Stochastic Differential Equations with Applications

December 4, 2012

Contents

1 Stochastic Differential Equations **7**

1.1 Introduction 7

1.1.1 Meaning of Stochastic Differential Equations 8

1.2 Some applications of SDEs 10

1.2.1 Asset prices 10

1.2.2 Population modelling 10

1.2.3 Multi-species models 11

2 Continuous Random Variables **13**

2.1 The probability density function 13

2.1.1 Change of independent deviates 13

2.1.2 Moments of a distribution 13

2.2 Ubiquitous probability density functions in continuous finance 14

2.2.1 Normal distribution 14

2.2.2 Log-normal distribution 14

2.2.3 Gamma distribution 14

2.3 Limit Theorems 15

2.3.1 The central limit results 17

2.4 The Wiener process 17

2.4.1 Covariance 18

2.4.2 Derivative of a Wiener process 18

2.4.3 Definitions 18

3 Review of Integration **21**

3.1 Bounded Variation . . . 21
3.2 Riemann integration . . . 23
3.3 Riemann-Stieltjes integration . . . 24
3.4 Stochastic integration of deterministic functions . . . 25

4 The Ito and Stratonovich Integrals 27
4.1 A simple stochastic differential equation . . . 27
4.1.1 Motivation for Ito integral . . . 28
4.2 Stochastic integrals . . . 29
4.3 The Ito integral . . . 31
4.3.1 Application . . . 33
4.4 The Stratonovich integral . . . 33
4.4.1 Relationship between the Ito and Stratonovich integrals . . . 35
4.4.2 Stratonovich integration conforms to the classical rules of integration . . . 37
4.5 Stratonovich representation on an SDE . . . 38

5 Differentiation of functions of stochastic variables 41
5.1 Ito's Lemma . . . 41
5.1.1 Ito's lemma in multi-dimensions . . . 44
5.1.2 Ito's lemma for products . . . 45
5.2 Further examples of SDEs . . . 45
5.2.1 Multi-dimensional Ornstein Uhlenbeck equation . . . 45
5.2.2 Logistic equation . . . 46
5.2.3 Square-root process . . . 47
5.2.4 Constant Elasticity of Variance Model . . . 48
5.2.5 SABR Model . . . 48
5.3 Heston's Model . . . 49
5.4 Girsanov's lemma . . . 49
5.4.1 Change of measure for a Wiener process . . . 51
5.4.2 Girsanov theorem . . . 52

6 The Chapman Kolmogorov Equation 55

6.1 Introduction . . . 55
6.1.1 Markov process . . . 56
6.2 Chapman-Kolmogorov equation . . . 57
6.2.1 Path continuity . . . 57
6.2.2 Drift and diffusion . . . 59
6.2.3 Drift and diffusion of an SDE . . . 60
6.3 Formal derivation of the Forward Kolmogorov Equation . . . 61
6.3.1 Intuitive derivation of the Forward Kolmogorov Equation . . . 66
6.3.2 The Backward Kolmogorov equation . . . 67
6.4 Alternative view of the forward Kolmogorov equation . . . 69
6.4.1 Jumps . . . 70
6.4.2 Determining the probability flux from an SDE - One dimension . . . 71

7 Numerical Integration of SDE **75**
7.1 Issues of convergence . . . 75
7.1.1 Strong convergence . . . 75
7.1.2 Weak convergence . . . 76
7.2 Deterministic Taylor Expansions . . . 77
7.3 Stochastic Ito-Taylor Expansion . . . 78
7.4 Stochastic Stratonovich-Taylor Expansion . . . 80
7.5 Euler-Maruyama algorithm . . . 81
7.6 Milstein scheme . . . 82
7.6.1 Higher order schemes . . . 83

8 Exercises on SDE **85**

Chapter 1

Stochastic Differential Equations

1.1 Introduction

Classical mathematical modelling is largely concerned with the derivation and use of ordinary and partial differential equations in the modelling of natural phenomena, and in the mathematical and numerical methods required to develop useful solutions to these equations. Traditionally these differential equations are deterministic by which we mean that their solutions are completely determined in the value sense by knowledge of boundary and initial conditions - identical initial and boundary conditions generate identical solutions. On the other hand, a Stochastic Differential Equation (SDE) is a differential equation with a solution which is influenced by boundary and initial conditions, but not predetermined by them. Each time the equation is solved under identical initial and boundary conditions, the solution takes different numerical values although, of course, a definite pattern emerges as the solution procedure is repeated many times.

Stochastic modelling can be viewed in two quite different ways. The optimist believes that the universe is still deterministic. Stochastic features are introduced into the mathematical model simply to take account of imperfections (approximations) in the current specification of the model, but there exists a version of the model which provides a perfect explanation of observation without redress to a stochastic model. The pessimist, on the other hand, believes that the universe is intrinsically stochastic and that no deterministic model exists. From a pragmatic point of view, both will construct the same model - its just that each will take a different view as to origin of the stochastic behaviour.

Stochastic differential equations (SDEs) now find applications in many disciplines including *inter alia* engineering, economics and finance, environmetrics, physics, population dynamics, biology and medicine. One particularly important application of SDEs occurs in the modelling of problems associated with water catchment and the percolation of fluid through porous/fractured structures.

In order to acquire an understanding of the physical meaning of a stochastic differential equation (SDE), it is beneficial to consider a problem for which the underlying mechanism is deterministic and fully understood. In this case the SDE arises when the underlying deterministic mechanism is not

fully observed and so must be replaced by a stochastic process which describes the behaviour of the system over a larger time scale. In effect, although the true mechanism is deterministic, when this mechanism cannot be fully observed it manifests itself as a stochastic process.

1.1.1 Meaning of Stochastic Differential Equations

A useful example to explore the mapping between an SDE and reality is consider the origin of the term "noise", now commonly used as a generic term to describe a zero-mean random signal, but in the early days of radio noise referred to an unwanted signal contaminating the transmitted signal due to atmospheric effects, but in particular, imperfections in the transmitting and receiving equipment itself. This unwanted signal manifest itself as a persistent background hiss, hence the origin of the term. The mechanism generating noise within early transmitting and receiving equipment was well understood and largely arose from the fact that the charge flowing through valves is carried in units of the electronic charge e (1.6×10^{-19} coulombs per electron) and is therefore intrinsically discrete.

Consider the situation in which a stream of particles each carrying charge q land on the plate of a leaky capacitor, the k^{th} particle arriving at time $t_k > 0$. Let $N(t)$ be the number of particles to have arrived on the plate by time t then

$$N(t) = \sum_{k=1}^{\infty} H(t - t_k)\,,$$

where $H(t)$ is Heaviside's function[1]. The noise resulting from the irregular arrival of the charged particles is called *shot noise*. The situation is illustrated in Figure 1.1

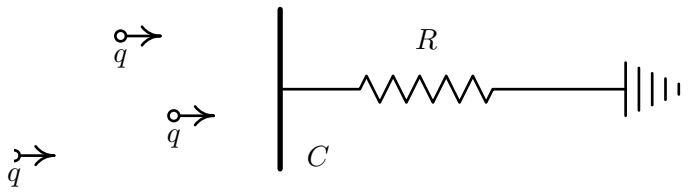

Figure 1.1: A model of a leaky capacitor receiving charges q

Let $V(t)$ be the potential across the capacitor and let $I(t)$ be the leakage current at time t then

[1]The Heaviside function, often colloquially called the 'Step Function", was introduced by Oliver Heaviside (1850-1925), an English electrical engineer, and is defined by the formula

$$H(x) = \int_{-\infty}^{x} \delta(s)\,ds \quad \rightarrow \quad H(x) = \begin{bmatrix} 1 & x > 0 \\ \frac{1}{2} & x = 0 \\ 0 & x < 0 \end{bmatrix}$$

Clearly the derivative of $H(x)$ is $\delta(x)$, Dirac's delta function

conservation of charge requires that

$$CV = q\,N(t) - \int_0^t I(s)\,ds$$

where $V(t) = RI(t)$ by Ohms law. Consequently, $I(t)$ satisfies the integral equation

$$CR\,I(t) = q\,N(t) - \int_0^t I(s)\,ds\,. \tag{1.1}$$

We can solve this equation by the method of Laplace transforms, but we avoid this temptation.

Another approach

Consider now the nature of $N(t)$ when the times t_k are unknown other than that electrons behave independently of each other and that the interval between the arrivals of electrons of the plate is Poisson distributed with parameter λ. With this understanding of the underlying mechanism in place, $N(t)$ is a Poisson deviate with parameter λt. At time $t + \Delta t$ equation (1.1) becomes

$$CR\,I(t+\Delta t) = q\,N(t+\Delta t) - \int_0^{t+\Delta t} I(s)\,ds\,.$$

After subtracting equation (1.1) from the previous equation the result is that

$$CR\,[I(t+\Delta t) - I(t)] = q\,[N(t+\Delta t) - N(t)] - \int_t^{t+\Delta t} I(s)\,ds\,. \tag{1.2}$$

Now suppose that Δt is sufficiently small that $I(t)$ does not change its value significantly in the interval $[t, t+\Delta t]$ then $\Delta I = I(t+\Delta t) - I(t)$ and

$$\int_t^{t+\Delta t} I(s)\,ds \approx I(t)\,\Delta t\,.$$

On the other hand suppose that Δt is sufficiently large that many electrons arrive on the plate during the interval $(t, t+\Delta t)$. So although $[N(t+\Delta t) - N(t)]$ is actually a Poisson random variable with parameter $\lambda\Delta t$, the central limit theorem may be invoked and $[N(t+\Delta t) - N(t)]$ may be approximated by a Gaussian deviate with mean value $\lambda\Delta t$ and variance $\lambda\Delta t$. Thus

$$N(t+\Delta t) - N(t) \approx \lambda\Delta t + \lambda\Delta W\,, \tag{1.3}$$

where ΔW is a Gaussian random variable with zero mean value and variance Δt. The sequence of values for ΔW as time is traversed in units of Δt define the independent increments in a Gaussian process $W(t)$ formed by summing the increments ΔW. Clearly $W(t)$ has mean value zero and variance t. The conclusion of this analysis is that

$$CR\Delta I = q\,(\lambda\Delta t + \sqrt{\lambda}\Delta W) - I(t)\Delta t \tag{1.4}$$

with initial condition $I(0) = 0$. Replacing ΔI, Δt and ΔW by their respective differential forms leads to the *Stochastic Differential Equation* (SDE)

$$dI = \frac{(q\lambda - I)}{CR}\,dt + \frac{q\sqrt{\lambda}}{CR}\,dW\,. \tag{1.5}$$

In this representation, dW is the increment of a *Wiener process*. Both I and W are functions which are continuous everywhere but differentiable nowhere. Equation (1.5) is an example of an *Ornstein-Uhlenbeck process*.

1.2 Some applications of SDEs

1.2.1 Asset prices

The most relevant application of SDEs for our purposes occurs in the pricing of risky assets and contracts written on these assets. One such model is Heston's model of stochastic volatility which posits that S, the price of a risky asset, evolves according to the equations

$$\begin{aligned} \frac{dS}{S} &= \mu\,dt + \sqrt{V}\,(\sqrt{1-\rho^2}\,dW_1 + \rho\,dW_2) \\ dV &= \kappa\,(\gamma - V)\,dt + \sigma\sqrt{V}\,dW_2 \end{aligned} \tag{1.6}$$

in which ρ, κ and γ take prescribed values and $V(t)$ is the instantaneous level of volatility of the stock, and dW_1 and dW_2 are differentials of uncorrelated Wiener processes. In the course of these lectures we shall meet other financial models.

1.2.2 Population modelling

Stochastic differential equations are often used in the modelling of population dynamics. For example, the *Malthusian* model of population growth (unrestricted resources) is

$$\frac{dN}{dt} = aN\,, \qquad N(0) = N_0\,, \tag{1.7}$$

where a is a constant and $N(t)$ is the size of the population at time t. The effect of changing environmental conditions is achieved by replacing $a\,dt$ by a Gaussian random variable with non-zero mean $a\,dt$ and variance $b^2\,dt$ to get the stochastic differential equation

$$dN = aN\,dt + bN\,dW\,, \qquad N(0) = N_0\,, \tag{1.8}$$

in which a and b (conventionally positive) are constants. This is the equation of a *Geometric Random Walk* and is identical to the risk-neutral asset model proposed by Black and Scholes for the evolution of the price of a risky asset. To take account of limited resources the Malthusian model of population growth is modified by replacing a in equation (1.8) by the term $\alpha(M - N)$ to get the well-known *Verhulst* or *Logistic* equation

$$\frac{dN}{dt} = \alpha\,N(M - N)\,, \qquad N(0) = N_0 \tag{1.9}$$

where α and M are constants with M representing the *carrying capacity* of the environment. The effect of changing environmental conditions is to make M a stochastic parameter so that the logistic model becomes

$$dN = aN(M - N)\,dt + b\,N\,dW \tag{1.10}$$

in which a and $b \geq 0$ are constants.

The Malthusian and Logistic models are special cases of the general stochastic differential equation

$$dN = f(N)\,dt + g(N)\,dW \tag{1.11}$$

where f and g are continuously differentiable in $[0, \infty)$ with $f(0) = g(0) = 0$. In this case, $N = 0$ is a solution of the SDE. However, the stochastic equation (1.11) is sensible only provided the boundary at $N = \infty$ is unattainable (a natural boundary). This condition requires that there exist $K > 0$ such that $f(N) < 0$ for all $N > K$. For example, $K = M$ in the logistic model.

1.2.3 Multi-species models

The classical multi-species model is the prey-predator model. Let $R(t)$ and $F(t)$ denote respectively the populations of rabbits and foxes is a given environment, then the Lotka-Volterra two-species model posits that

$$\begin{aligned} \frac{dR}{dt} &= R(\alpha - \beta\, F) \\ \frac{dF}{dt} &= F(\delta\, R - \gamma)\,, \end{aligned}$$

where α is the net birth rate of rabbits in the absence of foxes, γ is the natural death rate of foxes and β and δ are parameters of the model controlling the interaction between rabbits and foxes. The stochastic equations arising from allowing α and γ to become random variables are

$$\begin{aligned} dR &= R(\alpha - \beta\, F)\,dt + \sigma_r\, R\, dW_1\,. \\ dF &= F(\delta\, R - \gamma)\,dt + \sigma_f\, F\, dW_2\,. \end{aligned}$$

The generic solution to these equations is a cyclical process in which rabbits dominate initially, thereby increasing the supply of food for foxes causing the fox population to increase and rabbit population to decline. The increasing shortage of food then causes the fox population to decline and allows the rabbit population to recover, and so on.

There is a multi-species variant of the two-species classical Lotka-Volterra multi-species model with differential equations

$$\frac{dx_k}{dt} = (a_k + b_{kj}x_j)x_k\,, \qquad k \quad \text{n.s.}\,, \tag{1.12}$$

where the choice and algebraic signs of the a_k's and the b_{jk}'s distinguishes prey from predator. The simplest generalisation of the Lotka-Volterra model to stochastic a environment assumes the original a_k is stochastic to get

$$dx_k = (a_k + b_{kj}x_j)x_k\,dt + c_k\,x_k\,dW_k\,, \qquad k \quad \text{not summed.} \tag{1.13}$$

The generalised Lokta-Volterra models also have cyclical solutions. Is there an analogy with business cycles in Economics in which the variables x_k now denote measurable economic variables?

Chapter 2

Continuous Random Variables

2.1 The probability density function

Function $f(x)$ is a **probability density function** (PDF) with respect to a subset $\mathcal{S} \subseteq \mathbb{R}^n$ provided

$$(a) \quad f(x) \geq 0 \qquad \forall\, x \in \mathcal{S}\,, \qquad\qquad (ii) \quad \int_{\mathcal{S}} f(x)\, dx = 1\,. \tag{2.1}$$

In particular, if $E \subseteq \mathcal{S}$ is an event then the probability of E is

$$p(E) = \int_E f(x)\, dx\ .$$

2.1.1 Change of independent deviates

Suppose that $y = g(x)$ is an invertible mapping which associates $X \in \mathcal{S}_X$ with $Y \in \mathcal{S}_Y$ where $\mathcal{S}_Y$ is the image of the original sample space $\mathcal{S}_X$ under the mapping g. The probability density function $f_Y(y)$ of Y may be computed from the probability density function $f_X(x)$ of X by the rule

$$f_Y(y) = f_X(x) \left| \frac{\partial(x_1, \cdots, x_n)}{\partial(y_1, \cdots, y_n)} \right|. \tag{2.2}$$

2.1.2 Moments of a distribution

Let X be a continuous random variable with sample space $\mathcal{S}$ and PDF $f(\mathbf{x})$ then the mean value of the function $g(X)$ is defined to be

$$\bar{g} = \int_{\mathcal{S}} g(x) f(x)\, dx\,.$$

Important properties of the distribution itself are defined from this definition by assigning various scalar, vector of tensorial forms to g. The moment $M_{pq\cdots w}$ of the distribution is defined by the formula

$$M_{p\,q\cdots w} = \mathrm{E}\,[\,X_p X_q \cdots X_w\,] = \int_{\mathcal{S}} (x_p x_q \cdots x_w)\, f(x)\, dx\,. \tag{2.3}$$

2.2 Ubiquitous probability density functions in continuous finance

Although any function satisfying conditions (2.1) qualifies as a probability density function, there are several well-known probability density functions that occur frequently in continuous finance and with which one must be familiar. These are now described briefly.

2.2.1 Normal distribution

A continuous random variable X is Gaussian distributed with mean μ and variance σ^2 if X has probability density function

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\Big[-\frac{(x-\mu)^2}{2\sigma^2} \Big] .$$

Basic properties

- If $X \sim \mathrm{N}(\mu, \sigma^2)$ then $Y = (X-\mu)/\sigma \sim \mathrm{N}(0,1)$. This result follows immediately by change of independent variable from X to Y.
- Suppose that $X_k \sim \mathrm{N}(\mu_k, \sigma_k^2)$ for $k = 1, \ldots, n$ are n independent Gaussian deviates then mathematical induction may be used to establish the result

$$X = X_1 + X_2 + \ldots + X_n \sim \mathrm{N}\Big(\sum_{k=1}^{n} \mu_k \, , \, \sum_{k=1}^{n} \sigma_k^2 \Big) .$$

2.2.2 Log-normal distribution

The log-normal distribution with parameters μ and σ is defined by the probability density function

$$f(x \, ; \, \mu, \sigma) = \frac{1}{\sqrt{2\pi}\,\sigma} \frac{1}{x} \exp\Big[-\frac{(\log x - \mu)^2}{2\sigma^2} \Big] \qquad x \in (0, \infty) \tag{2.4}$$

Basic properties

- If X is log-normally distributed with parameters (μ, σ) then $Y = \log X \sim \mathrm{N}(\mu, \sigma^2)$. This result follows immediately by change of independent variable from X to Y.
- If X is log-normally distributed with parameters (μ, σ) independent Gaussian deviates then $\mathbb{E}[X] = e^{\mu + \sigma^2/2}$ and $\mathbb{V}[X] = e^{2\mu+\sigma^2}(e^{\sigma^2} - 1)$ and the median of X is e^{μ}.

2.2.3 Gamma distribution

The Gamma distribution with shape parameter α and scale parameter λ is defined by the probability density function

$$f(x) = \frac{1}{\lambda \Gamma(\alpha)} \Big(\frac{x}{\lambda} \Big)^{\alpha - 1} e^{-x/\lambda} ,$$

where λ and α are positive parameters.

Basic properties

- If X is Gamma distributed with parameters (α, λ) then $\mathbb{E}[X] = \alpha\lambda$ and $\mathbb{V}[X] = \alpha\lambda^2$.
- Suppose that $X_1, \cdots, X_n$ are n independent Gamma deviates such that X_k has shape parameter α_k and scale parameter λ, the same for all values of k, then mathematical induction may be used to establish the result that $X = X_1 + X_2 + \ldots + X_n$ is Gamma distributed with shape parameter $\alpha_1 + \alpha_2 + \cdots + \alpha_n$ and scale parameter λ.

2.3 Limit Theorems

"Limit Theorems" are arguably the most important theoretical results in probability. The results come in two flavours:

(i) **Laws of Large Numbers** where the aim is to establish convergence results relating sample and population properties. For example, how many times must one toss a coin to be 99% sure that the relative frequency of heads is within 5% of the true bias of the coin.

(ii) **Central Limit Theorems** where the aim is to establish properties of the distribution of the sample properties.

Subsequent statements will assume that $X_1, X_2, \ldots, X_n$ is a sample of n independent and identically distributed (i.i.d.) random variables. The sum of the random variables in the sample will be denoted by $S_n = X_1 + \cdots + X_n$. There are three different ways in which a random sequence Y_n can converge to, say T, as $n \to \infty$.

Definition

(a) We say that $Y_n \to Y$ **strongly** as $n \to \infty$ if

$$\text{Prob}\,(\,|\,Y_n - Y\,| \to 0 \text{ as } n \to \infty\,) = 1\,.$$

(b) We say that $Y_n \to Y$ as $n \to \infty$ in the **mean square** sense if

$$\|\,Y_n - Y\,\| = \sqrt{\mathrm{E}\,[\,(Y_n - Y)^2\,]} \to 0 \text{ as } n \to \infty\,.$$

(c) We say that $Y_n \to Y$ **weakly** as $n \to \infty$ if given $\varepsilon > 0$,

$$\text{Prob}\,(\,|\,Y_n - Y\,| \geq \varepsilon\,) \to 0 \text{ as } n \to \infty\,.$$

Weak convergence (sometimes called stochastic convergence) is the weakest convergence condition. Strong convergence and mean square convergence both imply weak convergence.

Before considering specific limit theorems, we establish Chebyshev's inequality.

Chebyshev's inequality

Let $g(x)$ be a non-negative function of a continuous random variable X then

$$\text{Prob}\,(\,g(X) > K\,) < \frac{\text{E}\,[\,g(X)\,]}{K}\,.$$

Justification Let X have pdf $f(x)$ then

$$\text{E}\,[\,g(X)\,] = \int g(x)\,f(x)\,dx \geq K \int_{g(x)\geq K} f(x)\,dx = K\,\text{Prob}\,(\,g(X) > K\,)\,.$$

Chebyshev's inequality follows immediately from this result.

