r_1}. Por razões análogas, vem $T_{r_2} = T_{r_2^-}$. Note-se que, se r_1 não estiver no espaço de estados, vem necessariamente $T_{r_1} = +\infty$.

Caso $S(r_1, x_0] = \int_{r_1}^{x_0} s(\xi) d\xi = +\infty$ para algum $x_0 \in (r_1, r_2)$, então

$$P_x \left[T_{r_1} < T_b\right] = 0 \quad \text{para todo o } r_1 < x < b < r_2 \tag{11.25}$$

e a fronteira r_1 diz-se *não-atractiva* pois o processo passa primeiro por qualquer ponto à esquerda do ponto de partida x antes de poder passar por r_1. Com efeito, de (11.20), vem $P_x[T_{r_1} < T_b] = \frac{S(b) - S(x)}{S(b) - S(r_1)} = \frac{S(x,b]}{S(r_1,b]} = 0$ pois $S(x, b] = \int_x^b s(\xi) d\xi > 0$ (visto $s(\xi) > 0$) e $S(r_1, b] = S(r_1, x_0] + S(x_0, b] = +\infty$. Aliás, se $S(r_1, x_0]$ é infinito [finito] para algum $x_0 \in (r_1, r_2)$, então é infinito [finito] para qualquer $x_0 \in (r_1, r_2)$.

O caso contrário ao anterior dá-se quando $S(r_1, x_0] = \int_{r_1}^{x_0} s(\xi) d\xi < +\infty$ para algum $x_0 \in (r_1, r_2)$, vindo então

$$P_x \left[T_{r_1} \leq T_b\right] > 0 \quad \text{para todo o } r_1 < x < b < r_2 \tag{11.26}$$

e a fronteira r_1 diz-se *atractiva*. Com efeito, $P_x[T_{r_1} < T_b] = \frac{S(x,b]}{S(r_1,b]} > 0$ pois $S(x, b] > 0$ e $S(r_1, b] < +\infty$.

Vimos assim que a condição necessária e suficiente para que a fronteira r_1 seja atractiva é que $s(x)$ seja integrável numa vizinhaça direita de r_1. Analogamente, poderíamos mostrar que uma condição necessária e suficiente para que a fronteira r_2 seja atractiva (a definição é seme-

lhante) é que $s(x)$ seja integrável numa vizinhança esquerda de r_2, isto é, $S[x_0, r_2) = \int_{x_0}^{r_2} s(\xi)d\xi < +\infty$ para algum $x_0 \in (r_1, r_2)$.

Uma fronteira atractiva pode não estar no espaço de estados e pode até suceder que a probabilidade de a atingir em tempo finito seja nula (fronteira inantigível). A fronteira r_1 diz-se *atingível* se a probabilidade de poder ser atingida em tempo finito for positiva, isto é, se

$$P_x[T_{r_1} < +\infty] > 0 \quad \text{para todo o } r_1 < x < r_2. \tag{11.27}$$

Se a fronteira r_1 é atingível, ela satisfaz a propriedade $\mathbb{E}_x[T_{r_1,b}] < +\infty$ mas nada garante que $\mathbb{E}_x[T_{r_1}]$ seja finita. Pode provar-se (ver [39]) que uma fronteira não-atractiva não é atingível (tendo-se $P_x[T_{r_1} < +\infty] = 0$ para qualquer $x \in (r_1, r_2)$) e que nem sequer é atingível em tempo infinito (isto é, $P_x[\lim_{t\uparrow+\infty} X(t) = r_1] = 0$ para qualquer $x \in (r_1, r_2)$). Já uma fronteira atractiva, seja ou não atingível, pode sempre ser atingida em tempo infinito (isto é, $P_x[\lim_{t\uparrow+\infty} X(t) = r_1] > 0$ para qualquer $x \in (r_1, r_2)$).

Uma fronteira não-atingível não faz obviamente parte do espaço de estados.

Uma condição necessária e suficiente para que a fronteira r_1 seja atingível é (ver [39]) que $\Sigma(r_1) := \int_{r_1}^{x_0} S(r_1, \xi]dM(\xi) < +\infty$ (depende de $x_0 \in (r_1, r_2)$ mas o valor de x_0 é irrelevante para saber se a quantidade é ou não finita). Note-se que, trocando a ordem de integração, se pode obter uma expressão alternativa para $\Sigma(r_1)$:

$$\begin{array}{rl} \Sigma(r_1) := & \int_{r_1}^{x_0} S(r_1, \xi]dM(\xi) = \int_{r_1}^{x_0} \int_{r_1}^{\xi} s(\eta)d\eta m(\xi)d\xi = \\ & \int_{r_1}^{x_0} \int_{\eta}^{x_0} m(\xi)d\xi s(\eta)d\eta = \int_{r_1}^{x_0} M(\eta, x_0]dS(\eta). \end{array} \tag{11.28}$$

A classificação das fronteiras envolve ainda outros aspectos interessantes. Em particular, se a fronteira é atingível, há que especificar, para além dos coeficientes de tendência e difusão, o comportamento na fronteira (absorção, reflexão, reflexão pegajosa, etc.). Para mais pormenores, consulte-se [39].

Caso r_1 e r_2 sejam ambas não-atractivas, podemos, em linguagem informal, dizer que há uma tendência para as trajectórias que se aproximam das fronteiras serem empurradas para o interior do espaço de estados, pelo que, como todos os estados interiores comunicam entre si, a distribuição transiente (distribuição de $X(t)$) poderá ter uma densidade $p(t, y) = f_{X(t)}(y)$ $(r_1 < y < r_2)$. A questão é se a distribuição transiente estabiliza no sentido de convergir quando $t \to +\infty$ para uma distribuição limite (seja $X_{+\infty}$ uma v.a. com essa distribuição) com f.d.p. $p(y) = f_{X_{+\infty}}(y)$ $(r_1 < y < r_2)$, a chamada *densidade estacionária*. Esta

é uma possibilidade real dado o processo de difusão ser homogéneo e, portanto, haver regras de transição entre estados independentes do tempo. Claro que o facto de as fronteiras serem não-atractivas é importante, caso contrário as trajectórias, ao serem atraídas para uma ou ambas as fronteiras, fariam com que a distribuição de probabilidade se concentrasse progressivamente numa ou em ambas as fronteiras, inviabilizando a existência de uma densidade limite no interior do espaço de estados. Porém, a condição de as fronteiras serem não-atractivas não é suficiente. Claro que, a existir densidade estacionária, espera-se intuitivamente que ela seja proporcional ao tempo de ocupação de cada estado, isto é, seja proporcional a $m(y)$, o que só será possível se $\int_{r_1}^{r_2} m(\xi)d\xi$ for finito.

Suponhamos então que r_1 e r_2 são fronteiras não-atractivas. A existir densidade estacionária $p(y)$, ela terá de ser densidade invariante e, portanto, satisfazer (5.20), que também podemos escrever na forma $\frac{d(a(y)p(y))}{dy} - \frac{1}{2}\frac{d^2(b(y)p(y))}{dy^2}$. Integrando esta equação em ordem a y, vem $-2a(y)p(y) + \frac{d(b(y)p(y))}{dy} = C$ (C constante). Multipliquemos membro a membro pelo factor integrante $s(y)$. Vem, após cálculos simples, $b(y)p(y)\frac{ds(y)}{dy} + \frac{d(b(y)p(y))}{dy}s(y) = Cs(y)$, pelo que $\frac{d(b(y)p(y)s(y))}{dy} = Cs(y)$. Integrando de novo, vem $b(y)p(y)s(y) = CS(y) + D$ (D constante), ou seja

$$p(y) = m(y)(CS(y) + D).$$

Como as fronteiras são não-atractivas, $s(y)$ é não-integrável na sua vizinhança e, portanto, $S(y) \uparrow +\infty$ quando $y \uparrow r_2$ e $S(y) \downarrow -\infty$ quando $y \downarrow r_1$ (daqui resulta, em particular, que a difusão de Itô $Y(t) = S(X(t))$, cujo coeficiente de tendência é nulo, tem como espaço de estados $(-\infty, +\infty)$). Só é possível manter $p(y)$ positivo para quase todo o $y \in (r_1, r_2)$ (indispensável para ser uma f.d.p.) se $C = 0$. Logo a existir densidade estacionária $p(y)$ ($y \in (r_1, r_2)$), ela terá de ser da forma

$$p(y) = Dm(y),$$

com D convenientemente escolhido de forma a que $\int_{r_1}^{r_2} p(\xi)d\xi = 1$. Para isso ser possível e $p(y)$ ser uma f.d.p. é indispensável que

$$M(r_1, r_2) := \int_{r_1}^{r_2} m(\xi)d\xi < +\infty,$$

caso em que vem

$$p(y) = \frac{m(y)}{\int_{r_1}^{r_2} m(\xi)d\xi} \qquad (r_1 < y < r_2). \tag{11.29}$$

Mas existir densidade invariante não garante que ela seja estacionária (densidade da distribuição limite quando $t \to +\infty$). Pode ver-se em [34] a demonstração de que, efectivamente, quando r_1 e r_2 são não-atractivas e $m(\xi)$ é integrável (isto é, $\int_{r_1}^{r_2} m(\xi)d\xi < +\infty$), então existe *densidade estacionária*, que é dada pela expressão (11.29), e o processo estocástico é mesmo *ergódico*. Por processo ergódico entende-se aquele em que as médias temporais de certas funções ao longo de uma trajectória convergem para a esperança matemática (média no conjunto das trajectórias) da mesma função, esperança essa calculada relativamente à distribuição estacionária. Em particular, temos que, se $h(x)$ for uma função mensurável-Borel limitada e se $X(0) = x \in (r_1, r_2)$, então (note que o limite é em média quadrática)

$$\underset{T\to+\infty}{\text{l.i.m.}} \frac{1}{T}\int_0^T h(X(t))dt = \mathbb{E}_x[h(X_{+\infty})] = \int_{r_1}^{r_2} h(y)p(y)dy. \qquad (11.30)$$

Para obter ergodicidade q.c. em vez de ergodicidade em m.q., isto é para que

$$\lim_{T\to+\infty} \frac{1}{T}\int_0^T h(X(t))dt = \mathbb{E}_x[h(X_{+\infty})] = \int_{r_1}^{r_2} h(y)p(y)dy \quad \text{q.c.}, \qquad (11.31)$$

nem é preciso exigir que h seja limitada, bastando que $\mathbb{E}_x[|h(X_{+\infty})|] = \int_{r_1}^{r_2} |h(y)|p(y)dy < +\infty$.

Um exemplo de EDE autónoma sem densidade estacionária é o *modelo de Black-Scholes* (8.1), em que supomos $X(0) = x > 0$ (vê-se que o espaço de estados é $(0, +\infty)$). De facto, se $R := r - \sigma^2/2 > 0$ [$R < 0$], vem $X(t) \to +\infty$ [$X(t) \to 0$, caso em que a distribuição limite está concentrada no ponto 0 e não tem densidade]. Se quisermos usar os métodos desta Secção, vemos que $a(x) = rx$, $b(x) = \sigma^2x^2$, $s(\xi) = \xi^{-2r/\sigma^2}$ (escolhendo $\xi_0 = 1$), $m(\xi) = \frac{1}{\sigma^2}\xi^{2r/\sigma^2-2}$. Vê-se que, se $R > 0$, vem $s(\xi)$ não-integrável na vizinhança direita da fronteira 0 e integrável na vizinhança esquerda da fronteira $+\infty$, pelo que 0 é não-atractiva e $+\infty$ é atractiva. Quando $R < 0$ passa-se precisamente o contrário. Quando $R = 0$, ambas as fronteiras são não-atractivas, mas $m(x)$ não é integrável e, portanto, não existe densidade estacionária; aliás, de (8.6), sabemos que $\ln(X(t)/x)$ tem distribuição normal com média $Rt = 0$ e variância $\sigma^2 t \to +\infty$ quando $t \to +\infty$.

Um exemplo de EDE autónoma com densidade estacionária (que já determinámos na Secção 11.1) é o *modelo de Vasicek*, cujo espaço de estados é $(-\infty, +\infty)$. Utilizando os métodos desta Secção, temos $a(x) =$

$-\alpha(x-R)$, $b(x)=\sigma^2$, $s(\xi)=\exp\left(\frac{\alpha}{\sigma^2}(\xi-R)^2\right)$ (escolhendo $\xi_0=R$) e $m(\xi)=\frac{1}{\sigma^2}\exp\left(-\frac{(\xi-R)^2}{2(\sigma^2/(2\alpha))}\right)$. Como $s(\xi)\to+\infty$ quando $\xi\to\pm\infty$, resulta que não é integrável em nenhuma das duas fronteiras $r_1=-\infty$ e $r_2=+\infty$, pelo que estas são não-atractivas. É óbvio que $m(y)$ é integrável pois, à parte uma constante multiplicativa, é a densidade de uma distribuição normal com média R e variância $\sigma^2/(2\alpha)$, distribuição que será assim a distribuição estacionária, como já se tinha visto na Secção 11.1.

11.4 Um exemplo de aplicação em crescimento populacional

O modelo de crescimento malthusiano $dX(t)/dt=rX(t)$ para o crescimento de uma população pressupõe que a taxa de crescimento *per capita* $\frac{1}{X(t)}\frac{dX(t)}{dt}$ é a constante r, ou seja, não há escassez de recursos a afectar a natalidade e a mortalidade mesmo quando a população é grande. Em ambiente aleatório, este modelo originou o modelo de crescimento populacional (8.1), conhecido na literatura financeira por modelo de Black-Scholes. No entanto, se os resursos forem limitados, é mais realista supor que os recursos disponíveis por indivíduo decresçam à medida que a população cresce, o que afectará a natalidade e/ou mortalidade, levando a que a taxa de crescimento *per capita* decresça quando a população aumenta. O modelo mais simples que incorpora este efeito é o conhecido *modelo logístico* ou de *Pearl-Verhulst*, em que a taxa de crescimento *per capita* decresce linearmente com o tamanho da população, isto é, $\frac{1}{X(t)}\frac{dX(t)}{dt}=r-aX(t)$ $(a>0)$. Agora r é a taxa máxima de crescimento *per capita*, só aproximada quando a população é muito pequena e não se faz, portanto, sentir a limitação dos recursos. Para dar a ideia de que representa a potencial taxa caso não houvesse limitação de recursos, chama-se a r *taxa intrínseca de crescimento.* Pondo $K=r/a$, vem

$$\frac{dX(t)}{dt}=rX(t)\left(1-\frac{X(t)}{K}\right),\qquad X(0)=x>0. \tag{11.32}$$

Supomos que $r>0$ e $K>0$. A taxa de crescimento *per capita* quando a população tem o tamanho y é agora $r(1-y/K)$. É fácil ver que os pontos de equilíbrio são $X=0$, que é instável, e $X=K$, que é globalmente assintoticamente estável, verificando-se que $X(t)\to K$ quando $t\to+\infty$, ou seja, K é a população que é sustentável em equilíbrio. Chama-se, por isso, a K, *capacidade de sustento do meio.* Se $0<x<K$, $X(t)$ é crescente e, se $0<x<K/2$, o gráfico de $X(t)$ como

função de t tem a conhecida forma sigmóide (curva logística) com ponto de inflexão de ordenada $K/2$. Se $x > K$, $X(t)$ é decrescente.

Se o ambiente está sujeito a flutuações aleatórias, que podemos aproximar por um ruído branco $\sigma\varepsilon(t)$ (com $\sigma > 0$ e ε um ruído branco padrão), e as flutuações afectam a taxa de crescimento *per capita*, vem a EDE de Stratonovich (estamos a supor que o ruído real é colorido e o aproximamos por um ruído branco)

$$(S) \quad \frac{1}{X(t)}\frac{dX(t)}{dt} = r\left(1 - \frac{X(t)}{K}\right) + \sigma\varepsilon(t).$$

Escrevendo na forma mais habitual, temos

$$(S) \quad dX(t) = rX(t)\left(1 - \frac{X(t)}{K}\right)dt + \sigma X(t)dW(t), \qquad X(0) = x > 0. \tag{11.33}$$

Esta EDE de Stratonovich é equivalente à EDE de Itô

$$dX(t) = \left(rX(t)\left(1 - \tfrac{X(t)}{K}\right) + \tfrac{\sigma^2}{2}X(t)\right)dt + \quad \sigma X(t)dW(t), \quad X(0) = x > 0.$$

Como é uma EDE autónoma em dimensão 1 e os coeficientes são de classe C^1, a solução forte existe e é única até a um possível instante de explosão, que iremos ver ser q.c. infinito, pelo que a solução existe e é única para todo o $t \geq 0$. A solução é um processo de difusão com coeficiente de tendência

$$a(x) = rx\left(1 - \frac{x}{K}\right) + \frac{\sigma^2}{2}x$$

e coeficiente de difusão

$$b(x) = \sigma^2 x^2.$$

Como $a(0) = 0$ e $b(0) = 0$, reconhece-se que $r_1 = 0$ e $r_2 = +\infty$ são as fronteiras.

As densidades de escala e velocidade obtêm-se facilmente, vindo

$$s(\xi) = C\xi^{-2r/\sigma^2 - 1}\exp\left(\frac{2r}{\sigma^2 K}\xi\right) \tag{11.34}$$

e

$$m(\xi) = \frac{1}{C\sigma^2}\xi^{2r/\sigma^2 - 1}\exp\left(-\frac{2r}{\sigma^2 K}\xi\right), \tag{11.35}$$

onde C é uma constante positiva arbitrária. Constata-se facilmente que

$s(\xi)$ não é integrável na vizinhamça direita de 0 nem na vizinhança esquerda de ∞, pelo que ambas as fronteiras são não-atractivas e, portanto, inatingíveis. Vemos assim que não há explosão (a solução nunca se torna infinita) e que o espaço de estados é $(0, +\infty)$. Como sabemos, as fronteiras são também inatingíveis em tempo infinito, pelo que $X(t)$ não tende para zero nem para $+\infty$ quando $t \to +\infty$. Em particular, não há *extinção matemática* da população, querendo com isto dizer que, q.c., não existe nenhum t tal que $X(t) = 0$ nem se verifica que $X(t)$ tenda para zero quando $t \to +\infty$.

Vamos ver se existe densidade estacionária $p(y)$ $(0 < y < +\infty)$, caso em que ela será proporcional a $m(y)$ (escolhendo a constante de proporcionalidade para que $\int_0^{+\infty} p(\xi)d\xi = 1$). Basta ver que $\int_0^{+\infty} m(\xi)d\xi < +\infty$. Reconhece-se facilmente que, à parte uma constante multiplicativa, $m(\xi)$ é a f.d.p. de uma *distribuição gama* com parâmetro de forma $2r/\sigma^2$ e parâmetro de escala $\sigma^2 K/(2r)$, pelo que é obviamente integrável. Claro que $p(y)$ será precisamente a densidade de uma distribuição gama com os parâmetros referidos, pelo que

$$p(y) = \frac{1}{\left(\frac{\sigma^2 K}{2r}\right)^{2r/\sigma^2} \Gamma(2r/\sigma^2)} y^{2r/\sigma^2 - 1} \exp\left(-\frac{y}{\frac{\sigma^2 K}{2r}}\right) \quad (0 < y < +\infty). \tag{11.36}$$

Note-se que a média da distribuição estacionária é precisamente K, o valor de equilíbrio do modelo determinístico. A moda da distribuição estacionária é $K(1 - \sigma^2/(2r))$, que está muito próxima de K no caso de a intensidade σ do ruído ser pequena. Claro que o processo é ergódico.

Note-se que $r(1 - y/K)$ deverá agora interpretar-se (ver Secção 9.2) como *taxa média geométrica de crescimento per capita* quando o tamanho da população é y. Alguns autores preferem usar o cálculo de Itô em vez de Stratonovich. Nesse caso, $r(1-y/K)$ deverá interpretar-se como sendo a *taxa média aritmética de crescimento per capita* quando o tamanho da população é y. Para o estudo deste modelo e de outros semelhantes, o leitor pode ver, por exemplo, [8], [9], [22], [27], [29] e [56]. Contudo, as conclusões são específicas da forma como a limitação de recursos afecta as taxas de natalidade e mortalidade, forma que é mal conhecida. Assim, para obter resultados que sejam robustos relativamente a essa forma (isto é relativamente ao modelo concreto escolhido), [16] usou modelos gerais em que a taxa média geométrica de crescimento *per capita* é uma função arbitrária $r(y)$. Há que fazer algumas suposições, claro, mas que sejam ditadas por considerações biológicas. Supõe-se, em particular, que a função $r(y)$ é de classe C^1 estritamente decrescente, que $r(0^+) > 0$ (podendo mesmo ser $+\infty$), que $r(+\infty) < 0$ (podendo mesmo ser $-\infty$) e que $\lim_{y \downarrow 0} yr(y) = 0$ (admite-se que não há imigração). Com

estas hipóteses, atingiram-se as mesmas conclusões, designadamente a existência de solução única para todo o $t \geq 0$, o espaço de estados ser $(0, +\infty)$, as fronteiras desse espaço serem não-atractivas, a inexistência de extinção matemática, a ergodicidade, a existência de densidade estacionária e o facto de a sua moda estar próxima do equilíbrio determinístico globalmente assintoticamente estável quando o ruído tem uma intensidade σ pequena. Claro que, se se tivesse suposto que $r(0^+) < 0$, a fronteira 0 seria atractiva e viria $X(t) \to 0$ quando $t \to +\infty$ (extinção matemática). Os resultados foram generalizados em [18] para intensidades de ruído $\sigma(y)$ que possam depender quase arbitrariamente (há pequenas restrições) do tamanho y da população.