Corollary

Let X be a random variable drawn from a distribution with finite mean μ and finite variance σ^2 then it follows directly from Chebyshev's inequality that

$$\text{Prob}\,(\,|\,X-\mu\,| > \epsilon) \leq \frac{\sigma^2}{\epsilon^2}\,.$$

Justification Take $g(x) = (X-\mu)^2/\sigma^2$ and $K = \varepsilon^2/\sigma^2$. Clearly $\text{E}\,[\,g(X)\,] = 1$ and Chebyshev's inequality now gives

$$\text{Prob}\,(\,(X-\mu)^2/\sigma^2 > \varepsilon^2/\sigma^2\,) < \frac{\sigma^2}{\varepsilon^2} \quad \rightarrow \quad \text{Prob}\,(\,|\,X-\mu\,| > \varepsilon\,) < \frac{\sigma^2}{\varepsilon^2}\,.$$

The weak law of large numbers

Let $X_1,\ X_2, \ldots, X_n$ be a sample of n i.i.d random variables drawn from a distribution with finite mean μ and finite variance σ^2, then for any positive ϵ

$$\text{Prob}\,\left(\left|\frac{S_n}{n} - \mu\right| > \epsilon\right) \rightarrow 0 \quad \text{as } n \rightarrow \infty\,.$$

Justification To establish this result, apply Chebyshev's inequality with $X = S_n/n$. In this case $E(X) = \mu$ and $\sigma_X^2 = \sigma^2/n$. It follows directly from Chebyshev's inequality that

$$p(|X-\mu| > \epsilon) \leq \frac{\sigma^2}{n\epsilon^2} \rightarrow 0 \text{ as } n \rightarrow \infty\,.$$

The strong law of large numbers

Let $X_1,\ X_2, \ldots, X_n$ be a sample of n i.i.d random variables drawn from a distribution with finite mean μ and finite variance σ^2, then

$$\frac{S_n}{n} \rightarrow \mu \text{ as } n \rightarrow \infty \text{ w.p. } 1\,.$$

[In fact, the finite variance condition is not necessary for the strong law of large numbers to be true.]

2.3.1 The central limit results

Let X_1, $X_2, \ldots X_n$ be a sample of n i.i.d. random variables from a distribution with finite mean μ and finite variance σ^2, then the random variable

$$Y_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}$$

has mean value zero and unit variance. In particular, the unit variance property means that Y_n does not degenerate as $n \to \infty$ by contrast with S_n/n. The law of the iterated logarithm controls the growth of Y_n as $n \to \infty$.

Central limit theorem

Under the conditions on X stated previously, the derived deviate Y_n satisfies

$$\lim_{n\to\infty} \text{Prob}\,(\,Y_n \leq y\,) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-t^2/2}\, dt\,.$$

The crucial point to note here is that the result is independent of distribution provided each deviate X_k is i.i.d. with finite mean and variance. For example, The strong law of large numbers would apply to samples drawn from the distribution with density $f(x) = 2(1+x)^{-3}$ with $x \in \mathbb{R}$ but not the central limit theorem.

2.4 The Wiener process

In 1827 the Scottish botanist Robert Brown (1773-1858) first described the erratic behaviour of particles suspended in fluid, an effect which subsequently came to be called "Brownian Motion". In 1905 Einstein explained Brownian motion mathematically[1] and this led to attempts by Langevin and others to formulate the dynamics of this motion in terms of stochastic differential equations. Norbert Wiener (1894-1964) introduced a mathematical model of Brownian motion based on a canonical process which is now called the *Wiener process*.

The Wiener process, denoted $W(t)$ $t \geq 0$, is a continuous stochastic process which takes the initial value $W(0) = 0$ with probability one and is such that the increment $[W(t) - W(s)]$ from $W(s)$ to $W(t)$ $(t > s)$ is an independent Gaussian process with mean value zero and variance $(t - s)$. If W_t is a Wiener process, then

$$\Delta W_t = W_{t+\Delta t} - W_t$$

is a Gaussian process satisfying

$$\mathbb{E}\,[\,\Delta W_t\,] = 0\,, \qquad \mathbb{V}\,[\,\Delta W_t\,] = \Delta t\,. \tag{2.5}$$

[1]The canonical equation is called the diffusion equation, written

$$\frac{\partial\theta}{\partial t} = \frac{\partial^2\theta}{\partial x^2}$$

in one dimension. It is classified as a parabolic partial differential equation and typically describes the evolution of temperature in heated material.

It is often convenient to write $\Delta W_t = \sqrt{\Delta t}\,\epsilon_t$ for the increment experienced by the Wiener process $W(t)$ in traversing the interval $[t, t+\Delta t]$ where $\epsilon_t \sim \mathrm{N}(0,1)$. Similarly, the infinitesimal increment dW in $W(t)$ during the increment dt is often written $dW = \sqrt{dt}\,\epsilon_t$. Suppose that $t = t_0 < \cdots < t_n = t+T$ is a dissection of $[t, t+T]$ then

$$W(t+T) - W(t) = \sum_{k=1}^{n} [\, W(t_k) - W(t_{k-1}) \,] = \sum_{k=1}^{n} \sqrt{(t_k - t_{k-1})}\,\epsilon_k\,, \qquad \epsilon_k \sim \mathrm{N}(0,1)\,. \tag{2.6}$$

Note that this incremental form allows the mean and variance properties of the Wiener process to be established directly from the properties of the Gaussian distribution.

2.4.1 Covariance

Let t and s be times with $t > s$ then

$$\mathrm{E}\,[\, W(t)W(s) \,] = \mathrm{E}\,[\, W(s)(W(t) - W(s)) + W^2(s) \,] = \mathrm{E}\,[\, W^2(s) \,] = s\,. \tag{2.7}$$

In the derivation of result (2.7) it has been recognised that $W(s)$ and the Wiener increment $W(t) - W(s)$ are independent Gaussian deviates. In general, $\mathrm{E}\,[\, W(t)W(s) \,] = \min(t, s)$.

2.4.2 Derivative of a Wiener process

The formal derivative of the Wiener process $W(t)$ is the limiting value of the ratio

$$\lim_{\Delta t \to 0} \frac{W(t+\Delta t) - W(t)}{\Delta t}\,. \tag{2.8}$$

However $W(t+\Delta t) - W(t) = \sqrt{\Delta t}\,\epsilon_t$ and so

$$\frac{W(t+\Delta t) - W(t)}{\Delta t} = \frac{\epsilon_t}{\sqrt{\Delta t}}\,.$$

Thus the limit (2.8) does not exist as $\Delta t \to 0^+$ and so $W(t)$ has no derivative in the classical sense. Intuitively this means that a particle with position $x(t) = W(t)$ has no well defined velocity although its position is a continuous function of time.

2.4.3 Definitions

1. A **martingale** is a gambling term for a fair[2] game in which the current information set I_t provides no advantage in predicting future winnings. A random variable X_t with finite expectation is a martingale with respect to the probability measure $\mathbb{P}$ if

$$\mathrm{E}_{\,\mathbb{P}}\,[\, X_{t+s} \,|\, I_t \,] = X_t\,, \qquad s > 0\,. \tag{2.9}$$

[2]Of course, martingale gambling strategies are an anathema to casinos; this is the primary reason for a house limit which, of course, precludes martingale strategies.

Thus the future expectation of X is its current value, or put another way, the future offers no opportunity for arbitrage.

2. A random variable X_t with finite expectation is said to be a **sub-martingale** with respect to the probability $\mathbb{Q}$ if

$$\mathrm{E}_{\mathbb{Q}}\left[\, X_{t+s} \,|\, I_t \,\right] \geq X_t \,, \qquad s > 0 \,. \tag{2.10}$$

3. Let $t_0 < t_1 \cdots < t \cdots$ be an increasing sequence of times with corresponding information sets $I_0, I_1, \cdots, I_t, \cdots$. These information sets are said to form a **filtration** if

$$I_0 \subseteq I_1 \subseteq \cdots \subseteq I_t \subseteq \cdots \tag{2.11}$$

For example, $W_t = W(t)$ is a martingale with respect to the information set I_0. Why? Because

$$\mathrm{E}\left[\, W_t \,|\, I_0 \,\right] = 0 = W(0) \,,$$

but W_t^2 is a sub-martingale since $\mathrm{E}\left[\, W_t^2 \,|\, I_0 \,\right] = t > 0$.

Chapter 3

Review of Integration

Before advancing to the discussion of stochastic differential equations, we need to review what is meant by integration. Consider briefly the one-dimensional SDE

$$dx_t = a(t, x_t)\, dt + b(t, x_t)\, dW_t\,, \qquad x(t_0) = x_0\,, \tag{3.1}$$

where dW_t is the differential of the Wiener process $W(t)$, and $a(t, x)$ and $b(t, x)$ are deterministic functions of x and t. The motivation behind this choice of direction lies in the observation that the formal solution to equation (3.1) may be written

$$x_t = x_0 + \int_{t_0}^{t} a(s, x_s)\, ds + \int_{t_0}^{t} b(s, x_s)\, dW_s\,. \tag{3.2}$$

Although this solution contains two integrals, to have any worth we need to know exactly what is meant by each integral appearing in this solution.

At a casual glance, the first integral on the right hand side of solution (3.2) appears to be a Riemann integral while the second integral appears to be a Riemann-Stieltjes integral. However, a complication arises from the fact that x_s is a stochastic process and so $a(s, x_s)$ and $b(s, x_s)$ behave stochastically in the integrals although a and b are themselves deterministic functions of t and x.

In fact, the first integral on the right hand side of equation (3.2) can always be interpreted as a Riemann integral. However, the interpretation of the second integral is problematic. Often it is an Ito integral, but may be interpreted as a Riemann-Stieltjes integral under some circumstances.

3.1 Bounded Variation

Let $\mathcal{D}_n$ denote the dissection $a = x_0 < x_1 < \cdots < x_n = b$ of the finite interval $[a, b]$. The *p-variation* of a function f over the domain $[a, b]$ is defined to be

$$V_p(f) = \lim_{n\to\infty} \sum_{k=1}^{n} |\, f(x_k) - f(x_{k-1})\,|^p\,, \qquad p > 0\,, \tag{3.3}$$

where the limit is taken over all possible dissections of $[a,b]$. The function f is said to be of *bounded p-variation* if $V_p(f) < M < \infty$. In particular, if $p = 1$, then f is said to be a function of *bounded variation* over $[a,b]$. Clearly if f is of bounded variation over $[a,b]$ then f is necessarily bounded in $[a,b]$. However the converse is not true; not all functions which are bounded in $[a,b]$ have bounded variation.

Suppose that f has a finite number of maxima and minima within $[a,b]$, say $a < \xi_1 < \cdots < \xi_m < b$. Consider the dissection formed from the endpoints of the interval and the m ordered maxima and minima of f. In the interval $[\xi_{k-1}, \xi_k]$, the function f either increases or decreases. In either event, the variation of f over $[\xi_{k-1}, \xi_k]$ is $|\, f(\xi_k) - f(\xi_{k-1})\,|$ and so the variation of f over $[a,b]$ is

$$\sum_{k=0}^{m} |\, f(\xi_{k+1}) - f(\xi_k)\,| < \infty\,.$$

Let $\mathcal{D}_n$ be any dissection of $[a,b]$ and suppose that $x_{k-1} \le \xi_j \le x_k$, then the triangle inequality

$$\begin{aligned} |\, f(x_k) - f(x_{k-1})\,| &= |\,[\, f(x_k) - f(\xi_j)\,] + [\, f(\xi_j) - f(x_{k-1})\,]\,| \\ &\le |\, f(x_k) - f(\xi_j)\,| + |\, f(\xi_j) - f(x_{k-1})\,| \end{aligned}$$

ensures that

$$V(f) = \lim_{n\to\infty} \sum_{k=1}^{n} |\, f(x_k) - f(x_{k-1})\,| \le \sum_{k=0}^{m} |\, f(\xi_{k+1}) - f(\xi_k)\,| < \infty\,. \tag{3.4}$$

Thus f is a function of bounded variation. Therefore to find functions which are bounded but are not of bounded variation, it is necessary to consider functions which have at least a countable family of maxima and minima. Consider, for example,

$$f(x) = \begin{cases} \sin(\pi/x) & x > 0 \\ 0 & x = 0 \end{cases}$$

and let $\mathcal{D}_n$ be the dissection with nodes $x_0 = 0$, $x_k = 2/(2n-2k+1)$ when $1 \le k \le n-1$ and with $x_n = 1$. Clearly $f(x_0) = f(x_n) = 0$ while $f(x_k) = \sin((n-k)\pi + \pi/2) = (-1)^{n-k}$. The variation of f over $\mathcal{D}_n$ is therefore

$$\begin{aligned} V_n(f) &= \sum_{k=1}^{n} |\, f(x_k) - f(x_{k-1})\,| \\ &= |\, f(x_0) - f(x_1)\,| + \cdots |\, f(x_k) - f(x_{k-1})\,| + \cdots |\, f(x_n) - f(x_{n-1})\,| \\ &= 1 + 2 \cdots (n-2) \text{ times} \cdots + 2 + 1 \\ &= 2(n-1) \end{aligned}$$

Thus f bounded in $[0,1]$ but is not a function of bounded variation in $[0,1]$.

3.2 Riemann integration

Let $\mathcal{D}_n$ denote the dissection $a = x_0 < x_1 < \cdots < x_n = b$ of the finite interval $[a, b]$. A function f is *Riemann integrable* over $[a, b]$ whenever

$$S(f) = \lim_{n\to\infty} \sum_{k=1}^{n} f(\xi_k)\,(x_k - x_{k-1})\,, \qquad \xi_k \in [x_{k-1}, x_k] \tag{3.5}$$

exists and takes a value which is independent of the choice of ξ_k. The value of this limit is called the *Riemann integral* of f over $[a, b]$ and has symbolic representation

$$\int_a^b f(x)\,dx = S(f)\,.$$

The dissection $\mathcal{D}_n$ and the interpretation of the summation on the right hand side of equation (3.5) are illustrated in Figure 3.1.

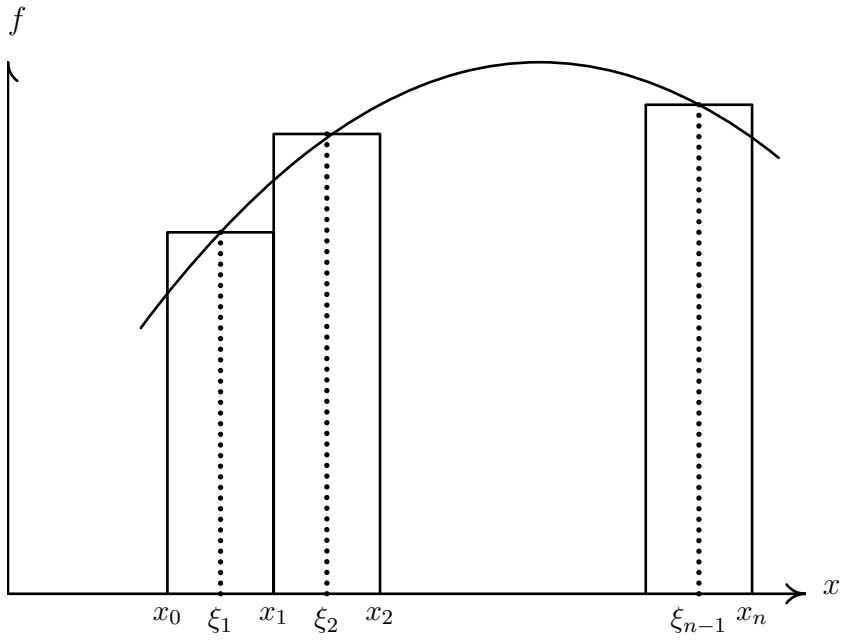

Figure 3.1: Illustration of a Riemann sum.

The main result is that for f to be Riemann integrable over $[a, b]$, it is sufficient that f is a function of bounded variation over $[a, b]$. There are two issues to address.

1. The existence of the limit (3.5) must be established;

2. It must be verified that the value of the limit is independent of the choice of ξ_k. This independence is a crucial property of Riemann integration not enjoyed in the integration of stochastic functions.

Independence of limit

Let $\eta_k^{(\mathrm{L})}$ and $\eta_k^{(\mathrm{U})}$ be respectively the values of $x \in [x_{k-1}, x_k]$ at which f attains its minimum and maximum values. Thus $f(\eta_k^{(\mathrm{L})}) \le f(\xi_k) \le f(\eta_k^{(\mathrm{U})})$. Let $S^{(\mathrm{L})}(f)$ and $S^{(\mathrm{U})}(f)$ be the values of the limit (3.5) in which ξ_k takes the respective values $\eta_k^{(\mathrm{L})}$ and $\eta_k^{(\mathrm{U})}$ then

$$S_n^{(\mathrm{L})}(f) = \sum_{k=1}^{n} f(\eta_k^{(\mathrm{L})})(x_k - x_{k-1}) \le \sum_{k=1}^{n} f(\xi_k)(x_k - x_{k-1}) \le \sum_{k=1}^{n} f(\eta_k^{(\mathrm{U})})(x_k - x_{k-1}) = S_n^{(\mathrm{U})}(f)\,.$$

Assuming the existence of limit (3.5) for all choices of ξ, it therefore follow that

$$S^{(\mathrm{L})}(f) \le S(f) \le S^{(\mathrm{U})}(f) \tag{3.6}$$

where $S(f)$ is the value of the limit (3.5) in which the sequence of ξ values are arbitrary. In particular,

$$\begin{aligned} |\, S_n^{(\mathrm{U})}(f) - S_n^{(\mathrm{L})}(f)\,| &= \Big|\, \sum_{k=1}^{n} f(\eta_k^{(\mathrm{U})})(x_k - x_{k-1}) - \sum_{k=1}^{n} f(\eta_k^{(\mathrm{L})})(x_k - x_{k-1}) \,\Big| \\ &= \Big|\, \sum_{k=1}^{n} \Big(f(\eta_k^{(\mathrm{U})}) - f(\eta_k^{(\mathrm{L})}) \Big)(x_k - x_{k-1}) \,\Big| \\ &\le \sum_{k=1}^{n} |\, f(\eta_k^{(\mathrm{U})}) - f(\eta_k^{(\mathrm{L})})|\,(x_k - x_{k-1}) \end{aligned}$$

Suppose now that Δ_n is the length of the largest sub-interval of $\mathcal{D}_n$ then it follows that

$$|\, S_n^{(\mathrm{U})}(f) - S_n^{(\mathrm{L})}(f)\,| \le \Delta_n \sum_{k=1}^{n} |\, f(\eta_k^{(\mathrm{U})}) - f(\eta_k^{(\mathrm{L})})| \le \Delta_n\, V(f)\,. \tag{3.7}$$

Since f is a function of bounded variation over $[a, b]$ then $V(f) < M < \infty$. On taking the limit of equation (3.7) as $n \to \infty$ and using the fact that $\Delta_n \to 0$, it is seen that $S^{(\mathrm{U})}(f) = S^{(\mathrm{L})}(f)$.

3.3 Riemann-Stieltjes integration

Riemann-Stieltjes integration is similar to Riemann integration except that summation is now performed with respect to g, a function of x, rather than x itself. For example, if F is the cumulative distribution function of the random variable X, then one definition for the expected value of X is the Riemann-Stieltjes integral

$$\mathrm{E}\,[\,X\,] = \int_a^b x\, dF \tag{3.8}$$

in which x is integrated with respect to F.

Let $\mathcal{D}_n$ denote the dissection $a = x_0 < x_1 < \cdots < x_n = b$ of the finite interval $[a, b]$. The *Riemann-Stieltjes* integral of f with respect to g over $[a, b]$ is defined to be the value of the limit

$$\lim_{n\to\infty} \sum_{k=1}^{n} f(\xi_k)\,[\,g(x_k) - g(x_{k-1})\,]\,, \qquad \xi_k \in [x_{k-1}, x_k] \tag{3.9}$$

when this limit exists and takes a value which is independent of the choice of ξ_k and the limiting properties of the dissection $\mathcal{D}_n$. Young has shown that the conditions for the existence of the Riemann-Stieltjes integral are met provided the points of discontinuity of f and g in $[a, b]$ are disjoint, and that f is of p-variation and g is of q variation where $p > 0$, $q > 0$ and $p^{-1} + q^{-1} > 1$.

Note that neither f nor g need be continuous functions, and of course, when g is a differentiable function of x then the Riemann-Stieltjes of f with respect to g may be replaced by the Riemann integral of fg' provided fg' is a function of bounded variation on $[a, b]$. So in writing down the prototypical Riemann-Stieltjes integral

$$\int f\, dg$$

it is usually assumed that g is not a differentiable function of x.

3.4 Stochastic integration of deterministic functions

The discussion of stochastic differential equations will involve the treatment of integrals of type

$$\int_a^b f(t)\, dW_t \tag{3.10}$$

in which f is a deterministic function of t. To determine what conditions must be imposed on f to guarantee the existence of this Riemann-Stieltjes integral, it is necessary to determine the value of p for which W_t is of bounded p-variation. Consider the interval $[t, t + h]$ and the properties of $|W(t + h) - W(t)|^p$.

Let $X = W_{t+h} - W_t \sim \mathrm{N}(0, h)$ and let $Y = |W_{t+h} - W_t|^p = |X|^p$. The sample space of Y is $[0, \infty)$ and $f_Y(y)$, the probability density function of Y, satisfies

$$f_Y(y) = 2\, f_X \frac{dx}{dy} = \frac{2y^{1/p-1}}{p\sqrt{2\pi\, h}}\, e^{-x^2/2h} = \sqrt{\frac{2}{\pi\, h}}\, \frac{y^{1/p-1}}{p}\, e^{-y^{2/p}/2h},$$

where $x = y^{1/p}$. If $p = 2$, then

$$f_Y(y) = \sqrt{\frac{1}{2\pi\, h}}\, y^{-1/2}\, e^{-y/2h}$$

making Y a gamma deviate with scale parameter $\lambda = 2h$ and shape parameter $\alpha = 1$. Otherwise,

$$\mathbb{E}[\,Y\,] = \mu = \int_0^\infty \sqrt{\frac{2}{\pi\, h}}\, \frac{y^{1/p}}{p}\, e^{-y^{2/p}/2h} = \frac{1}{p}\sqrt{\frac{2}{\pi\, h}} \int_0^\infty y^{1/p}\, e^{-y^{2/p}/2h} = \frac{(2h)^{p/2}}{\sqrt{\pi}}\, \Gamma\Big(\frac{p+1}{2}\Big),$$

and the variance of Y is

$$\begin{aligned}
\mathbb{V}[\,Y\,] &= \int_0^\infty \sqrt{\frac{2}{\pi\, h}}\, \frac{y^{1/p+1}}{p}\, e^{-y^{2/p}/2h} - \mu^2 \\
&= \frac{1}{p}\sqrt{\frac{2}{\pi\, h}} \int_0^\infty y^{1/p+1}\, e^{-y^{2/p}/2h} - \mu^2 \\
&= \frac{(2h)^p}{\sqrt{\pi}}\, \Gamma\Big(p + \frac{1}{2}\Big) - \mu^2 = \frac{(2h)^p}{\sqrt{\pi}} \Big[\, \Gamma\Big(p + \frac{1}{2}\Big) - \frac{1}{\sqrt{\pi}}\, \Gamma^2\Big(\frac{p+1}{2}\Big)\Big].
\end{aligned}$$

The process of forming the p-variation of the Wiener process W_t requires the summation of objects such as $|W_{t+h} - W_t|^p$ in which the values of h are simply the lengths of the intervals of the dissection, *i.e.* over the interval $[0, T]$ the p-variation of W_t requires the investigation of the properties of

$$V_n^{(p)}(W) = \sum_{k=1}^{n} |W(t_{k+1}) - W(t_k)|^p, \qquad t_{k+1} - t_k = h_k \qquad h_1 + h_2 + \cdots + h_n = T$$

and for W_t to be of p-variation on $[0, T]$, the limit of $V_n^{(p)}(W)$ must be finite for all dissections as $n \to \infty$. Consider however the uniform dissection in which $h_1 = \cdots = h_n = T/n$. In this case

$$\mathbb{E}[\, V_n^{(p)}(W)] = \sum_{k=1}^{n} \mathbb{E}[|W(t_{k+1}) - W(t_k)|^p] = \frac{(2T)^{p/2}}{\sqrt{\pi}} \Gamma\Big(\frac{p+1}{2} \Big) n^{1-p/2},$$

If $p < 2$, the expected value $\mathbb{E}[\, V_n^{(p)}(W)]$ is unbounded and so W_t cannot be of p-variation when $p < 2$. When $p = 2$ it is clear that $\mathbb{E}[\, V_n^{(p)}(W)] = T$. For general dissection $\mathcal{D}_n$ the distribution of $V_n^{(2)}(W)$ is unclear but when $h_1 = \cdots = h_n = T/n$ the function $V_n^{(2)}(W)$ is chi-squared distributed being the sum of the squares of n independent and identically distributed Gaussian random deviates.

The expected value of each term in the sum required to form the p-variation of W_t always exceeds a constant multiple of $1/n^{p/2}$ and therefore the sum of these expected values diverges for all $p \leq 2$ and converges for $p > 2$. The variance of the p-variation behaves like a sum of terms with behaviour $1/n^p$, and always converges. Thus the Wiener process is of bounded p-variation for $p > 2$.

Consequently, the properties of the Riemann-Stieltjes integral (3.8) guarantee that the stochastic integral (3.10) is defined as a Riemann-Stieltjes integral for all functions f with bounded q-variation where $q < 2$. Clearly functions of bounded variation ($q = 1$) satisfy this condition, and therefore if f is Riemann integrable over $[a, b]$, then the the stochastic integral (3.10) exists and can be treated as a Riemann-Stieltjes integral.

Chapter 4

The Ito and Stratonovich Integrals

4.1 A simple stochastic differential equation

To motivate the Ito and Stratonovich integrals, consider the generic one dimensional SDE

$$dx_t = a(t, x_t)\, dt + b(t, x_t)\, dW_t \tag{4.1}$$

where dW_t is the increment of the Wiener process, and $a(t, x)$ and $b(t, x)$ are deterministic functions of x and t. Of course, a and b behave randomly in the solution of the SDE by virtue of their dependence on x. The meanings of $a(t, x)$ and $b(t, x)$ become clear from the following argument.

Since $\mathbb{E}[\, dW_t \,] = 0$, it follows directly from equation (4.1) that $\mathrm{E}[\, dx \,] = a(t, x)\, dt$. For this reason, $a(t, x)$ is often called the **drift** component of the stochastic differential equation. One usually regards the equation $dx/dt = a(t, x)$ as the genitor differential equation from which the SDE (4.1) is born. This was precisely the procedure used to obtain SDEs from ODEs in the introductory chapter. Furthermore,

$$\mathbb{V}(dx) = \mathrm{E}[\, (dx - a(t, x)\, dt)^2 \,] = \mathbb{E}[\, b(t, x)\, dW_t^2 \,] = b(t, x)\, \mathbb{E}[\, dW_t^2 \,] = b^2(t, x)\, dt\,. \tag{4.2}$$

Thus $b^2(t, x)\, dt$ is the variance of the stochastic process dx and therefore $b^2(t, x)$ is the rate at which the variance of the stochastic process grows in time. The function $b^2(t, x)$ is often called the **diffusion** of the SDE and $b(t, x)$, being dimensionally a standard deviation, is usually called the **volatility** of the SDE. In effect, an ODE is an SDE with no volatility.

Note, however, that the definition of the terms $a(t, x)$ and $b(t, x)$ should not be confused with the idea that the solution to an SDE simply behaves like the underlying ODE $dx/dt = a(t, x)$ with superimposed noise of volatility $b(x, t)$ per unit time. In fact, $b(x, t)$ contributes to both drift and variance. Therefore, one **cannot** estimate the parameters of an SDE by treating the drift (deterministic behaviour of the solution) and volatility (local variation of solution) as separate processes.

4.1.1 Motivation for Ito integral

It has been noted previously that the solution to equation (4.1) has formal expression

$$x(t) = x(s) + \int_s^t a(r, x_r)\, dr + \int_s^t b(r, x_r)\, dW_r\,, \qquad t \geq s\,, \tag{4.3}$$

where the task is now to state how each integral on the right hand side of equation (4.3) is to be computed. Consider the computation of the integrals in (4.3) over the interval $[t, t+h]$ where h is small. Assuming that x_t is a continuous random function of t and that $a(t,x)$ and $b(t,x)$ are continuous functions of (t,x), then $a(r,x_r)$ and $b(r,x_r)$ in (4.3) may be approximated by $a(t,x_t)$ and $b(t,x_t)$ respectively to give

$$\begin{aligned} x(t+h) &\approx x(t) + a(t,x_t) \int_t^{t+h} dr + b(t,x_t) \int_t^{t+h} dW_r\,, \\ &= x(t) + a(t,x_t)\, h + b(t,x_t)\, [\, W(t+h) - W(t)\,]\,. \end{aligned} \tag{4.4}$$

Note, in particular, that $\mathbb{E}\,[\, b(t,x_t)\,(W(t+h) - W(t)\,] = 0$ and therefore the stochastic contribution to the solution (4.4) is a martingale[1], that is, it is non-anticipative. This solution intuitively agrees with ones conceptual picture of the future, namely that ones best estimate of the future is the current state plus a drift which, of course, is deterministic.

This simple approximation procedure (which will subsequently be identified as the Euler-Maruyama approximation) is consistent with the view that the first integral on the right hand side of (4.3) is a standard Riemann integral. On the other hand, the second integral on the right hand side of (4.3) is a *stochastic integral* with the martingale property, that is, it is a non-anticipative random variable with expected value zero - the value when $h = 0$. Let $s = t_0 < t_1 < \cdots < t_n = t$ be a dissection of $[s,t]$. The extension of the approximation procedure used in the derivation of (4.4) to the interval $[s,t]$ suggests the definition

$$\int_s^t b(r, x_r)\, dW_r = \lim_{n\to\infty} \sum_{k=0}^{n-1} b(t_k, x(t_k))\, [\, W(t_{k+1}) - W(t_k)\,]\,. \tag{4.5}$$

This definition automatically endows the stochastic integral with the martingale property - the reason is that $b(t_k, x_k)$ depends on the behaviour of the Wiener process in (s, t_k) alone and so the value of $b(t_k, x_k)$ is independent of $W(t_{k+1}) - W(t_k)$.

In fact, the computation of the stochastic integral in equation (4.5) is based on the mean square

[1] A *martingale* refers originally to a betting strategy popular in 18th century France in which the expected winnings of the strategy is the original stake. The simplest example of a martingale is the double or quit strategy in which a bet is doubled until a win is achieved. In this strategy the total loss after n unsuccessful bets is $(2^n - 1)S$ where S is the original stake. The bet at the $(n+1)$-th play is $2^n\, S$ so that the first win necessarily recovers all previous losses leaving a profit of S. Since a gambler with unlimited wealth eventually wins, devotees of the martingale betting strategy regarded it as a sure-fire winner. In reality, however, the exponential growth in the size of bets means that gamblers rapidly become bankrupt or are prevented from executing the strategy by the imposition of a maximum bet.

limiting process. More precisely, the *Ito Integral* of $b(t, x_t)$ over $[a, b]$ is defined by

$$\lim_{n\to\infty} \mathrm{E}\Big[\sum_{k=1}^{n} b(t_{k-1}\,, x_{k-1})\,(\,W_k - W_{k-1}\,) - \int_a^b b(t, x_t)\,dW_t\,\Big]^2 = 0\,, \tag{4.6}$$

where $x_k = x(t_k)$ and $W_k = W(t_k)$. The limiting process in (4.6) defines the *Ito integral* of $b(t, x_t)$ (assuming that x_t is known), and the computation of integrals defined in this way is called *Ito Integration*. Most importantly, the Ito integral is a martingale, and stochastic differential equations of the type (4.1) are more correctly called Ito stochastic differential equations.

In particular, the rules for Ito integration can be different from the rules for Riemann/Riemann-Stieltjes integration. For example,

$$\int_s^t dW_r = W(t) - W(s)\,, \qquad \text{but} \qquad \int_s^t W_r\,dW_r \neq \frac{W^2(t) - W^2(s)}{2}\,.$$

To appreciate why the latter result is false, simply note that

$$\mathbb{E}\Big[\int_s^t W_r\,dW_r\Big] = 0 \neq \mathbb{E}\Big[\frac{W^2(t) - W^2(s)}{2}\Big] = \frac{t-s}{2} > 0\,.$$

Clearly the Riemann-Stieltjes integral is a sub-martingale, but what has gone wrong with the Riemann-Stieltjes integral in the previous illustration is the crucial question. Suppose one considers

$$\int_s^t W_r\,dW_r$$

within the framework of Riemann-Stieltjes integration with $f(t) = W(t)$ and $g(t) = W(t)$. The key observation is that $W(t)$ is of p-variation with $p > 2$ and therefore $p^{-1} + q^{-1} < 1$ in this case. Although this observation is not helpful in evaluating this integral (if indeed it has a value), the fact that inequality $p^{-1} + q^{-1} > 1$ fails indicates that the integral cannot be interpreted as a Riemann-Stieltjes integral. Clearly the rules of Ito-Calculus are quite different from those of Liebnitz's Calculus.

4.2 Stochastic integrals

The findings of the previous section suggest that special consideration must be given to the computation of the generic integral

$$\int_a^b f(t, W_t)\,dW_t\,. \tag{4.7}$$

While superficially taking the form of a Riemann-Stieltjes, both components of the integrand have p-variation and q-variation such that $p > 2$ and $q > 2$ and therefore $p^{-1} + q^{-1} < 1$. Nevertheless, expression (4.7) may be assigned a meaning in the usual way. Let $\mathcal{D}_n$ denote the dissection $a = t_0 < t_1 < \cdots < t_n = b$ of the finite interval $[a, b]$. Formally,

$$\int_a^b f(t, W_t)\,dW_t = \lim_{n\to\infty} \sum_{k=1}^{n} f(\xi_k, W(\xi_k))\,\Big(\,W(t_k) - W(t_{k-1})\,\Big)\,, \qquad \xi_k \in [t_{k-1}, t_k] \tag{4.8}$$

whenever this limit exist. Of course, if $\xi_k \in (t_{k-1}, t_k)$ there in general the integral cannot be a martingale because $f(\xi_k, W(\xi_k))$ encapsulates the behaviour of W_t in the interval $[t_{k-1}, \xi_k]$, and therefore the random variable $f(\xi_k, W(\xi_k))$ is not in general independent of $W(t_k) - W(t_{k-1})$. Thus

$$\mathbb{E}\,[\, f(\xi_k, W(\xi_k))\,(\,W(t_k) - W(t_{k-1})\,)\,] \neq 0\,. \qquad t_{k-1} < \xi_k <= t_k\,.$$

Pursuing this idea to its ultimate conclusion indicates that the previous integral is a martingale if and only if $\xi_k = t_{k-1}$, because in this case $f(\xi_k, W(\xi_k))$ and $W(t_k) - W(t_{k-1})$ are independent random variables thereby making the expectation of the integral zero.

To provide an explicit example of this idea, consider the case $f(t, W_t) = W_t$, that is, compute the stochastic integral

$$I = \int_a^b W_t\, dW_t\,. \tag{4.9}$$

Let $W_k = W(t_k)$, let $\xi_k = t_{k-1} + \lambda\,(t_k - t_{k-1})$ where $\lambda \in [0,1]$ then $W(\xi_k) = W_{k-1+\lambda}$ and

$$\int_a^b W_t\, dW_t = \lim_{n\to\infty} \sum_{k=1}^{n} W(\xi_k)\,\Big(\,W(t_k) - W(t_{k-1})\,\Big) = \lim_{n\to\infty} \sum_{k=1}^{n} W_{k-1+\lambda}(\,W_k - W_{k-1})\,, \tag{4.10}$$

It is straightforward algebra to show that

$$\sum_{k=1}^{n} W_{k-1+\lambda}(\,W_k - W_{k-1}) = \sum_{k=1}^{n} \frac{1}{2}\Big[W_k^2 - W_{k-1}^2 - (W_k - W_{k-1+\lambda})^2 + (W_{k-1+\lambda} - W_{k-1})^2\Big]$$

from which it follows that

$$\begin{aligned}\sum_{k=1}^{n} W_{k+\lambda}(\,W_k - W_{k-1}) &= \frac{1}{2}\Big[W_k^2 - W_{k-1}^2 - (W_k - W_{k-1+\lambda})^2 + (W_{k-1+\lambda} - W_{k-1})^2\Big] \\ &= \frac{1}{2}\Big[W_n^2 - W_0^2\Big] + \frac{1}{2}\sum_{k=1}^{n}\Big[(W_{k-1+\lambda} - W_{k-1})^2 - (W_k - W_{k-1+\lambda})^2\Big]\,.\end{aligned}$$

Now observe that

$$\mathbb{E}\,[\,(W_k - W_{k-1+\lambda})^2] = (1-\lambda)(t_k - t_{k-1})\,, \quad \mathbb{E}\,[\,(W_{k-1+\lambda} - W_{k-1})^2] = \lambda(t_k - t_{k-1})$$

and also that $\mathbb{E}[\,(W_k - W_{k-1+\lambda})(W_{k-1+\lambda} - W_{k-1})\,]$ although property will not be needed in what follows. Evidently

$$\mathbb{E}\Big[\,\sum_{k=1}^{n} W_{k+\lambda}(\,W_k - W_{k-1})\,\Big] = \frac{b-a}{2} + \frac{1}{2}\sum_{k=1}^{n}(2\lambda - 1)(t_k - t_{k-1}) = \lambda(b-a)\,, \tag{4.11}$$

from which it follows immediately that the integral is a martingale provided $\lambda = 0$. The choice $\lambda = 0$ (left hand endpoint of dissection intervals) defines the *Ito integral*. In particular, the value of the limit defining the integral would appear to depend of the choice of ξ_k within the dissection $\mathcal{D}_n$.

The choice $\lambda = 1/2$ (midpoint of dissection intervals) is called the *Stratonovich integral* and leads to a sub-martingale interpretation of the limit (4.8).

4.3 The Ito integral

The *Ito Integral* of $f(t, W_t)$ over $[a, b]$ is defined by

$$\int_a^b f(t, W_t)\, dW_t = \lim_{n\to\infty} \sum_{k=1}^{n} f(t_{k-1}\,, W_{k-1})\,(\,W_k - W_{k-1}\,)\,, \tag{4.12}$$

where $W_k = W(t_k)$ and convergence of the limiting process is measured in the mean square sense, *i.e.* in the sense that

$$\lim_{n\to\infty} \mathrm{E}\Big[\sum_{k=1}^{n} f(t_{k-1}\,, W_{k-1})\,(\,W_k - W_{k-1}\,) - \int_a^b f(t, W_t)\, dW_t\,\Big]^2 = 0\,. \tag{4.13}$$

The direct computation of Ito integrals from the definition (4.13) is now illustrated with reference to the Ito integral (4.9). Recall that

$$S_n = \sum_{k=1}^{n} W_{k-1}\,(\,W_k - W_{k-1}\,) = \frac{1}{2}\,(W^2(b) - W^2(a)) - \frac{1}{2}\sum_{k=1}^{n}(W_k - W_{k-1})^2\,. \tag{4.14}$$

It is convenient to note that $W_k - W_{k-1} = \sqrt{t_k - t_{k-1}}\,\varepsilon_k$ where $\varepsilon_1, \cdots, \varepsilon_n$ are n uncorrelated standard normal deviates. When expressed in this format

$$S_n = \sum_{k=1}^{n} W_{k-1}\,(\,W_k - W_{k-1}\,) = \frac{1}{2}\,(W^2(b) - W^2(a)) - \frac{1}{2}\sum_{k=1}^{n}(t_k - t_{k-1})\,\varepsilon_k^2\,,$$

which may in turn be rearranged to give

$$S_n - \frac{1}{2}[W^2(b) - W^2(a) - (b-a)] = \frac{1}{2}\sum_{k=1}^{n}(t_k - t_{k-1})\,(\varepsilon_k^2 - 1)\,. \tag{4.15}$$

To finalise the calculation of the Ito Integral and establish the mean square convergence of the right hand side of (4.15) we note that

$$\begin{aligned}
&\lim_{n\to\infty} \mathbb{E}\Big[\,S_n - \frac{1}{2}[W^2(b) - W^2(a) - (b-a)]\,\Big]^2 \\
&\quad= \frac{1}{4}\lim_{n\to\infty} \mathbb{E}\Big[\,\Big(\sum_{k=1}^{n}(t_k - t_{k-1})\,(\varepsilon_k^2 - 1)\Big)^2\,\Big] \\
&\quad= \frac{1}{4}\lim_{n\to\infty} \mathbb{E}\Big[\,\sum_{j,k=1}^{n}(t_j - t_{j-1})(t_k - t_{k-1})\,(\varepsilon_k^2 - 1)(\varepsilon_j^2 - 1)\,\Big] \\
&\quad= \frac{1}{4}\lim_{n\to\infty} \sum_{k=1}^{n}(t_k - t_{k-1})^2\ \mathbb{E}\,[\,(\varepsilon_k^2 - 1)^2\,]
\end{aligned} \tag{4.16}$$

where the last line reflects the fact that $,(\varepsilon_k^2 - 1)$ and $(\varepsilon_j^2 - 1)$ are independent zero-mean random deviates for $j \neq k$. Clearly

$$\mathbb{E}\,[\,(\varepsilon_k^2 - 1)^2\,] = \mathbb{E}\,[\,\varepsilon_k^4\,] - 2\mathbb{E}\,[\,\varepsilon_k^2\,] + 1 = 3 - 2 + 1 = 2\,.$$

In conclusion,

$$\lim_{n\to\infty} \mathbb{E}\Big[S_n - \frac{1}{2}[W^2(b) - W^2(a) - (b-a)]\Big]^2 = \frac{1}{2}\lim_{n\to\infty}\sum_{k=1}^{n}(t_k - t_{k-1})^2 = 0\,. \tag{4.17}$$

The mean square convergence of the Ito integral is now established and so

$$\int_a^b W_s\, dW_s = \frac{1}{2}\Big[W^2(b) - W^2(a) - (b-a)\Big]\,. \tag{4.18}$$

As has already been indicated, the Ito's definition of a stochastic integral does not obey the rules of classical Calculus - the term $(b-a)/2$ would be absent in classical Calculus. The important advantage enjoyed by the Ito stochastic integral is that it is a martingale.

The correlation formula for Ito integration

Another useful property of Ito integration is the **correlation formula** which states that if $f(t, W_t)$ and $g(t, W_t)$ are two stochastic functions then

$$\mathbb{E}\Big[\int_a^b f(t, W_t)\, dW_t \int_a^b g(s, W_s)\, dW_s \Big] = \int_a^b \mathbb{E}\,[\, f(t, W_t)\, g(t, W_t)\,]\, dt\,,$$

and in the special case $f = g$, this result reduces to

$$\mathbb{E}\Big[\int_a^b f(t, W_t)\, dW_t \int_a^b f(s, W_s)\, dW_s \Big] = \int_a^b \mathbb{E}\,[\, f^2(t, W_t)\,]\, dt\,.$$

This is a useful result in practice as we shall see later.

Proof The justification of the correlation formula begins by considering the partial sums

$$S_n^{(f)} = \sum_{k=1}^{n} f(t_{k-1}\,, W_{k-1})\,(\,W_k - W_{k-1}\,)\,, \qquad S_n^{(g)} = \sum_{k=1}^{n} g(t_{k-1}\,, W_{k-1})\,(\,W_k - W_{k-1}\,)\,, \tag{4.19}$$

from which the Ito integral is defined. It follows that

$$\begin{aligned} &\mathbb{E}\Big[\int_a^b f(t, W_t)\, dW \int_a^b g(s, W_s)\, dW_s \Big] \\ &\quad = \lim_{n\to\infty} \sum_{j,k=1}^{n} \mathbb{E}\,[\, f(t_{k-1}\,, W_{k-1})\, g(t_{j-1}\,, W_{j-1})(\,W_k - W_{k-1}\,)\,(\,W_j - W_{j-1}\,)\,] \end{aligned} \tag{4.20}$$

The products $f(t_{k-1}\,, W_{k-1})g(t_{j-1}\,, W_{j-1})$ and $(W_k - W_{k-1})(W_j - W_{j-1})$ are uncorrelated when $j \neq k$, and so the contributions to the double sum in equation (4.20) arise solely from the case $k = j$ to get

$$\mathbb{E}\Big[\int_a^b f(t, W_t)\, dW_t \int_a^b g(t, W_t)\, dW_t \Big] = \lim_{n\to\infty} \sum_{k=1}^{n} \mathbb{E}\Big[f(t_{k-1}\,, W_{k-1})\, g(t_{k-1}\,, W_{k-1})(\,W_k - W_{k-1}\,)^2 \Big]\,.$$

From the fact that $f(t_{k-1}\,, W_{k-1})g(t_{j-1}\,, W_{k-1})$ is uncorrelated with $(W_k - W_{k-1})^2$, it follows that

$$\mathbb{E}\Big[\, f(t_{k-1}\,, W_{k-1})\, g(t_{k-1}\,, W_{k-1})(\, W_k - W_{k-1}\,)^2 \Big] = \mathbb{E}\,[\, f(t_{k-1}\, g(t_{k-1}\,, W_{k-1})\,]\, \mathbb{E}\,[\,(\, W_k - W_{k-1}\,)^2\,]$$

After replacing $\mathbb{E}\,[\,(\, W_k - W_{k-1}\,)^2\,]$ by $(t_k - t_{k-1})$, it is clear that

$$\mathbb{E}\Big[\, \int_a^b f(t, W_t)\, dW_t \int_a^b g(t, W_t)\, dW_t \Big] = \lim_{n\to\infty} \sum_{k=1}^{n} \mathrm{E}\,[\, f(t_{k-1}\,, W_{k-1})\, g(t_{k-1}\,, W_{k-1})\,](t_k - t_{k-1})\,.$$

The right hand side of the previous equation is by definition the limiting process corresponding to the Riemann integral of $\mathrm{E}\,[\, f(t_{k-1}\,, W_{k-1})\, g(t_{k-1}\,, W_{k-1})(t_k - t_{k-1})\,]$ over the interval $[a, b]$, which in turn establishes the stated result that

$$\mathbb{E}\Big[\, \int_a^b f(t, W_t)\, dW_t \int_a^b g(t, W_t)\, dW_t \Big] = \int_a^b \mathrm{E}\,[\, f(t\,, W_t)\, g(t\,, W_t)\,]\, dt\,. \tag{4.21}$$

4.3.1 Application

A frequently occurring application of the correlation formula arises when the Wiener process $W(t)$ is absent from f, *i.e.* the task is to assign meaning to the integral

$$\xi = \int_a^b f(t)\, dW_t\,. \tag{4.22}$$

in which $f(t)$ is a function of bounded variation. The integral, although clearly interpretable as an integral of Riemann-Stieltjes type, is nevertheless a random variable with $\mathbb{E}[\,\xi\,] = 0$ because $\mathbb{E}[\, dW_t\,] = 0$. Furthermore, by recognising that the value of the integral is the limit of a weighted sum of independent Gaussian random variables, it is clear that ξ is simply a Gaussian deviate with zero mean value. To complete the specification of ξ it therefore remains to compute the variance of the integral in equation (4.22). The correlation formula is useful in this respect and indicates immediately that

$$\mathbb{V}[\,\xi\,] = \int_a^b f^2(t)\, dt \tag{4.23}$$

thereby completing the specification of ξ.