O facto de não haver extinção matemática nos modelos referidos com as suposições feitas, não significa que a população esteja livre da extinção no sentido biológico do termo. De facto, a variável de estado X é contínua, mas os tamanhos reais da população são discretos, pelo que pode suceder que, no modelo, X se possa aproximar muito de zero sem ocorrer a extinção matemática, enquanto que, numa população real, valores positivos abaixo de um indivíduo são impossíveis. Por outro lado, as hipóteses do modelo excluem flutuações aleatórias de natureza demográfica (só se consideram flutuações ambientais), que são particularmente relevantes para populações pequenas. Em populações reais, podem ainda ocorrer outros fenómenos não considerados no modelo, como os efeitos de Allee. Assim, para estudar a extinção é preferível não usar o conceito de extinção matemática, mas o conceito de *extinção realista*, em que se considera extinta qualquer população que atinja um *limiar de extinção* $a > 0$ adequadamente escolhido (por exemplo, pode ser $a = 1$ ou $a = 2$ em populações sexuadas, ou um valor a abaixo do qual se instalam efeitos de Allee), considerando-se que, quando o tamanho da população atinge esse valor, a extinção é inevitável.

Para que o problema tenha interesse isto é, para que a população não esteja extinta à partida, vamos supor $a < X(0) = x$. A *probabilidade de extinção* $P[\text{existir } t > 0 \text{ tal que } X(t) \leq a]$ é, para os modelos considerados acima, igual a 1 dado que a ergodicidade implica que todos os estados interiores (neste caso todos são interiores) sejam visitados, mais cedo ou mais tarde, com probabilidade um. Sendo a extinção (realista) inevitável, a questão relevante é estudar o *tempo de extinção*, que não é mais que o *tempo de primeira passagem* T_a pelo limiar de extinção a.

Na Secção 11.2, estudámos $T_{a,b}$ com $a < x < b$. Como o processo passa por a com probabilidade um, isto é, $T_a < +\infty$ q.c. e como não passa por $r_2 = +\infty$ visto ser fronteira inacessível, temos $T_a = T_{a,+\infty} = \lim_{b \uparrow +\infty} T_{a,b}$, pelo que basta fazer $b \uparrow +\infty$ no estudo anteriormente feito. Pode ver-se em [23] a aplicação aos modelos logístico e de Gompertz.

Aqui vamos apenas apresentar o caso do modelo logístico (11.33).

Podíamos recorrer ao último teorema da Secção 11.2 com $h(x) \equiv 1$ e $q(\xi) = \exp(-\lambda\xi)$ para obter uma equação diferencial para a transformada de Laplace $\mathbb{E}_x[\exp(-\lambda T_{a,b})]$ pois, uma vez resolvida, podíamos determinar a f.d.p. de $T_{a,b}$ por inversão e os momentos $\mathbb{E}_x[(T_{a,b})^n]$ por derivação em ordem a λ no ponto $\lambda = 0$. Infelizmente, não conseguimos resolver a equação diferencial explicitamente e os métodos numéricos de inversão são um tanto instáveis. Recorremos assim a (11.21) com $h(x) \equiv 1$ para obter $\mathbb{E}_x[T_{a,b}]$. Vem

$$U_1(x) = \mathbb{E}_x[T_{a,b}] =$$
$$2\left(u(x)\int_x^b (S(b) - S(\xi))m(\xi)d\xi + (1-u(x))\int_a^x (S(\xi) - S(a))m(\xi)d\xi\right),$$

com $u(x)$ dado por (11.20). Usando (11.34) e (11.35), obtemos

$$u(x) = \frac{\int_\alpha^\gamma z^{-\frac{2r}{\sigma^2}-1}e^z dz}{\int_\alpha^\beta z^{-\frac{2r}{\sigma^2}-1}e^z dz},$$

com

$$\alpha = \frac{2r}{\sigma^2 K}a,\ \beta = \frac{2r}{\sigma^2 K}b,\ \gamma = \frac{2r}{\sigma^2 K}x,$$

e, após algumas simplificações,

$$\mathbb{E}_x[T_{a,b}] =$$
$$\frac{2}{\sigma^2\int_\alpha^\beta y^{-\frac{2r}{\sigma^2}-1}e^y dy}\left(\int_\alpha^\gamma y^{-\frac{2r}{\sigma^2}-1}e^y dy\int_\gamma^\beta\int_v^\beta y^{-\frac{2r}{\sigma^2}-1}e^y dy v^{\frac{2r}{\sigma^2}-1}e^{-v}dv\right.$$
$$\left. + \int_\gamma^\beta y^{-\frac{2r}{\sigma^2}-1}e^y dy\int_\alpha^\gamma\int_\alpha^v y^{-\frac{2r}{\sigma^2}-1}e^y dy v^{\frac{2r}{\sigma^2}-1}e^{-v}dv\right).$$

Fazendo $b \uparrow +\infty$ (que implica $\beta \uparrow +\infty$), obtemos, após levantamento de indeterminações,

$$\mathbb{E}_x[T_a] = \frac{2}{\sigma^2}\int_\alpha^\gamma \Gamma\left(\frac{2r}{\sigma^2}, y\right)y^{-\frac{2r}{\sigma^2}-1}e^y dy,$$

onde $\Gamma(c,x) = \int_x^{+\infty} t^{c-1}e^{-t}dt$ é a função gama incompleta.

Apresentamos na Figura 11.1 alguns gráficos do tempo médio de extinção. Para isso, há vantagem em trabalhar com grandezas adimensionais, para o que vamos considerar $R = \frac{r}{\sigma^2}$ (espécie de relação sinal/ruído), $d = a/K$ (limiar de extinção expresso como fracção da capacidade de sustento do meio) e $z = x/a$ (medida do afastamento re-

lativo da população inicial relativamente ao limiar de extinção). Como r é uma taxa de renovação quando o tamanho da população é pequeno, $1/r$ é o tamanho médio duma geração nas mesmas circunstâncias, pelo que rT_a é o tempo de extinção medido em número dessas gerações; o gráfico mostra-nos o seu valor esperado $\mathbb{E}_x[rT_a]$ como função de z para alguns valores de R e d. Note-se que

$$\mathbb{E}_x[rT_a] = 2R \int_{2Rd}^{2Rdz} \Gamma(2R, y) y^{-2R-1} e^y dy.$$

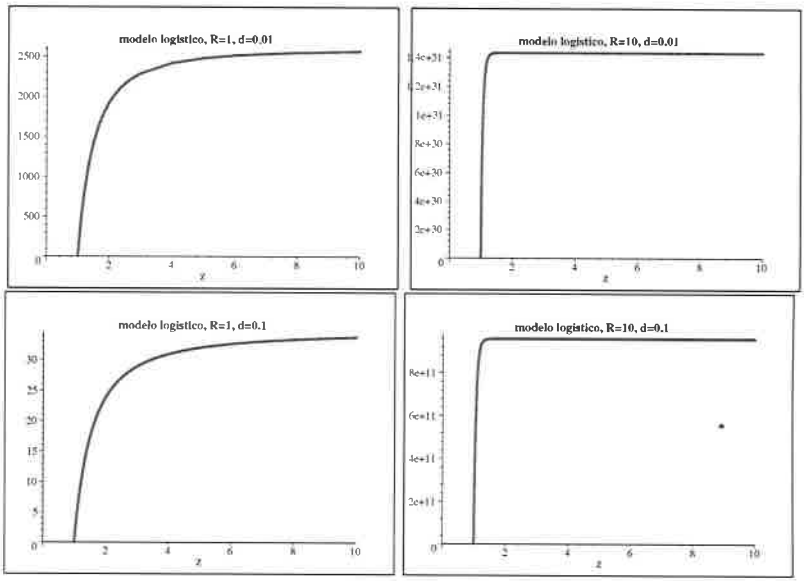


Figura 11.1: Comportamento do tempo médio de extinção para uma população descrita pelo modelo logístico. A abcissa é $z = x/a$ e a ordenada é $r\mathbb{E}_x[T_a]$. Note-se que a escala vertical difere entre os gráficos.

Note-se que o tempo médio de extinção aumenta quando R aumenta (logo aumenta quando a intensidade σ das flutuações aleatórias diminui) e diminui quando d aumenta (isto é, quando a capacidade de sustento se aproxima do limiar de extinção). Também se vê que o tempo médio de extinção aumenta quando z aumenta (logo aumenta quando a população inicial x aumenta ou quando o limiar de extinção a diminui), embora esse aumento seja muito ténue a partir de certa altura (para os casos

ilustrados na Figura 11.1, vemos que, quando $z > 2$, o tempo médio de extinção está praticamente estabilizado).

Em [23] também se obtém o devio-padrão do tempo de extinção, que é da mesma ordem de grandeza da média. Para isso, sendo $U_n(x) = \mathbb{E}_x\left[(T_{a,b})^n\right]$, resolveu-se a equação diferencial (ver último teorema da Secção 11.2) $\mathcal{D}U_n(x) = -nU_{n-1}(x)$ com $U_n(a) = U_n(b) = 0$ e $n = 2$, recorrendo à expressão já obtida de $U_1(x)$, e fez-se $b \uparrow +\infty$.

Para o estudo de modelos de crescimento populacional com pesca pode consultar [10], [12], [17], [19] e referências neles contidas.

A este propósito, a gestão de explorações de pesca, caça ou floresta pode beneficiar de uma área de estudo muito interessante, o controlo óptimo estocástico, de forma a regular o esforço de capturas no sentido de optimizar determinada função objectivo (por exemplo, relacionada com os lucros) sujeita a certas restrições (por exemplo, de natureza económica, ambiental ou social). As aplicações do controlo óptimo estocástico estendem-se a outras áreas, como a gestão de carteiras de investimento e consumo ou a exploração espacial (por exemplo, no consumo de combustível em correcção de trajectória de satélites). O leitor interessado em controlo óptimo estocástico pode consultar [30]. Para algumas aplicações à gestão de recursos naturais, pode consultar [2] e [44] ou, para uma conexão entre esta matéria e as opções financeiras, [1].

Capítulo 12

Teorema de Girsanov

12.1 Introdução através de um exemplo

Consideremos um espaço de probabilidade $(\Omega, \mathcal{F}, P)$ e um processo de Wiener $W(t)$ nele definifdo. Consideremos uma EDE autónoma

$$dX(t) = f(X(t))dt + g(X(t))dW(t), \quad X(0) = x \quad t \in [0, d], \qquad (12.1)$$

com f e g satisfazendo uma condição de Lipschitz, de forma a garantir a existência e unicidade de solução. Também podemos escrever a EDE na forma integral

$$X(t) = x + \int_0^t f(X(s))ds + \int_0^t g(X(s))dW(s), \quad t \in [0, d]. \qquad (12.2)$$

Como em EDE autónomas a condição de Lipschitz implica a condição de restrição ao crescimento, podemos aplicar o teorema de existência e unicidade (ver Secção 7.1). Assim, a solução $X(t)$ é um processo de difusão homogéneo com coeficiente de tendência $a(y) = f(y)$ e coeficiente de difusão $b(y) = g^2(y)$ e a solução está em $H^2[0, d]$. Uma vez que estamos a supor uma condição inicial determinística (automaticamente independente do processo de Wiener), podemos trabalhar com a filtração natural do processo de Wiener $\mathcal{M}_s = \sigma(W(u) : 0 \leq u \leq s)$, como suporemos neste Capítulo, salvo menção em contrário. Aliás, como estamos a trabalhar exclusivamente no intervalo de tempo $[0, d]$ e $\mathcal{M}_s \subset \mathcal{M}_d$ para $s \in [0, d]$, podemos restringir o processo de Wiener a esse intervalo e trabalhar com o espaço de probabilidade $(\Omega, \mathcal{M}_d, P)$, onde utilizamos a mesma letra P para designar a restrição da probabilidade P a $\mathcal{M}_d$. Como podemos e como é conveniente, será isso que faremos daqui em

diante.

Se o coeficiente de tendência fosse nulo, viria $X(t) = x + \int_0^t g(X(s))dW(s)$ e $X(t)$ seria uma martingala (relativamente à filtração $\mathcal{M}_s$). A vantagem de se trabalhar com martingalas e de tirar partido das suas propriedades é inegável, quer do ponto de vista da teoria, quer das aplicações.

Na Secção 11.2 vimos um método de tornar nulo o coeficiente de tendência através de uma mudança de variável $Y(t) = S(X(t))$, onde S é uma função de escala. Mas, naturalmente, o processo inicial $X(t)$ mantém o seu coeficiente de tendência inicial. Aqui, vamos ver um outro método, em que não precisamos de alterar o processo $X(t)$. O método permite não apenas mudar o coeficiente de tendência para um coeficiente nulo mas também para outro coeficiente de tendência por nós escolhido. Isto é particularmente interessante em aplicações financeiras, em que podemos estar interessados em trabalhar com preços constantes ou normalizados (preços em que se desconte alguma taxa com variação determinística, por exemplo a taxa de rendimento de certas obrigações).

Dizíamos que não era preciso alterar o processo $X(t)$. Mas alguma coisa terá que mudar porque alterar apenas o coeficiente de tendência alteraria a solução, que já não seria $X(t)$. A ideia será alterar o coeficiente de tendência e, para compensar, alterar o processo de Wiener $W(t)$ que aparece na EDE de forma a que solução se mantenha. Lembramos que $W(t)$ era um processo de Wiener com respeito à probabilidade P. É natural que o processo alterado $W^*(t)$ já não seja um processo de Wiener com resprito a P. Felizmente, porém, $W^*(t)$ é um processo de Wiener com respeito a uma probabilidade modificada P^* e, além disso, a probabilidade P^* é *equivalente* a P. [1] A mudança de P para P^* compensa a mudança do coeficiente de tendência.

No caso particular de a alteração se fazer para um coeficiente de

[1] As probabilidades P e P^* dizem-se *equivalentes* se P^* for absolutamente contínua com respeito a P e vice-versa. O conceito aplica-se também a medidas. Uma medida μ (que pode, em particular, ser uma medida de probabilidade) definida num espaço mensurável $(\Omega, \mathcal{F})$ diz-se *absolutamente contínua* com respeito a outra medida ν definida no mesmo espaço mensurável se $\nu(N) = 0$ implicar $\mu(N) = 0$; escreve-se $\mu << \nu$. Suponha que ν é *medida finita-σ* (isto é, qualquer conjunto mensurável é união contável de conjuntos com medida finita); as probabilidades são medidas finitas e, portanto, também são medidas finitas-σ. O *teorema de Radon-Nikodym* mostra que $\mu << \nu$ sse μ tem uma *função densidade-ν* $f(\omega)$ não-negativa, chamada *derivada de Radon-Nikodym* e representada por $f = \frac{d\mu}{d\nu}$. Quer-se com isso dizer que $\mu(A) = \int_A f d\nu$ para todo o $A \in \mathcal{F}$. Essa derivada está quase sempre (relativamente a ν) univocamente definida e, para funções mensuráveis arbitrárias $X(\omega)$, vem $\int_\Omega X d\mu = \int_\Omega X \frac{d\mu}{d\nu} d\nu$ sempre que um dos integrais esteja definido. Note que a f.d.p. de uma v.a. absolutamente contínua X é a derivada de Radon-Nikodym da sua distribuição de probabilidade P_X.

tendência nulo, a solução $X(t)$ da EDE, que não era martingala para a probabilidade P, torna-se uma martingala para a nova probabilidade P^*. A propriedade de ser martingala é particularmente poderosa em aplicações financeiras; assim, se $Y(t)$ é uma martingala para P^*, vem, para $t \in [0,d]$, $Y(t) = \mathbb{E}^{P^*}[Y(d)|\mathcal{M}_t]$ (representamos por $\mathbb{E}^{P^*}$ a esperança matemática com respeito à prrobabilidade P^*), e podemos obter $Y(t)$ ($t \in [0,d]$) à custa do conhecimento de $Y(d)$. Isso é particularmente útil para a determinação do valor de alguns produtos financeiros, por exemplo opções de compra europeias (ver Capítulo 13). Note-se que, enquanto P pode ser interpretada como dando as probabilidades efectivas dos diferentes cenários de mercado, a nova probabilidade P^* dará probabilidades fictícias especificamente construídas para que $X(t)$ não tenha tendência (isto é, seja uma martingala, o que significa ser neutra do ponto de vista do risco financeiro).

Para melhor entender o método, vamos utilizá-lo num caso particular muito simples, precisamente o caso em que os coeficientes de tendência e difusão são constantes (*movimento browniano com tendência*). Mais precisamente, vamos considerar o caso $a(x) = f(x) \equiv \alpha$ e $b(x) = g^2(x) = 1$. Temos a EDE $dX(t) = \alpha dt + 1dW(t)$, com $X(0) = x$ ($x \in \mathbb{R}$), cuja solução é um processo em tempo contínuo $X(t)$ ($t \in [0,d]$).

Para melhor acompanhar o que está a suceder, vamos simultaneamente considerar uma discretização temporal deste problema com dt substituído por $\Delta t > 0$ e, para facilitar os cálculos, vamos supor $\Delta t = 1$. Em tempo discreto, temos os instantes $t = 0, 1, ..., n$ e, para facilitar a analogia, vamos pôr $d = n$. Vamos designar o processo em tempo discreto correspondente a $X(t)$ ($t \in [0,n]$) por X_t ($t = 0, 1, ..., n$). O processo X_t será a solução da equação às diferenças estocástica $\Delta X_t = \alpha \Delta t + 1 \Delta W(t)$ com $X_0 = x$, onde $\Delta X_t := X_t - X_{t-1}$ e $\Delta W(t) := W(t) - W(t-1)$. Podemos escrever a equação às diferenças estocástica na forma $X_t - X_{t-1} = \alpha + Z_t$ ($t = 0, 1, ..., n$), onde $Z_t := \Delta W(t)$ ($t = 1, 2, ..., n$) é um *ruído branco em tempo discreto* gaussiano, isto é, uma sequência de v.a. i.i.d. normais padrão.

Então $V_t = Z_1 + Z_2 + ... + Z_t$ ($t = 1, ..., n$), que tem distribuição normal com média 0 e variância t, representa em tempo discreto um papel análogo ao papel que $W(t)$ representa em tempo contínuo (aliás, para $t = 0, 1, ..., n$, vem $V_t = W(t)$). Repare-se que $W(t)$ é uma martingala em tempo contínuo (para a filtração natural $\mathcal{M}_t$) e que V_t é uma martingala em tempo discreto (para a filtração natural $\mathcal{N}_t = \sigma(V_u : 0 \leq u \leq t)$). À semelhança do que fizémos em tempo contínuo, só precisamos de trabalhar (e, portanto, só vamos trabalhar) no espaço de probabilidade $(\Omega, \mathcal{N}_n, P)$, onde utilizamos a mesma letra P para designar a restrição da probabilidade P a $\mathcal{N}_n$.

Note que, neste exemplo muito simples em que os coeficientes de tendência e difusão são constantes, a discretização temporal concorda sem erros com o problema inicial em tempo contínuo. Com efeito, a solução da EDE é $X(t) = x + \alpha t + W(t)$, com distribuição normal de média $x+\alpha t$ e variância t. Ora, para t inteiro não-negativo, a solução X_t da equação às diferenças estocástica é $X_t = x+\alpha t+V_t$, com distribuição normal de média $x + \alpha t$ e variância t.

Seja $Z_t^* = \alpha + Z_t$ (que queremos corresponda a $dW^*(t)$); note-se que a média já não é zero mas antes α. Seja $V_t^* = Z_1^*+Z_2^*+...+Z_t^* = \alpha t+V_t$, que tem distribuição $\mathcal{N}(\alpha t, t)$. O correspondente em tempo contínuo será $W^*(t) = \alpha t + W(t) \frown \mathcal{N}(\alpha t, t)$, mas é óbvio que já não é um processo de Wiener, nem é martingala, com respeito à probabilidade P com que estamos a trabalhar. Da mesma forma, V_t^* também já não é martingala.

Agora vem a nova equação às diferenças estocástica $X_t - X_{t-1} = 0 + Z_t^*$, em que o novo coeficiente de tendência é zero. Corresponde-lhe, em tempo contínuo, a nova EDE $dX(t) = 0dt + dW^*(t)$, também com novo coeficiente de tendência nulo.

Repare-se que $M_t = \exp\left(-\alpha V_t - \frac{1}{2}\alpha^2 t\right)$ $(t = 0, 1, 2, ..., n)$ é uma martingala em tempo discreto. Também $M(t) = \exp\left(-\alpha W(t) - \frac{1}{2}\alpha^2 t\right)$ $(t \in [0, n])$ é martingala em tempo contínuo.

Seja P^* a nova probabilidade definida pela *derivada de Radon-Nikodym* $dP^*/dP = M_n$ no espaço mensurável $(\Omega, \mathcal{N}_n)$ (o mesmo que foi usado para P). Dizer que $dP^*/dP = M_n$ significa que, para qualquer $A \in \mathcal{N}_n$,

$$P^*(A) = \int_{\omega\in A} dP^*(\omega) = \int_{\omega\in A} \frac{dP^*}{dP} dP(\omega) = \int_{\omega\in A} M_n dP(\omega).$$

Em particular, a f.d. conjunta de $Z_1^*, Z_2^*, ..., Z_n^*$ com respeito à nova probabilidade P^* é

$$\begin{aligned}
F^{P^*}_{Z_1^*,Z_2^*,...,Z_n^*}(z_1^*, z_2^*, ..., z_n^*) &= P^*[Z_1^* \le z_1^*, Z_2^* \le z_2^*, ..., Z_n^* \le z_n^*] \\
&= \int_{[Z_1^*\le z_1^*, Z_2^*\le z_2^*, ..., Z_n^*\le z_n^*]} dP^*(\omega) \\
&= \int_{[Z_1^*\le z_1^*, Z_2^*\le z_2^*, ..., Z_n^*\le z_n^*]} \exp\left(-\alpha V_n - \tfrac{1}{2}\alpha^2 n\right) dP(\omega) \\
&= \int_{-\infty}^{z_n^*} ... \int_{-\infty}^{z_2^*} \int_{-\infty}^{z_1^*} \exp\left(-\alpha \textstyle\sum_{t=1}^n y_t - \frac{1}{2}\alpha^2 n\right) \\
&\qquad f^P_{Z_1^*,Z_2^*,...,Z_n^*}(y_1^*, y_2^*, ..., y_n^*) dy_1^* dy_2^* ... dy_n^* \\
&= \int_{-\infty}^{Z_n^*} ... \int_{-\infty}^{Z_2^*} \int_{-\infty}^{Z_1^*} \exp\left(-\alpha \textstyle\sum_{t=1}^n (y_t^* - \alpha) - \frac{1}{2}\alpha^2 n\right) \\
&\qquad \frac{1}{(\sqrt{2\pi})^n} \exp\left(-\tfrac{1}{2}\textstyle\sum_{t=1}^n (y_t^* - \alpha)^2\right) dy_1^* dy_2^* ... dy_n^* \\
&= \int_{-\infty}^{Z_n^*} ... \int_{-\infty}^{Z_2^*} \int_{-\infty}^{Z_1^*} \tfrac{1}{(\sqrt{2\pi})^n} \exp\left(-\tfrac{1}{2}\textstyle\sum_{t=1}^n y_t^{*2}\right) dy_1^* dy_2^* ... dy_n^*.
\end{aligned}$$

Tabela 1: Exemplo de mudança de tendência: correspondência entre o modelo em tempo contínuo e em tempo discreto.