4.4 The Stratonovich integral

The fact that Ito integration does not conform to the traditional rules of Riemann and Riemann-Stieltjes integration makes Ito integration an awkward procedure. One way to circumvent this awkwardness is to introduce the Stratonovich integral which can be demonstrated to obey the traditional rules of integration under quite relaxed condition but, of course, the martingale property of the Ito integral must be sacrificed in the process. The Stratonovich integral is a sub-martingale. The overall strategy is that Ito integrals and Stratonovich integrals are related, but that the latter frequently conforms to the rules of traditional integration in a way to be made precise at a later time.

Given a suitably differentiable function $f(t, W_t)$, the Stratonovich integral of f with respect to the Wiener process $W(t)$ (denoted by "$\circ$") is defined by

$$\int_a^b f(t, W_t) \circ dW_t = \lim_{n\to\infty} \sum_{k=1}^{n} f(\xi_k, W(\xi_k))\,(\,W_k - W_{k-1}\,)\,, \qquad \xi_k = \frac{t_{k-1}+t_k}{2}\,, \tag{4.24}$$

in which the function f is sampled at the midpoints of the intervals of a dissection, and where the limiting procedure (as with the Ito integral) is to be interpreted in the mean square sense, *i.e.* the value of the integral requires that

$$\lim_{n\to\infty} \mathbb{E}\Big[\sum_{k=1}^{n} f(\xi_k\,, W(\xi_k)\,)\,(\,W_k - W_{k-1}\,) - \int_a^b f(t, W_t) \circ dW_t\,\Big]^2 = 0\,. \tag{4.25}$$

To appreciate how Stratonovich integration might differ from Ito integration, it is useful to compute

$$\int_a^b W_t \circ dW_t\,.$$

The calculation mimics the procedure used for the Ito integral and begins with the Riemann-Stieltjes partial sum

$$S_n = \sum_{k=1}^{n} W_{k-1/2}\,(\,W_k - W_{k-1}\,) \tag{4.26}$$

where the notation $W_{k-1+\lambda} = W(t_{k-1} + \lambda(t_k - t_{k-1})\,)$ has been used for convenience. This sum in now manipulated into the equivalent algebraic form

$$\frac{1}{2}\Big[\,\sum_{k=1}^{n}(W_k^2 - W_{k-1}^2) + \sum_{k=1}^{n}(W_{k-1/2} - W_{k-1})^2 - \sum_{k=1}^{n}(W_k - W_{k-1/2})^2\,\Big]\,,$$

where we note that $(W_{k-1/2} - W_{k-1})$ and $(W_k - W_{k-1/2})$ are uncorrelated Gaussian deviates with mean zero and variance $(t_k - t_{k-1})/2$. Consequently we may write $(W_{k-1/2} - W_{k-1}) = \sqrt{(t_k - t_{k-1})/2}\;\varepsilon_k$ and $(W_k - W_{k-1/2}) = \sqrt{(t_k - t_{k-1})/2}\;\eta_k$ in which ε_k and η_k are uncorrelated standard normal deviates for each value of k. Thus the partial sum underlying the definition of the Stratonovich integral is

$$S_n = \frac{1}{2}\Big[\,W^2(b) - W^2(a) + \frac{1}{2}\sum_{k=1}^{n}(t_k - t_{k-1})\,(\varepsilon_k^2 - \eta_k^2)\,\Big]\,. \tag{4.27}$$

The argument used to establish the mean square convergence of the Ito integral may be repeated again, and in this instance the argument will require the consideration of

$$\begin{aligned}
\lim_{n\to\infty} \mathbb{E}\Big[S_n - \frac{W^2(b) - W^2(a)}{2}\,\Big]^2 &= \lim_{n\to\infty} \mathbb{E}\Big[\,\frac{1}{4}\sum_{k=1}^{n}(t_k - t_{k-1})\,(\varepsilon_k^2 - \eta_k^2)\,\Big]^2 \\
&= \lim_{n\to\infty} \mathbb{E}\Big[\,\frac{1}{4}\sum_{k=1}^{n}(t_k - t_{k-1})(t_j - t_{j-1})\,(\varepsilon_k^2 - \eta_k^2)(\varepsilon_j^2 - \eta_j^2)\,\Big] \\
&= \lim_{n\to\infty} \frac{1}{4}\sum_{k=1}^{n}(t_k - t_{k-1})^2\,\mathbb{E}\,[\,(\varepsilon_k^2 - \eta_k^2)^2\,]\,,
\end{aligned}$$

where the last line of the previous calculation has recognised the fact that ε_k and η_k are uncorrelated deviates for distinct values of k. Furthermore, the independence of ε_k and η_k for each value of k indicates that

$$\mathbb{E}\,[\,(\varepsilon_k^2 - \eta_k^2)^2\,] = \mathbb{E}\,[\,\varepsilon_k^4\,] - 2\mathbb{E}\,[\,\varepsilon_k^2\eta_k^2\,] + \mathbb{E}\,[\,\eta_k^4\,] = 3 - 2 \times 1 \times 1 + 3 = 4\,.$$

In conclusion,

$$\lim_{n\to\infty}\ \mathbb{E}\left[S_n - \frac{(W^2(b) - W^2(a))}{2}\,\right]^2 = \lim_{n\to\infty}\ \sum_{k=1}^{n}(t_k - t_{k-1})^2 = 0\,, \tag{4.28}$$

thereby establishing the result

$$I = \int_a^b\ W_t\ \circ\ dW_t = \frac{W^2(b) - W^2(a)}{2}\,. \tag{4.29}$$

This example suggests several properties of the Stratonovich integral:-

1. Unlike the Ito integral, the Stratonovich integral is in general anticipatory - in the previous example $\mathbb{E}\,[\,I\,] = (b-a)/2 > 0$ and so I is a sub-martingale;

2. On the basis of this example it would seem that the Stratonovich integral conforms to the traditional rules of Integral Calculus;

3. There is a suggestion that the Stratonovich integral differs from the Ito integral through a "mean value" which in this example is interpretable as a drift.

4.4.1 Relationship between the Ito and Stratonovich integrals

The connection between the Ito and Stratonovich integrals suggested by the illustrative example, namely that the value of the Stratonovich integral is the sum of the values of the Ito integral and a deterministic drift term, is correct for the class of functions $f(t, W_t)$ satisfying the reasonable conditions

$$\int_a^b\ \mathbb{E}\,[\,f(t,W_t)\,]^2\,dt < \infty\,, \qquad \int_a^b\ \mathbb{E}\left[\,\frac{\partial f(t,W_t)}{\partial W_t}\,\right]^2 dt < \infty\,. \tag{4.30}$$

The primary result is that the Ito integral of $f(t, W_t)$ and the Stratonovich integral of $f(t, W_t)$ are connected by the identity

$$\int_a^b\ f(t,W_t)\ \circ\ dW_t = \int_a^b\ f(t,W_t)\,dW_t + \frac{1}{2}\int_a^b\ \frac{\partial f(t\,,W_t)}{\partial W_t}\,dt\,. \tag{4.31}$$

Of course, the first of conditions (4.30) is also necessary for the convergence of the Ito and Stratonovich integrals, and so it is only the second condition which is new. Its role is to ensure the existence of the second integral on the right hand side of (4.31). The derivation of identity (4.31) is based on the idea that $f(t, W_t)$ can be expanded locally as a Taylor series. The strategy of the proof focusses

on the construction of an identity connecting the partial sums from which the Stratonovich and Ito integrals take their values. Thus

$$\int_a^b f(t,W_t)\circ dW_t = \lim_{n\to\infty} S_n^{\text{Stratonovich}}, \qquad \int_a^b f(t,W_t)\,dW_t = \lim_{n\to\infty} S_n^{\text{Ito}},$$

where

$$S_n^{\text{Stratonovich}} = \sum_{k=1}^{n} f(t_{k-1/2},W_{k-1/2})\,(W_k - W_{k-1}), \qquad S_n^{\text{Ito}} = \sum_{k=1}^{n} f(t_{k-1},W_{k-1})\,(W_k - W_{k-1}).$$

The difference $S_n^{\text{Stratonovich}} - S_n^{\text{Ito}}$ is first constructed, and the expression then simplified by replacing $f(t_{k-1/2},W_{k-1/2})$ by a two-dimensional Taylor series about $t = t_{k-1}$ and $W = W_{k-1}$. The steps in this calculation are as follows.

$$\begin{aligned}
S_n^{\text{Stratonovich}} - S_n^{\text{Ito}} &= \sum_{k=1}^{n} \Big[f(t_{k-1/2},W_{k-1/2}) - f(t_{k-1},W_{k-1})\Big](W_k - W_{k-1}) \\
&= \sum_{k=1}^{n} \Big[\, f(t_{k-1},W_{k-1}) + (t_{k-1/2}-t_{k-1})\,\frac{\partial f(t_{k-1}\,W_{k-1})}{\partial t} \\
&\qquad + (W_{k-1/2}-W_{k-1})\,\frac{\partial f(t_{k-1}\,W_{k-1})}{\partial W} + \cdots - f(t_{k-1},W_{k-1})\Big](W_k - W_{k-1}) \\
&= \sum_{k=1}^{n} (t_{k-1/2}-t_{k-1})\,\frac{\partial\, f(t_{k-1}\,W_{k-1})}{\partial t}\,(W_k - W_{k-1}) \\
&\qquad + \sum_{k=1}^{n} (W_{k-1/2}-W_{k-1})\,\frac{\partial f(t_{k-1}\,W_{k-1})}{\partial W}(\,W_k - W_{k-1}\,) + \cdots
\end{aligned}$$

The first summation on the right hand side of this computation is $O(t_k - t_{k-1})^{3/2}$, and therefore this takes the value zero as $n\to\infty$ being of order $O(t_k - t_{k-1})^3$ in the limiting process of mean squared convergence. This simplification of the previous equation yields

$$S_n^{\text{Stratonovich}} - S_n^{\text{Ito}} = \sum_{k=1}^{n} (W_{k-1/2}-W_{k-1})\,\frac{\partial f(t_{k-1}\,W_{k-1})}{\partial W}(\,W_k - W_{k-1}\,) + \cdots$$

which is now restructured using the identity $(\,W_k - W_{k-1}\,) = (\,W_k - W_{k-1/2}\,) + (\,W_{k-1/2} - W_{k-1}\,)$ to give

$$\begin{aligned}
S_n^{\text{Stratonovich}} &= S_n^{\text{Ito}} + \frac{1}{2}\sum_{k=1}^{n} \frac{\partial f(t_{k-1},W_{k-1})}{\partial W}(t_k - t_{k-1}) \\
&\qquad + \sum_{k=1}^{n} \frac{\partial f(t_{k-1}\,W_{k-1})}{\partial W}\Big[(W_{k-1/2}-W_{k-1})(W_k - W_{k-1}) - \frac{t_k - t_{k-1}}{2}\Big].
\end{aligned} \tag{4.32}$$

Taking account of the fact that $\mathrm{E}\,[\,(W_{k-1/2}-W_{k-1})\,(W_k - W_{k-1}) - (t_k - t_{k-1})/2\,] = 0$, the second of conditions (4.30) now guarantees mean square convergence, and the identity

$$\int_a^b f(t,W_t)\,o\,dW_t = \int_a^b f(t,W_t)\,dW_t + \frac{1}{2}\int_a^b \frac{\partial f(t\,,W_t)}{\partial W}\,dt \tag{4.33}$$

connecting the Ito and Stratonovich integrals for a large range of functions is recovered. Of course, the second integral is to be interpreted as a Riemann integral.

4.4.2 Stratonovich integration conforms to the classical rules of integration

The efficacy of the Stratonovich approach to Ito integration is based on the fact that the value of a large class of Stratonovich integrals may be computed using the rules of standard integral calculus. To be specific, suppose $g(t, W_t)$ is a deterministic function of t and W_t then the main result is that

$$\int_a^b \frac{\partial g(t, W_t)}{\partial W_t} \circ dW_t = g(b, W(b)\,) - g(a, W(a)\,) - \int_a^b \frac{\partial g}{\partial t}\, dt\,. \tag{4.34}$$

In particular, if $g = g(W_t)$ (g does not make explicit reference to t) then identity (4.34) becomes

$$\int_a^b \frac{\partial g(t, W_t)}{\partial W_t} \circ dW_t = g(b, W(b)\,) - g(a, W(a)\,)\,. \tag{4.35}$$

Thus the Stratonovich integral of a function of a Wiener process conforms to the rules of standard integral calculus. Identity (4.34) is now justified.

Proof

Let $\mathcal{D}_n$ denote the dissection $a = t_0 < t_1 < \cdots < t_n = b$ of the finite interval $[a, b]$. The Taylor series expansion of $g(t, W_t)$ about $t = t_{k-1/2}$ gives

$$\begin{aligned}
g(t_k\,, W_k) &= g(t_{k-1/2}\,, W_{k-1/2}) + \frac{\partial g(t_{k-1/2}\,, W_{k-1/2})}{\partial t}\,(t_k - t_{k-1/2}) \\
&\quad + \frac{\partial g(t_{k-1/2}\,, W_{k-1/2})}{\partial W_t}\,(W_k - W_{k-1/2}) + \cdots \\
&\quad + \frac{1}{2}\,\frac{\partial^2 g(t_{k-1/2}\,, W_{k-1/2})}{\partial W_t^2}\,(W_k - W_{k-1/2})^2 + \cdots \\
g(t_{k-1}\,, W_{k-1}) &= g(t_{k-1/2}\,, W_{k-1/2}) + \frac{\partial g(t_{k-1/2}\,, W_{k-1/2})}{\partial t}\,(t_{k-1} - t_{k-1/2}) \\
&\quad + \frac{\partial g(t_{k-1/2}\,, W_{k-1/2})}{\partial W_t}\,(W_{k-1} - W_{k-1/2}) + \cdots \\
&\quad + \frac{1}{2}\,\frac{\partial^2 g(t_{k-1/2}\,, W_{k-1/2})}{\partial W_t^2}\,(W_{k-1} - W_{k-1/2})^2 + \cdots
\end{aligned} \tag{4.36}$$

Equations (4.36) provide the building blocks for the derivation of identity (4.34). The equations are first subtracted and then summed from $k = 1$ to $k = n$ to obtain

$$\begin{aligned}
\sum_{k=1}^{n} g(t_k\,, W_k) - g(t_{k-1}\,, W_{k-1}) &= \sum_{k=1}^{n} \frac{\partial g(t_{k-1/2}\,, W_{k-1/2})}{\partial t}\,(t_k - t_{k-1}) \\
&\quad + \sum_{k=1}^{n} \frac{\partial g(t_{k-1/2}\,, W_{k-1/2})}{\partial W_t}\,(W_k - W_{k-1}) + \cdots \\
&\quad + \frac{1}{2}\sum_{k=1}^{n} \frac{\partial^2 g(t_{k-1/2}\,, W_{k-1/2})}{\partial W_t^2}\,\Big[\,(W_k - W_{k-1/2})^2 - (W_{k-1} - W_{k-1/2})^2\,\Big] + \cdots
\end{aligned}$$

The left hand side of this equation is $g(b, W(b)\,) - g(a, W(a)\,)$. Now take the limit of the right hand

side of this equation to get

$$\begin{aligned} \lim_{n\to\infty} \sum_{k=1}^{n} \frac{\partial g(t_{k-1/2}\,, W_{k-1/2})}{\partial t}\,(t_k - t_{k-1}) &= \int_a^b \frac{\partial g(t, W_t)}{\partial t}\,dt \\ \lim_{n\to\infty} \sum_{k=1}^{n} \frac{\partial g(t_{k-1/2}\,, W_{k-1/2})}{\partial W_t}\,(W_k - W_{k-1}) &= \int_a^b \frac{\partial g(t, W_t)}{\partial W_t} \circ dW_t \end{aligned} \tag{4.37}$$

while the mean square limit of the third summation may be demonstrated to be zero. It therefore follows directly that

$$\int_a^b \frac{\partial g(t, W_t)}{\partial W_t}\, o\, dW_t = g(b, W(b)\,) - g(a, W(a)\,) - \int_a^b \frac{\partial g}{\partial t}\,dt\,. \tag{4.38}$$

Example Evaluate the Ito integral

$$\int_a^b W_t\, dW_t\,.$$

Solution The application of identity (4.33) with $f(t, W_t) = W_t$ gives

$$\int_a^b W_t\, o\, dW_t = \int_a^b W_t\, dW_t + \frac{1}{2}\int_a^b dt = \int_a^b W_t\, dW_t + \frac{b-a}{2}\,.$$

The application of identity (4.38) with $g(t, W_t) = W_t^2/2$ now gives

$$\int_a^b W_t\, o\, dW_t = \frac{1}{2}\Big[\, W^2(b) - W^2(a)\,\Big]\,.$$

Combining both results together gives the familiar result

$$\int_a^b W_t\, dW_t = \frac{1}{2}\Big[\, W^2(b) - W^2(a) - (b-a)\,\Big]\,.$$

4.5 Stratonovich representation on an SDE

The discussion of stochastic integration began by observing that the generic one-dimensional SDE

$$dx_t = a(t, x_t)\,dt + b(t, x_t)\,dW_t \tag{4.39}$$

had formal solution

$$x_b = x_a + \int_a^b a(s, x_s)\,ds + \int_a^b b(s, x_s)\,dW_s\,, \qquad b \geq a\,, \tag{4.40}$$

in which the second integral must be interpreted as an Ito stochastic integral. Because every Ito integral has an equivalent Stratonovich representation, then the Ito SDE (4.39) will have an equivalent Stratonovich representation which may be obtained by replacing the Ito integral in the formal solution by a Stratonovich integral and auxiliary terms. The objective of this section is to find the Stratonovich SDE corresponding to (4.39).

Let $f(t, x_t)$ be a function of (t, x) where x_t is the solution of the SDE (4.39). The key idea is to expand $f(t, x)$ by a Taylor series about (t_{k-1}, x_{k-1}). By definition,

$$\begin{aligned}
\int_a^b f(t, x_t) \circ dW_t &= \lim_{n\to\infty} \sum_{k=1}^{n} f(t_{k-1/2}, x_{k-1/2}) \, (W_k - W_{k-1}) \\
&= \lim_{n\to\infty} \sum_{k=1}^{n} \Big[f(t_{k-1}, x_{k-1}) + \frac{\partial f(t_{k-1}, x_{k-1})}{\partial t} \, (t_{k-1/2} - t_{k-1}) \\
&\qquad + \frac{\partial f(t_{k-1}, x_{k-1})}{\partial x} \, (x_{k-1/2} - x_{k-1}) + \cdots \Big] \, (W_k - W_{k-1}) \\
&= \lim_{n\to\infty} \sum_{k=1}^{n} \Big[f(t_{k-1}, x_{k-1}) \, (W_k - W_{k-1}) \\
&\qquad + \frac{\partial f(t_{k-1}, x_{k-1})}{\partial t} \, (t_{k-1/2} - t_{k-1}) \, (W_k - W_{k-1}) \\
&\qquad + \frac{\partial f(t_{k-1}, x_{k-1})}{\partial x} \, (x_{k-1/2} - x_{k-1}) \, (W_k - W_{k-1}) + \cdots .
\end{aligned}$$

Taking account of the fact that the limit of the second sum has value zero the previous analysis gives

$$\int_a^b f(t, x_t) \circ dW_t = \int_a^b f(t, X_t) \, dW_t + \lim_{n\to\infty} \sum_{k=1}^{n} \frac{\partial f(t_{k-1}, x_{k-1})}{\partial x} \, (x_{k-1/2} - x_{k-1})(W_k - W_{k-1}) . \quad (4.41)$$

Now

$$x_{k-1/2} - x_{k-1} = a(t_{k-1}, x_{k-1}) \, (t_{k-1/2} - t_{k-1}) + b(t_{k-1}, x_{k-1}) \, (W_{k-1/2} - W_{k-1}) + \cdots$$

and therefore the summation in expression (4.41) becomes

$$\begin{aligned}
&\lim_{n\to\infty} \sum_{k=1}^{n} \frac{\partial f(t_{k-1}, x_{k-1})}{\partial x} \Big[a(t_{k-1}, x_{k-1}) \, (t_{k-1/2} - t_{k-1}) \\
&\qquad + b(t_{k-1}, x_{k-1}) \, (W_{k-1/2} - W_{k-1}) + \cdots \Big] \, (W_k - W_{k-1}) \\
&= \lim_{n\to\infty} \sum_{k=1}^{n} \frac{\partial f(t_{k-1}, x_{k-1})}{\partial x} \, b(t_{k-1}, x_{k-1}) \, (W_{k-1/2} - W_{k-1}) \, (W_k - W_{k-1}) \\
&= \frac{1}{2} \int_a^b \frac{\partial f(t, x_t)}{\partial x_t} \, b(t, x_t) \, dt .
\end{aligned}$$

Omitting the analytical details associated with the mean square limit, the final conclusion is

$$\int_a^b f(t, x_t) \, o \, dW_t = \int_a^b f(t, x_t) \, dW_t + \frac{1}{2} \int_a^b \frac{\partial f(t, x_t)}{\partial x_t} \, b(t, x_t) \, dt . \quad (4.42)$$

We apply this identity immediately with $f(t, x) = b(t, x)$ to obtain

$$\int_a^b b(t, x_t) \, o \, dW_t = \int_a^b b(t, x_t) \, dW_t + \frac{1}{2} \int_a^b b(t, x_t) \, \frac{\partial b(t, x_t)}{\partial x_t} \, dt$$

which may in turn be used to remove the Ito integral in equation (4.40) and replace it with a Stratonovich integral to get

$$x_b = x_a + \int_a^b \Big[a(t, x_t) - \frac{b(t, x_t)}{2} \, \frac{\partial b(t, x_t)}{\partial x_t} \, b(t, x_t) \Big] \, dt + \int_a^b b(t, x_t) \circ dW_t . \quad (4.43)$$

Solution (4.43) or, equivalently, the stochastic differential equation

$$dx_t = \Big[\, a(t, x_t) - \frac{b(t, x_t)}{2} \, \frac{\partial b(t\,, x_t)}{\partial x_t} \,\Big]\, dt + b(t, x_t) \circ dW_t \,, \tag{4.44}$$

is called the *Stratonovich* stochastic differential equation.

Chapter 5

Differentiation of functions of stochastic variables

Previous discussion has focussed on the interpretation of integrals of stochastic functions. The differentiability of stochastic functions is now examined. It has been noted previously that the Wiener process has no derivative in the sense of traditional Calculus, and therefore differentiation for stochastic functions involves relationships between differentials rather than derivatives. The primary result is commonly called Ito's Lemma.

5.1 Ito's Lemma

Ito's lemma provides the basic rule by which the differential of composite functions of deterministic and random variables may be computed, *i.e.* Ito's lemma is the stochastic equivalent of the chain rule[1] in traditional Calculus. Let $F(x,t)$ be a suitably differentiable function then Taylor's theorem states that

$$\begin{aligned} F(t+dt, x+dx) &= F(t,x) + F_x(t,x)\,dx + F_t(t,x)\,dt \\ &\quad + \frac{1}{2}\left[\, F_{xx}\,(dx)^2 + 2F_{tx}\,(dt)(dx) + F_{tt}\,(dt)^2 \,\right] + o(dt)\,. \end{aligned} \tag{5.1}$$

The differential dx is now assigned to a stochastic process, say the process defined by the simple stochastic differential equation

$$dx = a(t,x)\,dt + b(t,x)\,dW_t\,. \tag{5.2}$$

[1] The chain rule is the rule for differentiating functions of a function.

Substituting expression (5.2) into (5.1) gives

$$\begin{aligned} F(t+dt,x+dx)-F(t,x) &= \frac{\partial F}{\partial t}\,dt+\frac{\partial F}{\partial x}\Big[a(t,x)\,dt+b(t,x)\,dW_t\Big] \\ &+\frac{1}{2}\,\frac{\partial^2 F}{\partial x^2}\Big[a^2(t,x)(dt)^2+2a(t,x)b(t,x)\,dt\,dW_t+b^2(t,x)\,(dW_t)^2\Big] \\ &+\frac{\partial^2 F}{\partial t\partial x}\,(dt)\Big[a(t,x)\,dt+\frac{1}{2}\,b(t,x)\,dW_t\Big]+\frac{\partial^2 F}{\partial t^2}\,(dt)^2+o(dt)\,. \end{aligned}$$

The differential $dF = F(t+dt, x+dx) - F(t,x)$ is expressed to first order in the differentials dt and dW_t, bearing in mind that $(dW_t)^2 = dt + o(dt)$ and $(dt)\,(dW_t) = O(dt^{3/2})$. After ignoring terms of higher order than dt (which would vanish in the limit as $dt \to 0$), the final result of this operation is **Ito's Lemma** in two dimensions, namely

$$dF = \Big[\,\frac{\partial F(t,x)}{\partial t}+\frac{\partial F(t,x)}{\partial x}\,a(t,x)+\frac{b^2(t,x)}{2}\,\frac{\partial^2 F(t,x)}{\partial x^2}\,\Big]\,dt+\frac{\partial F}{\partial x}\,b(t,x)\,dW_t\,. \tag{5.3}$$

Ito's lemma is now used to solve some simple stochastic differential equations.

Example Solve the stochastic differential equation

$$dx = a\,dt + b\,dW_t$$

in which a and b are constants and dW_t is the differential of the Wiener process.

Solution The equation $dx = a\,dt + b\,dW_t$ may be integrated immediately to give

$$x(t)-x(s)=\int_s^t dx=\int_s^t a\,dt+\int_s^t b\,dW=a(t-s)+b(W_t-W_s)\,.$$

The initial value problem therefore has solution $x(t) = x(s) + a(t-s) + b\,(W_t - W_s)$.

Example Use Ito's lemma to solve the first order SDE

$$dS = \mu S\,dt + \sigma S\,dW_t$$

where μ and σ are constant and dW is the differential of the Wiener process. This SDE describes *Geometric Brownian motion* and was used by Black and Scholes to model stock prices.

Solution The formal solution to this SDE obtained by direct integration has no particular value. While it might seem beneficial to divide both sides of the original SDE by S to obtain

$$\frac{dS}{S} = \mu\,dt + \sigma\,dW_t\,,$$

thereby reducing the problem to that of the previous problem, the benefit of this procedure is illusory since $d(\log S) \neq dS/S$ in a stochastic environment. In fact, if $dS = a(t,S)\,dt + b(t,S)\,dW$ then Ito's lemma gives

$$d(\log S)=\Big[\,\frac{a(t,S)}{S}-\frac{b^2(t,S)}{2S^2}\,\Big]\,dt+\frac{b(t,S)}{S}\,dW_t$$

When $a(t,S) = \mu\, S$ and $b(t,S) = \sigma\, S$, it follows from the previous equation that

$$d(\log S) = (\mu - \sigma^2/2)\, dt + \sigma\, dW_t\,.$$

This is the SDE encountered in the previous example. It can be integrated to give the solution

$$S(t) = S(s) \exp\Big[(\mu - \sigma^2/2)(t-s) + \sigma(W_t - W_s)\Big]\,.$$

Another approach Another way to solve this SDE is to recognise that if $F(t,S)$ is any function of t and S then Ito's lemma states that

$$dF = \Big[\, \frac{\partial F}{\partial t} + \mu S\, \frac{\partial F}{\partial S} + \frac{\sigma^2\, S^2}{2}\, \frac{\partial^2 F}{\partial S^2}\Big]\, dt + \sigma S\, \frac{\partial F}{\partial S}\, dW_t$$

A simple class of solutions arises when the coefficients of the dt and dW are simultaneously functions of time t only. This idea suggest that we consider a function F such that $S\,\partial F/\partial S = c$ with c constant. This occurs when $F(t,S) = c \log S + \psi(t)$. With this choice,

$$\frac{\partial F}{\partial t} + \mu S\, \frac{\partial F}{\partial S} + \frac{\sigma^2\, S^2}{2}\, \frac{\partial^2 F}{\partial S^2} = \frac{d\psi}{dt} + c\Big[\, \mu - \frac{\sigma^2}{2}\, \Big]$$

and the resulting expression for F is

$$F(t, S(t)\,) - F(s, S(s)\,) = \psi(t) - \psi(s) + c\Big[\, \mu - \frac{\sigma^2}{2}\, \Big](t-s) + c\sigma(W_t - W_s)\,.$$

Example Use Ito's lemma to solve the first order SDE

$$dx = ax\, dt + b\, dW_t$$

in which a and b are constants and dW_t is the differential of the Wiener process.

Solution The linearity of the drift specification means that the equation can be integrated by multiplying with the classical integrating factor of a linear ODE. Let $F(t,x) = x\, e^{-at}$ then Ito's lemma gives

$$dF = dx\, e^{-at} - ax\, e^{-at}\, dt = e^{-at}(ax\, dt + b\, dW_t - ax\, dt\,) = b\, e^{-at}\, dW_t\,.$$

Integration now gives

$$d(\, xe^{-at}\,) = be^{-at}\, dW_t \quad \rightarrow \quad x(t) = x(0)\, e^{at} + \int_0^t e^{a(t-s)}\, dW_s\,. \tag{5.4}$$

The solution in this case is a Riemann-Stieltjes integral. Evidently the value of this integral is a Gaussian deviate with mean value zero and variance

$$\int_0^t e^{2a(t-s)}\, ds = \frac{e^{2at} - 1}{2a}\,.$$

5.1.1 Ito's lemma in multi-dimensions

Ito's lemma in many dimensions is established in a similar way to the one-dimensional case. Suppose that $x_1, \cdots, x_N$ are solutions of the N stochastic differential equations

$$dx_i = a_i(t, x_1 \cdots x_N)\, dt + \sum_{p=1}^{M} \sigma_{ip}(t, x_1 \cdots x_N)\, dW_p\,, \tag{5.5}$$

where σ_{ip} is an $N \times M$ array and $dW_1, \cdots, dW_M$ are increments in the M ($\leq N$) Wiener processes $W_1, \cdots, W_M$. Suppose further that these Wiener processes have correlation structure characterised by the $M \times M$ array Q in which $Q_{ij}\, dt = \mathbb{E}\,[dW_i dW_j]$. Clearly Q is a positive definite array, but not necessarily a diagonal array.

Let $F(t, x_1 \cdots x_N)$ be a suitably differentiable function then Taylor's theorem states that

$$\begin{aligned} F(t+dt, x_1+dx_1 \cdots x_N+dx_N) &= F(t, x_1 \cdots x_N) + \frac{\partial F}{\partial t}\, dt + \sum_{j=1}^{N} \frac{\partial F}{\partial x_j}\, dx_j \\ &+ \frac{1}{2}\Big[\, \frac{\partial^2 F}{\partial t^2}(dt)^2 + 2\sum_{j=1}^{N} \frac{\partial^2 F}{\partial x_j \partial t}\,(dx_j)\, dt + \sum_{j,k=1}^{N} \frac{\partial^2 F}{\partial x_j \partial x_k}(dx_j)(dx_k)\Big] + o(dt)\,. \end{aligned} \tag{5.6}$$

The differentials $dx_1, \cdots, dx_N$ are now assigned to the stochastic process in (5.5) to get

$$\begin{aligned} F(t+dt, x_1+dx_1 \cdots x_N+dx_N) - F(t, x_1 \cdots x_N) &= \frac{\partial F}{\partial t}\, dt + \sum_{j=1}^{N} \frac{\partial F}{\partial x_j}\Big[a_j\, dt + \sum_{p=1}^{M} \sigma_{jp}\, dW_p\Big] \\ + \frac{1}{2}\frac{\partial^2 F}{\partial t^2}(dt)^2 + \sum_{j=1}^{N} \frac{\partial^2 F}{\partial x_j \partial t}\Big[a_j\, dt + \sum_{p=1}^{M} \sigma_{jp}\, dW_p\Big]\, dt & \\ + \frac{1}{2}\sum_{j,k=1}^{N} \frac{\partial^2 F}{\partial x_j \partial x_k}\Big[a_j\, dt + \sum_{p=1}^{M} \sigma_{jp}\, dW_p\Big]\Big[a_k\, dt + \sum_{q=1}^{M} \sigma_{kq}\, dW_q\Big] + o(dt)\,. & \end{aligned} \tag{5.7}$$

Equation (5.7) is now simplified by eliminating all terms of order $(dt)^{3/2}$ and above to obtain in the first instance.

$$\begin{aligned} dF &= \frac{\partial F}{\partial t}\, dt + \sum_{j=1}^{N} \frac{\partial F}{\partial x_j}\Big[a_j\, dt + \sum_{p=1}^{M} \sigma_{jp}\, dW_p\Big] \\ &+ \frac{1}{2}\sum_{j,k=1}^{N} \frac{\partial^2 F}{\partial x_k \partial x_k} \sum_{p=1}^{M}\sum_{q=1}^{M} \sigma_{jp}\sigma_{kq}\, dW_p\, dW_q + o(dt)\,. \end{aligned} \tag{5.8}$$

Recall that $Q_{ij}\, dt = \mathbb{E}\,[dW_i dW_j]$, which in this context asserts that $dW_p dW_q = Q_{pq}\, dt + o(dt)$. Thus the multi-dimensional form of Ito's lemma becomes

$$dF = \Big[\, \frac{\partial F}{\partial t} + \sum_{j=1}^{N} \frac{\partial F}{\partial x_j} a_j + \frac{1}{2}\sum_{j,k=1}^{N} \frac{\partial^2 F}{\partial x_j \partial x_k}\, g_{jk}\Big]\, dt + \sum_{j=1}^{N}\sum_{p=1}^{M} \frac{\partial F}{\partial x_j}\sigma_{jp}\, dW_p\,, \tag{5.9}$$

where g is the $N \times N$ array with $(j,k)^{th}$ entry

$$g_{jk} = \sum_{p=1}^{M}\sum_{q=1}^{M} \sigma_{jp}\sigma_{kq}\, Q_{pq}\,. \tag{5.10}$$

5.1.2 Ito's lemma for products

Suppose that $x(t)$ and $y(t)$ satisfy the stochastic differential equations

$$dx = \mu_x\,dt + \sigma_x\,dW_x\,, \qquad\qquad dy = \mu_y\,dt + \sigma_y\,dW_y$$

then one might reasonably wish to compute $d(xy)$. In principle, this computation will require an expansion of $x(t+dt)y(t+dt) - x(t)y(t)$ to order dt. Clearly

$$\begin{aligned} d(xy) &= x(t+dt)y(t+dt) - x(t)y(t) \\ &= [x(t+dt) - x(t)]y(t+dt) + x(t)[y(t+dt) - y(t)] \\ &= [x(t+dt) - x(t)][y(t+dt) - y(t)] + [x(t+dt) - x(t)]y(t) + x(t)[y(t+dt) - y(t)] \\ &= (dx)\,y(t) + x(t)\,(dy) + (dx)\,(dy)\,. \end{aligned}$$

The important observation is that there is an additional contribution to $d(xy)$ in contrast to classical Calculus where this contribution vanishes because it is $O(dt^2)$.

5.2 Further examples of SDEs

We now consider various SDEs that arise in Finance.

5.2.1 Multi-dimensional Ornstein Uhlenbeck equation

The Ornstein Uhlenbeck (OU) equation in N dimensions posits that the N dimensional state vector X satisfies the SDE

$$dX = (B - AX)\,dt + C\,dW$$

where B is an $N \times 1$ vector, A is a nonsingular $N \times N$ matrix and C is an $N \times M$ matrix while dW is the differential of the M dimensional Wiener process $W(t)$. Classical Calculus suggests that Ito's lemma should be applied to $F = e^{At}(X - A^{-1}B)$. The result is

$$dF = Ae^{At}(X - A^{-1}B)\,dt + e^{At}\,dX = e^{At}((AX - B)\,dt + dX) = e^{At}C\,dW\,.$$

Integration over $[0,t]$ now gives

$$\begin{aligned} F(t) - F(0) = \int_0^t e^{As}C\,dW \quad &\to \quad e^{At}(X - A^{-1}B) = (X_0 - A^{-1}B) + \int_0^t e^{As}C\,dW_s \\ &\to \quad X = (I - e^{-At})A^{-1}B + e^{-At}X_0 + \int_0^t e^{-A(t-s)}C\,dW_s \end{aligned}$$

The solution is a multivariate Gaussian process with mean value $(I - e^{-At})A^{-1}B + e^{-At}X_0$ and covariance

$$\begin{aligned}
\Sigma &= \mathbb{E}\Big[\Big(\int_0^t e^{-A(t-s)}C\,dW_s\Big)\Big(\int_0^t e^{-A(t-s)}C\,dW_s\Big)^T\Big] \\
&= \int_0^t\int_0^t \mathbb{E}\Big[\Big(e^{-A(t-s)}C\,dW_s\Big)\Big(e^{-A(t-v)}C\,dW_v\Big)^T\Big] \\
&= \int_0^t\int_0^t e^{-A(t-s)}C\,\mathbb{E}[\,dW_s dW_v^T\,]C^T\,e^{-A^T(t-v)} \\
&= \int_0^t\int_0^t e^{-A(t-s)}C\,\mathbb{E}\,Q\,\delta(s-v)\,C^T\,e^{-A^T(t-v)}\,ds\,dv \\
&= \int_0^t e^{-A(t-s)}CQC^T\,e^{-A^T(t-s)}\,ds \\
&= \int_0^t e^{-Au}CQC^T\,e^{-A^Tu}\,du\,.
\end{aligned}$$

5.2.2 Logistic equation

Recall how the Logistic model was a modification of the Malthusian model of population to take account of limited resources. The Logistic model captured the effect of changing environmental conditions by proposing that the maximum supportable population was a Gaussian random variable with mean value, and in so doing gave rise to the stochastic differential equation

$$dN = aN(M-N)\,dt + b\,N\,dW_t\,, \qquad N(0) = N_0 \tag{5.11}$$

in which a and $b \geq 0$ are constants. Although there is no foolproof way of treating SDEs with state-dependent diffusions, two possible strategies are:-

(a) Introduce a change of dependent variable that will eliminate the dependence of the diffusive term on state variables;

(b) Use the change of dependent variable that would be appropriate for solving the underlying ODE obtained by ignoring the diffusive term.

Note also that any change of dependent variable that involves either multiplying the original unknown by a function of t or adding a function of t to the original unknown necessarily obeys the rules of classical Calculus. The reason is that the second derivative of such a change of variable with respect to the state variable is zero.

In the case of equation (5.11) strategy (a) would suggest that the appropriate change of variable is $\psi = \log N$ while strategy (b) would suggest the change of variable $\phi = 1/N$ because the underlying ODE is a Bernoulli equation. The resulting SDE in each case can be rearranged into the format

$$\begin{aligned}
\text{Case (a)} \quad d\psi &= [\,(aM - b^2/2)\,dt + b\,dW_t\,] - a\,e^{\psi}\,dt\,, \qquad \psi(0) = \log N_0\,, \\
\text{Case (b)} \quad d\phi &= -\phi\,[\,(aM - b^2)\,dt + b\,dW_t\,] + a\,dt\,, \qquad \phi(0) = 1/N_0\,.
\end{aligned} \tag{5.12}$$

In either case it seems convenient to define

$$\xi(t) = (aM - b^2/2)t + b\,W_t\,, \qquad \xi(0) = 0\,. \tag{5.13}$$

In case (a) the equation satisfied by $z(t) = \psi(t) - \xi(t)$ is

$$dz = -a\,e^{z+\xi}\,dt = -a\,e^{z}\,e^{\xi}\,dt \quad \rightarrow \quad e^{-z}\,dz = -a\,e^{\xi}\,dt$$

with initial condition $z(0) = \psi(0) = \log N_0$. Clearly the equation satisfied by z has solution

$$\int_{z(0)}^{z(t)} e^{-z}\,dz = -a\int_0^t e^{\xi(s)}\,ds = -e^{-z(t)} + e^{-z(0)} \quad \rightarrow \quad e^{-z(t)} = e^{-z(0)} + a\int_0^t e^{\xi(s)}\,ds \tag{5.14}$$

in which all integrals are Riemann-Stieltjes integrals. The definition of z and initial conditions are now substituted into solution (5.14) to get the final solution

$$\frac{e^{\xi(t)}}{N(t)} = \frac{1}{N_0} + a\int_0^t e^{\xi(s)}\,ds$$

which may be rearranged to give

$$N(t) = \frac{N_0\,\exp\left[\,(aM - b^2/2)t + bW_t\,\right]}{1 + aN_0\displaystyle\int_0^t \exp\left[\,(aM - b^2/2)s + bW_s\,\right]ds}\,. \tag{5.15}$$

If approach (b) is used, then the procedure is to solve for $z(t) = \log\phi(t)$ obtaining in the process a differential equation for z that is effectively equivalent to equation (5.14).

5.2.3 Square-root process

The process $X(t)$ obeys a square-root process if

$$dX = \alpha(\theta - X)\,dt + \sigma\sqrt{X}\,dW_t$$

in which θ is the mean process, α controls the rate of restoration to the mean and σ is the volatility control. While the square-root and OU process have identical drift specifications, there the similarity ends. The square-root process, unlike the OU process, has no known closed-form solution although its moments of all orders can be calculated in closed form as can its transitional probability density function. The square-root process has been proposed by Cox, Ingersoll and Ross as the prototypical model of interest rates in which context it is commonly known as the CIR process. An extension of the CIR process proposed by Chan, Karoyli, Longstaff and Saunders introduces a "levels" parameter γ to obtain the SDE

$$dX = \alpha(\theta - X)\,dt + \sigma\,X^{\gamma}\,dW_t\,.$$

In this case only the mean process is known in closed form. Neither higher order moments nor the transitional probability density function have known closed-form expressions.

5.2.4 Constant Elasticity of Variance Model

The Constant Elasticity of Variance (CEV) model proposes that S, the spot price of an asset, evolves according to the stochastic differential equation

$$dS = \mu S\,dt + \sigma S^{\alpha}\,dW_t \tag{5.16}$$

where μ is the mean stock return (sum of a risk-free rate and an equity premium), dW is the differential of the Wiener process, σ is the volatility control and α is the elasticity factor or levels parameter. This model generalises the Black-Scholes model through the inclusion of the elasticity factor α which need not take the value 1. Let $\psi = S^{2-2\alpha}$ then Ito's lemma gives

$$\begin{aligned} d\psi &= (2-2\alpha)\,S^{1-2\alpha}\,dS + \frac{(2-2\alpha)(1-2\alpha)}{2}\,S^{-2\alpha}(\sigma^2 S^{2\alpha})\,dt \\ &= (1-\alpha)(1-2\alpha)\sigma^2\,dt + (2-2\alpha)\,S^{1-2\alpha}\,(\mu S dt + \sigma S^{\alpha}\,dW_t) \\ &= (1-\alpha)[(1-2\alpha)\sigma^2 - 2\mu\,S^{2-2\alpha}]\,dt + 2\sigma(1-\alpha)\,S^{1-\alpha}\,dW_t\,. \end{aligned}$$

S is now eliminated from the previous equation to get the final SDE

$$d\psi = (1-\alpha)[(1-2\alpha)\sigma^2 - 2\mu\,\psi]\,dt + 2\sigma(1-\alpha)\,\sqrt{\psi}\,dW_t\,.$$

Interestingly ψ satisfies a model equation that is structurally similar to the CIR class of model, particularly when $\alpha \in (1/2, 1)$.