Modelo em tempo contínuo	Modelo correspondente em tempo discreto
Instantes $t \in [0, n]$	Instantes $t = 0, 1, ..., n$
Processo estocástico $X(t)$ em tempo contínuo	Processo estocástico X_t em tempo discreto
$dX(t)$	$\Delta X_t = X_t - X_{t-1}$
dt	$\Delta t = 1$
$dW(t)$	$Z_t := \Delta W(t) = W(t) - W(t-1)$ (v.a. normais padrão i.i.d.)
$W(t) = \Delta W(1) + ... + \Delta W(t)$ $\frown \mathcal{N}(0, t)$ é martingala em tempo contínuo	$V_t := Z_1 + ... + Z_t (= W(t))$ $\frown \mathcal{N}(0, t)$ é martingala em tempo discreto
equação diferencial estocástica $dX(t) = \alpha dt + dW(t)$ com $X_0 = x$	equação às diferenças estocástica $\Delta X_t = \alpha \Delta t + \Delta W(t)$ ou $X_t - X_{t-1} = \alpha + Z_t$ com $X_0 = x$
a solução é $X(t) = x + \alpha t + W(t)$ $\frown \mathcal{N}(x + \alpha t, t)$	a solução é $X_t = x + \alpha t + V_t$ $\frown \mathcal{N}(x + \alpha t, t)$
$dW^*(t)$	$Z_t^* = \alpha + Z_t \frown \mathcal{N}(\alpha, 1)$
$W^*(t) = \alpha t + W(t) \frown \mathcal{N}(\alpha t, t)$ não é martingala nem processo de Wiener com respeito a P	$V_t^* = Z_1^* + Z_2^* + ... + Z_t^* =$ $\alpha t + V_t \frown \mathcal{N}(\alpha t, t)$ não é martingala com respeito a P
equação diferencial estocástica nova $dX(t) = 0dt + dW^*(t)$ coeficiente de tendência $= 0$ solução: $X(t) = x + W^*(t)$	equação às diferenças estocástica nova $X_t - X_{t-1} = 0 + Z_t^*$ coeficiente de tendência $= 0$ solução: $X_t = x + V_t^*$
$M(t) = \exp\left(-\alpha W(t) - \frac{1}{2}\alpha^2 t\right)$ é martingala	$M_t = \exp\left(-\alpha V_t - \frac{1}{2}\alpha^2 t\right)$ é martingala
nova probabilidade P^* definida por $dP^*/dP = M(n)$	nova probabilidade P^* definida por $dP^*/dP = M_n$
	com respeito a P^* $Z_1^*, Z_2^*, ..., Z_n^*$ são i.i.d. $\mathcal{N}(0, 1)$
com respeito a P^* $W^*(t) \frown \mathcal{N}(0, t)$ e é processo de Wiener (logo martingala)	com respeito a P^* $V_t^* \frown \mathcal{N}(0, t)$ e é martingala

Logo, com respeito a P^*, as v.a. $Z_1^*, Z_2^*, ..., Z_n^*$ são i.i.d. com distribuição normal padrão, pelo que têm média zero. Portanto, com respeito a P^*, $V_t^* = Z_1^* + Z_2^* + ... + Z_t^*$ tem distribuição $\mathcal{N}(0,t)$ e facilmente se reconhece que, relativamente a P^*, é uma martingala em tempo discreto.

Analogamente, $W^*(t)$ será, com respeito a uma probabilidade P^* apropriada, um processo de Wiener e, portanto, uma martingala. A probabilidade P^* apropriada para tempo contínuo é definida pela derivada de Radon-Nikodym $dP^*/dP = M(n) = M(d)$ no espaço mensurável $(\Omega, \mathcal{M}_n)$ (o mesmo que foi usado para P). Isso significa que, para qualquer $A \in \mathcal{M}_n$,

$$P^*(A) = \int_A dP^*(\omega) = \int_A \frac{dP^*}{dP} dP(\omega) = \int_A M(n) dP(\omega).$$

Note-se ainda que, sendo $M(d) > 0$, vem $dP/dP^* = 1/M(d)$ e, portanto, tendo qualquer das probabilidades derivada de Radon-Nikodym relativamente à outra, elas são absolutamente contínuas relativamente à outra, logo *equivalentes*. Têm, assim, os mesmos conjuntos de probabilidade nula.

Vamos mostrar que $W^*(t)$ é processo de Wiener relativamente à probabilidade P^*.

Note-se que $W^*(t) = \alpha t + W(t)$ $(t \in [0,d])$ tem a mesma filtração natural $\mathcal{M}_t$ que $W(t)$. Reconhece-se que, para $s < t$, o incremento $W^*(t) - W^*(s) = \alpha(t-s) + W(t) - W(s)$ é, tal como o incremento $W(t) - W(s)$ do processo $W(t)$, independente de $\mathcal{M}_s$ e, portanto, independente de qualquer incremento no intervalo $[0,s]$. Assim, concluímos que o processo $W^*(t)$ tem incrementos independentes.

Também é óbvio que $W^*(0) = 0$ q.c.-P^* (pois P e P^* têm os mesmos conjuntos de probabilidade nula).

Para mostrar que $W^*(t)$ é processo de Wiener relativamente à probabilidade P^* basta só mostrar que, para $0 \leq s < t \leq d$, a distribuição com respeito a P^* de $W^*(t) - W^*(s)$ é $\mathcal{N}(0, t-s)$. Designemos por $F_Z^Q(z) = Q[Z \leq z]$ e por f_Z^Q a f.d. e a f.d.p. (se existir), respectivamente, da v.a Z com respeito à probabilidade Q. Seja $A_x = \{\omega \in \Omega : W^*(t) - W^*(s) \leq x\} = \{\omega \in \Omega : W(t) - W(s) + \alpha(t-s) \leq x\}$. A f.d. de $W^*(t) - W^*(s)$

com respeito a P^* é

$$F^{P^*}_{W^*(t)-W^*(s)}(x) = P^*(A_x) = \int_{A_x} dP^*(\omega) = \int_{A_x} \frac{dP^*(\omega)}{dP(\omega)} dP(\omega) =$$
$$\int_\Omega \frac{dP^*(\omega)}{dP(\omega)} I_{A_x}(\omega) dP(\omega) = \mathbb{E}^P\left[\exp\left(-\alpha W(d) - \tfrac{1}{2}\alpha^2 d\right) I_{A_x}(\omega)\right] =$$
$$e^{-\frac{1}{2}\alpha^2 d}\mathbb{E}^P\left[e^{-\alpha(W(d)-W(t))} e^{-\alpha(W(t)-W(s))} I_{A_x}(\omega) e^{-\alpha W(s)}\right] =$$
$$e^{-\frac{1}{2}\alpha^2 d}\mathbb{E}^P\left[e^{-\alpha(W(d)-W(t))}\right] \cdot$$
$$\cdot \mathbb{E}^P\left[e^{-\alpha(W(t)-W(s))} I_{A_x}(\omega)\right] \mathbb{E}^P\left[e^{-\alpha W(s)}\right] =$$
$$e^{-\frac{1}{2}\alpha^2 d} e^{\frac{1}{2}\alpha^2(d-t)} \mathbb{E}^P\left[e^{-\alpha(W(t)-W(s))} I_{A_x}(\omega)\right] e^{\frac{1}{2}\alpha^2 s} =$$
$$e^{-\frac{1}{2}\alpha^2(t-s)} \int_{A_x} e^{-\alpha(W(t)-W(s))} dP(\omega) =$$
$$e^{-\frac{1}{2}\alpha^2(t-s)} \int_{-\infty}^{x-\alpha(t-s)} e^{-\alpha z} f^P_{W(t)-W(s)}(z) dz =$$
$$e^{-\frac{1}{2}\alpha^2(t-s)} \int_{-\infty}^{x-\alpha(t-s)} e^{-\alpha z} \frac{1}{\sqrt{2\pi(t-s)}} \exp\left(-\frac{z^2}{2(t-s)}\right) dz =$$
$$\int_{-\infty}^{x-\alpha(t-s)} \frac{1}{\sqrt{2\pi(t-s)}} \exp\left(-\frac{(z+\alpha(t-s))^2}{2(t-s)}\right) dz =$$
$$\int_{-\infty}^{x} \frac{1}{\sqrt{2\pi(t-s)}} \exp\left(-\frac{v^2}{2(t-s)}\right) dv,$$

que é precisamente a f.d. de uma v.a. $\mathcal{N}(0, t-s)$. Está concluída a demonstração de que $W^*(t)$ é um processo de Wiener com respeito à nova probabilidade P^*.

Claro que a solução de $dX(t) = 0dt + dW^*(t)$ é $X(t) = x + \int_0^t dW^*(s)$ $= x + W^*(t)$ e é uma martingala para a probabilidade P^* (e a filtração $\mathcal{M}_t$).

A remoção da tendência e construção da nova probabilidade a que acabámos de proceder para este exemplo, quer em tempo contínuo quer em versão discreta, encontram-se resumidos na Tabela 1.

O resultado obtido para o modelo em tempo contínuo pode ser estendido ao caso geral em que f e g não têm que ser constantes e também se pode mudar para um coeficiente de tendência diferente do inicial mas não necessariamente nulo. É esse resultado que constitui o teorema de Girsanov, a ser estudado na Secção 12.2.

12.2 Teorema de Girsanov

Seja $0 < d < \infty$. Consideremos um espaço de probabilidade $(\Omega, \mathcal{F}, P)$ e nele um processo de Wiener $W(t)$ $(t \in [0, d])$. Seja $\mathcal{M}_t = \sigma(W(u) : 0 \leq u \leq t)$ a filtração natural do processo de Wiener. Vamos restringir P a $\mathcal{M}_d$, isto é, vamos trabalhar com o espaço de probabilidade $(\Omega, \mathcal{M}_d, P)$.

Teorema de Girsanov

Seja $X(t)$ a solução da EDE

$$dX(t) = f(X(t))dt + g(X(t))dW(t), \quad X(0) = x \quad t \in [0, d], \tag{12.3}$$

com f e g satisfazendo uma condição de Lipschitz (o que garante existência e unicidade de solução).

Suponhamos que $f^*(x)$ também satisfaz uma condição de Lipschitz. Suponhamos que existe uma função $\theta(x)$ tal que [2]

$$g(x)\theta(x) = f(x) - f^*(x) \tag{12.4}$$

e seja

$$M(t) = \exp\left(-\int_0^t \theta(X(s))dW(s) - \frac{1}{2}\int_0^t \theta^2(X(s))ds\right). \tag{12.5}$$

Suponhamos que $\theta(X(t)) \in M^2[0, d]$ [3] *e que $M(t)$ ($t \in [0, d]$) é uma martingala com respeito a P (relativamente à filtração $\mathcal{M}_t$). Uma condição suficiente para que isto se verifique é que seja válida a condição de Novikov*

$$\mathbb{E}^P\left[\exp\left(\frac{1}{2}\int_0^d \theta^2(X(s))ds\right)\right] < +\infty. \tag{12.6}$$

Seja P^* uma nova medida de probabilidade no espaço mensurável $(\Omega, \mathcal{M}_d)$ definida pela derivada de Radon-Nikodym

$$dP^*/dP = M(d)$$

e seja

$$W^*(t) = W(t) + \int_0^t \theta(X(s))ds. \tag{12.7}$$

Então P^* é equivalente a P e $W^*(t)$ é um processo de Wiener relativamente a P^*. [4]

Além disso, o processo $X(t)$ satisfaz a EDE

$$dX(t) = f^*(X(t))dt + g(X(t))dW^*(t), \quad X(0) = x \quad t \in [0, d] \tag{12.8}$$

e, portanto, as distribuições de dimensão finita de $X(t)$ relativamente a

---

[2] Caso $g(x) \neq 0$, vem $\theta(x) = \frac{f(x) - f^*(x)}{g(x)}$.

[3] Ou seja (ver Capítulo 6) $\theta(t, \omega) := \theta(X(t, \omega))$ é função conjuntamente mensurável adaptada à filtração $\mathcal{M}_t$ aqui escolhida e tal que $\int_0^d \theta^2(t, \omega)dt < +\infty$ q.c.-P.

[4] Em geral, não é um processo de Wiener relativamente a P.

P^* coincidem com as distribuições de dimensão finita relativamente a P da solução $Y(t)$ da EDE [5]

$$dY(t) = f^*(Y(t))dt + g(Y(t))dW(t), \quad Y(0) = x \quad t \in [0, d]. \tag{12.9}$$

Nota. No caso particular de $f^* \equiv 0$, tem-se $X(t) = x + \int_0^t g(X(s)dW^*(s)$. Como g satisfaz uma condição de Lipschitz, sabemos que $g(s,\omega) := g(X(s,\omega)) \in H^{2P^*}[0,d]$ ($H^{2P^*}[0,d]$ é a classe das funções conjuntamente mensuráveis adaptadas à filtração $\mathcal{M}_s$ tais que $\mathbb{E}^{P^*}\left[\int_0^d g^2(s)ds\right] < +\infty$). Logo, devido às propriedades do integral estocástico como função do limite superior de integração, $X(t)$ será, neste caso particular, uma martingala relativamente a P^* (com respeito à filtração $\mathcal{M}_t$). Temos assim um processo de transformar a solução de uma EDE autónoma, que em geral não é martingala relativamente à probabilidade original, numa martingala, mas relativamente a uma nova medida de probabilidade. O exemplo tratado na Secção 12.1 é o caso particular de $f(x) \equiv \alpha$ e $g(x) \equiv 1$, vindo $\theta(x) = \frac{f(x)-0}{g(x)} = \alpha$.

Nota. Não é difícil generalizar ao caso de EDE multidimensionais nos termos referidos na Secção 7.3. Naturalmente que $\boldsymbol{\theta}$ é agora uma vector coluna com a mesma dimensão m do vector do processo de Wiener e que $M(t) = \exp\left(-\int_0^t \boldsymbol{\theta}^T(\mathbf{X}(s))d\mathbf{W}(s) - \frac{1}{2}\int_0^t \boldsymbol{\theta}^T(\mathbf{X}(s))\boldsymbol{\theta}(\mathbf{X}(s))ds\right)$. Na condição de Novikov, deve substituir-se θ^2 por $\boldsymbol{\theta}^T\boldsymbol{\theta}$. No caso de $n = m$ e de a matriz $\mathbf{g}$ ser invertível, pode escolher-se $\boldsymbol{\theta} = \mathbf{g}^{-1}(\mathbf{f} - \mathbf{f}^*)$.

Nota. Outra generalização relativamente trivial consiste em permitir a substituição de $f(X(s))$ por $F(s,\omega) \in M^2[0,d]$ (agora a EDE inicial pode não ser autónoma), desde que exista $\theta(s,\omega)$ tal que $g(X(s,\omega))\theta(s,\omega) = F(s,\omega) - f^*(x)$. Exige-se agora que $\theta(s,\omega) \in M^2[0,d]$ e que, relativamente a P, $M(t) = \exp\left(-\int_0^t \theta(s,\omega)dW(s) - \frac{1}{2}\int_0^t \theta^2(s,\omega))ds\right)$ seja martingala, sendo condição suficiente para que tal suceda que se verifique a condição de Novikov $\mathbb{E}\left[\exp\left(\frac{1}{2}\int_0^d \theta^2(s,\omega)ds\right)\right] < +\infty$. Agora vem $W^*(t) = W(t) + \int_0^t \theta(s,\omega)ds$.

Antes de demonstrar o teorema, vamos demonstrar dois lemas.

[5] Como só as distribuições de dimensão finita são relevantes para o cálculo de probabilidades relativas aos valores do processo, é indiferente trabalhar com as distribuições de $Y(t)$ relativamente a P ou com as distribuições de $X(t)$ relativamente a P^*. Isso pode ser útil na determinação de soluções fracas de EDE.

Lema. *Sejam P e P^* probabilidades num espaço mensurável $(\Omega, \mathcal{G})$ tal que $\frac{dP^*}{dP} = h(\omega)$, seja Y uma v.a. em $(\Omega, \mathcal{G})$ tal que $\mathbb{E}^{P^*}[|Y|] = \int_\Omega |Y(\omega)|h(\omega)dP(\omega) < +\infty$ e seja $\mathcal{H}$ uma sub-álgebra-σ de $\mathcal{G}$. Então*

$$\mathbb{E}^P[hY|\mathcal{H}] = \mathbb{E}^{P^*}[Y|\mathcal{H}]\mathbb{E}^P[h|\mathcal{H}] \text{ q.c.-}P.$$

Demonstração

Seguimos de perto a demonstração de [51]. Seja $H \in \mathcal{H}$. Vem, usando a definição de probabilidade condicional e as propriedades das esperanças condicionais,

$$\int_H \mathbb{E}^{P^*}[Y|\mathcal{H}]hdP = \int_H \mathbb{E}^{P^*}[Y|\mathcal{H}]dP^* = \int_H YdP^* = \int_H YhdP = \int_H \mathbb{E}^P[Yh|\mathcal{H}]dP.$$

Por outro lado,

$$\int_H \mathbb{E}^{P^*}[Y|\mathcal{H}]hdP = \int_\Omega \mathbb{E}^{P^*}[Y|\mathcal{H}]hI_H dP = \mathbb{E}^P\left[\mathbb{E}^{P^*}[Y|\mathcal{H}]hI_H\right] = \mathbb{E}^P\left[\mathbb{E}^P\left[\mathbb{E}^{P^*}[Y|\mathcal{H}]hI_H|\mathcal{H}\right]\right] = \mathbb{E}^P\left[I_H\mathbb{E}^{P^*}[Y|\mathcal{H}]\mathbb{E}^P[h|\mathcal{H}]\right] = \int_H \mathbb{E}^{P^*}[Y|\mathcal{H}]\mathbb{E}^P[h|\mathcal{H}]dP.$$

Logo

$$\int_H \mathbb{E}^{P^*}[Y|\mathcal{H}]\mathbb{E}_P[h|\mathcal{H}]dP = \int_H \mathbb{E}^P[Yh|\mathcal{H}]dP$$

para qualquer $H \in \mathcal{H}$, o que prova o pretendido. ■

O facto de no teorema de Girsanov se ter $\frac{dP^*}{dP} = M(d)$ em $\mathcal{M}_d$ significa que, dado qualquer $A \in \mathcal{M}_d$, vem

$$P^*(A) = \int_A dP^*(\omega) = \int_A \frac{dP^*}{dP}dP(\omega) = \int_A M(d)dP(\omega).$$

Será que se pode ir mais longe caso $A \in \mathcal{M}_t$ com $t \in [0, d]$? Vejamos o:

Lema. *Nas condições do teorema de Girsanov e para $t \in [0, d]$, vem $\frac{dP^*}{dP} = M(t)$ em $\mathcal{M}_t$, querendo com isto significar que, para $A \in \mathcal{M}_t$, se tem*

$$P^*(A) = \int_A dP^*(\omega) = \int_A M(t)dP(\omega).$$

Demonstração

Seguimos de perto a demonstração de [51]. De facto, atendendo a que $M(t)$ é uma martingala relativamente a P (com respeito à filtração

$\mathcal{M}_t$), vem

$$\begin{aligned} P^*(A) = \quad & \int_A dP^* = \int_A M(d)dP = \int_\Omega I_A M(d)dP = \mathbb{E}^P\left[I_A M(d)\right] = \\ & \mathbb{E}^P\left[\mathbb{E}^P\left[I_A M(d)|\mathcal{M}_t\right]\right] = \mathbb{E}^P\left[I_A \mathbb{E}^P\left[M(d)|\mathcal{M}_t\right]\right] = \\ & \mathbb{E}^P\left[I_A M(t)\right] = \int_A M(t)dP. \quad \blacksquare \end{aligned}$$

Esboço da demonstração do teorema de Girsanov

Condição de Novikov

Seja

$$Z(t) = -\int_0^t \theta(X(s)dW(s) - \frac{1}{2}\int_0^t \theta^2(X(s))ds.$$

Aplicando o teorema de Itô a $M(t) = \exp(Z(t))$, e atendendo a que $\theta(X(t)) \in M^2[0,t]$, vem

$$\begin{aligned} dM(t) = \quad & e^{Z(t)}dZ(t) + \tfrac{1}{2}e^{Z(t)}(dZ(t))^2 = \\ & M(t)\left(-\theta dW(t) - \tfrac{1}{2}\theta^2 dt\right) + \tfrac{1}{2}M(t)\theta^2 dt = \\ & -\theta(X(t))M(t)dW(t), \end{aligned}$$

donde, como $M(0) = 1$,

$$M(t) = 1 - \int_0^t \theta(X(s))M(s)dW(s). \tag{12.10}$$

Se a função integranda $\theta(X(s))M(s)$ estiver em $H^2[0,d]$, as propriedades do integral estocástico como função do limite superior de integração implicam que $M(t)$ seja uma martingala relativamente a P (com respeito à filtração $\mathcal{M}_t$). Mas nada garante que $\theta(X(s))M(s) \in H^2[0,d]$, daí se ter exigido no enunciado que $M(t)$ seja martingala relativamente a P.