Applications of equation (5.16) often set $\mu = 0$ and re-express equation (5.17) in the form

$$dF = \sigma F^{\alpha}\,dW_t\,, \tag{5.17}$$

where $F(t)$ is now the forward price of the underlying asset. When $\alpha < 1$, the volatility of stock price (*i.e.* $\sigma F^{\alpha-1}$) increases after negative returns are experienced resulting in a shift of probability to the left hand side of the distribution and a fatter left hand tail.

5.2.5 SABR Model

The Stochastic Alpha, Beta, Rho model, or SABR model for short, was proposed by Hagan *et al.* [**?**] in order to rectify the difficulty that the volatility control parameter σ in the CEV model is constant, whereas returns from real assets are known to be negatively correlated with volatility. The SABR model incorporates this correlation by treating σ in the CEV model as a stochastic process in its own right and appending to the CEV model a separate evolution equation for σ to obtain

$$\begin{aligned} dF &= \sigma F^{\beta}\,(\,\sqrt{1-\rho^2}\,dW_1 + \rho\,dW_2\,) \\ d\sigma &= \sigma\alpha\,dW_2\,, \end{aligned} \tag{5.18}$$

where dW_1 and dW_2 are uncorrelated Wiener processes and the levels parameter $\beta \in [0, 1]$. Thus σ of the CEV model is driven by a Wiener process correlated with parameter ρ to the Wiener process in the model for the forward price F. The choice $\alpha = 0$ in the SABR model gives the CEV model.

The volatility in the SABR model follows a geometric walk with zero drift and solution

$$\sigma(t) = \sigma_0 \, \exp\Big(-\frac{\alpha^2 t}{2} + \alpha \, W_2(t) \Big) \, ,$$

where σ_0 is the initial value of σ. Despite the closed-form expression for $\sigma(t)$ the SABR model has no closed-form solutions except in the cases $\beta = 0$ (solve for F) and $\beta = 1$ (solve for $\log F$). In overview, the values of the parameters ρ and β control the curvature of the volatility smile whereas the value of α, the volatility of volatility, controls the skewness of the lognormal distribution of volatility.

5.3 Heston's Model

The SABR model is a recent example of a stochastic volatility options pricing model of which there exists many. The most renowned and widely used of these is Heston's model. The model was proposed in 1993 and when reformulated in terms of $y = \log S$ posits that

$$\begin{aligned} dy &= (r + \beta \, h) \, dt + \sqrt{h} \, (\rho \, dW_t + \sqrt{1 - \rho^2} \, dZ_t) \, , \\ dh &= \kappa(\gamma - h) \, dt + \sigma \sqrt{h} \, dW_t \, , \end{aligned} \tag{5.19}$$

where h is volatility and dW_t and dZ_t are increments in the independent Wiener processes W_t and Z_t, the parameter $\beta = \lambda(1 - \rho^2) - 1/2$ contains the equity premium, r is the risk-free force of interest and $\rho \in [-1, 1]$ is the correlation between innovations in asset returns and volatility.

However, Heston's model is an example of a bivariate affine process. While the transitional probability density function cannot be computed in closed form, it is possible to determine the characteristic function of Heston's model in closed form. This calculation requires the solution of partial differential equations and will be the focus of our attention in a later chapter.

5.4 Girsanov's lemma

Girsanov's lemma is concerned with changes of measure and the Radon-Nikodym[2] derivative. A simple way to appreciate what is meant by a change in measure is to consider the evolution of asset prices as described by a binomial with transitions of fixed duration. Suppose the spot price of an asset is S, and that at the next transition the price of the asset is either S_u ($> S$) with probability q or S_d ($< S$) with probability $(1 - q)$, and take $\mathbb{Q}$ to be the measure under which the expected one-step-ahead asset price is S, *i.e.* the process is a $\mathbb{Q}$-Martingale. Clearly the value of q satisfies

$$S_u q + S_d(1 - q) = S \quad \rightarrow \quad q = \frac{S - S_d}{S_u - S_d}$$

leading to upward and downward movements of the asset price with respective probabilities

$$q = \frac{S - S_d}{S_u - S_d} \, , \qquad\qquad 1 - q = \frac{S_u - S}{S_u - S_d} \, .$$

[2]The Radon-Nikodym derivative is essentially a Jacobian.

Let $\mathbb{P}$ be another measure in which the asset price at each transition moves upwards with probability p and downwards with probability $(1-p)$ and consider a pathway (or filtration) through the binomial tree consisting of N transitions in which the asset price moved upwards n times and downwards $N-n$ times attaining the sequence of states $S_1, S_2, \cdots, S_N$. The likelihoods associated with this path under the $\mathbb{P}$ and $\mathbb{Q}$ measures are respectively

$$\mathcal{L}^{(P)}(S_1, \cdots, S_N) = p^n (1-p)^{N-n}\,, \qquad \mathcal{L}^{(Q)}(S_1, \cdots, S_N) = q^n (1-q)^{N-n}\,.$$

The likelihood ratio of the $\mathbb{Q}$ measure to the $\mathbb{P}$ measure is therefore

$$L_N(S_1, \cdots, S_N) = \frac{q^n (1-q)^{N-n}}{p^n (1-p)^{N-n}} = \Big(\frac{q}{p}\Big)^n \Big(\frac{1-q}{1-p}\Big)^{N-n}\,.$$

Clearly the value of L_N is a function of the path (or filtration - say $\mathcal{F}_N$) - straightforward in this case because it is only the number of upward and downward movements that matter and not when these movements occur. The quantity L_N is called the "Radon-Nikodym" derivative of $\mathbb{Q}$ with respect to $\mathbb{P}$ and is commonly written

$$L_N = \left.\frac{d\mathbb{Q}}{d\mathbb{P}}\right|_{\mathcal{F}_N}\,.$$

Moreover

$$\frac{L_{N+1}}{L_N} = \begin{cases} \dfrac{q}{p} & S_{N+1} > S_N\,, \\ \dfrac{1-q}{1-p} & S_{N+1} < S_N\,. \end{cases}$$

The expectation of L_{N+1} under the $\mathbb{P}$ measure conditional on the filtration $\mathcal{F}_N$ gives

$$\mathbb{E}^{\mathbb{P}}[L_{N+1}\,|\,\mathcal{F}_N\,] = p\Big(L_N \times \frac{q}{p}\Big) + (1-p)\Big(L_N \times \frac{1-q}{1-p}\Big) = L_N\,.$$

Thus L_N, or in general the Radon-Nikodym derivative, is a martingale with respect to the $\mathbb{P}$ measure. But $L_0 = 1$ and so iteration of the identity $\mathbb{E}^{\mathbb{P}}[L_{N+1}\,|\,\mathcal{F}_N\,] = L_N$ indicates that $\mathbb{E}^{\mathbb{P}}[L_N] = 1$. In this illustration we associate the $\mathbb{P}$ measures with risk in our (risk-averse) world and for this reason it is usually called the "Physical" measure. By contrast, the measure $\mathbb{Q}$ is commonly called the "Risk-Neutral" measure.

Now consider $\mathbb{E}^{\mathbb{P}}[\psi\, L_N\,|\,\mathcal{F}_N\,]$. The value is

$$\psi_U \left(\frac{q}{p}\, L_N\right) p + \psi_D \left(\frac{1-q}{1-p}\, L_N\right)(1-p) = (\psi_U\, q + \psi_D\,(1-q)) L_N = \mathbb{E}^{\mathbb{Q}}[\psi\,|\,\mathcal{F}_N\,]$$

and so the Radon-Nikodym derivative enjoys the key property that it allows expectations with respect to the $\mathbb{Q}$ measure to be expressed as expectations with respect to the $\mathbb{P}$ via the formula

$$\mathbb{E}^{\mathbb{Q}}\,[\,\Psi\,] = \mathbb{E}^{\mathbb{P}}\,[\,\Psi\, L_N\,]$$

where Ψ, for example, is the payoff from a call option.

5.4.1 Change of measure for a Wiener process

The illustrative example involving the binomial tree expressed changes in measure in terms of re-weighting the probability of upward and downward movement, and in so doing the probabilities associated with pathways through the tree. The same procedure can be extended to continuous processes, and in particular the Wiener process. The natural measure $\mathbb{P}$ on the Wiener process is associated with the probability density function

$$p(x) = \frac{1}{\sqrt{2\pi}} \exp\Big[-\frac{x^2}{2} \Big] .$$

Suppose, for example, that the Wiener process $W(t)$ attains the values $(x_1, \ldots, x_n)$ at the respective times $(t_1, \ldots, t_n)$ in which $\mathcal{D} = \{t_0, \cdots, t_n\}$ is a dissection of the finite interval $[0, T]$. Because increments of $W(t)$ are independent, then the likelihood $\mathcal{L}^{(n)}(P)$ associated with this path is

$$\mathcal{L}^{(P)}(x_1, \cdots, x_n) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\Delta t_i}} \exp\left(-\frac{(\Delta x_i)^2}{2\Delta t_i} \right) ,$$

where $\Delta x_i = x_i - x_{i-1}$ and $\Delta t_i = t_i - t_{i-1}$. The limit of this expression may be interpreted as the measure $\mathbb{P}$ for the continuous process $W(t)$. The question is now to determine how $W(t)$ changes under a different measure, say $\mathbb{Q}$, and of course what restrictions (if any) must be imposed on $\mathbb{Q}$.

Equivalent measure

Equivalence in the case of the binomial tree is the condition $p > 0 \iff q > 0$. In effect, the measures $\mathbb{P}$ and $\mathbb{Q}$ must both refer to binomial trees and not a binomial tree in one measure and a cascading process in the other measure. Equivalence between the measures $\mathbb{P}$ and $\mathbb{Q}$ in the binomial tree is a particular instance of the following definition of equivalence.

Two measures $\mathbb{P}$ and $\mathbb{Q}$ are equivalent (*i.e.* one can be distorted into the other) if and only if the measures agree on which events have zero probability, *i.e.*

$$\mathbb{P}(X) > 0 \iff \mathbb{Q}(X) > 0 .$$

Given two equivalent measures $\mathbb{P}$ and $\mathbb{Q}$, the "Radon-Nikodym" derivative of $\mathbb{Q}$ with respect to $\mathbb{P}$ is defined by the likelihood ratio

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = \lim_{n\to\infty} \frac{\mathcal{L}^{(Q)}(x_1, \cdots, x_n)}{\mathcal{L}^{(P)}(x_1, \cdots, x_n)} .$$

Alternatively one may write

$$\mathbb{Q}(X) = \int_X d\mathbb{Q}(x) = \int_X \phi \, d\mathbb{P}(x)$$

where $\phi = d\mathbb{Q}/d\mathbb{P}$ is the Radon-Nikodym derivative of $\mathbb{Q}$ with respect to $\mathbb{P}$. Clearly ϕ simply states what distortion should be applied to the measure $\mathbb{P}$ to get the measure $\mathbb{Q}$. Given and function Ψ, then

$$\mathbb{E}^{\mathbb{Q}}[\Psi] = \mathbb{E}^{\mathbb{P}}[\Psi \, \phi] .$$

5.4.2 Girsanov theorem

Let $W(t)$ be a $\mathbb{P}$-Wiener process under the natural filtration $\mathcal{F}_t$, namely that for which the probabilities of paths are determined directly from the $\mathrm{N}(0,1)$ probability density function, and define $Z(t)$ to be a process adapted from $W(t)$ for which Novikov's condition

$$\mathbb{E}\left[\exp\left(\frac{1}{2}\int_0^T Z^2(s)\,ds\right)\right] < \infty$$

is satisfied for some finite time horizon T. Let $\mathbb{Q}$ to be a measure related to $\mathbb{P}$ with Radon-Nikodym derivative

$$\phi(t) = \frac{d\mathbb{Q}}{d\mathbb{P}} = \exp\Big(-\int_0^t Z(s)\,dW(s) - \frac{1}{2}\int_0^t Z^2(s)\,ds\Big).$$

Under the probability measure $\mathbb{Q}$, the process

$$W^{(\phi)}(t) = W(t) + \int_0^t Z(s)\,ds$$

is a standard Wiener process.

Justification

From the definition of $\mathbb{Q}$ it is clear that $\mathbb{P}$ and $\mathbb{Q}$ are equivalent measures. One strategy is use the moment generating function to demonstrate that $W^{(\phi)}(t)$ is a natural Wiener process under the $\mathbb{Q}$ measure, namely that $\mathbb{E}^{\mathbb{Q}}[\,W^{(\phi)}(t)\,] = 0$ and that $\mathbb{V}^{\mathbb{Q}}[\,W^{(\phi)}(t)\,] = t$. We begin by recalling that the moment generating function of the Gaussian distribution $\mathrm{N}(\mu,\sigma^2)$ is

$$M(\theta) = \mathbb{E}^{\mathbb{P}}[\,e^{\theta X}\,] = \frac{1}{\sqrt{2\pi}\,\sigma}\int_{\mathbb{R}} e^{\theta X} e^{-(X-\mu)^2/2\sigma^2}\,dX = e^{\theta\mu+\theta^2\sigma^2/2}.$$

The formal proof of Girsanov's theorem starts with the Radon-Nikodym identity

$$\mathbb{E}^{\mathbb{Q}}[\,\Psi\,] = \mathbb{E}^{\mathbb{P}}\left[\frac{d\mathbb{Q}}{d\mathbb{P}}\,\Psi\right]$$

when particularized by replacing Ψ with $e^{\theta W^{(\phi)}(t)}$ to get

$$\begin{aligned}
\mathbb{E}^{\mathbb{Q}}[\,e^{\theta W^{(\phi)}(t)}\,] &= \mathbb{E}^{\mathbb{P}}\left[\frac{d\mathbb{Q}}{d\mathbb{P}}\,e^{\theta W^{(\phi)}(t)}\right] \\
&= \mathbb{E}^{\mathbb{P}}\left[\exp\Big(-\int_0^t Z(s)\,dW(s) - \frac{1}{2}\int_0^t Z^2(s)\,ds\Big)\exp\Big(\theta W(t) + \theta\int_0^t Z(s)\,ds\Big)\right] \\
&= \exp\Big(\theta\int_0^t Z(s)\,ds - \frac{1}{2}\int_0^t Z^2(s)\,ds\Big)\mathbb{E}^{\mathbb{P}}\left[\exp\Big(\theta W(t) - \int_0^t Z(s)\,dW(s)\Big)\right]
\end{aligned}$$

The focus is now on the computation of $\mathbb{E}^{\mathbb{P}}\left[\exp\left(\theta W(t) - \int_0^t Z(s)\,dW(s)\right)\right]$. For convenience let

$$\xi(t) = \theta W(t) - \int_0^t Z(s)\,dW(s)$$

then $\mathbb{E}^{\mathbb{P}}[\,\xi(t)\,] = 0$ and

$$\begin{aligned}\mathbb{V}^{\mathbb{P}}[\,\xi(t)\,] &= \mathbb{E}^{\mathbb{P}}[\,\theta^2 W^2(t)\,] - 2\theta\mathbb{E}^{\mathbb{P}}\Big[W(t)\int_0^t Z(s)\,dW(s)\Big] + \mathbb{E}^{\mathbb{P}}\Big[\Big(\int_0^t Z(s)\,dW(s)\Big)^2\Big] \\ &= \theta^2\,t - 2\theta\int_0^t Z(s)\,\mathbb{E}^{\mathbb{P}}\,[W(t)\,dW(s)\,] + \int_0^t Z^2(s)\,ds\end{aligned}$$

Clearly the Novikov condition simply guarantees that $e^{\xi(t)}$ has finite mean and variance. Moreover, it is obvious that $\mathbb{E}^{\mathbb{P}}\,[W(t)\,dW(s)\,] = ds$ and therefore

$$\begin{aligned}\mathbb{E}^{\mathbb{P}}[\,\xi(t)\,] &= 0\,, \\ \mathbb{V}^{\mathbb{P}}[\,\xi(t)\,] &= \theta^2\,t - 2\theta\int_0^t Z(s)\,ds + \int_0^t Z^2(s)\,ds\,.\end{aligned}$$

It now follows from the moment generating function of $\mathrm{N}(\mu,\sigma^2)$ with $\mu = 0$ and $\theta = 1$ that

$$\mathbb{E}^{\mathbb{P}}[\,e^{\,\xi(t)}\,] = \exp\Big(\frac{\theta^2}{2}\,t - \theta\int_0^t Z(s)\,ds + \frac{1}{2}\int_0^t Z^2(s)\,ds\Big)\,.$$

Consequently

$$\begin{aligned}\mathbb{E}^{\mathbb{Q}}\,[\,e^{\,\theta W^{(\phi)}(t)}\,] &= \exp\Big(\theta\int_0^t Z(s)\,ds - \frac{1}{2}\int_0^t Z^2(s)\,ds\Big)\mathbb{E}^{\mathbb{P}}\Big[\exp\Big(\theta W(t) - \int_0^t Z(s)\,dW(s)\Big) \\ &= \exp\Big(\theta\int_0^t Z(s)\,ds - \frac{1}{2}\int_0^t Z^2(s)\,ds\Big)\exp\Big(\frac{\theta^2}{2}\,t - \theta\int_0^t Z(s)\,ds + \frac{1}{2}\int_0^t Z^2(s)\,ds\Big) \\ &= \exp\Big(\frac{\theta^2}{2}\,t\Big)\,.\end{aligned}$$

which is just the moment generating function of a standard Wiener process, thereby establish the claim of the Girsanov's theorem.

Chapter 6

The Chapman Kolmogorov Equation

6.1 Introduction

Let the state of a system at time t after initialisation be described by the n-dimensional random variable $\mathbf{x}(t)$ with sample space Ω. Let $\mathbf{x}_1, \mathbf{x}_2, \cdots$ be the measured state of the system at times $t_1 > t_2 > \cdots \geq 0$, then $\mathbf{x}(t)$ is a *stochastic process* provided the joint probability density function

$$f(\mathbf{x}_1, t_1; \mathbf{x}_2, t_2; \cdots) \tag{6.1}$$

exists for all possible sequences of measurements $\mathbf{x}_k$ and times t_k. When the joint probability density function (6.1) exists, it can be used to construct the conditional probability density function

$$f(\mathbf{x}_1, t_1; \mathbf{x}_2, t_2; \cdots \,|\, \mathbf{y}_1, \tau_1; \mathbf{y}_2, \tau_2; \cdots) \tag{6.2}$$

which describes the likelihood of measuring $\mathbf{x}_1, \mathbf{x}_2, \cdots$ at times $t_1 > t_2 > \cdots \geq 0$, **given** that the system attained the states $\mathbf{y}_1, \mathbf{y}_2, \cdots$ at times $\tau_1 > \tau_2 > \cdots \geq 0$ where, of necessity, $t_1 > t_2 > \cdots > \tau_1 > \tau_2 > \cdots \geq 0$. Kolmogorov's axioms of probability yield

$$f(\mathbf{x}_1, t_1; \mathbf{x}_2, t_2; \cdots \,|\, \mathbf{y}_1, \tau_1; \mathbf{y}_2, \tau_2; \cdots) = \frac{f(\mathbf{x}_1, t_1; \mathbf{x}_2, t_2; \cdots; \mathbf{y}_1, \tau_1; \mathbf{y}_2, \tau_2; \cdots)}{f(\mathbf{y}_1, \tau_1; \mathbf{y}_2, \tau_2; \cdots)}\,, \tag{6.3}$$

i.e. the conditional density is the usual ratio of the joint density of $\mathbf{x}$ and $\mathbf{y}$ (numerator) and the marginal density of $\mathbf{y}$ (denominator). In the same notation there is a hierarchy of identities which must be satisfied by the conditional probability density functions of any stochastic process. The first identity is

$$f(\mathbf{x}_1, t_1) = \int_\Omega f(\mathbf{x}_1, t_1; \mathbf{x}, t)\, d\mathbf{x} = \int_\Omega f(\mathbf{x}_1, t_1 \,|\, \mathbf{x}, t)\, f(\mathbf{x}, t)\, d\mathbf{x} \tag{6.4}$$

which apportions the probability with which the state $\mathbf{x}_1$ at future time t_1 is accessed at a previous time t from each $\mathbf{x} \in \Omega$. The second group of conditions take the generic form

$$\begin{aligned} f(\mathbf{x}_1, t_1 \,|\, \mathbf{x}_2, t_2; \cdots) &= \int_\Omega f(\mathbf{x}_1, t_1; \mathbf{x}, t; \mathbf{x}_2, t_2; \cdots)\, d\mathbf{x} \\ &= \int_\Omega f(\mathbf{x}_1, t_1 \,|\, \mathbf{x}, t; \mathbf{x}_2, t_2; \cdots)\, f(\mathbf{x}, t \,|\, \mathbf{x}_2, t_2; \cdots)\, d\mathbf{x} \end{aligned} \tag{6.5}$$

in which it is taken for granted that $t_1 > t > t_2 > \cdots \geq 0$. Similarly, a third group of conditions can be constructed from identities (6.4) and (6.5) by regarding the pair $(\mathbf{x}_1, t_1)$ as a token and replacing it with an arbitrary sequence of pairs involving the attainment of states at times after t.

Of course, equations (6.1-6.5) are also valid when Ω is a discrete sample space in which case the joint probability density functions are sums of delta functions and integrals over Ω are replaced by summations over Ω.

As might be anticipated the structure presented in equations (6.1-6.5) is too general for making useful progress, and so it is desirable to introduce further simplifying assumptions. Two obvious simplifications are possible.

(a) Assume that $(\mathbf{x}_k, t_k)$ are independent random variables. The joint probability density function is therefore the product of the individual probabilities $f(\mathbf{x}_k, t_k)$, *i.e.*

$$f(\mathbf{x}_1, t_1; \mathbf{x}_2, t_2; \cdots \,|\, \mathbf{y}_1, \tau_1; \mathbf{y}_2, \tau_2; \cdots) = f(\mathbf{x}_1, t_1; \mathbf{x}_2, t_2; \cdots) = \prod f(\mathbf{x}_k, t_k)\,.$$

In effect, there is no conditioning whatsoever in the resulting stochastic process and consequently historical states contain no useful information.

(b) Assume that the current state perfectly captures all historical information. This is the simplest paradigm to incorporate historical information within a stochastic process. Processes with this property are called *Markov* processes.

Assumption (a) is too strong. It is taken "as read" that in practice it is observations of historical states of a system in combination with time that conditions the distribution of future states, and not just time itself as would be the case if assumption (a) is adopted.