A condição de Novikov implica

$$\exp\left(\mathbb{E}^P\left[\frac{1}{2}\int_0^d \theta^2(X(s))ds\right]\right) \leq \mathbb{E}^P\left[\exp\left(\frac{1}{2}\int_0^d \theta^2(X(s))ds\right)\right] < +\infty,$$

e, portanto, implica $\theta \in H^2[0,d] \subset M^2[0,d]$. A condição de Novikov também implica que $M(t)$ seja martingala relativamente a P pois permite aproximar (12.10) por uma técnica de truncatura sofisticada que garante a propriedade de martingala, mas que está fora do âmbito deste livro. O leitor interessado pode vê-la em [37].

Equivalência

Tal como se fez na Secção 12.1, basta atender a que $M(d) > 0$.

$X(t)$ satisfaz (12.8)

Basta notar que $dW^*(t) = dW(t) + \theta(X(t))dt$ e substituir $dW(t)$ por $dW^*(t) - \theta(X(t))dt$ em (12.3) para obter (12.8).

$W^*(t)$ é processo de Wiener relativamente a P^*

Seguimos de perto a demonstração de [51]. Seja

$$R(t) = W^{*2}(t) - t.$$

Atendendo à *caracterização de Lévy* (ver Secção 4.2) do processo de Wiener, basta mostrar que $W^*(t)$ e $R(t)$ são martingalas relativamente a P^* (e com respeito à filtração $\mathcal{M}_t$).

Sejam

$$K(t) = M(t)W^*(t)$$

e

$$L(t) = M(t)R(t).$$

Vem $K(0) = 0$ e

$$\begin{aligned} &dK(t) = M(t)dW^*(t) + W^*(t)dM(t) + dM(t)dW^*(t) = \\ &M(t)(\theta(X(t))dt + dW(t)) + W^*(t)(-\theta(X(t))M(t)dW(t)) - \\ &\quad \theta(X(t))M(t)dt = \\ &M(t)(1 - \theta(X(t))W^*(t))dW(t), \end{aligned}$$

donde

$$K(t) = \int_0^t M(s)(1 - \theta(X(s))W^*(s))dW(s)$$

é martingala relativamente a P.

Para $0 \leq s < t \leq d$ e atendendo a que $\mathcal{M}_s \subset \mathcal{M}_t$, vem, usando a propriedade de martingala e os dois lemas acima,

$$\begin{aligned} &M(s)W^*(s) = K(s) = \mathbb{E}^P[K(t)|\mathcal{M}_s] = \mathbb{E}^P[M(t)W^*(t)|\mathcal{M}_s] = \\ &\mathbb{E}^P\left[\tfrac{dP^*}{dP}W^*(t)|\mathcal{M}_s\right] = \mathbb{E}^{P^*}\left[W^*(t)|\mathcal{M}_s\right]\mathbb{E}^P\left[\tfrac{dP^*}{dP}|\mathcal{M}_s\right] = \\ &\mathbb{E}^{P^*}\left[W^*(t)|\mathcal{M}_s\right]\mathbb{E}^P\left[M(t)|\mathcal{M}_s\right] = \mathbb{E}^{P^*}\left[W^*(t)|\mathcal{M}_s\right]M(s). \end{aligned}$$

Como $M(s) \neq 0$, vem

$$W^*(s) = \mathbb{E}^{P^*}\left[W^*(t)|\mathcal{M}_s\right],$$

o que prova que $W^*(t)$ é martingala relativamente a P^*.

A demonstração de que $R(t) = W^{*2}(t) - t$ é martingala relativamente a P^* é semelhante, trabalhando-se agora com $L(t)$ em vez de $K(t)$. ■

Capítulo 13

Opções e fórmula de Black-Scholes

13.1 Introdução

Uma *opção* confere ao seu detentor o direito, mas não a obrigação, de efectuar determinada actividade financeira. O detentor exercerá esse direito conforme o que lhe for mais vantajoso. Vejamos dois exemplos:

- Um importador recebeu uma encomenda cujo custo é de um milhão de dólares a pagar em 60 dias. Suponhamos que actualmente o preço de compra do dólar é de 0,81 Euros mas que, se o importador tiver de os comprar daqui a 60 dias a um preço superior a 0,85 Euros, terá prejuízos consideráveis. Para resolver o problema, poderá adquirir uma opção de compra de um milhão de dólares com data de expiração daqui a 60 dias ao preço de exercício de 0,85 Euros por dólar. Se, daqui a 60 dias, o preço de compra do dólar for superior a 0,85 Euros, ele exercerá a opção e comprará o milhão de dólares ao preço de exercício de 0,85 Euros por dólar. Se o preço de compra do dólar for inferior a 0,85 Euros, ele prescinde de exercer a sua opção e compra os dólares no mercado.

- Uma empresa tem de pagar mensalmente juros de um empréstimo à taxa EURIBOR mais um "spread" de 1,5% e quer fazer o seu orçamento. Para evitar desvios negativos desagradáveis relativamente às previsões orçamentais, gostaria de ter a garantia de que a taxa EURIBOR, actualmente de 2%, não ultrapassará 2,5%, mas gostaria de poder beneficiar se porventura ela não subisse tanto

ou até descesse. Pode adquirir um cabaz de opções a vencer mensalmente que o reembolsem da diferença caso a taxa EURIBOR ultrapasse na ocasião os $2,5\%$.

Estes foram apenas dois exemplos de uma grande variedade de opções, que também podem ser adquiridas para fins meramente especulativos. Há opções sobre acções, sobre obrigações, sobre mercadorias (café, soja, barrigas de porcos, ouro, zinco, etc.), sobre cotações de moedas estrangeiros, sobre índices de acções, sobre cotações de fundos financeiros, sobre catástrofes. As opçoes podem ser de *compra* (em inglês *"call option"*) ou *venda* (em inglês *"put option"*) de um produto financeiro. As opções podem ser *europeias* (caso a data de exercício seja rígida e fixada previamemnte), *americanas* (caso possa ser exercida em qualquer altura até a uma data pré-estabelecida). Existem ainda outras variantes de produtos financeiros derivados ("warrants", opções exóticas de vátios tipos, etc.).

Aqui vamos concentrar-nos no problema clássico de opções europeias de compra de acções tratado por Black, Scholes e Merton (ver [6] e [50]), que foi a primeira pedra de um novo e emocionante capítulo da matemática financeira.

Considere dois activos financeiros, um que é investimento sem risco, suponhamos uma obrigação com taxa instantânea de rendimento μ fixa (pode também ser uma conta bancária permitindo depóstitos ou empréstimos com taxa de juro fixa μ), e o outro é uma acção cotada na bolsa.

Designemos por $B(t)$ o valor da obrigação no instante t (isto é, daqui a t unidades de tempo), que será naturalmente dado por

$$B(t) = B(0)\exp(\mu t) \tag{13.1}$$

pois $dB(t)/dt = \mu B(t)$. Podemos escrever

$$dB(t) = \mu B(t)dt. \tag{13.2}$$

Supomos $B(0)$ um valor conhecido (determinístico).

No caso de ser obrigação, pressupõe-se que é transaccionável em qualquer instante (transacções em contínuo) e que a transacção, seja compra ou venda (com o mesmo valor da obrigação para ambos os casos), se faz sem encargos (impostos, taxas ou comissões). Estamos a supor que o rendimento será sempre pago na integra no instante da transacção, seja ele qual for; não é esse o caso de certos depósitos a prazo, por exemplo a 30 dias, em que os juros são pagos na totalidade se o depósito for levantado em instantes múltiplos inteiros de 30 dias mas se perdem os

juros desde o último final de prazo em caso de levantamento intercalar. Supõe-se também que qualquer interventor no mercado pode comprar ou emitir obrigações, pelo que o número de obrigações que detém pode ser positivo ou negativo. Se for negativo, isso significa que as emitiu para outros interventores e que está obrigado a entregar-lhes o seu valor $B(t)$ por cada obrigação se os outros interventores decidirem resgatá-las no instante t. Costuma-se falar em *posições longas* (do inglês "long") ou *posições curtas* (do inglês "short"), respectivamente para a detenção de um número negativo ou positivo de obrigações. Também podem ser vendidos e comprados números não inteiros de obrigações (na prática, se cada obrigação for de pequeno valor, esta hipótese é uma razoável aproximação). Caso se trate de uma conta bancária (em vez de obrigações), as nossas suposições significam que pode ter saldo positivo ou negativo e a taxa de juro é a mesma em ambos os casos, podendo o detentor levantar e depositar na conta a quantia (mesmo não-inteira) que desejar em qualquer instante sem perda de juros, não havendo encargos com essas transacções. Mas, para fixar a linguagem, falaremos daqui em diante de obrigações (sem risco) com rendimento fixo.

Designemos por $S(t)$ a cotação da acção no instante t e suponhamos que ela tem uma taxa média (aritmética) de rendimento r e uma volatilidade $\sigma > 0$, seguindo um *modelo de Black-Scholes* (ver Capítulo 3 e Capítulo 8 e ainda, para a interpretação de r, o Capítulo 9). Seja $S(0) > 0$ conhecido (determinístico). Temos assim a EDE de Itô

$$dS(t) = rS(t)dt + \sigma S(t)dW(t), \tag{13.3}$$

cuja solução é, como sabemos,

$$S(t) = S(0)\exp\left((r - \sigma^2/2)t + \sigma W(t)\right). \tag{13.4}$$

Também as acções se supõem transaccionáveis em qualquer instante (transacções em contínuo), com o mesmo valor para a compra ou a venda e sem encargos de transacção. Supões-se que não há dividendos (embora também se possam incluir com uma modificação do modelo de Black-Scholes). Supõe-se também que pode haver posições longas ou curtas. Se um interventor tem posição curta (número negativo de acções), isso significa que ele vendeu acções que não tinha, ficando obviamente obrigado a comprá-las de volta a quem as detenha em qualquer instante t ao valor de mercado $S(t)$. Também podem ser vendidos e comprados números não inteiros de acções (na prática, se cada acção for de pequeno valor, esta hipótese é uma razoável aproximação).

Uma *opção europeia de compra* com *data de expiração* $d > 0$ (instante determinístico) e *valor de exercício* $K > 0$ dá ao seu detentor o direito,

mas não a obrigação, de adquirir uma acção no instante d pagando o preço de exercício K. Caso a cotação $S(d)$ no instante d seja superior a K, o detentor da opção exercê-la-á comprando uma acção ao preço K e ficando com um benefício $S(d) - K > 0$ (pode se quiser vendê-la ao preço de mercado $S(d) > K$). Caso $S(d)$ seja inferior a K, o detentor da opção não irá exercê-la (se o fizesse teria de pagar K pela acção, o que lhe traria um prejuízo já que pode comprá-la no mercado ao preço $S(d) < K$); logo o seu benefício será nulo. Conclui-se que uma opção de compra dá ao seu detentor um benefício

$$(S(d) - K)^+ := \max(S(d) - K, 0).$$

Claro que a opção não pode ser gratuita pois quem a fornecesse, na melhor das hipótese, nada ganharia e, na pior, teria um prejuízo. A questão posta por Black, Merton e Scholes foi a de saber qual deveria ser, em cada instante t, o preço ou custo justo $Y(t)$ da opção. Normalmente interessa saber o preço $Y(0)$ da opção no instante $t = 0$ da sua compra, mas, como o seu detentor poderá querer vendê-la num instante posterior $t \leq d$ (depois de d a opção deixa de ter qualquer interesse ou valor), interessa também saber o seu valor em qualquer desses instantes.

Para além das hipóteses simplificadoras atrás referidas (e que podem ser depois relaxadas com modificações adequadas do modelo), supuseram também o princípio de *não-arbitragem*. A tradução matemática do termo é complexa mas, em linguagem corrente, significa que não é possível, com um investimento nulo, obter um lucro sem risco. Outra definição equivalente é a de que, se dois *portfolios* (carteiras de investimentos) têm, em certo instante futuro, o mesmo valor quaisquer que sejam os cenários do mercado ω, então devem ter o mesmo valor agora. Com efeito, se assim não fosse, quem tivesse o portfolio mais valorizado agora, poderia trocá-lo pelo menos valorizado, ficando no futuro com o mesmo valor que teria se não tivesse feito a troca mas embolsando a diferença entre o valor actual dos dois portfolios; teria assim um lucro com um investimento nulo e sem correr qualquer risco. A hipótese de não-arbitragem é, de certa forma, consequência da ideia de mercado eficiente e de operadores de mercado igualmente informados e de comportamento inteligente, pelo que qualquer situação de arbitragem seria rapidamente desfeita pela maior procura e subida do valor dos activos subvalorizados; não sobrevive obviamente se houver "inside trading".

Vejamos agora o que se entende por *estratégia de transacção* $(b(t), s(t))$ $(t \geq 0)$ de um interventor no mercado. Aqui $b(t)$ representa o número de obrigações e $s(t)$ o número de acções que ele detém no instante t. Note-se que $b(t)$ e $s(t)$ podem tomar valores reais positivos (posições longas) ou negativos (posições curtas). O conhecimento da es-

tratégia indica-nos quais as compras e vendas efectuadas. Por exemplo, sempre que $s(t)$ desce de uma certa quantidade, isso significa que naquele instante o inverventor vendeu tantas acções quanto a descida observada. Se sobe, houve compra. Claro que $(b(0), s(0))$ é o portfolio inicial de obrigações e acções e o seu valor será

$$b(0)B(0) + s(0)S(0).$$

A estratégia diz-se *autofinanciada* se o seu valor em cada instante t for o valor inicial mais os rendimentos acumulados gerados pelas acções $\int_0^t s(u)dS(u)$ e pelas obrigações $\int_0^t b(u)dB(u)$ detidas em cada momento. Isto é, as compras e vendas de obrigações e acções são livres mas só podem recorrer a meios financeiros contidos no portfolio, não podendo haver saídas ou entradas de meios financeiros. Em termos matemáticos, a estratégia é autofinanciada se, para todo o $t \geq 0$, se tiver

$$b(t)B(t)+s(t)S(t) = b(0)B(0)+s(0)S(0)+\int_0^t b(u)dB(u)+\int_0^t s(u)dS(u). \tag{13.5}$$

13.2 Fórmula de Black-Scholes e estratégia hedging

Olhemos agora para o valor da opção $Y(t)$. Como no instante d ela confere o benefício $(S(d) - K)^+$, esse deverá ser o seu justo valor (custo) no instante d, isto é

$$Y(d) = (S(d) - K)^+.$$

Suponhamos que existe uma estratégia autofinanciada $(b(t), s(t))$ cujo valor no instante d é precisamente $Y(d)$, ou seja

$$b(d)B(d) + s(d)S(d) = Y(d).$$

Então a hipótese de não-arbitragem obriga a que

$$b(0)B(0) + s(0)S(0) = Y(0).$$

Com efeito, caso fosse $b(0)B(0) + s(0)S(0) < Y(0)$, alguém poderia vender uma opção (são permitidas posições longas e curtas) por $Y(0)$, utilizar parte desse dinheiro para adquirir o portfolio $(b(0), s(0))$ de valor $b(0)B(0) + s(0)S(0)$ (isto é, adquirir $b(0)$ obrigações e $s(0)$ acções) e guardar a diferença $Y(0) - (b(0)B(0) + s(0)S(0)) > 0$. Usando o portfolio e adoptando a estratégia de transacções $(b(t), s(t))$ até ao instante d (como é autofinanciada, as transacções fazem-se com os meios gerados pelo próprio portfolio), no instante d o portfolio vale $b(d)B(d) +$

$s(d)S(d) = Y(d)$, podendo ser liquidado o portfolio para obter a quantia $Y(d)$, que será precisamente o benefício que deverá entregar ao detentor da opção. Quem seguisse estes procedimentos ficaria a ganhar precisamente a diferença guardada no início $Y(0) - (b(0)B(0) + s(0)S(0)) > 0$ sem ter feito qualquer investimento nem corrido qualquer risco, pelo que haveria arbitragem. Caso fosse $b(0)B(0) + s(0)S(0) > Y(0)$, poderia seguir um procedimento simétrico, adquirindo o portfolio $(-b(0), -s(0))$ de valor $-b(0)B(0) - s(0)S(0)$ e simultaneamente comprando uma opção por $Y(0)$, guardando a diferença $-Y(0) + (-b(0)B(0) - s(0)S(0)) > 0$. Depois investiria de acordo com a estratégia (que também é autofinanciada) $(-b(t), -s(t))$, liquidando o portfolio no instante d pelo valor $-b(d)B(d) - s(d)S(d) = -Y(d)$, precisamente oque irá receber de benefício pela opção. Quem seguisse estes procedimentos ficaria a ganhar precisamente a diferença guardada no início $-Y(0) + (-b(0)B(0) - s(0)S(0)) > 0$ sem ter feito qualquer investimento nem corrido qualquer risco, pelo que haveria arbitragem.

O mesmo raciocínio aplicado ao instante $t \in [0, d]$ levaria a concluir que o princípio da não-arbitragem implica

$$b(t)B(t) + s(t)S(t) = Y(t). \tag{13.6}$$

Claro que o preço $Y(t)$ da opção no instante t, dado o carácter markoviano e não antecipativo da informação, só pode depender do instante t e da cotação $S(t)$ da acção nesse instante (e não em instantes anteriores ou posteriores). Note-se que o valor da obrigação em cada momento é determinístico. Logo, existe uma função $C(t, x)$ tal que

$$Y(t) = C(t, S(t)).$$

Usando a fórmula de Itô, vem, usando (13.2) e (13.3),

$$\begin{aligned} dY(t) = & \frac{\partial C(t,S(t))}{\partial t}dt + \frac{\partial C(t,S(t))}{\partial x}dS(t) + \frac{1}{2}\frac{\partial^2 C(t,S(t))}{\partial x^2}(dS(t))^2 = \\ & \left(\frac{\partial C}{\partial t} + \frac{\partial C}{\partial x}rS(t) + \frac{1}{2}\frac{\partial^2 C}{\partial x^2}\sigma^2 S^2(t)\right)dt + \frac{\partial C}{\partial x}\sigma S(t)dW(t). \end{aligned} \tag{13.7}$$

Atendendo a (13.5) e (13.6), vem

$$\begin{aligned} & Y(t) = b(t)B(t) + s(t)S(t) = \\ & b(0)B(0) + s(0)S(0) + \int_0^t b(u)dB(u) + \int_0^t s(u)dS(u), \end{aligned} \tag{13.8}$$

donde resulta

$$\begin{aligned} &dY(t) = \\ &b(t)dB(t) + s(t)dS(t) = b(t)\mu B(t)dt + s(t)\left(rS(t)dt + \sigma S(t)dW(t)\right) = \\ &\left(b(t)\mu B(t) + s(t)rS(t)\right)dt + s(t)\sigma S(t)dW(t). \end{aligned} \tag{13.9}$$

Como $Y(t)$ é um processo de difusão e, portanto, tem um só coeficiente de tendência e um só coeficiente de difusão, podemos igualar os coeficientes entre (13.7) e (13.9). Vem

$$b(t)\mu B(t) + s(t)rS(t) = \frac{\partial C}{\partial t} + \frac{\partial C}{\partial x}rS(t) + \frac{1}{2}\frac{\partial^2 C}{\partial x^2}\sigma^2 S^2(t) \tag{13.10}$$

$$s(t)\sigma S(t) = \frac{\partial C}{\partial x}\sigma S(t). \tag{13.11}$$

Como $S(t) > 0$, de (13.11) vem

$$s(t) = \frac{\partial C(t, S(t))}{\partial x}, \tag{13.12}$$

pelo que $C(t, S(t)) = Y(t) = b(t)B(t) + s(t)S(t) = b(t)B(t) + S(t)\frac{\partial C(t,S(t))}{\partial x}$, donde

$$b(t) = \frac{1}{B(t)}\left(C(t, S(t)) - \frac{\partial C(t, S(t))}{\partial x}S(t)\right). \tag{13.13}$$

Substituindo em (13.10) $b(t)$ e $s(t)$ pelas suas expressões (13.12) e (13.13), obtemos, usando x em lugar de $S(t)$ para simplificar a notação, para $t \in [0, d)$ e $x > 0$, a equação às derivadas parciais

$$-\frac{\partial C(t,x)}{\partial t} = \frac{\partial C(t,x)}{\partial x}\mu x + \frac{1}{2}\frac{\partial^2 C(t,x)}{\partial x^2}\sigma^2 x^2 - \mu C(t,x). \tag{13.14}$$

A condição de fronteira é obviamente

$$C(d, x) = (x - K)^+. \tag{13.15}$$

Se invertermos a ordem do tempo, pondo $\tau = d - t$ e $D(\tau, x) = C(d-\tau, x)$ (donde $Y(t) = D(d-t, S(t))$), vem $\frac{\partial D}{\partial \tau} = \frac{\partial D}{\partial x}\mu x + \frac{1}{2}\frac{\partial^2 D}{\partial x^2}\sigma^2 x^2 - \mu D$ com condição terminal $D(0, x) = (x-K)^+$. Trata-se de um caso particular da equação de Feynman-Kac (10.11) para $h(x) = (x-K)^+$ (não é propriamente de suporte compacto mas pode ser aproximada por funções desse tipo), $q(x) \equiv \mu$, um coeficiente de tendência μx e um coeficiente de difusão $\sigma^2 x^2$. Assim, atendendo a (10.10), a solução $D(\tau, x)$ será da

forma

$$D(\tau,x)=\mathbb{E}_x\left[e^{-\int_0^\tau \mu ds}(X(\tau)-K)^+\right]=\mathbb{E}_x\left[e^{-\mu\tau}(X(\tau)-K)^+\right], \tag{13.16}$$

onde $X(t)$ é um processo de difusão com valor inicial $X(0)=x$ e os coeficientes de tendência e difusão já referidos. Logo $dX(t)=\mu X(t)dt+\sigma X(t)dW(t)$, pelo que vem $X(\tau)=x\exp\left((\mu-\sigma^2/2)\tau+\sigma W(\tau)\right)$. Como τ é determinístico e $Z(\tau)=\ln X(\tau)\frown\mathcal{N}(\ln x+(\mu-\sigma^2/2)\tau,\sigma^2\tau)$, vem, com $\alpha=\mu-\sigma^2/2$ e $\beta=\mu+\sigma^2/2$,

$$\begin{array}{l}
D(\tau,x)=e^{-\mu\tau}\mathbb{E}_x\left[\left(e^{Z(\tau)}-K\right)^+\right]= \\
e^{-\mu\tau}\int_{-\infty}^{+\infty}(e^z-K)^+\frac{1}{\sigma\sqrt{2\pi\tau}}\exp\left(-\frac{(z-\ln x-\alpha\tau)^2}{2\sigma^2\tau}\right)dz \\
e^{-\mu\tau}\int_{\ln K}^{+\infty}(e^z-K)\frac{1}{\sigma\sqrt{2\pi\tau}}\exp\left(-\frac{(z-\ln x-\alpha\tau)^2}{2\sigma^2\tau}\right)dz= \\
\int_{\ln K}^{+\infty}e^{-\mu\tau}e^z\frac{1}{\sigma\sqrt{2\pi\tau}}\exp\left(-\frac{(z-\ln x-\alpha\tau)^2}{2\sigma^2\tau}\right)dz- \\
\int_{\ln K}^{+\infty}e^{-\mu\tau}K\frac{1}{\sigma\sqrt{2\pi\tau}}\exp\left(-\frac{(z-\ln x-\alpha\tau)^2}{2\sigma^2\tau}\right)dz= \\
\int_{\ln K}^{+\infty}x\frac{1}{\sigma\sqrt{2\pi\tau}}\exp\left(-\frac{(z-\ln x-\beta\tau)^2}{2\sigma^2\tau}\right)dz- \\
\int_{\ln K}^{+\infty}e^{-\mu\tau}K\frac{1}{\sigma\sqrt{2\pi\tau}}\exp\left(-\frac{(z-\ln x-\alpha\tau)^2}{2\sigma^2\tau}\right)dz= \\
x\int_{\frac{\ln K-\ln x-\beta\tau}{\sigma\sqrt{\tau}}}^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-u^2/2}du-Ke^{-\mu\tau}\int_{\frac{\ln K-\ln x-\alpha\tau}{\sigma\sqrt{\tau}}}^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-v^2/2}dv \\
x\left(1-\Phi\left(\frac{\ln K-\ln x-\beta\tau}{\sigma\sqrt{\tau}}\right)\right)-Ke^{-\mu\tau}\left(1-\Phi\left(\frac{\ln K-\ln x-\alpha\tau}{\sigma\sqrt{\tau}}\right)\right),
\end{array}$$

onde $\Phi(u)=\int_{-\infty}^{u}\frac{1}{\sqrt{2\pi}}e^{-v^2/2}dv$ é a função de distribuição de uma v.a. normal padrão.

Daqui resulta, após algumas contas,

$$\begin{array}{rl}
C(t,x)= & x\Phi\left(\theta_t\right)-K\exp(-\mu(d-t))\Phi\left(\theta_t-\sigma\sqrt{d-t}\right), \\
\theta_t= & \frac{\ln(x/K)+(\mu+\sigma^2/2)(d-t)}{\sigma\sqrt{d-t}}.
\end{array} \tag{13.17}$$

Esta é a célebre *fórmula de Black-Scholes.*

Para obter o valor da opção num instante t basta saber o valor $x=S(t)$ da cotação da acção nesse instante:

$$Y(t)=C(t,S(t)), \tag{13.18}$$

pelo que

$$Y(t) = C(t, S(t)) = \\ S(t)\Phi\left(\frac{\ln\left(\frac{S(t)}{K}\right)+(\mu+\frac{\sigma^2}{2})(d-t)}{\sigma\sqrt{d-t}}\right) - Ke^{-\mu(d-t)}\Phi\left(\frac{\ln\left(\frac{S(t)}{K}\right)+(\mu-\frac{\sigma^2}{2})(d-t)}{\sigma\sqrt{d-t}}\right). \tag{13.19}$$

O valor inicial da opção facilmente se obtém pondo $t = 0$.

Note-se que a fórmula não depende da taxa média de rendimento r das acções mas antes da taxa de rendimento μ das obrigações sem risco, o que é particularmente importante porque r é um parâmetro muito difícil de estimar com precisão a partir da observação das cotações da acção. Note que $C(d, S(d)) = (S(d) - K)^+$; com efeito, com $x = S(d)$, $\theta_d = +\infty\ [-\infty]$ conforme $S(d) > K\ [< K]$ e note que, para $t = d$, $\theta_d - \sigma\sqrt{d-t} = \theta_d$, pelo que $C(d, S(d)) = S(d) - K$ se $S(d) > K$ e $C(d, S(d)) = 0$ se $S(d) < K$.

A validade da fórmula de Black-Scholes dependia, porém, da existência de uma estratégia autofinanciada $(b(t), s(t))$ tal que $Y(T) = b(T)B(T) + s(T)S(T)$. Será que ela existe? A resposta é positiva. É precisamente a estratégia dada pelas expressões (13.12) e (13.13), de que se obtém, utilizando (13.19) e após alguns cálculos,

$$s(t) = \Phi\left(\frac{\ln\left(\frac{S(t)}{K}\right) + (\mu + \frac{\sigma^2}{2})(d-t)}{\sigma\sqrt{d-t}}\right) \tag{13.20}$$

e

$$b(t) = -\frac{1}{B(t)}Ke^{-\mu(d-t)}\Phi\left(\frac{\ln\left(\frac{S(t)}{K}\right) + (\mu - \frac{\sigma^2}{2})(d-t)}{\sigma\sqrt{d-t}}\right). \tag{13.21}$$

Note-se que $s(t) > 0$ e $b(t) < 0$, o que significa que a estratégia autofinanciada referida tem sempre uma posição longa em acções e curta em obrigações. O capital da estratégia autofinanciada investido em acções e obrigações deve ser, no instante t, respectivamente, a primeira e a segunda parcela da segunda linha da expressão (13.19). É fácil ver que a estratégia é de facto autofinanciada.

Uma conclusão extremamente importante é que existe uma estratégia autofinanciada dada por (13.20) e (13.21) cujo valor da carteira reproduz, em cada instante t, o valor da opção $Y(t)$, isto é, reproduz os resultados que se podem obter com a compra de uma opção cujo preço seja calculado de acordo com a fórmula de Black-Scholes. Podemos chamá-la *estratégia hedging* (a palavra inglesa "hedging" refere-se a colocação de uma cerca à

volta da casa), que poderíamos traduzir por *estratégia de protecção*. Em princípio, a aquisição de uma opção europeia de compra com preço dado pela fórmula de Black-Scholes é desnecessária pois o que se iria gastar com a sua aquisição pode ser gasto na compra do portfolio $(b(0), s(0))$ e seguido da adopção da estratégia autofinanciada hedging, que resulta em obter-se no instante d o mesmo benefício que se obteria com a opção, isto seja qual for o cenário do mercado. Também o operador financeiro que venda uma opção pelo valor determinado pela fórmula de Black-Scholes fica isento de qualquer benefício ou prejuízo, pois o que recebe da venda pode ser investido na compra do portfolio $(b(0), s(0))$ e seguido da adopção da estratégia autofinanciada hedging, resultando na obtenção no instante d de resultados financeiros iguais ao benefício a pagar ao comprador da opção.

Há, porém, uma questão relevante, que é a de, sempre que $S(t)$ varia, variam os valores de $b(t)$ e de $s(t)$, obrigando à compra ou venda de acções e obrigações para se manter na estratégia hedging, o que não é prático para operadores individuais. Teoricamente, $S(t)$ varia continuamente ao longo do intervalo $[0, d]$, o que significa transacções contínuas. Tal, naturalmente, só é viável sem custos de transacção (com custos, há que fazer ajustamentos).

13.3 Um exemplo numérico

Consideremos o exemplo da cotação das acções do banco BCP entre 8 de Abril de 1991 e 30 de Junho de 1997, cuja trajectória observada, em escala logatítmica, aparece na Figura 2.1 (Secção 2.3). Da trajectória foi possível estimar os parâmetros $\hat{\sigma} = 0,1932/\sqrt{\text{ano}}$ (um intervalo de confiança a 95% é $0,1932 \pm 0,0070$) e $\hat{R} = 0,084/\text{ano}$ (um intervalo de confiança a 95% é $0,084 \pm 0,152$), com $R = r - \sigma^2/2$.

Vamos supor que a taxa instantânea de rendimento das obrigações sem risco é de $\mu = 4\%$ ao ano (é um valor um tanto elevado para os dias de hoje mas não para aquela época). Suponhamos que, no dia 30 de Junho de 1997, que vamos tomar como o nosso instante 0, alguém adquiriu uma opção europeia de compra de uma acção com data de exercício em 30 de Setembro de 1997 (o nosso instante $d = 0,2518$ anos) e valor de exercício de $K = 3500$ escudos (a moeda era então o escudo). Note-se que, no dia 30 de Junho de 1997, a cotação da acção era de $S(0) = 3350$ escudos e o intervalo de confiança a 95% da previsão do valor da acção a 30 de Setembro de 1997 era de 3422 ± 732 escudos. Quem comprou a opção não podia saber que, no dia 30 de Setembro de 1997, a cotação da acção no mercado seria $S(d) = 3720$ escudos, pelo que iria ter um benefício de 220 escudos (podia mesmo não ter qualquer

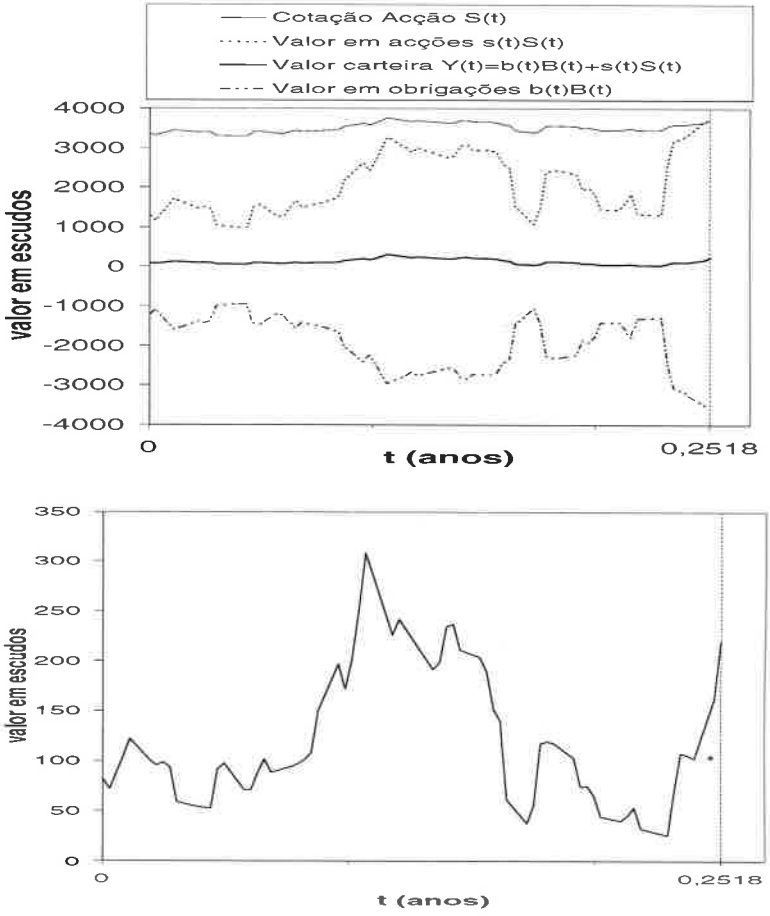


Figura 13.1: Estratégia hedging para uma opção de compra europeia de uma acção do BCP, adquirida em 30 de Junho de 1997 ($t = 0$) com data de exercício 30 de Setembro de 1997 ($d = 0,2518$ anos) e valor de exercício $K = 3500$ escudos. A taxa de rendimento instantânea das obrigações sem risco é $\mu = 4\%$ ao ano. Na parte superior da figura, encontramos, de cima para baixo, a evolução de:
a) cotação $S(t)$ da acção ($S(0) = 3350$ escudos, $S(d) = 3720$ escudos);
b) valor $s(t)S(t)$ da componente em acções da estratégia hedging;
c) valor total da estratégia hedging e também valor da opção $Y(t) = b(t)B(t) + s(t)S(t)$ ($Y(0) = 82,21$ escudos e $Y(d) = 3720 - 3500 = 220$ escudos), que está repetido e ampliado na parte inferior da figura;
d) valor $b(t)B(t)$ da componente em obrigações da estratégia hedging.

benefício se a cotação na data de expiração da opção estivesse abaixo de 3500, como até era mais provável de acordo com as previsões).

De acordo com a fórmula de Black-Scholes, o custo em 30 de Junho de 1997 da opção era $Y(0) = 82,21$ escudos. A Figura 13.1 apresenta a evolução da estratégia autofinanciada hedging, dicriminando o valor global da carteira e o valor em acções e obrigações.

13.4 Obtenção da fórmula de Black-Scholes via teorema de Girsanov

Vamos chegar à fórmula de Black-Scholes por outra via, que tira partido do teorema de Girsanov e que facilita o tratamento deste problema quando se utilizam modelos mais complexos que o modelo de Black-Scholes.

Dá jeito trabalhar com preços deflaccionados, usando como referência o preço do investimento (neste caso as obrigações) sem risco $B(t) = B(0)\exp(\mu t)$. Os preços deflaccionados das obrigações, das acções e das opções são, respectivamente,

$$\tilde{B}(t) = B(t)/B(t) \equiv 1, \tag{13.22}$$

$$\tilde{S}(t) = S(t)/B(t) = S(t)\exp(-\mu t)/B(0) \tag{13.23}$$

e

$$\tilde{Y}(t) = Y(t)/B(t). \tag{13.24}$$

Vem

$$d\tilde{S}(t) = \frac{1}{B(0)}\{-\mu S(t)\exp(-\mu t)dt + \exp(-\mu t)dS(t) + 0(dS(t))^2\},$$

donde

$$d\tilde{S}(t) = (r-\mu)\tilde{S}(t)dt + \sigma\tilde{S}(t)dW(t), \tag{13.25}$$

pelo que $\tilde{S}(t)$ é um processo de difusão homogémeo com coeficiente de tendência

$$f(x) = (r-\mu)x \tag{13.26}$$

e coeficiente de difusão

$$g^2(x) = (\sigma x)^2. \tag{13.27}$$

Vamos querer mudar o seu coeficiente de tendência para

$$f^*(x) = 0 \tag{13.28}$$

por aplicação do *teorema de Girsanov*. A função (que representa o *preço*

de mercado do risco)

$$\theta(x) = \frac{r-\mu}{\sigma} \tag{13.29}$$

satisfaz a equação (12.4), que neste caso toma a forma $\sigma x\theta(x) = (r-\mu)x - 0$.

De (12.5) obtemos a martingala

$$M(t) = \exp\left(-\frac{r-\mu}{\sigma}W(t) - \frac{(r-\mu)^2}{2\sigma^2}t\right) \tag{13.30}$$

e definimos a nova medida de probabilidade P^* através da sua derivada de Radon-Nikodym

$$\frac{dP^*}{dP} = M(d) = \exp\left(-\frac{r-\mu}{\sigma}W(d) - \frac{(r-\mu)^2}{2\sigma^2}d\right). \tag{13.31}$$

De (12.7) obtemos

$$W^*(t) = W(t) + \frac{r-\mu}{\sigma}t, \tag{13.32}$$

que é um processo de Wiener relativamente à probabilidade P^*. De (12.8) vem

$$d\tilde{S}(t) = \sigma\tilde{S}(t)dW^*(t), \tag{13.33}$$

cuja solução é

$$\tilde{S}(t) = \tilde{S}(0) + \int_0^t \sigma\tilde{S}(u)dW^*(u). \tag{13.34}$$

Pelas propriedades do integral estocástico, $\tilde{S}(t)$ é uma martingala com respeito a P^*. Pelo facto de P^* transformar em martingala o processo das cotações da acção desinflaccionadas pelo valor do investimento sem risco e pelo facto de ser uma medida de probabilidade equivalente a P, a literatura financeira designa P^* por *medida martingala equivalente.*

Usando o teorema de Itô, vem

$$dS(t) = d(\tilde{S}(t)B(t)) = \mu S(t)dt + \sigma S(t)dW^*(t). \tag{13.35}$$

Repare-se que, relativamente a $W^*(t)$ (processo de Wiener relativamente a P^*), ou seja relativamente à medida martingala equivalente P^*, a taxa média (aritmética) de rendimento das acções é igual a μ, isto é, igual à taxa de rendimento do investimento sem risco. A nova probabilidade P^* modifica as probabilidades dos cenários de mercado de maneira a que a economia pareça neutral ao risco.

Se houver uma *estratégia hedging* $(b(t), s(t))$, isto é, uma estratégia autofinanciada que reproduza o valor $Y(t)$ da opção, então

$$Y(t) = b(t)B(t) + s(t)S(t) =$$
$$b(0)B(0) + s(0)S(0) + \int_0^t b(u)dB(u) + \int_0^t s(u)dS(u).$$

Logo

$$\begin{aligned} dY(t) = \; & b(t)dB(t) + s(t)dS(t) = \\ & b(t)\mu B(t)dt + s(t)\left(\mu S(t)dt + \sigma S(t)dW^*(t)\right) = \\ & \mu(b(t)B(t) + s(t)S(t)) + \sigma s(t)S(t)dW^*(t) = \\ & \mu Y(t)dt + \sigma s(t)S(t)dW^*(t) \end{aligned}$$

e

$$\begin{aligned} d\tilde{Y}(t) = \; & \tfrac{1}{B(0)} d\left(e^{-\mu t}Y(t)\right) = \\ & \tfrac{1}{B(0)}\left(\mu e^{-\mu t}Y(t)dt + e^{-\mu t}dY(t) + 0(dY(t))^2\right) = \\ & \tfrac{1}{B(t)}\left(-\mu Y(t)dt + dY(t)\right) = \\ & \tfrac{1}{B(t)}\left(-\mu Y(t) + \mu Y(t)dt + \sigma s(t)S(t)dW^*(t)\right) = \\ & \sigma s(t)\tilde{S}(t)dW^*(t). \end{aligned}$$

Daqui resulta que

$$\tilde{Y}(t) = \tilde{Y}(0) + \int_0^t \sigma s(u)\tilde{S}(u)dW^*(u) \tag{13.36}$$

é uma martingala relativamente a P^*.

Logo temos, para $0 \leq t \leq d$,

$$\tilde{Y}(t) = \mathbb{E}^{P^*}[\tilde{Y}(d)|\mathcal{M}_t]. \tag{13.37}$$

"Traduzindo" directamente (13.37) vem, pondo $Y(t) = C(t, S(t))$,

$$\frac{C(t,S(t))}{B(0)e^{\mu t}} = \mathbb{E}^{P^*}\left[\left(\frac{S(d)-K}{B(0)e^{\mu d}}\right)^+ \middle| S(t)\right],$$

donde

$$C(t,x) = \mathbb{E}^{P^*}\left[\left(S(d)e^{-\mu(d-t)} - Ke^{-\mu(d-t)}\right)^+ \middle| S(t) = x\right]. \tag{13.38}$$

De (13.35), vem

$$d\ln S(t) = (\mu - \sigma^2/2)dt + \sigma dW^*(t) \tag{13.39}$$

e, portanto,

$$\ln S(d) = \ln S(t) + (\mu - \sigma^2/2)(d-t) + \sigma(W^*(d) - W^*(t)).$$

Logo, condicionalmente a $S(t) = x$ e relativamente à probabilidade P^*, a v.a $Z = \ln S(d)$ tem distribuição

$$\mathcal{N}\left(\ln x + (\mu - \sigma^2/2)(d-t), \sigma^2(d-t)\right),$$

a v.a.

$$U = \frac{Z - \ln x - (\mu - \sigma^2/2)(d-t)}{\sigma\sqrt{d-t}}$$

é normal padrão e vem

$$\begin{aligned}&\left(S(d)e^{-\mu(d-t)} - Ke^{-\mu(d-t)}\right)^+ = \\ &\left(xe^{-\sigma^2(d-t)/2+\sigma\sqrt{d-t}\,U} - Ke^{-\mu(d-t)}\right)^+.\end{aligned}$$

Logo, de (13.38), resulta que $C(t,x)$ é a esperança matemática de

$$\left(xe^{-\sigma^2(d-t)/2+\sigma\sqrt{d-t}\,U} - Ke^{\mu(d-t)}\right)^+,$$

com U v.a. normal padrão. Logo, pondo

$$\gamma = \frac{\ln K - \ln x - (\mu - \sigma^2/2)(d-t)}{\sigma\sqrt{d-t}},$$

vem

$$\begin{aligned}&C(t,x) = \int_{-\infty}^{+\infty}\left(xe^{-\sigma^2(d-t)/2+\sigma\sqrt{d-t}\,u} - Ke^{-\mu(d-t)}\right)^+ \frac{1}{\sqrt{2\pi}}e^{-u^2/2}du = \\ &\int_{\gamma}^{+\infty}\left(xe^{-\sigma^2(d-t)/2+\sigma\sqrt{d-t}\,u} - Ke^{-\mu(d-t)}\right)\frac{1}{\sqrt{2\pi}}e^{-u^2/2}du = \\ &x\int_{\gamma}^{+\infty}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(u-\sigma\sqrt{d-t})^2}{2}\right)du - Ke^{-\mu(d-t)}\int_{\gamma}^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-u^2/2}du = \\ &x\int_{\gamma-\sigma\sqrt{d-t}}^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-v^2/2}dv - Ke^{-\mu(d-t)}\int_{\gamma}^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-u^2/2}du = \\ &x\left(1 - \Phi\left(\gamma - \sigma\sqrt{d-t}\right)\right) - Ke^{-\mu(d-t)}\left(1 - \Phi(\gamma)\right) = \\ &x\Phi\left(\frac{\ln(x/K)+(\mu+\sigma^2/2)(d-t)}{\sigma\sqrt{d-t}}\right) - Ke^{-\mu(d-t)}\Phi\left(\frac{\ln(x/K)+(\mu-\sigma^2/2)(d-t)}{\sigma\sqrt{d-t}}\right).\end{aligned}$$

Obtivémos por outra via a fórmula de Black-Scholes (13.17).