6.1.1 Markov process

A stochastic process possesses the Markov property if the historical information in the process is perfectly captured by the most recent observation of the process. In effect,

$$f(\mathbf{x}_1, t_1; \mathbf{x}_2, t_2; \cdots \,|\, \mathbf{y}_1, \tau_1; \mathbf{y}_2, \tau_2; \cdots) = f(\mathbf{x}_1, t_1; \mathbf{x}_2, t_2; \cdots \,|\, \mathbf{y}_1, \tau_1) \tag{6.6}$$

where it is again assumed that $t_1 > t_2 > \cdots > \tau_1 > \tau_2 > \cdots \geq 0$. Suppose that the sequence of states under discussion are $(\mathbf{x}_k, t_k)$ where $k = 1, \cdots, m$ then it follows directly from equation (6.3) that

$$f(\mathbf{x}_1, t_1; \mathbf{x}_2, t_2; \cdots; \mathbf{x}_m, t_m) = f(\mathbf{x}_1, t_1; \,|\, \mathbf{x}_2, t_2; \cdots \mathbf{x}_m, t_m)\, f(\mathbf{x}_2, t_2; \cdots; \mathbf{x}_m, t_m) \tag{6.7}$$

and from equation (6.6) that

$$f(\mathbf{x}_1, t_1 \,|\, \mathbf{x}_2, t_2; \cdots \mathbf{x}_m, t_m) = f(\mathbf{x}_1, t_1 \,|\, \mathbf{x}_2, t_2)\,. \tag{6.8}$$

Consequently, for a Markov process, results (6.7) and (6.8) may be combined to give

$$f(\mathbf{x}_1, t_1; \mathbf{x}_2, t_2; \cdots; \mathbf{x}_m, t_m) = f(\mathbf{x}_1, t_1 \,|\, \mathbf{x}_2, t_2)\, f(\mathbf{x}_2, t_2; \cdots; \mathbf{x}_m, t_m)\,. \tag{6.9}$$

Equation (6.9) forms the basis of an induction procedure from which it is follows directly that

$$f(\mathbf{x}_1, t_1; \mathbf{x}_2, t_2; \cdots; \mathbf{x}_m, t_m) = f(\mathbf{x}_m, t_m) \prod_{k=1}^{k=m-1} f(\mathbf{x}_k, t_k \,|\, \mathbf{x}_{k+1}, t_{k+1}) \,. \tag{6.10}$$

6.2 Chapman-Kolmogorov equation

Henceforth the stochastic process will be assumed to possess the Markov property. Identity (6.4) remains unchanged. However, the identities characterised by equation (6.5) are now simplified since the Markov property requires that

$$\begin{aligned} f(\mathbf{x}_1, t_1 \,|\, \mathbf{x}, t; \mathbf{x}_2, t_2; \cdots) &= f(\mathbf{x}_1, t_1 \,|\, \mathbf{x}, t) \,, \\ f(\mathbf{x}, t \,|\, \mathbf{x}_2, t_2; \cdots) &= f(\mathbf{x}, t \,|\, \mathbf{x}_2, t_2) \,. \end{aligned}$$

Consequently, the general condition (6.5) simplifies to

$$f(\mathbf{x}_1, t_1 \,|\, \mathbf{x}_2, t_2; \cdots) = \int_\Omega f(\mathbf{x}_1, t_1 \,|\, \mathbf{x}, t) \, f(\mathbf{x}, t \,|\, \mathbf{x}_2, t_2) \, d\mathbf{x} = f(\mathbf{x}_1, t_1 \,|\, \mathbf{x}_2, t_2) \tag{6.11}$$

for Markov stochastic processes. This identity is called the *Chapman-Kolmogorov equation*. Furthermore, the distinction between the second and third groups of identities vanishes for Markov stochastic processes.

Summary The probability density function $f(\mathbf{x}_1, t_1)$ and the conditional probability density function $f(\mathbf{x}_1, t_1 \,|\, \mathbf{x}, t)$ (often called the transitional probability density function) for a Markov stochastic process satisfy the identities

$$\begin{aligned} f(\mathbf{x}_1, t_1) &= \int_\Omega f(\mathbf{x}_1, t_1 \,|\, \mathbf{x}, t) \, f(\mathbf{x}, t) \, d\mathbf{x} \,, \\ f(\mathbf{x}_1, t_1 \,|\, \mathbf{x}_2, t_2) &= \int_\Omega f(\mathbf{x}_1, t_1 \,|\, \mathbf{x}, t) \, f(\mathbf{x}, t \,|\, \mathbf{x}_2, t_2) \, d\mathbf{x} \,. \end{aligned}$$

In these identities time is regarded as a parameter which can be either discrete or continuous.

When time is a continuous parameter the integral form of the Chapman-Kolmogorov property can be expressed as a partial differential equation under relatively weak conditions. The continuity of the process path with respect to time now becomes an issue to be addressed. The requirement of path continuity imposes restrictions on the form of the transitional density function $f(\mathbf{x}_1, t_1 \,|\, \mathbf{x}_2, t_2)$.

6.2.1 Path continuity

Suppose that a stochastic process starts at $(\mathbf{y}, t)$ and passes through $(\mathbf{x}, t+h)$. For a Markov process, the distribution of possible states $\mathbf{x}$ is perfectly captured by the transitional probability density

function $f(\mathbf{x}, t+h \,|\, \mathbf{y}, t)$ where $h \geq 0$. Given $\varepsilon > 0$, the probability that a path is more distant than ε from $\mathbf{y}$ is just

$$\int_{|\mathbf{x}-\mathbf{y}|>\varepsilon} f(\mathbf{x}, t+h \,|\, \mathbf{y}, t)\, d\mathbf{x}\,. \tag{6.12}$$

A process is said to be *path-continuous* if for any $\varepsilon > 0$

$$\int_{|\mathbf{x}-\mathbf{y}|>\varepsilon} f(\mathbf{x}, t+h \,|\, \mathbf{y}, t)\, d\mathbf{x} = o(h) \tag{6.13}$$

with probability one, and uniformly with respect to t, h and the state $\mathbf{y}$. In effect, the probability that a finite fraction of paths extend more than ε from y shrinks to zero faster than h. The concept of path-continuity is often captured by the definition

Path-Continuity Given any $\mathbf{x}, \mathbf{y} \in \Omega$ satisfying $|\mathbf{x}-\mathbf{y}| > \varepsilon$ and $t > 0$, assume that the limit

$$\lim_{h\to 0^+} \frac{f(\mathbf{x}, t+h \,|\, \mathbf{y}, t)}{h} = \phi(\mathbf{x} \,|\, \mathbf{y}, t\,)$$

exists and is uniform with respect to $\mathbf{x}$, $\mathbf{y}$ and t.
Path-continuity is characterised by the property $\phi(\mathbf{x} \,|\, \mathbf{y}, t\,) = 0$ when $\mathbf{x} \neq \mathbf{y}$.

Example

Investigate the path continuity of the one dimensional Markov processes defined by the transitional density functions

$$(a) \quad f(x, t+h \,|\, y, t) = \frac{1}{\sqrt{2\pi}\, h}\, e^{-(x-y)^2/2h^2}\,, \qquad (b) \quad f(x, t+h \,|\, y, t) = \frac{h}{\pi}\, \frac{1}{h^2 + (x-y)^2}\,.$$

Solution

It is demonstrated easily that both expressions are non-negative and integrate to unity with respect to x for any t, h and y. Each expression therefore defines a valid transitional density function. Note, in particular, that both densities satisfy automatically the consistency condition

$$\lim_{h\to 0^+} f(\mathbf{x}, t+h \,|\, \mathbf{y}, t) = \delta(x-y)\,.$$

Density (a) Consider

$$\begin{aligned}
\int_{|x-y|>\varepsilon} f(x, t+h \,|\, y, t)\, dx &= \int_{y+\varepsilon}^{\infty} \frac{1}{\sqrt{2\pi}\, h}\, e^{-(x-y)^2/2h^2}\, dx + \int_{-\infty}^{y-\varepsilon} \frac{1}{\sqrt{2\pi}\, h}\, e^{-(x-y)^2/2h^2}\, dx \\
&= \frac{2}{\sqrt{2\pi}\, h} \int_{\varepsilon}^{\infty} e^{-x^2/2h^2}\, dx \\
&= \frac{1}{\sqrt{\pi}} \int_{\varepsilon^2/2h^2}^{\infty} y^{-1/2}\, e^{-y}\, dy
\end{aligned}$$

Since $y \geq \varepsilon^2/2h^2$ in this integral then $y^{-1/2} \leq \sqrt{2}\,(h/\varepsilon)$. Thus

$$\frac{1}{h}\int_{|x-y|>\varepsilon} f(x,t+h\,|\,y,t)\,dx < \sqrt{\frac{2}{\pi}}\,\frac{1}{\varepsilon}\int_{\varepsilon^2/2h^2}^{\infty} e^{-y}\,dy = \sqrt{\frac{2}{\pi}}\,\frac{1}{\varepsilon}\left[1-e^{-\varepsilon^2/2h^2}\right] \to 0$$

as $h \to 0^+$, and therefore the stochastic process defined by the transitional probability density function (a) is path-continuous.

Density (b) Consider

$$\begin{aligned}
\int_{|x-y|>\varepsilon} f(x,t+h\,|\,y,t)\,dx &= \int_{y+\varepsilon}^{\infty}\frac{h}{\pi}\,\frac{dx}{h^2+(x-y)^2} + \int_{-\infty}^{y-\varepsilon}\frac{h}{\pi}\,\frac{dx}{h^2+(x-y)^2} \\
&= \frac{2h}{\pi}\int_{\varepsilon}^{\infty}\frac{dx}{h^2+x^2} \\
&= \frac{2}{\pi}\left[\,\tan^{-1}\frac{x}{h}\,\right]_{\varepsilon}^{\infty} \\
&= \frac{2}{\pi}\left[\,\frac{\pi}{2}-\tan^{-1}\frac{\varepsilon}{h}\,\right].
\end{aligned}$$

For fixed $\varepsilon > 0$, define $h = \varepsilon\tan\theta$ so that the limiting process $h \to 0^+$ may be replaced by the limiting process $\theta \to 0^+$. Thus

$$\frac{1}{h}\int_{|x-y|>\varepsilon} f(x,t+h\,|\,y,t)\,dx = \frac{2\theta}{\pi\varepsilon\tan\theta} \to \frac{2}{\pi\varepsilon} \neq 0$$

as $h \to 0^+$, and therefore the stochastic process defined by the transitional probability density function (b) is **not** path-continuous.

Remarks

The transitional density function (a) describes a *Brownian* stochastic process. Consequently, a Brownian path is continuous although it has no well-defined velocity. The transitional density function (b) describes a *Cauchy* stochastic process which, by comparison with the Brownian stochastic process, is seen to be discontinuous.

6.2.2 Drift and diffusion

Suppose that for a particular stochastic process the change of state in moving from $\mathbf{y}$ to $\mathbf{x}$, *i.e.* $(\mathbf{x}-\mathbf{y})$, occurs during the time interval $[t, t+h]$ with probability $f(\mathbf{x}, t+h\,|\,\mathbf{y}, t\,)$. The drift and diffusion of this process characterise respectively the expected rate of change in state, calculated over all paths terminating within distance ε of $\mathbf{y}$, and the rate of generation of the covariance of these changes. The drift and diffusion are defined as follows.

Drift The drift of a (Markov) process at $\mathbf{y}$ is defined to be the instantaneous rate of change of state for all processes close to $\mathbf{y}$ and is the value of the limit

$$\lim_{\varepsilon\to 0^+}\Big[\lim_{h\to 0^+}\frac{1}{h}\int_{|\mathbf{x}-\mathbf{y}|<\varepsilon}(\,\mathbf{x}-\mathbf{y}\,)\,f(\mathbf{x},t+h\,|\,\mathbf{y},t)\,d\mathbf{x}\Big]=\mathbf{a}(\mathbf{y},t)\,, \tag{6.14}$$

which is assumed to exist, and which requires that the inner limit converges uniformly with respect to $\mathbf{y}$, t and ε.

Diffusion The instantaneous covariance of a (Markov) process at $\mathbf{y}$ is defined to be the instantaneous rate of change of covariance for all processes close to $\mathbf{y}$ and is the value of the limit

$$\lim_{\varepsilon\to 0^+}\Big[\lim_{h\to 0^+}\frac{1}{h}\int_{|\mathbf{x}-\mathbf{y}|<\varepsilon}(\,\mathbf{x}-\mathbf{y}\,)\oplus(\,\mathbf{x}-\mathbf{y}\,)\,f(\mathbf{x},t+h\,|\,\mathbf{y},t)\,d\mathbf{x}\Big]=\mathbf{g}(\mathbf{y},t)\,, \tag{6.15}$$

which is assumed to exist, and which requires that the inner limit converges uniformly with respect to $\mathbf{y}$, t and ε.

Finally, all third and higher order moments computed over states $\mathbf{x},\mathbf{y}\in\Omega$ satisfying $|\mathbf{x}-\mathbf{y}|>\varepsilon$ necessarily vanish as $\varepsilon\to 0^+$ because the integral defining the k-th moment is $O(\varepsilon^{k-2})$ for $k\geq 2$.

6.2.3 Drift and diffusion of an SDE

Suppose that the time course of the process $\mathbf{x}(t)=\{x_1(t),\cdots,x_N(t)\}$ follows the stochastic differential equation

$$dx_j=\mu_j(\mathbf{x},t)\,dt+\sum_{\alpha=1}^{M}\sigma_{j\alpha}(\mathbf{x},t)\,dW_\alpha \tag{6.16}$$

where $1\leq j\leq N$ and $dW_1,\cdots,dW_M$ are increments in the correlated M Wiener processes $W_1,\cdots,W_M$ such that $\mathbb{E}^{\mathbb{P}}[\,dW_j\,dW_k\,]=Q_{jk}\,dt$. The issue is now to use the general definition of drift and diffusion to determine the drift and diffusion specifications in the case of the SDE 6.16.

The expected drift at $\mathbf{x}$ during $[t,t+h]$ is

$$\begin{aligned}a_j=\lim_{h\to 0^+}\frac{1}{h}\,\mathrm{E}\,[\,dx_j\,] &= \lim_{h\to 0^+}\frac{1}{h}\,\mathrm{E}\Big[\,\mu_j(\mathbf{x},t)\,h+\sum_{\alpha=1}^{M}\sigma_{j\alpha}(\mathbf{x},t)\,dW_\alpha\Big]\\ &= \lim_{h\to 0^+}\frac{1}{h}\,\mathbb{E}\,[\,\mu_j(\mathbf{x},t)\,h\,]=\mu_j(\mathbf{x},t)\,.\end{aligned} \tag{6.17}$$

The expected covariance at $\mathbf{x}$ during $[t, t+h]$ is

$$\begin{aligned} g_{jk} &= \lim_{h\to 0^+} \frac{1}{h}\,\mathrm{E}\,[\,(dx_j - \mu_j\,h)((dx_j - \mu_j\,h)\,] \\ &= \lim_{h\to 0^+} \frac{1}{h} \sum_{\alpha,\beta=1}^{M} \mathbb{E}\,[\,\sigma_{j\alpha}\,dW_\alpha\,\sigma_{k\beta}\,dW_\beta\,] \\ &= \lim_{h\to 0^+} \sum_{\alpha,\beta=1}^{M} \frac{\sigma_{j\alpha}\,\sigma_{k\beta}}{h}\,\mathbb{E}\,[\,dW_\alpha\,dW_\beta\,] = \sum_{\alpha,\beta=1}^{M} \sigma_{j\alpha}\,Q_{\alpha\beta}\,\sigma_{k\beta}\,. \end{aligned} \tag{6.18}$$

Evidently g_{jk} is a symmetric array. Furthermore, since Q is positive definite then $Q = F^{\mathrm{T}}F$ and $\Sigma\,Q\,\Sigma^{\mathrm{T}} = \Sigma\,F^{\mathrm{T}}\,F\Sigma^{\mathrm{T}} = (F\,\Sigma^{\mathrm{T}})^{\mathrm{T}}\,(F\Sigma^{\mathrm{T}})$ and therefore $G = [g_{jk}]$ is also positive definite, as required.

6.3 Formal derivation of the Forward Kolmogorov Equation

The construction of the differential equation representation of the Chapman-Kolmogorov condition follows closely the discussion of path-continuity, but admits the possibility of discontinuous paths by extending the analysis to include Markov processes for which

$$\int_{|\mathbf{x}-\mathbf{y}|>\varepsilon} f(\mathbf{x}, t+h\,|\,\mathbf{y}, t)\,d\mathbf{x} = O(h) \tag{6.19}$$

as $h \to 0^+$. The Cauchy stochastic process gives an example of a qualifying transitional probability density function. Suppose that $\varepsilon > 0$ is given together with a Markov process with transitional probability density function $f(\mathbf{x}_1, t_1\,|\,\mathbf{x}_2, t_2\,)$.

Let $\psi(\mathbf{x})$ be an arbitrary twice continuously differentiable function of state, then the rate of change of the expected value of ψ is defined by

$$\begin{aligned} \frac{\partial}{\partial t}\int_\Omega \psi(\mathbf{x})\,f(\mathbf{x}, t\,|\,\mathbf{y}, T)\,d\mathbf{x} &= \lim_{h\to 0^+} \frac{1}{h}\Big[\int_\Omega \psi(\mathbf{x})\,[\,f(\mathbf{x}, t+h\,|\,\mathbf{y}, T) - f(\mathbf{x}, t\,|\,\mathbf{y}, T)\,]\,d\mathbf{x}\Big] \\ &= \lim_{h\to 0^+} \frac{1}{h}\Big[\int_\Omega \psi(\mathbf{x})\Big(\int_\Omega f(\mathbf{x}, t+h\,|\,\mathbf{z}, t)\,f(\mathbf{z}, t\,|\,\mathbf{y}, T)\,d\mathbf{z}\Big)\,d\mathbf{x} \\ &\qquad - \int_\Omega \psi(\mathbf{z})\,f(\mathbf{z}, t\,|\,\mathbf{y}, T)\,d\mathbf{z}\Big] \\ &= \lim_{h\to 0^+} \frac{1}{h}\Big[\int_\Omega\int_\Omega \psi(\mathbf{x}) f(\mathbf{z}, t\,|\,\mathbf{y}, T)\,f(\mathbf{x}, t+h\,|\,\mathbf{z}, t)\,d\mathbf{x}\,d\mathbf{z} \\ &\qquad - \int_\Omega \psi(\mathbf{z})\,f(\mathbf{z}, t\,|\,\mathbf{y}, T)\,d\mathbf{z}\Big] \end{aligned} \tag{6.20}$$

where the Chapman-Kolmogorov property has been used on the first integral on the right hand side of equation (6.20). The sample space Ω is now divided into the regions $|\mathbf{x}-\mathbf{z}| < \varepsilon$ and $|\mathbf{x}-\mathbf{z}| > \varepsilon$

where ε is sufficiently small that $\psi(\mathbf{x})$ may be represented by the Taylor series expansion

$$\begin{aligned}\psi(\mathbf{x}) &= \psi(\mathbf{z}) + \sum_{k=1}^{N} (x_k - z_k)\,\frac{\partial \psi(\mathbf{z})}{\partial\, z_k} + \frac{1}{2}\,\sum_{j,k=1}^{N} (\,x_j - z_j\,)\,(x_k - z_k)\,\frac{\partial^2 \psi(\mathbf{z})}{\partial\, z_j\,\partial\, z_k} \\ &\qquad + \sum_{j,k=1}^{N} (x_j - z_j)(x_k - z_k)\,R_{jk}(\mathbf{x},\mathbf{z})\end{aligned} \tag{6.21}$$

within the region $|\mathbf{x} - \mathbf{z}| < \varepsilon$ and in which $R_{jk}(\mathbf{x},\mathbf{z}) \to 0$ as $|\mathbf{x} - \mathbf{z}| \to 0$. The right hand side of equation (6.21) is now manipulated through a number of steps after first replacing ψ(x) by its Taylor series and rearranging expression

$$\lim_{h\to 0^+} \frac{1}{h}\Big[\, \int_\Omega \int_\Omega \psi(\mathbf{x}) f(\mathbf{z},t\,|\,\mathbf{y},T)\, f(\mathbf{x},t+h\,|\,\mathbf{z},t)\, d\mathbf{x}\, d\mathbf{z} - \int_\Omega \psi(\mathbf{z})\, f(\mathbf{z},t\,|\,\mathbf{y},T)\, d\mathbf{z}\Big] \tag{6.22}$$

by means of the decomposition

$$\begin{aligned}\lim_{\varepsilon\to 0^+} \Big[\, &\lim_{h\to 0^+} \frac{1}{h} \int_\Omega \int_{|\mathbf{x}-\mathbf{z}|<\varepsilon} \psi(\mathbf{x})\, f(\mathbf{z},t\,|\,\mathbf{y},T)\, f(\mathbf{x},t+h\,|\,\mathbf{z},t)\, d\mathbf{x}\, d\mathbf{z} \\ &+ \lim_{h\to 0^+} \frac{1}{h} \int_\Omega \int_{|\mathbf{x}-\mathbf{z}|>\varepsilon} \psi(\mathbf{x})\, f(\mathbf{z},t\,|\,\mathbf{y},T)\, f(\mathbf{x},t+h\,|\,\mathbf{z},t)\, d\mathbf{x}\, d\mathbf{z} \\ &- \lim_{h\to 0^+} \frac{1}{h} \int_\Omega \psi(\mathbf{z})\, f(\mathbf{z},t\,|\,\mathbf{y},T)\, d\mathbf{z}\Big]\,.\end{aligned} \tag{6.23}$$