A dedução anterior pressuponha a existência da estratégia hedging. Da expressão de $C(t,x)$ podemos obter essa estratégia e mostrar que reproduz o valor da opção pelos mesmos métodos que usámos na Secção

13.2. Noutros modelos mais complexos do que o de Black-Scholes em que não seja possível obter a expressão explícita de $C(t,s)$, a prova da existência e a determinação da estratégia hedging poderá ter de recorrer ao teorema da representação de Yor, que está fora do âmbito deste texto. Nesses modelos mais complexos, é possível obter valores aproximados de $Y(t)$, calculando aproximadamente $C(t,x)$ por simulação de Monte Carlo, uma vez que a esperança matemática de (13.38) pode ser simulada. Essa a vantagem deste método do teorema de Girsanov.

Vamos na próxima Secção ver um método numérico alternativo para obter aproximadamente $C(t,x)$, baseado na aproximação pelo modelo binomial.

13.5 O modelo binomial

O *modelo binomial*, também conhecido como *modelo de Cox-Ross-Rubinstein* ou *modelo CRR*, é uma discretização no tempo e no espaço de estados para o processo $\ln S(t)$, trabalhando com a medida martingala equivalente P^*, que nesta técnica é, em cada passo, aproximada por uma v.a. de Bernoulli. Esta técnica pode também ajudar o leitor a compreender melhor o que se está a passar. Mas o seu objectivo é o de fazer simulações dos fenómenos de uma forma muito rápida e obter aproximações das quantidades de interesse, como seja o valor de uma opção. Vamos exemplificar com o modelo de Black-Scholes, usando o exemplo da Secção 13.2, mas a técnica aplica-se, com as necessárias adaptações, a qualquer outro modelo. Pela sua flexibilidade e rapidez, trata-se de um instrumento muito útil.

Partamos da equação (13.39), em que vemos que o processo $\ln S(t)$ é, relativamente à medida martingala equivalente P^*, um processo de difusão com coefciente de tendência $\mu - \sigma^2/2$ e coeficiente de difusão σ^2. A ideia é substituir este processo por um passeio aleatório adequado $\ln \bar{S}(t)$, que é um processo em tempo discreto e com espaços de estados discreto tal que, em cada passo do tempo, o processo se move para um dos dois estados imediatamente contíguos àquele em que estava. O intervalo de tempo $[0,d]$ é discretizado em n passos de igual duração $\Delta t = d/n$.

Note-se que $\ln S(t+\Delta t) - \ln S(t) = \int_t^{t+\Delta t} (\mu - \sigma^2/2) ds + \int_t^{t+\Delta t} \sigma dW(s)$ $= (\mu - \sigma^2/2)\Delta t + \sigma \Delta W(t) \frown \mathcal{N}((\mu - \sigma^2/2)\Delta t, \sigma^2 \Delta t)$, isto é, no passo temporal que vai de t para $t + \Delta t$, o processo original é incrementado de uma quantidade aleatória com distribuição normal de média $(\mu - \sigma^2/2)\Delta t$ e desvio-padrão $\sigma\sqrt{\Delta t}$. [1] O que se faz é, em cada passo temporal, apro-

[1] Para outros modelos, a distribuição seria aproximadamente normal com média igual ao coeficiente de tendência multiplicado por Δt e desvio-padrão igual à raíz

ximar a distribuição normal por uma distribuição de Bernoulli aproximadamente com a mesma média e desvio-padrão. Suponhamos que, no instante t, o processo $\ln \bar{S}(t)$ está no estado z e seja $\Delta z = \sigma\sqrt{\Delta t}$. Então ele pode, no passo temporal discreto que vai de t para $t + \Delta t$, transitar para o estado $z + \Delta z$ ou $z - \Delta z$, com probabilidades p^* e $q^* = 1 - p^*$, respectivamente. Se p^* fosse igual a $1/2$, o salto médio seria 0, diferente do pretendido, e o desvio padrão seria $\Delta z = \sigma\sqrt{\Delta t}$, precisamente o pretendido. Há portanto que fazer uma pequena correcção do valor de $p*$ de modo a acertar a média sem, no entanto, alterar significativamente o desvio-padrão. Como o salto médio é $p^*\Delta z - q^*\Delta z = (2p^* - 1)\Delta z$, obtém-se exactamente a média desejada $(\mu - \sigma^2/2)\Delta t$ escolhendo

$$p^* = \frac{1}{2}\left(1 + \frac{(\mu - \sigma^2/2)\Delta t}{\sigma\sqrt{\Delta t}}\right). \tag{13.40}$$

O desvio-padrão, porém, vem agora $= \sigma\sqrt{\Delta t}$+termos de ordem inferior, o que é uma aproximação satisfatória. Claro que, ao fim de alguns saltos, o incremento global terá (neste exemplo) distribuição binomial (soma de v.a. de Bernoulli i.i.d.). Facilmente se reconhece que, para $t = 0, \Delta t, 2\Delta t, ..., d = n\Delta t$, a distribuição binomial de $\ln \bar{S}(t)$ se aproxima, quando $n \to +\infty$, da distribuição normal (relativamente à medida de probabilidade P^*) de $\ln S(t)$.

A Figura 13.2 ilustra uma trajectória do modelo binomial.

Apliquemos o método no caso do exemplo da Secção 13.3, onde $S(0) = 3350$, $\sigma = 0,1932/\sqrt{\text{ano}}$ (estimativa), $\mu = 0,04$ por ano, $d = 0,2518$ anos. Para simplificar as contas, e porque se trata só de uma ilustração, vamos pôr $n = 1$, mas seria desejável um n relativamente grande para que a aproximação fosse razoável. Logo $\Delta t = d/n = 0,2518$ anos e $\Delta z = \sigma\sqrt{\Delta t} = 0,09695$. Pode-se então saltar de $\ln S(0) = 8,116716$ para $8,116716 + 0,09695 = 8,213666$ (a que corresponde $\bar{S}(0 + \Delta t) = \bar{S}(d) = 3691,05$ escudos) ou para $8,116716 - 0,09695 = 8,019766$ (a que corresponde $\bar{S}(0 + \Delta t) = \bar{S}(d) = 3040,47$ escudos). As probabilidades de cada salto são, respectivamente, $p* = 0,52771$ (usou-se (13.40)) e $q* = 0,47229$.

Note-se que como se usou $K = 3500$ escudos, se fosse $S(d) = 3040,47$ viria $Y(d) = 0$ e, se fosse $S(d) = 3691,05$ escudos, viria $Y(d) = 191,05$ escudos. Supondo que cada obrigação vale 1 escudo no instante $t = 0$, ela vale $B(d) = B(0)\exp(\mu d) = \exp(0,04 \times 0,2518) = 1,010123$ no instante d e o valor desinflaccionado da opção será $\tilde{Y}(d) = Y(d)/1,010123$, ou seja, será 0 ou 189,14. O valor esperado de $\tilde{Y}(d)$ com a informação disponível no instante 0 dependerá da probabilidade p^*. Usando o valor

quadrada do coeficiente de difusão multiplicado por $\sqrt{\Delta t}$.

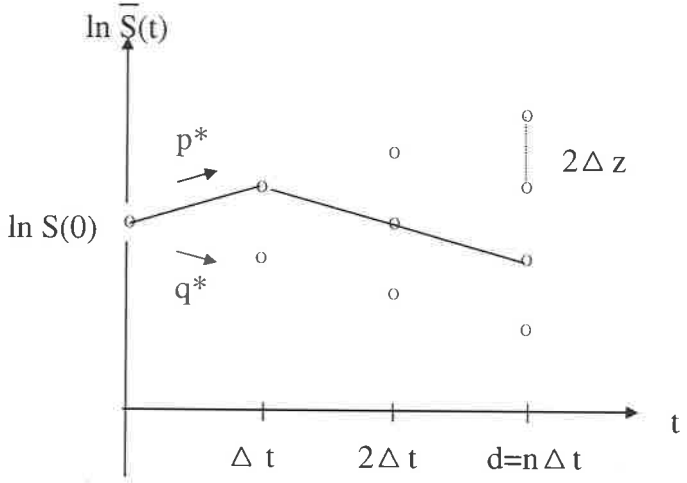


Figura 13.2: Ilustração do modelo binomial. Note-se que $\Delta z = \sigma\sqrt{\Delta t}$ e que o número de intervalos n em que é partido o intervalo $[0, d]$ deve ser relativamente elevado (nesta ilustração, pôs-se $n = 3$). Ilustra-se uma possível trajectória do processo $\ln \bar{S}(t)$, que aproxima $\ln S(t)$. As probabilidades de, em cada passo temporal, se mover para cima ou para baixo são, respectivamente, p^* (dada por (13.40)) e $q^* = 1 - p^*$.

aproximado $p^* = 0,52271$ acima obtido, o valor esperado (que aproxima o valor esperado segundo a medida martingala equivalente) de $\bar{Y}(d)$ é $0,52271 \times 189,14 + 0,47229 \times 0 = 98,87$. Esse deveria ser aproximadamente o valor de $\bar{Y}(0)$, que, como $B(0) = 1$, seria também $Y(0)$. Como vimos na Secção 13.3, o valor exacto de $Y(0)$ é de $82,21$ escudos. A aproximação não é famosa mas, tendo usado um só passo, não estaríamos à espera de milagres.

Como obter aproximadamente a estratégia hedging? Neste caso, como só há um passo temporal que começa em 0, a estratégia consiste na única transacção que se faz no início para constituir a carteira. Se houvesse vários passos temporais, em cada um deles haveria compras e vendas apropriadas de acções e obrigações. Neste exemplo com um só passo temporal, indicar a estratégia é indicar o número de obrigações $b = b(0) = b(d)$ e o número de acções $s = s(0) = s(d)$ a incluir inicialmente na carteira. Se, no único passo temporal deste exemplo,

saltarmos para $\ln \bar{S}(d) = 3691,05$ escudos, vem que o valor da carteira no instante d é $b \times B(d) + s \times 3691,05 = b\exp(0,04 \times 0,2518) + s \times 3691,05 = 1,010123b + 3691,05s$, mas este valor deve igualar o benefício de 191,05 escudos do detentor da opção, pelo que temos a equação $1,010123b + 3691,05s = 191,05$. Caso saltemos para $\bar{S}(d) = 3040,47$ escudos, obtemos semelhantemente a equação $1,010123b + 3040,47s = 0$. Resolvendo o sistema das duas equações, vemos que a estratégia será adquirir $s = 0,29366$ acções e $b = -883,92$ obrigações. Claro que $bB(0) + sS(0) = -883,92 + 0,29366 \times 3350 = 98,84$ dá (à parte um pequeno erro de cálculo) o valor já obtido de $Y(0)$. Claro que esta estratégia em um só acto, em que não se dá oportunidade de ir transaccionando ao longo do tempo, dará uma aproximação muito crua.

Naturalmente, obteríamos aproximações muito melhores de $Y(0)$ (e de $Y(t)$ para valores de t pertencentes à grelha temporal considerada na discretização) e também melhores aproximações da estratégia hedging (agora com vários períodos de transacção em vez de um só) se tivessemos usado um n elevado em vez de $n = 1$.

13.6 Opções europeias de venda

Uma *opção europeia de venda* com *data de expiração* $d > 0$ (instante determinístico) e *valor de exercício* $K > 0$ dá ao seu detentor o direito, mas não a obrigação, de vender uma acção no instante d pelo preço de exercício K. Caso a cotação $S(d)$ no instante d seja inferior a K, o detentor da opção exercê-la-á vendendo uma acção ao preço K e ficando com um benefício $K - S(d) > 0$. Caso $S(d)$ seja superior a K, o detentor da opção não irá obviamente exercê-la e o seu benefício será nulo. Conclui-se que uma opção de venda dá ao seu detentor um benefício

$$(K - S(d))^+ := \max(K - S(d), 0).$$

O tratamento destas opções reduz-se ao tratamento já feito das opções de compra mediante fórmulas simples que permitem passar de um caso para outro, as quais vamos apresentar de seguida.

Seja $Y_P(t)$ (usamos o índice "P" para indicar que é uma opção de venda, que, em inglês, se diz "put option") o valor de uma opção de venda no instante t e seja $(b_P(t), s_P(t))$ $(t \in [0, d])$ uma estratégia hedging, isto é uma estratégia autofinanciada que reproduza o valor $Y_P(d)$ da opção no instante d (e, portanto, pelo princípio da não-arbitragem, também reproduza o valor da opção em todos os instantes $t \in [0, d]$). Facilmente se vê que, sendo $Y(t)$ o valor de uma opção de compra com a mesma

data e valor de exercício, vem

$$Y_P(d) - Y(d) = -S(d) + K = -S(d) + \frac{K}{B(0)\exp(\mu d)} B(d).$$

Logo, um portfolio com 1 opção de venda e -1 opção de compra tem, no instante d, o mesmo valor que um portfolio com -1 acção e $\frac{K}{B(0)\exp(\mu d)}$ obrigações. Pelo princípio da não-arbitragem, os dois portfolios devem ter o mesmo valor em qualquer instante $t \in [0, d]$, pelo que vem

$$\begin{aligned} Y_P(t) - Y(t) = -S(t) + K = & \quad -S(t) + \frac{K}{B(0)\exp(\mu d)} B(t) = \\ & -S(t) + Ke^{-\mu(d-t)}, \end{aligned}$$

donde

$$Y_P(t) = Y(t) - S(t) + \frac{K}{B(0)\exp(\mu d)} B(t) = Y(t) - S(t) + Ke^{-\mu(d-t)}. \tag{13.41}$$

Seja $(b(t), s(t))$ a estratégia hedging da opção de compra. Então temos $b(t)B(t) + s(t)S(t) = Y(t)$ e $b_P(t)B(t) + s_P(t)S(t) = Y_P(t)$, o que, substituído em (13.41), dá

$$\left(b_P(t) - b(t) - \frac{K}{B(0)\exp(\mu d)}\right) B(t) + (s_P(t) - s(t) + 1)\, S(t) = 0.$$

Isso sugere que a estratégia hedging para a opção de venda seja dada por

$$\begin{aligned} b_P(t) &= b(t) + \tfrac{K}{B(0)\exp(\mu d)} \\ s_P(t) &= s(t) - 1. \end{aligned} \tag{13.42}$$

É trivial verificar que a sugestão está correcta, o que fica como exercício para o leitor.

Já as opções americanas de compra e venda são de mais difícil tratamento.

13.7 Outros modelos e opções

O que se fez nas Secções anteriores pode ser consideravelmente generalizado. Para começar, podem existir várias acções ou activos financeiros (até taxas de juro, taxas de câmbio, etc.) com cotações/valores $S_1(t), S_2(t), ..., S_n(t)$. Estes activos podem seguir modelos diferentes (com outros comportamentos da taxa média de rendimento e da volatilidade) do modelo de Black-Scholes, o investimento sem risco pode ser

agora apenas um investimento seguro com uma taxa de rendimento variável e podem considerar-se outros tipos de opções mais complicadas.

Daqui em diante, supomos que t varia no intervalo $[0, d]$, com d real positivo. Consideremos um processo de Wiener a m dimensões $\mathbf{W}(t) = [W_1(t), ..., W_m(t)]^T$ num espaço de probabilidade $(\Omega, \mathcal{F}, P)$ e seja $\mathcal{M}_t$ a sua filtração natural.

Em vez de $B(t)$, vamos designar por $S_0(t)$ o valor do investimento seguro e supor que ele satisfaz

$$dS_0(t) = \mu(t, \omega) S_0(t) dt$$

com condição inicial $S_0(0)$ determinística e com a taxa de rendimento μ função limitada e adaptada à filtração $\mathcal{M}_t$.

Para $i = 1, ..., n$, vamos supor que $S_i(t)$ satisfaz

$$dS_i(t) = F_i(t, \omega) dt + \sum_{j=1}^{m} G_{ij}(t, \omega) dW_j(t)$$

com $S_i(0)$ determinístico. Se designarmos por $\mathbf{F}$ o vector coluna dos F_i, por $\mathbf{G}$ a matriz $n \times m$ dos G_{ij} e por $\mathbf{G}_i$ a sua i-ésima linha, podemos escrever $dS_i = F_i dt + \mathbf{G}_i d\mathbf{W}_i(t)$ $(i = 1, ..., n)$. Supomos que $\mathbf{F}$ e $\mathbf{G}$ são adaptadas à filtração $\mathcal{M}_t$ e satisfazem as condições para que o vector coluna

$$\mathbf{S}(t) = [S_0(t), S_1(t), ..., S_n(t)]^T$$

seja um processo de Itô. Claro que, se, em particular, se tiver $\mu(s, \omega) = \mu(t, S_0(t, \omega), S_1(t, \omega), ..., S_n(t, \omega))$ e semelhantemente para os F_i e os G_{ij}, temos uma EDE multidimensional mas em que $S_0(t) = S_0(0) \exp\left(\int_0^t \mu(u, \omega) du\right)$ não é impulsionado pelo processo de Wiener.

A $\mathbf{S}(t)$ chama-se um *mercado*. Chamamos *mercado normalizado*

$$\tilde{\mathbf{S}}(t) = [\tilde{S}_0(t), \tilde{S}_1(t), ..., \tilde{S}_n(t)]^T = [1, S_1(t)/S_0(t), ..., S_n(t)/S_0(t)]^T$$

ao que resulta de usar como deflactor o processo $S_0(t)$. Vem

$$d\tilde{\mathbf{S}}(t) = \frac{1}{S_0(t)} (d\mathbf{S}(t) - \mu(t)\mathbf{S}(t)dt).$$

Um *portfolio* $\mathbf{s}(t, \omega)$ $(t \in [0, d])$, que abreviamos para $\mathbf{s}(t)$, é um vector linha

$$\mathbf{s}(t) = [s_0(t), s_1(t), ..., s_n(t)]$$

adaptado à filtração $\mathcal{M}_t$, que representa a quantidade de cada um dos $n + 1$ activos financeiros detidos em cada momento por um investidor.

O seu valor é

$$V(t) = V_s(t,\omega) = \mathbf{s}(t)\mathbf{S}(t) = \sum_{i=0}^{n} s_i(t)S_i(t).$$

O portfolio diz-se *autofinanciado* se

$$V(t) = V(0) + \int_0^t \mathbf{s}(u)d\mathbf{S}(u),$$

ou seja $dV(t) = \mathbf{s}(t)d\mathbf{S}(t)$, além de satisfazer uma condição técnica adequada para que a expressão definidora faça sentido (por exemplo, embora não seja precisa condição tão forte, que μ, $\mathbf{F}$, $\mathbf{G}$, $\mathbf{s}$ e $\mathbf{S}$ estejam em $M^2[0,t]$). Um portfolio autofinanciado diz-se um *portfolio admissível* se o seu valor for limitado inferiormente (isso impede endividamentos sem limite, evitando situações semelhantes à da estratégia sempre ganhadora em jogos de azar que consiste em duplicar a aposta até ganhar).

Um portfolio diz-se uma *arbitragem* se for admissível, se o seu valor for nulo no instante 0 e não-negativo no instante d e se houver uma probabilidade positiva de ter valor positivo no instante d. Isto significa que, com um investimento nulo e sem qualquer risco, há uma probabilidade positiva de lucro. A hipótese habitual de trabalho é a de que o mercado não permite a existência de arbitragens.

Uma medida de probabilidade P^* definida em $(\Omega, \mathcal{M}_d)$ equivalente a P diz-se uma *medida martingala equivalente* se o mercado normalizado $\tilde{\mathbf{S}}(t)$ for uma martingala relativamente a P^* (com respeito à filtração $\mathcal{M}_t$).

Um resultado importante, que não demonstramos aqui (pode ver, por exemplo, [51]) é o de que, se existe uma medida martingala equivalente, então não há arbitragens no mercado. O recíproco não é verdadeiro; é precisa uma propriedade um pouco mais forte que a ausência de arbitragem para garantir a existência de medida martingala equivalente. O teorema de Girsanov dá-nos condições suficientes para isso.

Embora ainda se possa generalizar mais, referimos aqui o caso de uma opção europeia mais geral, que pode ser qualquer v.a. C mensurável-$\mathcal{M}_d$ limitada inferiormente, correspondente ao benefício a que o seu detentor tem direito na data de expiração d. Ela diz-se *atingível* no mercado $\mathbf{S}(t)$ se existir pelo menos um *portfolio hedging*, isto é, um portfolio admissível $\mathbf{s}(t)$ tal que $C = z + \int_0^d \mathbf{s}(t)d\mathbf{S}(t)$ para algum número real z. Aqui z representa a riqueza inicial e o integral representa a valorização do portfolio. O mercado diz-se *completo* se todas as opções acima consideradas forem atingíveis. Pode provar-se que, para saber se uma opção é atingível é

indiferente trabalhar com o mercado original ou com o mercado normalizado, pelo que o mercado original é completo sse o mercado normalizado também for.

Em [51] pode ver-se que, se existir um vector coluna m-dimensional $\boldsymbol{\theta}(t,\omega)$ em $H^2[0,d]$ tal que [2]

$$\mathbf{G}(t)\boldsymbol{\theta}(t) = \mathbf{F}(t) - \mu(t)[S_1(t),...,S_n(t)]^T$$

e se verifique a *condição de Novikov* $\mathbb{E}^P\left[\exp\left(\frac{1}{2}\int_0^d \boldsymbol{\theta}^T(u)\boldsymbol{\theta}(u)du\right)\right] < \infty$, então o mercado é completo sse a matriz $\mathbf{G}(t)$ tiver característica m para quase todo o par de valores (t,ω). Claro que só podemos ter um mercado completo se $n \geq m$. O número de fontes de ruído independentes é a dimensão m do processo de Wiener. O facto de a característica de $\mathbf{G}(t)$ ser m significa que, entre os n activos não-seguros do mercado que podem ser utilizados para fazer hedging, há m activos financeiros sem redundâncias. Consideramos que haverá redundância entre vários activos não-seguros se o valor de um deles depender apenas do valor dos restantes e do do activo seguro. Por exemplo, uma acção e uma opção europeia de compra sobre ela formam um conjunto redundante de activos, porque o valor da opção é função da cotação da acção e do valor da obrigação sem risco. Também, um conjunto fixo de 20 acções e um índice bolsista baseado na média das suas cotações é necessariamente redundante, pois a informação contida no índice está já contida nas acções componentes.

Ao leitor interessado nestas e noutras questões relacionadas, recomendamos a leitura de [38].

[2] Corresponde a (12.4) com f^* escolhido de modo a que o processo normalizado tenha tendência nula.

Capítulo 14

Síntese

Façamos uma breve síntese dos principais aspectos tratados neste texto, com uma certa informalidade que leva a não mencionar as questões mais técnicas.

As equações diferenciais estocásticas (EDE) são úteis nos mais diversos ramos da Ciência e da Tecnologia para modelar fenómenos dinâmicos que seriam descritos por equações diferenciais ordinárias não fora o facto de a sua dinâmica ser perturbada por um ruído. O ruído, processo estocástico em tempo contínuo, será, por razões de comodidade matemática, aproximado por um ruído branco $\varepsilon(t) = \varepsilon(t, \omega)$, que é um processo estocástico generalizado gaussiano estacionário em que os valores do processo em instantes distintos são independentes. Ele será a derivada (no sentido das funções generalizadas, já que a derivada ordinária não existe) do processo de Wiener (também conhecido por movimento browniano) $W(t, \omega)$ $(t \geq 0)$. É costume abreviar para $W(t)$, isto é, não referir explicitamente a dependência do acaso ω.

O processo de Wiener, que supomos definido num espaço de probabilidade $(\Omega, \mathcal{F}, P)$, é um processo estocástico (propriamente dito) em tempo contínuo com valores em $\mathbb{R}$, com trajectórias quase certamente (q.c.) contínuas, com incrementos independentes, com $X(0) = 0$ e tal que o incremento $W(t) - W(s)$ $(s < t)$ tem distribuição $\mathcal{N}(0, t - s)$ (distribuição normal com média 0 e variância $t - s$). Ele foi usado inicialmente por Bachelier em 1900 para decrever a cotação de uma acção na bolsa e (na versão multidimensional) por Einstein em 1905 para descrever o movimento browniano de uma partícula suspensa num fluído (a primeira EDE, cuja solução é o processo de Ornstein-Uhlenbeck, surge precisamente para melhorar este modelo do movimento browniano). Wiener e Lévy são responsáveis pela teoria matemática. O pro-

cesso de Wiener é um processo de Markov, isto é, conhecido o valor presente, o futuro é independente do passado. É até um processo de Markov forte, isto é, a propriedade de Markov funciona mesmo quando se toma para instante presente um instante aleatório, desde que seja um tempo de Markov τ, isto é, desde que seja possível determinar se $\tau \leq t$, com t fixo, apenas conhecendo os valores do processo até ao instante t, inclusivé. A f.d.p. de transição do processo de Wiener, isto é, para $s < t$, a f.d.p. de $W(t)$ dado $W(s) = x$ é $\mathcal{N}(x, t-s)$. O processo de Wiener é também uma martingala (com repeito à sua filtração natural), isto é, para $s < t$, $\mathbb{E}[W(t)|W(u) : 0 \leq u \leq s] = W(s)$. As suas trajectórias são contínuas mas muito irregulares, de variação ilimitada q.c. e, portanto, q.c. sem derivada.

Estudaram-se EDE da forma

$$\frac{dX(t)}{dt} = f(t, X(t)) + g(t, X(t))\varepsilon(t), \quad X(0) = X_0, \quad t \in [0, d],$$

onde a condição inicial X_0 é uma v.a. independente do processo de Wiener (que pode ser degenerada numa constante determinística). É habitual escrevê-las na forma

$$dX(t) = f(t, X(t))dt + g(t, X(t))dW(t), \quad X(0) = X_0, \quad t \in [0, d],$$

ou, para lhe dar verdadeiramente significado, na forma integral (abreviamos, como usualmente, $X(t, \omega)$ para $X(t)$)

$$X(t) = X_0 + \int_0^t f(s, X(s))ds + \int_0^t g(s, X(s))dW(s), \quad t \in [0, d].$$

O primeiro integral pode ser definido para cada ω (isto é para cada trajectória) como integral de Riemann ou de Lebesgue, mas o segundo integral não pode ser definido como integral de Riemann-Stieltjes porque a variação ilimitada da função integradora $W(t)$ provoca (salvo nos casos mais simples) que os limites (por exemplo, em média quadrática) das somas de Riemann-Stieltjes dependam da escolha dos pontos intermédios da função integranda. É, portanto, necessário, definir os integrais estocásticos. Só vamos considerar funções integrandas $G(s, \omega)$ não-antecipativas, que são funções que, além de conjuntamente mensuráveis nas suas duas variáveis, não "adivinham" o futuro, isto é, são independentes dos futuros valores do ruído perturbador, o mesmo é dizer independentes dos incrementos futuros do processo de Wiener. Embora a definição seja um pouco mais complexa para funções integrandas $G(s, \omega)$ mais gerais, no caso de elas serem não-antecipativas e contínuas

em média quadrática, o integral estocástico pode definir-se por

$$\int_0^t G(s)dW(s) = \int_0^t G(s,\omega)dW(t,\omega) = \text{l.i.m.}_{n\to+\infty} \sum_{k=1}^n G(t_{n,k-1})(W(t_{n,k}) - W(t_{n,k-1})).$$

Aqui $0 = t_{n,0} < t_{n,1} < ... < t_{n,n} = t$ $(n = 1, 2, ...)$ são decomposições do intervalo de integração com diâmetro a tender para zero quando $n \to +\infty$ e l.i.m. representa o limite em média quadrática. Este integral, em que se faz a escolha não-antecipativa para os pontos intermédios da função integranda, é o integral de Itô. A escolha não-antecipativa traduz a ideia de "não adivinhação" do futuro e dá ao integral boas propriedades probabilísticas. Porém, ele não segue as regras usuais de cálculo. Outras escolhas ou combinações de escolhas dos pontos intermédios dão outros integrais estocásticos, entre os quais se destaca o de Stratonovich, que se representa por $(S)\int_0^t G(s)dW(s)$ ou $\int_0^t G(s) \circ dW(s)$ e que segue as regras usuais de cálculo, embora não tenha tão boas propriedades probabilísticas.

Trabalhámos com o integral de Itô e estendemos a definição para funções integrandas pertencentes ao espaço de Hilbert $H^2[0,t]$ de funções não-antecipativas tais que $(\|G\|_{2*})^2 = \int_0^t \mathbb{E}[G^2]ds < +\infty$ ($\|\cdot\|_{2*}$ é a norma). O integral é uma v.a. e está no espaço de Hilbert L^2 das v.a. com norma finita (em L^2 a norma de uma v.a. X é $\mathbb{E}[X^2])^{1/2}$). Verificámos que o integral tem esperança matemática nula e preserva a norma, isto é $\mathbb{E}\left[\left(\int_0^t G(s)dW(s)\right)^2\right] = \int_0^t \mathbb{E}[G^2(s)]ds$. Também vericámos, que como função do limite superior de integração, o integral de Itô é uma função contínua e uma martingala. No caso de G ser determinística, o integral $\int_0^t G(s)dW(s)$ tem distribuição $\mathcal{N}(0, \int_0^t G^2(s)ds)$.

O integral de Itô foi depois estendido a funções da classe $M^2[0,t]$, isto é, funções não-antecipativas tais que $\int_0^t G(s)ds < +\infty$ q.c., mas para estas funções já não há garantia de ele ter momentos nem de ser martingala. Definiu-se um processo de Itô como um processo da forma $X(t) = X_0 + \int_0^t F(s,\omega)ds + \int_0^t G(s,\omega)dW(s,\omega)$ (ou seja, $dX(t) = F(t)dt + G(t)dW(t)$ em notação diferencial) com F não-antecipativa tal que $\int_0^d |F(s)|ds < +\infty$ q.c. e $G \in M^2[0,d]$. O integral de Itô segue regras especiais de cálculo, o cálculo de Itô, que são consubstanciadas no teorema/fórmula de Itô. Esta diz que, se $h(t,x)$ for uma função real de classe $C^{1,2}$, e se $X(t)$ for o processo de Itô acima, então $Y(t) = h(t, X(t))$ é um processo de Itô e vem

$$dY(t) = \frac{\partial h(t,X(t))}{\partial t}dt + \frac{\partial h(t,X(t))}{\partial x}dX(t) + \frac{1}{2}\frac{\partial^2 h(t,X(t))}{\partial x^2}(dX(t))^2,$$

onde $(dX(t))^2$ deve ser calculado usando as regras $(dt)^2 = dtdW(t) = 0$ e $(dW(t)) = dt$ (esta última é que fura as regras usuais).

Definido o integral estocástico, a forma integral da EDE tem sentido e podemos procurar condições para a existência e unicidade de solução. Apresentámos um teorema de existência e unicidade de solução para a EDE $dX(t) = f(t, X(t))dt + g(t, X(t))dW(t)$ $(t \in [0, d])$ com $X(0) = X_0$ independente do processo de Wiener tal que $\mathbb{E}[X_0^2] < +\infty$, que exige que $f(t, x)$ e $g(t, x)$ satisfaçam, relativamente a x, uma condição de Lipschitz e uma restrição ao crescimento. Nestas condições verificámos que a solução está em $H^2[0, d]$, é contínua e contínua em média quadrática, tem média e variância finitas e é um processo de Markov. Verificámos ainda que, se f e g forem contínuas em t, a solução é também um processo de difusão com coeficiente de tendência $a(s, x) = f(s, x)$ e coeficiente de difusão $b(t, x) = g^2(t, x)$, isto é, um processo de Markov de trajectórias q.c contínuas que verifica as propriedades (5.1), (5.2) e (5.3). Os coeficientes de tendência e difusão caracterizam o processo de difusão e dão-nos as velocidades com que variam a média e a variância do processo, respectivamente. A vantagem é que, para estes processos são válidas as equações de Kolmogorov (ver Secção 5.2).

No caso particular das EDE autónomas, aquelas em que $f(t, x) \equiv f(x)$ e $g(t, x) \equiv g(x)$ não dependem do tempo t, a condição de Lipschitz é suficiente pois implica a restrição ao crescimento e a solução é um processo de difusão homogéneo (os coeficientes de tendência e difusão também não dependem do tempo) com coeficiente de tendência $a(x) = f(x)$ e coeficiente de difusão $b(x) = g^2(x)$. Neste caso, quando a condição inicial $X_0 \equiv x$ é determinística, chamamos à sua solução difusão de Itô. Neste caso, a equação de Kolmogorov regressiva para funcionais $u(t, x) = \mathbb{E}_x[h(X(t))] := \mathbb{E}[h(X(t))|X(0) = x]$, com $h(x)$ função mensurável-Borel contínua limitada, toma a forma $\frac{\partial u}{\partial t} = \mathcal{D}u$, com $\lim_{u \downarrow 0} u(t, x) = h(x)$. Aqui $\mathcal{D} = a(x)\frac{\partial}{\partial x} + \frac{1}{2}b(x)\frac{\partial^2}{\partial x^2}$ é o operador de difusão.

Para as EDE autónomas, no caso unidimensional (isto é, não podemos extrapolar quando generalizarmos a sistemas de EDE), basta até que f e g sejam de classe C^1 para existir solução única até um instante de explosão e, sempre que seja possível mostrar (por outras técnicas) que o instante de explosão é q.c. infinito, pode concluir-se que existe solução única para todo o $t \geq 0$.

Se usarmos integrais de Stratonovich em vez de integrais de Itô na forma integral da EDE, teremos EDE de Stratonovich, que representamos por

$$(S) \quad dX(t) = f(t, X(t))dt + g(t, X(t))dW(t), \quad X(0) = X_0.$$

Na forma integral é

$$X(t) = X_0 + \int_0^t f(s, X(s))ds + \ (S) \ \int_0^t g(s, X(s))dW(s),$$

onde podemos, sob condições de regularidade adequadas, definir o integral de Stratonovich por

$$(S)\int_0^t g(s, X(s))dW(s) = \int_0^t g(s, X(s)) \circ dW(s) =$$
$$\text{l.i.m.} \ \sum_{k=1}^n g\left(t_{n,k-1}, \frac{X(t_{n,k-1})+X(t_{n,k})}{2}\right)(W(t_{n,k}) - W(t_{n,k-1})).$$

Então conclui-se que a EDE de Stratonovich referida é equivalente (tem a mesma solução) que a EDE de Itô

$$dX(t) = \left(f(t, X(t)) + \frac{1}{2}\frac{\partial g(t, X(t))}{\partial x}g(t, Xt))\right)dt + g(t, X(t))dW(t).$$

A fórmula, usada em sentido inverso, permite reduzir uma EDE de Itô a uma de Stratonovich com a mesma solução, o que é muito útil pois, para resolver esta, podemos usar as regras de cálculo a que estamos habituados em vez do cálculo de Itô que nos é estranho.

Tudo isto pode ser generalizado ao caso multidimensional, o que permite tratar sistemas de EDE.

No Capítulo 8, tratámos com algum detalhe o modelo de Black-Scholes $dX(t) = rX(t)dt + \sigma X(t)dW(t)$, com $X(0) = x_0$ determinístico. Ele é usado para modelar a dinâmica da cotação de uma acção na bolsa e a dinâmica de crescimento de uma população sem restrição de recursos em ambiente aleatório. Aqui o acaso ω representa um cenário do mercado ou um estado da natureza escolhido ao acaso no conjunto Ω de todos os possíveis cenários do mercado ou estados da natureza. O modelo pode escrever-se na forma $dX(t)/dt = (r + \sigma\varepsilon(t))X(t)$ que traduz a ideia de que a taxa de rendimento r, em vez de ser constante, é perturbada por um ruído branco. O parâmetro r pode interpretar-se agora como uma taxa média de rendimento. O parâmetro σ, que mede a intensidade das flutuações aleatórias, é conhecido por volatilidade na literatura financeira. A solução $X(t) = x_0 \exp\left((r - \sigma^2/2)t + \sigma W(t)\right)$ foi facilmente obtida, quer recorrendo à fórmula de Itô aplicada a $\ln X(t)$, quer convertendo-a numa EDE de Stratonovich equivalente e resolvendo-a pelas regras usuais de cálculo.

EDE idênticas dão soluções diferentes conforme se adopte o cálculo de Itô ou o de Stratonovich e isso tem dado controvérsias na literatura. As diferenças podem mesmo ser de carácter qualitativo. Por exemplo, no modelo de Black-Scholes aplicado ao crescimento de uma população,

o cálculo de Stratonovich garante que a população não se extingue se a taxa média de crescimento *per capita* r for positiva, mas no cálculo de Itô ela extingue-se para $r < \sigma^2/2$, o que inclui valores positivos de r. É habitual dar-se a receita, baseada em certos teoremas limite, de que, se o processo decorre intinsecamente em tempo contínuo na natureza e o ruído perturbador é colorido, quando nós o aproximamos pela ficção matemática de um ruído branco, o cálculo que dá melhor aproximação é o de Stratonovich. Em contrapartida, se o processo decorre intrinsecamente em tempo discreto (com um ruído branco em tempo discreto, que é um processo estocástico propriamente dito) e nós aproximamos a equação às diferenças estocástica por uma EDE (agora com a ficção matemática de ruído branco em tempo contínuo), a melhor aproximação é o cálculo de Itô. Porém, em muitas aplicações, é duvidoso qual das situações se aplica.

A controvérsia foi por nós resolvida (ver Capítulo 9) notando que o parâmetro r, ao contrário do que implicitamente se supõe na literatura, tem significados diferentes em termos da dinâmica populacional, significando diferentes tipos de média. De facto, se usarmos o cálculo de Itô, r significa a taxa média aritmética e, se usarmos o cálculo de Stratonovich, significa a taxa média geométrica (rigorosamente, não deveríamos usar a mesma letra para quantidades com significados diferentes) e, se atendermos à diferença $\sigma^2/2$ entre as duas médias, os dois cálculos coincidem completamente nos seus resultados. E isso é uma situação mais geral que se aplica a uma ampla classe de modelos.

Estudámos (Secção 11.1) o modelo de Ornstein-Uhlenbeck e uma sua variante, o modelo de Vasicek, verificando nestes dois casos a existência de densidade estacionária. Isto é, nestes casos a solução $X(t)$ da EDE tem distribuição de probabilidade que converge quando $t \to +\infty$ para uma distribuição limite cuja f.d.p. é a densidade estacionária. Esta verificação foi possível porque conseguimos resolver explicitamente a EDE, o que nem sempre sucede. Para processos de difusão unidimensionais regulares (todos os estados do interior do espaço de estados comunicam entre si) com coeficientes de tendência $a(x)$ e de difusão $b(x)$ contínuos em que o coeficiente de difusão é positivo no interior do espaço de estados, existem métodos de verificação de existência e determinação da densidade estacionária que não exigem a resolução explícita da EDE correspondente. Esses métodos baseiam-se na classificação das duas fronteiras $r_1 \geq -\infty$ e $r_2 \leq +\infty$ (com $r_1 < r_2$) do espaço de estados.

Para tal usam-se as densidades de escala $s(\xi) = \exp\left(-\int_{\xi_0}^{\xi} \frac{2a(\eta)}{b(\eta)} d\eta\right)$ e de velocidade $m(\xi) = \frac{1}{s(\xi)b(\xi)}$, onde ξ_0 é um ponto arbitrário do interior do espaço de estados, pelo que estas funções estão definidas a menos de uma constante multiplicativa. As funçoes de escala e de velocidade,

definidas a menos de constantes aditiva e multiplicativa, são $S(x) = \int_{x_0}^{x} s(\xi)d\xi$ e $M(x) = \int_{x_0}^{x} m(\xi)d\xi$, com x_0 escolhido arbitrariamente no interior do espaço de estados. Elas definem as medidas de escala e de velocidade $S(c,d] = S(d) - S(c)$ e $M(c,d] = M(d) - M(c)$. Verifica-se que, se $h(x)$ é de classe C^2, vem $\mathcal{D}h(x) = \frac{1}{2}\frac{d}{dM(x)}\left(\frac{dh(x)}{dS(x)}\right)$.

Se $r_1 < a < x < b < r_2$ e partirmos de $X(0) = x$, obtemos (ver Secção 11.2) equações diferenciais ordinárias para certos funcionais do tempo de Markov $T_{a,b} = \min\{T_a, T_b\}$ (tempo de primeira saída do intervalo (a,b), onde $T_y = \inf\{t \geq 0 : X(t) = y\}$ é o tempo de primeira passagem por y). Em particular, resolvendo a equação apropriada, mostrámos que $u(x) = P_x[T_b < T_a] := P[T_b < T_a | X(0) = x] = \frac{S(x)-S(a)}{S(b)-S(a)}$ (probabilidade de a primeira saída de (a,b) ser no ponto b). Daqui resulta que, para $r_1 < x < b < r_2$, vem $P_x[T_{r_1} < T_b] = 0$, isto é, a fronteira r_1 é não-atractiva, sse $S(r_1, b] = +\infty$ (ou seja, sse $s(x)$ é não integrável numa vizinhaça de r_1). Conclusão semelhante pode tirar-se para a fronteira r_2. Pode provar-se que as fronteiras não-atractivas não são atingíveis nem em tempo finito nem no limite quando $t \to +\infty$.

Caso ambas as fronteiras sejam não-atractivas, a probabilidade é "empurrada" para o interior do espaço de estados e há esperança de que exista densidade estacionária. De facto assim sucede desde que também $m(x)$ seja integrável, isto é $M = \int_{r_1}^{r_2} m(\xi)d\xi < +\infty$, vindo para densidade estacionária $p(y) = m(y)/M$ $(r_1 < y < r_2)$, expressão que se obtém usando a equação de Kolmogorov progressiva. Neste caso, o processo é também ergódico, isto é certos momentos da solução podem ser estimados usando os correspondentes momentos amostrais ao longo de uma única trajectória do processo, o que é extremamente útil pois, em muitas aplicações, como as financeiras, só temos uma trajectória observada e não podemos fazer o tempo andar para trás e repetir a "experiência" com outros cenários de mercado.

O estudo que fizémos na Secção 11.2 sobre tempos de primeira passagem foi utilizado num exemplo para estudar a extinção de populações (ver Secção 11.4).

O teorema de Girsanov foi tratado no Capítulo 12 e permite mudar o coeficiente de tendência numa EDE $dX(t) = f(X(t))dt + g(X(t))dW(t)$ $(t \in [0,d])$ para outro coeficiente f^*. Isso tem particular vantagem quando se muda para um coeficiente nulo, pois uma EDE da forma $dY(t) = 0dt + g(Y(t))dW(t)$ $(t \in [0,d])$ com $Y(0) = x$ tem como solução $Y(t) = x + \int_0^t g(Y(t))dW(t)$, que, se a função integranda estiver em $H^2[0,d]$, é, devido às propriedades do integral estocástico, uma martingala. A vantagem de uma martingala é que o seu valor em qualquer instante pode ser obtido como uma esperança matemática condicional

do seu valor no instante terminal d e há aplicações financeiras em que isso é muito útil. Para mudar o coeficiente de tendência de f para f^* sem alterar a solução precisamos de alterar o processo de Wiener. Porém, ao fazê-lo, este deixa de ser processo de Wiener relativamente à probabilidade inicial P, mas existe uma outra probabilidade P^* equivalente a P relativamente à qual o processo alterado passa a ser um processo de Wiener. Temos uma nova EDE com respeito a um novo processo de Wiener num novo espaço de probabilidade. O leitor deverá consultar o enunciado do Teorema da Secção 12.2 para ver os pormenores deste procedimento.

No Capítulo 13 estuda-se a célebre fórmula de Black-Scholes para uma opção europeia de compra, mas julgamos preferível não apresentar aqui um resumo e remeter o leitor para a leitura do próprio Capítulo. Outras aplicações financeiras são também mencionadas, embora muito sucintamente.

Ao longo deste texto fomos dando indicações para o leitor interessado noutros tópicos importantes que a natureza introdutória deste trabalho não permite tratar. Mas outras áreas como o cálculo de Malliavin ou as equações às derivadas parciais estocásticas são desenvolvimentos relativamente recentes extremamente promissores. As EDE têm também grande importância no estudo teórico de certas equações diferenciais parciais.

Em termos de aplicações, demos apenas alguns exemplos em áreas científicas nuito delimitadas, principalmente nas áreas financeiras e de dinâmica de populações, mas é rara a área científica e tecnológica onde modelos baseados em EDE não tenham sido aplicados. Mesmo nas áreas focadas, apenas abordámos algumas das muitas questões que têm sido tratadas. Por exemplo, a nível financeiro, não tratámos as opções americanas, não considerámos a influência de dividendos e custos de transacção, nem o caso em que a volatilidade (e, portanto, o coeficiente de difusão) pode ser ela própria perturbada por um ruído. O caso de processos não contínuos, em que se admitem também saltos poissonianos para modelar ocasionais "catástrofes", não foi aqui estudado, embora tenha bastante interesse em diversas aplicações.

O estudo que fizémos das EDE é de carácter introdutório, mas procura proporcionar ao leitor uma entrada neste mundo, fornecendo-lhe alguns dos instrumentos mais relevantes para a utilização das EDE em modelação.

Referências

[1] Alvarez, L. H. R. (2000). On the option interpretation of rational harvesting planning. *J. Mathematical Biology* **40**: 383-405.

[2] Alvarez, L. H. R. e Shepp, L. A. (1997). Optimal harvesting in stochastically fluctuating populations. *J. Mathematical Biology* **37**: 155-177.

[3] Arnold, L. (1974). *Stochastic Differential Equations. Theory and Applications.* Wiley, New York.

[4] Bachelier, L. (1900). Théorie de la speculation. *Annales Scientifiques de l'École Normal Supérieure 3éme Série* **17**: 21:88.

[5] Basawa, I. V. e Prakasa Rao, B. L. S. (1980) *Statistical Inference for Stochastic Processes.* Academic Press, Londom.

[6] Black, F. e Scholes, M. (1973). The pricing of options and corporate liabilities. *J. Political Economy* **81**: 637-654.

[7] Bouleau, N. e Lépingle, D. (1994) *Numerical Methods for Stochastic Processes.* Wiley, New York.

[8] Braumann, C. A. (1983). Population growth in random environments. *Bull. Mathematical Biology* **45**: 635-641.

[9] Braumann, C. A. (1983). Population extinction probabilities and methods of estimation for population stochastic differential equation models. Em *Nonlinear Stochastic Problems* (Bucy, R. S. e Moura, J. M. F., eds.), p. 553-559, D. Reidel, Dordrecht.

[10] Braumann, C. A. (1985). Stochastic differential equation models of fisheries in an uncertain world: extinction probabilities, optimal fishing effort, and parameter estimation. Em *Mathematics in Biology and Medicine* (Capasso, V., Grosso, E. e Paveri-Fontana, L. S., eds.), p. 201-206, Springer, Berlin.

[11] Braumann, C. A. (1993). Model fitting and prediction in stochastic population growth models in random environments. *Bull. Intl. Statistical Institute* **LV-CP1**: 163-164.

[12] Braumann, C. A. (1993). General models of fishing with random growth parameters. Em *Mathematics Applied to Biology and Medicine* (Demongeot, J. e Capasso, V., eds.), p. 155-161, Wuerz, Winnipeg.

[13] Braumann, C. A. (1996). Estimação de parâmetros em modelos de crescimento e pesca em ambientes aleatórios. Em *Bom Senso e Sensibilidade, Traves Mestras da Estatística* (Branco, J., Gomes, P. e Prata, J., eds.), p. 103-107, SPE e Edições Salamandra, Lisboa.

[14] Braumann, C. A. (1997). Parameter estimation in population growth and fishing models in random environments. *Bull. Intl. Statistical Institute* **LVII-CP1**: 21-22.

[15] Braumann, C. A. (1999). Comparison of geometric Brownian motions and applications to population growth and finance. *Bull. Intl. Statistical Institute* **LVIII-CP1**: 125-126.

[16] Braumann, C. A. (1999). Applications of stochastic differential equations to population growth. Em *Proceedings of the Ninth International Colloquium on Differential Equations* (Bainov, D., ed.), p. 47-52, VSP, Utrecht.

[17] Braumann, C. A. (1999). Variable effort fishing models in random environments. *Mathematical Biosciences* **156**: 1-19.

[18] Braumann, C. A. (2001). Crescimento de populações em ambiente aleatório: generalização a densidades de ruído dependentes da densidade da população. Em *A Estatística em Movimento* (Neves, M. M., Cadima, J., Martins, M. J. e Rosado, F., eds.), p. 119-128, Edições SPE.

[19] Braumann, C. A. (2002). Variable effort harvesting models in random environments: generalization to density-dependent noise intensities. *Mathematical Biosciences* **177&178**: 229-245.

[20] Braumann, C. A. (2003). Modeling population growth in random environments: Ito or Stratonovich calculus? *Bull. Intl. Statistical Institute* **LX-CP1**: 119-120.

[21] Braumann, C. A. (2005). Ito or Stratonovich calculus in random population growth. *Mathematical Biosciences* (no prelo).

[22] Capocelli, R. M. e Ricciardi, L. M. (1974). A diffusion model for population growth in a random environment. *Theoretical Population Biol.* **5**: 28-41.

[23] Carlos, C. e Braumann, C. A. (2005). Tempos de extinção para populações em ambiente aleatório. Em *Estatística Jubilar* (Braumann, C. A., Infante, P., Oliveira, M. M., Alpízar-Jara, R. e Rosado, F., eds.), p. 133-142, Edições SPE.

[24] Clark, J. M. C. (1966). *The Representation of Nonlinear Stochastic Systems with Applications to Filtering*, Ph. D. Thesis. Imperial College, London.

[25] Cox, D. R. e Miller, H. D. (1965). *The Theory of Stochastic Processes.* Chapman and Hall, London.

[26] Dacunha -Castelle, D. e Florens-Zmirou, D. (1986). Estimators of coefficients of a diffusion from discrete observations. *Stochastics* **19**: 263-284.

[27] Dennis, B. e Patil, G. P. (1984). The gamma distribution and weighted multimodal gamma distributions as models of population abundance. *Mathematical Biosciences* **68**: 187-212.

[28] Einstein, A. (1956). *Investigations on the Theory of Brownian Movement.* Dover, New York.

[29] Feldman, M. W. e Roughgarden, J. (1975). A population's stationary distribution and chance of extinction in a stochastic environment with remarks on the theory of species packing. *Theoretical Population Biol.* **7**: 197-207.

[30] Fleming, W. H. e Rishel, R.W. (1975). *Deterministic and Stochastic Optimal Control.* Springer, New York.

[31] Freedman, D. (1983). *Brownian Motion and Diffusion.* Springer, New York.

[32] Gard, T. C. (1988). *Introduction to Stochastic Differential Equations.* Marcel Dekker, New York.

[33] Gikhman, I. I. e Skorohod, A. V. (1969). *Introduction to the Theory of Random Processes.* W. B. Saunders, Philadelphia.

[34] Gikhman, I. I. e Skorohod, A. V. (1972). *Stochastic Differential Equations.* Springer, Berlin.

[35] Gray, A. H. (1964). *Stability and Related Problems in Randomly Excited Systems*, Ph. D. Thesis. California Institute of Technology Press.

[36] Itô, K. (1951). *On Stochastic Differential Equations.* American Mathematical Society Memoirs No. 4, New York.

[37] Karatzas, I. e Shreve, S. E. (1991). *Brownian Motion and Stochastic Calculus*, 2nd. edition. Springer, New York.

[38] Karatzas, I. e Shreve, S. E. (1998). *Methods of Mathematical Finance.* Springer, New York.

[39] Karlin, S. e Taylor, H. M. (1981). *A Second Course in Stochastic Processes.* Academic Press, New York.

[40] Khashminsky, R. Z. (1969). *Ustoychivost' Sistem Diffentsual'nykh Uraveniy pri Sluchaynykh Vozmushcheniyskh* (em russo), Estabilidade de Sistemas de Equações Diferenciais na Presença de Perturbações Aleatórias. Nauka Press, Moscovo.

[41] Kloeden, P. E. e Platen, E. (1992). *Numerical Solution of Stochastic Differential Equations.* Springer, Berlin.

[42] Kloeden, P. E., Platen, E. e Schurz, H. (1994). *Numerical Solution of SDE Through Computer Experiments.* Springer, Berlin.

[43] Küchler, U. e Sørensen, M. (1997). *Exponential Families of Stochastic Processes.* Springer, New York.

[44] Lungu, E. e Øksendal, B. (1997). Optimal harvesting from a population in a stochastic crowded environment. *Mathematical Biosciences* **145**: 47-75.

[45] May, R. M. (1973). Stability in randomly fluctuating versus deterministic environments. *American Naturalist* **107**: 621-650.

[46] May, R. M. (1974). *Stability and Complexity in Model Ecosystems* (2nd edition). Princeton University Press, New Jersey.

[47] May, R. M. e MacArthur, R. H. (1972). Niche overlap as a function of environmental variability. *Proc. Natl. Acad. Sci. USA* **69**: 1109-1113.

[48] McKean, H. P. (1969). *Stochastic Integrals.* Academic Press, New York.

[49] Merton, R. (1971). Optimum consumption and portfolio rules in a continuous time model. *J. Economic Theory* **3**: 373-413.

[50] Merton, R. (1973). Theory of rational option pricing. *Bell J. Economics and Management Science* **4**: 141-183.

[51] Øksendal, B. (2003). *Stochastic Differential Equations. An Introduction with Applications* (sixth edition). Springer, Berlin.

[52] Rudin, W. (1974). *Real and Complex Analysis* (second edition). McGraw-Hill, New York.

[53] Schuss, Z. (1980). *Theory and Applications of Stochastic Differential Equations.* John Wiley & Sons, New York.

[54] Stratonovich, R. L. (1966). A new representation of stochastic integrals and equations. *SIAM J. Control* **4**: 362-371.

[55] Stroock, D. W. e Varadhan, S. R. S. (1979). *Multidimensional Diffusion Processes.* Springer, Berlin.

[56] Tuckwell, H. C. (1974). A study of some diffusion models of population growth. *Theoretical Population Biol.* **3**: 345-357.

[57] Turelli, M. (1978). A reexamination of stability in randomly varying environments with comments on the stochastic theory of limiting similarity. *Theoretical Population Biol.* **13**: 244-267.

[58] Uhlenbeck, G. E. e Ornstein, L. S. (1930). On the theory of Brownian motion. *Physical Review* **36**: 823-841.

[59] Wong, E. e Hajek, B. (1985). *Stochastic Processes in Engineering Systems.* Springer, New York.

[60] Wong, E. e Zakai, M. (1965). The oscillationms of stochastic integrals. *Z. Wahrscheinlichkeitstheorie verw. Geb.*, **4**: 103-112.

[61] Wong, E. e Zakai, M. (1969). Riemann-Stieltjes approximation of stochastic integrals. *Z. Wahrscheinlichkeitstheorie verw. Geb.*, **12**: 87-97.

Índice Remissivo

acaso, 8, 9, 13
acontecimentos, 7, 8
acontecimentos independentes, 32
aditividade-σ, 7, 56
aleatoriedade ambiental, 99
aleatoriedade demográfica, 99
álgebra-σ, 7
álgebra-σ de Borel, 9
álgebra-σ gerada, 9
álgebras-σ independentes, 32
arbitragem, 152, 170
átomos, 10

Bachelier, 2
Black, 3, 26, 150
Brown, 1

cálculo de Itô, 72
cálculo de Stratonovich, 77, 78
call option, 150
capacidade de sustento do meio, 130
caracterização de Lévy, 35, 148
cenário do mercado, 8, 13, 26
classes de equivalência, 10
coeficiente de difusão, 42
coeficiente de tendência, 42
com probabilidade um, 10
condição de Lipschitz, 84
condição de Lipschitz local, 94
condição de Novikov, 144, 171
conjunto de índices, 13
conjunto mensurável, 7
conjunto negligenciável, 9
conjunto universal, 7
continuidade em média quadrática, 55
convergência com probabilidade um, 69
convergência em distribuição, 46
convergência em média quadrática, 35, 69
convergência em probabilidade, 69
convergência quase certa, 35, 69
cotação de acção, 26, 99
covariância, 11
Cox, 164
crescimento populacional, 25, 99, 130
critério de Kolmogorov, 17

data de expiração, 151, 167
densidade com respeito a uma medida, 138
densidade de equilíbrio, 46
densidade de escala, 122
densidade de transição, 23
densidade de velocidade, 122
densidade espectral, 18
densidade estacionária, 46, 119, 127, 129
densidade invariante, 46
derivada de Radon-Nikodym, 138, 140
desigualdade de Gronwall, 87

desigualdade de Schwarz, 66
desigualdade de Tchebyshev, 70
desigualdades maximais das martingalas, 19
difusão de Itô, 96, 111, 118
distribuição estacionária, 119
distribuição gama, 132
distribuição normal, 10, 27
distribuição normal multivariada, 11
distribuição transiente, 119
distribuições de dimensão finita, 15
domínio do operador infinitésimal, 112
Dynkin, 4

Einstein, 2, 27, 117
equação às diferenças estocástica, 26
equação de difusão, 45
equação de Fokker-Planck, 45
equação de Itô, 102
equação de Kolmogorov progressiva, 45, 102
equação de Kolmogorov regressiva, 43, 112, 113
equação de ordem superior, 97
equação de Stratonovich, 78, 102, 105
equação diferencial estocástica, 1, 26, 51, 83, 84
equação diferencial estocástica autónoma, 96, 100, 111
equação diferencial estocástica multidimensional, 80, 97
equação do calor, 47
equação integral estocástica, 52, 84
equações de Chapman-Kolmogorov, 23
escala natural, 123
espaço amostral, 7
espaço de Banach, 11
espaço de estados, 15
espaço de Hilbert, 11, 56
espaço de medida, 56
espaço de probabilidade, 7
espaço de probabilidade completo, 8
espaço mensurável, 7
espaço mensurável produto, 17
espaços-L^p, 10, 11
esperança matemática, 10, 11
esperança matemática condicional, 12
estado da natureza, 8, 26
estimadores de máxima verosimilhança, 103
estratégia autofinanciada, 153
estratégia de protecção, 158
estratégia de transacção, 152
estratégia hedging, 157, 162
existência, 84
extinção de população, 101
extinção matemática, 132
extinção realista, 133

Feynman, 4
filtração, 18
filtração não-antecipativa, 55
filtração natural, 18
fórmula de Black-Scholes, 156
fórmula de Dynkin, 113
fórmula de Feynman-Kac, 114
fórmula de Itô, 72, 100
fórmula de Itô multidimensional, 81
fronteira atingível, 127
fronteira atractiva, 126
fronteira não-atractiva, 126
função característica, 10
função de auto-covariância, 18, 29
função de distribuição, 9

função de distribuição conjunta, 11
função de distribuição espectral, 18
função de distribuição imprópria, 20
função de escala, 122
função de velocidade, 122
função delta de Dirac, 29
função densidade de probabilidade, 9, 11
função em escada, 57
função indicatriz de um conjunto, 10
função integranda com valores matriciais, 80
função massa de probabilidade, 9
função mensurável, 8, 62
função não-antecipativa, 55, 57
função simples, 10

Girsanov, 5

iguais com probabilidade um, 10
incremento, 26
incrementos estacionários, 32
incrementos independentes, 32
incrementos não-correlacionados, 68
independência, 32
integral como aplicação contínua, 60, 61
integral como aplicação linear, 59, 61
integral de função em escada, 57
integral de Itô, 53, 55, 60, 84, 100
integral de Itô como função do limite superior de integração, 65
integral de Itô de função contínua m.q., 55, 62
integral de Itô de função determinística, 62
integral de Itô multidimensional, 80
integral de Lebesgue, 10
integral de Riemann-Stieltjes, 52
integral de Stratonovich, 55, 78
integral não-antecipativo, 53
Itô, 1

Kac, 4

lei do arcoseno, 39
lei do logaritmo iterado, 36
lei forte dos grandes números, 36
lema de Bellman-Gronwall, 87
lema de Borel-Cantelli, 64, 67
Lévy, 2, 31
limiar de extinção, 133

martingala, 19, 66
matriz de variância-covariância, 11
máximo do processo de Wiener, 39
média, 10
média infinitésimal, 42
medida, 56
medida absolutamente contínua, 138
medida de escala, 122
medida de Lebesgue, 9, 56
medida de velocidade, 122
medida finita-σ, 138
medida martingala equivalente, 161, 170
mercado, 169
mercado completo, 170
mercado normalizado, 169
Merton, 2, 150
modelo binomial, 164
modelo CRR, 164
modelo de Black-Scholes, 26, 51, 99, 129, 151

modelo de Cox-Ingersoll-Ross, 119
modelo de Cox-Ross-Rubinstein, 164
modelo de Ornstein-Uhlenbeck, 117
modelo de Pearl-Verhulst, 130
modelo de Vasicek, 118, 129
modelo logístico, 130
modelo malthusiano, 25
momentos amostrais, 46
momentos de conjunto, 46
momentos infinitésimais, 42
movimento browniano, 1, 27, 31, 117
movimento browniano com tendência, 43, 47, 139
movimento browniano geométrico, 100

não-arbitragem, 152
Nikodym, 138
norma-L^p, 11
norma-L^{2*}, 56, 80

opção, 149
opção americana, 150
opção atingível, 170
opção de compra, 150
opção de venda, 150
opção europeia, 150
opção europeia de compra, 151
opção europeia de venda, 167
operador de difusão, 43, 112
operador infinitésimal, 112
Ornstein, 1
Ornstein-Ulhenbeck, 1

passeio aleatório, 34
portfolio, 152, 169
portfolio admissível, 170
portfolio autofinanciado, 170
portfolio hedging, 170
posição curta, 151
posição longa, 151
preço de mercado do risco, 160
preservação da norma, 60, 61
probabilidade, 7
probabilidade condicional, 12
probabilidade de extinção, 133
probabilidades de transição, 23
probabilidades equivalentes, 138, 142
problema de Cauchy, 51
processo adaptado a uma filtração, 19
processo contínuo, 16
processo contínuo em média quadrática, 18
processo de difusão, 41
processo de difusão homogéneo, 42, 44, 45
processo de difusão multidimensional, 48
processo de Itô, 71
processo de Itô multidimensional, 81
processo de Markov, 21, 22
processo de Markov forte, 24, 37, 96
processo de Markov homogéneo, 23
processo de Ornstein-Uhlenbeck, 118
processo de Wiener, 32, 42
processo de Wiener multidimensional, 39
processo de Wiener padrão, 27, 31
processo ergódico, 46, 119, 129
processo estacionário, 18
processo estacionário de segunda ordem, 18
processo estacionário em sentido lato, 18
processo estocástico, 13
processo estocástico em tempo

contínuo, 13
processo estocástico em tempo discreto, 13
processo estocástico estritamente estacionário, 18
processo estocástico generalizado, 29
processo estocástico multidimensional, 17
processo gaussiano, 18
processo mensurável, 17
processo regular, 120
processo separável, 16, 17
processos com trajectórias idênticas, 16
processos estocásticos equivalentes, 16
produto interno, 11, 57, 80
propriedade de compatibilidade, 15
propriedade de Markov, 22
propriedade de Markov forte, 24
propriedade de simetria, 15
put option, 150

quase certamente, 10

Radon, 138
regra da cadeia, 72
regra de diferenciação de uma função composta, 72
restrição ao crescimento, 84
reversão para a média, 118
Ross, 164
Rubinstein, 164
ruído branco, 30
ruído branco em tempo discreto, 29
ruído branco padrão, 29
ruído branco padrão em tempo discreto, 139
ruído colorido, 29

Scholes, 2, 3, 26, 150
sistema de equações diferenciais estocásticas, 80
sistema de equações integrais estocásticas, 80
solução de equação diferencial estocástica, 51, 84
solução forte, 95
solução fraca, 95
solução fundamental, 44, 45
somas de Riemann-Stieltjes, 52
Stratonovich, 4
subespaço denso, 63
submartingala, 19
sucessão aproximadora, 60, 63, 68
supermartingala, 19

taxa de crescimento *per capita*, 25
taxa de juro, 25, 118
taxa de rendimento, 25
taxa intrínseca de crescimento, 130
taxa média aritmética de rendimento, 109
taxa média geométrica de rendimento, 109
tempo de extinção, 133
tempo de Markov, 20
tempo de paragem, 20
tempo de primeira passagem, 20, 38, 133
teorema da extensão de Kolmogorov, 15
teorema de Fubini, 56
teorema de Girsanov, 160
teorema de Itô, 72, 73, 81
teorema de Lusin, 64
teorema de Radon-Nikodym, 12, 138
teste de Feller, 97
trajectória, 13
trajectórias contínuas, 67

transformada de Fourier, 18, 29

Uhlenbeck, 1
unicidade, 84
unicidade forte, 95
unicidade fraca, 95
universo, 7

valor de exercício, 151, 167
valor esperado, 10
valor médio, 10
variação ilimitada, 35
variância, 10
variância infinitésimal, 42
variáveis aleatórias independentes, 32
variáveis aleatórias equivalentes, 10
variáveis aleatórias quase iguais, 10
variável aleatória, 8
variável aleatória absolutamente contínua, 9
variável aleatória discreta, 9
variável aleatória estendida, 20
variável aleatória gaussiana, 10
variável aleatória multidimensional, 11
variável aleatória normal, 10
vector aleatório, 11
vector aleatório absolutamente contínuo, 11
vector aleatório discreto, 11
vector aleatório gaussiano, 11
vector aleatório normal, 11
vector médio, 11
versão de um processo estocástico, 16
versão separável, 17
volatilidade, 27, 99

Wiener, 2, 31



ISBN 972-8890-06-0
Depósito Legal nº229951/05