Substituting the Taylor series for $\psi(\mathbf{x})$ into expression (6.23) gives

$$\begin{aligned}\lim_{\varepsilon\to 0^+} \lim_{h\to 0^+} \frac{1}{h}\Big[\, &\int_\Omega \int_{|\mathbf{x}-\mathbf{z}|<\varepsilon} \psi(\mathbf{z})\, f(\mathbf{z},t\,|\,\mathbf{y},T)\, f(\mathbf{x},t+h\,|\,\mathbf{z},t)\, d\mathbf{x}\, d\mathbf{z} \\ &+ \int_\Omega \int_{|\mathbf{x}-\mathbf{z}|<\varepsilon} (x_k - z_k)\,\frac{\partial \psi(\mathbf{z})}{\partial\, z_k}\, f(\mathbf{z},t\,|\,\mathbf{y},T)\, f(\mathbf{x},t+h\,|\,\mathbf{z},t)\, d\mathbf{x}\, d\mathbf{z} \\ &+ \frac{1}{2}\sum_{j,k=1}^{N} \int_\Omega \int_{|\mathbf{x}-\mathbf{z}|<\varepsilon} (x_j - z_j)(x_k - z_k)\,\frac{\partial^2 \psi(\mathbf{z})}{\partial\, z_j\,\partial\, z_k}\, f(\mathbf{z},t\,|\,\mathbf{y},T)\, f(\mathbf{x},t+h\,|\,\mathbf{z},t)\, d\mathbf{x}\, d\mathbf{z} \\ &+ \sum_{j,k=1}^{N} \int_\Omega \int_{|\mathbf{x}-\mathbf{z}|<\varepsilon} (x_j - z_j)(x_k - z_k)\, R_{jk}(\mathbf{x},\mathbf{z})\, f(\mathbf{z},t\,|\,\mathbf{y},T)\, f(\mathbf{x},t+h\,|\,\mathbf{z},t)\, d\mathbf{x}\, d\mathbf{z} \\ &+ \int_\Omega \int_{|\mathbf{x}-\mathbf{z}|>\varepsilon} \psi(\mathbf{x})\, f(\mathbf{z},t\,|\,\mathbf{y},T)\, f(\mathbf{x},t+h\,|\,\mathbf{z},t)\, d\mathbf{x}\, d\mathbf{z} - \int_\Omega \psi(\mathbf{z})\, f(\mathbf{z},t\,|\,\mathbf{y},T)\, d\mathbf{z}\Big]\,.\end{aligned} \tag{6.24}$$

The definition of instantaneous drift in formula (6.14) is now used to deduce that

$$\begin{aligned}\lim_{\varepsilon\to 0^+} \lim_{h\to 0^+} \frac{1}{h} \sum_{k=1}^{N} \int_\Omega \int_{|\mathbf{x}-\mathbf{z}|<\varepsilon} (x_k - z_k)\,\frac{\partial \psi(\mathbf{z})}{\partial\, z_k}\, f(\mathbf{z},t\,|\,\mathbf{y},T)\, f(\mathbf{x},t+h\,|\,\mathbf{z},t)\, d\mathbf{x}\, d\mathbf{z} \\ = \sum_{k=1}^{N} \int_\Omega a_k(\mathbf{z},t)\,\frac{\partial \psi(\mathbf{z})}{\partial\, z_k}\, f(\mathbf{z},t\,|\,\mathbf{y},T)\, d\mathbf{z}\,.\end{aligned} \tag{6.25}$$

Similarly, the definition of instantaneous covariance in formula (6.15) is used to deduce that

$$\begin{aligned}\lim_{\varepsilon\to 0^+} \lim_{h\to 0^+} \frac{1}{2h} \sum_{j,k=1}^{N} \int_\Omega \int_{|\mathbf{x}-\mathbf{z}|<\varepsilon} (x_j - z_j)(x_k - z_k)\,\frac{\partial^2 \psi(\mathbf{z})}{\partial\, z_j\,\partial\, z_k}\, f(\mathbf{z},t\,|\,\mathbf{y},T)\, f(\mathbf{x},t+h\,|\,\mathbf{z},t)\, d\mathbf{x}\, d\mathbf{z} \\ = \frac{1}{2}\sum_{j,k=1}^{N} \int_\Omega g_{jk}(\mathbf{z},t)\,\frac{\partial^2 \psi(\mathbf{z})}{\partial\, z_j\,\partial\, z_k}\, f(\mathbf{z},t\,|\,\mathbf{y},T)\, d\mathbf{z}\,.\end{aligned} \tag{6.26}$$

Moreover, since $R_{jk}(\mathbf{x},\mathbf{z}) \to 0$ as $|\mathbf{x}-\mathbf{z}| \to 0$ then it is obvious that the contribution from the remainder of the Taylor series becomes arbitrarily small as $\varepsilon \to 0^+$ and therefore

$$\lim_{\varepsilon\to 0^+} \lim_{h\to 0^+} \frac{1}{h} \sum_{j,k=1}^{N} \int_\Omega \int_{|\mathbf{x}-\mathbf{z}|<\varepsilon} (x_j - z_j)(x_k - z_k) R_{jk}(\mathbf{x},\mathbf{z})\, f(\mathbf{z},t\,|\,\mathbf{y},T)\, f(\mathbf{x},t+h\,|\,\mathbf{z},t)\, d\mathbf{x}\, d\mathbf{z} = 0\,.$$

This result, together with (6.25) and (6.26) enable expression (6.24) to be further simplified to

$$\begin{aligned}\lim_{\varepsilon\to 0^+} \lim_{h\to 0^+} \frac{1}{h} \Big[& \int_\Omega \int_{|\mathbf{x}-\mathbf{z}|<\varepsilon} \psi(\mathbf{z})\, f(\mathbf{z},t\,|\,\mathbf{y},T)\, f(\mathbf{x},t+h\,|\,\mathbf{z},t)\, d\mathbf{x}\, d\mathbf{z} \\ & + \int_\Omega \int_{|\mathbf{x}-\mathbf{z}|>\varepsilon} \psi(\mathbf{x})\, f(\mathbf{z},t\,|\,\mathbf{y},T)\, f(\mathbf{x},t+h\,|\,\mathbf{z},t)\, d\mathbf{x}\, d\mathbf{z} - \int_\Omega \phi(\mathbf{z})\, f(\mathbf{z},t\,|\,\mathbf{y},T)\, d\mathbf{z} \Big] \\ & + \sum_{k=1}^{N} \int_\Omega a_k(\mathbf{z},t) \frac{\partial \psi(\mathbf{z})}{\partial\, z_k} f(\mathbf{z},t\,|\,\mathbf{y},T)\, d\mathbf{z} + \frac{1}{2} \sum_{j,k=1}^{N} \int_\Omega g_{jk}(\mathbf{z},t) \frac{\partial^2 \psi(\mathbf{z})}{\partial\, z_j\, \partial\, z_k} f(\mathbf{z},t\,|\,\mathbf{y},T)\, d\mathbf{z}\,.\end{aligned} \tag{6.27}$$

However, $f(\mathbf{x},t+h\,|\,\mathbf{z},t)$ is itself a probability density function and therefore satisfies

$$\int_\Omega f(\mathbf{x},t+h\,|\,\mathbf{z},t)\, d\mathbf{x} = 1\,.$$

Consequently, the first integral in expression (6.27) can be manipulated into the form

$$\begin{aligned}& \int_\Omega \int_{|\mathbf{x}-\mathbf{z}|<\varepsilon} \psi(\mathbf{z})\, f(\mathbf{z},t\,|\,\mathbf{y},T)\, f(\mathbf{x},t+h\,|\,\mathbf{z},t)\, d\mathbf{x}\, d\mathbf{z} \\ & \quad = \int_\Omega \psi(\mathbf{z})\, f(\mathbf{z},t\,|\,\mathbf{y},T) \Big(\int_\Omega f(\mathbf{x},t+h\,|\,\mathbf{z},t)\, d\mathbf{x} - \int_{|\mathbf{x}-\mathbf{z}|>\varepsilon} f(\mathbf{x},t+h\,|\,\mathbf{z},t)\, d\mathbf{x} \Big)\, d\mathbf{z} \\ & \quad = \int_\Omega \psi(\mathbf{z})\, f(\mathbf{z},t\,|\,\mathbf{y},T)\, d\mathbf{z} - \int_\Omega \int_{|\mathbf{x}-\mathbf{z}|>\varepsilon} \psi(\mathbf{z})\, f(\mathbf{z},t\,|\,\mathbf{y},T)\, f(\mathbf{x},t+h\,|\,\mathbf{z},t)\, d\mathbf{x}\, d\mathbf{z}\,.\end{aligned} \tag{6.28}$$

The first integral appearing in expression (6.27) is now replaced by equation (6.28). After some obvious cancellation, the expression becomes

$$\begin{aligned}\lim_{\varepsilon\to 0^+} \lim_{h\to 0^+} \frac{1}{h} \Big[& \int_\Omega \int_{|\mathbf{x}-\mathbf{z}|>\varepsilon} \psi(\mathbf{x})\, f(\mathbf{z},t\,|\,\mathbf{y},T)\, f(\mathbf{x},t+h\,|\,\mathbf{z},t)\, d\mathbf{x}\, d\mathbf{z} \\ & - \int_\Omega \int_{|\mathbf{x}-\mathbf{z}|>\varepsilon} \psi(\mathbf{z})\, f(\mathbf{z},t\,|\,\mathbf{y},T)\, f(\mathbf{x},t+h\,|\,\mathbf{z},t)\, d\mathbf{x}\, d\mathbf{z} \Big] \\ & + \sum_{k=1}^{N} \int_\Omega a_k(\mathbf{z},t) \frac{\partial \psi(\mathbf{z})}{\partial\, z_k} f(\mathbf{z},t\,|\,\mathbf{y},T)\, d\mathbf{z} + \frac{1}{2} \sum_{j,k=1}^{N} \int_\Omega g_{jk}(\mathbf{z},t) \frac{\partial^2 \psi(\mathbf{z})}{\partial\, z_j\, \partial\, z_k} f(\mathbf{z},t\,|\,\mathbf{y},T)\, d\mathbf{z}\,.\end{aligned} \tag{6.29}$$

The path-continuity condition (6.12) is now used to treat the limits in expression (6.29). Clearly

$$\begin{aligned}\lim_{\varepsilon\to 0^+} \lim_{h\to 0^+} \frac{1}{h} \Big[& \int_\Omega \int_{|\mathbf{x}-\mathbf{z}|>\varepsilon} \psi(\mathbf{x})\, f(\mathbf{z},t\,|\,\mathbf{y},T)\, f(\mathbf{x},t+h\,|\,\mathbf{z},t)\, d\mathbf{x}\, d\mathbf{z} \\ & - \int_\Omega \int_{|\mathbf{x}-\mathbf{z}|>\varepsilon} \psi(\mathbf{z})\, f(\mathbf{z},t\,|\,\mathbf{y},T)\, f(\mathbf{x},t+h\,|\,\mathbf{z},t)\, d\mathbf{x}\, d\mathbf{z} \Big] \\ & = \int_\Omega \int_\Omega \psi(\mathbf{x})\, f(\mathbf{z},t\,|\,\mathbf{y},T)\, \phi(\mathbf{x},\mathbf{z})\, d\mathbf{x}\, d\mathbf{z} - \int_\Omega \int_\Omega \psi(\mathbf{z})\, f(\mathbf{z},t\,|\,\mathbf{y},T)\, \phi(\mathbf{x},\mathbf{z})\, d\mathbf{x}\, d\mathbf{z} \Big] \\ & = \int_\Omega \int_\Omega \psi(\mathbf{z}) \Big[f(\mathbf{x},t\,|\,\mathbf{y},T)\, \phi(\mathbf{z},\mathbf{x}) - f(\mathbf{z},t\,|\,\mathbf{y},T)\, \phi(\mathbf{x},\mathbf{z}) \Big]\, d\mathbf{x}\, d\mathbf{z}\,,\end{aligned} \tag{6.30}$$

and consequently expression (6.24) is finally simplified to

$$\begin{aligned}&\int_\Omega\int_\Omega \psi(\mathbf{z})\,\Big[\,f(\mathbf{x},t\,|\,\mathbf{y},T)\,\phi(\mathbf{z},\mathbf{x})-f(\mathbf{z},t\,|\,\mathbf{y},T)\,\phi(\mathbf{x},\mathbf{z})\,\Big]\,d\mathbf{x}\,d\mathbf{z}\\&\quad+\sum_{k=1}^{N}\int_\Omega a_k(\mathbf{z},t)\,\frac{\partial\psi(\mathbf{z})}{\partial\,z_k}\,f(\mathbf{z},t\,|\,\mathbf{y},T)\,d\mathbf{z}+\frac{1}{2}\sum_{j,k=1}^{N}\int_\Omega g_{jk}(\mathbf{z},t)\,\frac{\partial^2\psi(\mathbf{z})}{\partial\,z_j\,\partial\,z_k}\,f(\mathbf{z},t\,|\,\mathbf{y},T)\,d\mathbf{z}\,.\end{aligned}\tag{6.31}$$

Expression (6.31) is now incorporated into equation (6.20) to obtain the final form

$$\begin{aligned}\int_\Omega \psi(\mathbf{z})\,\frac{\partial\,f(\mathbf{z},t\,|\,\mathbf{y},T)}{\partial t}\,d\mathbf{z}&=\int_\Omega\int_\Omega \psi(\mathbf{z})\,\Big[\,f(\mathbf{x},t\,|\,\mathbf{y},T)\,\phi(\mathbf{z},\mathbf{x})-f(\mathbf{z},t\,|\,\mathbf{y},T)\,\phi(\mathbf{x},\mathbf{z})\,\Big]\,d\mathbf{x}\,d\mathbf{z}\\&\quad+\sum_{k=1}^{N}\int_\Omega a_k(\mathbf{z},t)\,\frac{\partial\psi(\mathbf{z})}{\partial\,z_k}\,f(\mathbf{z},t\,|\,\mathbf{y},T)\,d\mathbf{z}+\frac{1}{2}\sum_{j,k=1}^{N}\int_\Omega g_{jk}(\mathbf{z},t)\,\frac{\partial^2\psi(\mathbf{z})}{\partial\,z_j\,\partial\,z_k}\,f(\mathbf{z},t\,|\,\mathbf{y},T)\,d\mathbf{z}\,.\end{aligned}\tag{6.32}$$

Let the k-th component of $\mathbf{a}(\mathbf{z},t)$ be denoted by $a_k(\mathbf{z},t)$, then Gauss's theorem gives

$$\begin{aligned}&\int_\Omega\sum_{k=1}^{N} a_k(\mathbf{z},t)\,\frac{\partial\psi(\mathbf{z})}{\partial\,z_k}\,f(\mathbf{z},t\,|\,\mathbf{y},T)\,d\mathbf{z}\\&\quad=\int_\Omega\sum_{k=1}^{N}\frac{\partial}{\partial z_k}\Big(\,a_k\,\psi(\mathbf{z})\,f(\mathbf{z},t\,|\,\mathbf{y},T)\Big)\,dV-\int_\Omega\psi(\mathbf{z})\,\frac{\partial}{\partial z_k}\Big(a_k\,f(\mathbf{z},t\,|\,\mathbf{y},T)\Big)\,dV\\&\quad=\int_{\partial\Omega}\sum_{k=1}^{N} n_k a_k\,\psi(\mathbf{z})\,f(\mathbf{z},t\,|\,\mathbf{y},T)\,dV-\int_\Omega\psi(\mathbf{z})\,\frac{\partial}{\partial z_k}\Big(a_k\,f(\mathbf{z},t\,|\,\mathbf{y},T)\Big)\,dV\,.\end{aligned}\tag{6.33}$$

Similarly, let $\mathbf{g}(\mathbf{z},t)$ denote the symmetric $N\times N$ array with $(j,k)^{th}$ component $g_{jk}(\mathbf{z},t)$, then Gauss's theorem gives

$$\begin{aligned}&\int_\Omega\sum_{j,k=1}^{N} g_{jk}(\mathbf{z},t)\,\frac{\partial^2\psi(\mathbf{z})}{\partial\,z_j\,\partial\,z_k}\,f(\mathbf{z},t\,|\,\mathbf{y},T)\,d\mathbf{z}\\&\quad=\int_\Omega\sum_{j,k=1}^{N}\frac{\partial}{\partial z_j}\Big(\frac{\partial\psi}{\partial z_k}\,g_{jk}(\mathbf{z},t)\,f(\mathbf{z},t\,|\,\mathbf{y},T)\Big)\,dV-\int_\Omega\sum_{j,k=1}^{N}\frac{\partial\psi}{\partial z_k}\,\frac{\partial}{\partial x_j}\Big(g_{jk}(\mathbf{z},t)\,f(\mathbf{z},t\,|\,\mathbf{y},T)\Big)\,dV\\&\quad=\int_{\partial\Omega}\sum_{j,k=1}^{N} n_j\frac{\partial\psi}{\partial z_k}\,g_{jk}(\mathbf{z},t)\,f(\mathbf{z},t\,|\,\mathbf{y},T)\,dA-\int_\Omega\sum_{j,k=1}^{N}\frac{\partial}{\partial z_k}\Big(\psi(\mathbf{z},t)\frac{\partial}{\partial x_j}\Big(g_{jk}(\mathbf{z},t)\,f(\mathbf{z},t\,|\,\mathbf{y},T)\Big)\Big)\,dV\\&\qquad+\int_\Omega\sum_{j,k=1}^{N}\psi(\mathbf{z})\,\frac{\partial^2}{\partial z_j\partial z_k}\Big(g_{jk}(\mathbf{z},t)\,f(\mathbf{z},t\,|\,\mathbf{y},T)\Big)\,dV\\&\quad=\int_{\partial\Omega}\sum_{j,k=1}^{N} n_j\Big[\frac{\partial\psi}{\partial z_k}\,g_{jk}(\mathbf{z},t)\,f(\mathbf{z},t\,|\,\mathbf{y},T)-\psi(\mathbf{z},t)\frac{\partial}{\partial x_k}\Big(g_{jk}(\mathbf{z},t)\,f(\mathbf{z},t\,|\,\mathbf{y},T)\Big)\Big]\,dA\\&\qquad+\int_\Omega\sum_{j,k=1}^{N}\psi(\mathbf{z})\,\frac{\partial^2}{\partial z_j\partial z_k}\Big(g_{jk}(\mathbf{z},t)\,f(\mathbf{z},t\,|\,\mathbf{y},T)\Big)\,dV\,.\end{aligned}\tag{6.34}$$

Identities (6.33) and (6.34) are now introduced into equation (6.31) to obtain on re-arrangement

$$\begin{aligned}
&\int_\Omega \psi(\mathbf{z}) \Big[\frac{\partial\, f(\mathbf{z},t\,|\,\mathbf{y},T)}{\partial t}\, d\mathbf{z} + \frac{\partial}{\partial z_k}\Big(\sum_{k=1}^{N} a_k\, f(\mathbf{z},t\,|\,\mathbf{y},T)\Big) - \frac{1}{2}\sum_{j,k=1}^{N} \frac{\partial^2}{\partial z_j \partial z_k}\Big(g_{jk}\, f(\mathbf{z},t\,|\,\mathbf{y},T)\Big) \\
&\qquad - \int_\Omega \Big(f(\mathbf{x},t\,|\,\mathbf{y},T)\,\phi(\mathbf{z},\mathbf{x}) - f(\mathbf{z},t\,|\,\mathbf{y},T)\,\phi(\mathbf{x},\mathbf{z})\Big)\, d\mathbf{x}\Big]\, d\mathbf{z} \\
&= \int_{\partial\Omega} \psi(\mathbf{z})\, n_k \Big[f(\mathbf{z},t\,|\,\mathbf{y},T)\, a_k - \frac{1}{2}\frac{\partial}{\partial z_j}\Big(g_{jk}\, f(\mathbf{z},t\,|\,\mathbf{y},T)\Big)\Big]\, dA \\
&\qquad + \frac{1}{2}\int_{\partial\Omega} f(\mathbf{z},t\,|\,\mathbf{y},T)\, \frac{\partial \psi}{\partial z_j}\, g_{jk}\, n_k\, dA\,.
\end{aligned} \tag{6.35}$$

However, equation (6.35) is an identity in the sense that it is valid for all choices of $\phi(\mathbf{z})$. The discussion of the surface contributions in equation (6.35) is generally problem dependent. For example, a well known class of problem requires the stochastic process to be restored to the interior of Ω when it strikes $\partial\Omega$. Cash-flow problems fall into this category. Such problems need special treatment and involve point sources of probability.

In the absence of restoration and assuming that the stochastic process is constrained to the region Ω, the transitional density function and its normal derivative are both zero on $\partial\Omega$, that is,

$$\frac{\partial\, f(\mathbf{x},t\,|\,\mathbf{y},T)}{\partial\, \mathbf{x}} = f(\mathbf{x},t\,|\,\mathbf{y},T) = 0\,, \qquad \mathbf{x}\in\partial\Omega\,.$$

The surface terms in (6.35) make no contribution and therefore

$$\begin{aligned}
&\int_\Omega \phi(\mathbf{z}) \Big[\frac{\partial\, f(\mathbf{z},t\,|\,\mathbf{y},T)}{\partial t} + \sum_{k=1}^{N} \frac{\partial}{\partial z_k}\Big(a_k\, f(\mathbf{z},t\,|\,\mathbf{y},T)\Big) - \frac{1}{2}\sum_{j,k=1}^{N} \frac{\partial^2}{\partial z_j \partial z_k}\Big(g_{jk}\, f(\mathbf{z},t\,|\,\mathbf{y},T)\Big)\Big]\, d\mathbf{z} \\
&\qquad = \int_\Omega \psi(\mathbf{z}) \Big[\int_\Omega \Big(f(\mathbf{x},t\,|\,\mathbf{y},T)\,\phi(\mathbf{z},\mathbf{x}) - f(\mathbf{z},t\,|\,\mathbf{y},T)\,\phi(\mathbf{x},\mathbf{z})\Big)\, d\mathbf{x}\Big]\, d\mathbf{z}
\end{aligned} \tag{6.36}$$

for all twice continuously differentiable functions $f(\mathbf{z})$. Thus $f(\mathbf{z},t\,|\,\mathbf{y},T)$ satisfies the partial differential equation

$$\begin{aligned}
\frac{\partial\, f(\mathbf{z},t\,|\,\mathbf{y},T)}{\partial t} &= \sum_{k=1}^{N} \frac{\partial}{\partial z_k}\Big(\frac{1}{2}\sum_{j=1}^{N} \frac{\partial\,[\, g_{jk}(\mathbf{z},t)\, f(\mathbf{z},t\,|\,\mathbf{y},T)\,]}{\partial\, z_j} - a_k(\mathbf{z},t)\, f(\mathbf{z},t\,|\,\mathbf{y},T)\Big) \\
&\quad + \int_\Omega \Big(f(\mathbf{x},t\,|\,\mathbf{y},T)\,\phi(\mathbf{z},\mathbf{x}) - f(\mathbf{z},t\,|\,\mathbf{y},T)\,\phi(\mathbf{x},\mathbf{z})\Big)\, d\mathbf{x}\,.
\end{aligned} \tag{6.37}$$

Equation (6.37) is the differential form of the Chapman-Kolmogorov condition for a Markov process which is continuous in space and time. The equation is solved with the initial condition

$$f(\mathbf{z},T\,|\,\mathbf{y},T) = \delta(\mathbf{z}-\mathbf{y})$$

and the boundary conditions

$$\frac{\partial\, f(\mathbf{x},t\,|\,\mathbf{y},T)}{\partial\, \mathbf{x}} = f(\mathbf{x},t\,|\,\mathbf{y},T) = 0\,, \qquad \mathbf{x}\in\partial\Omega\,.$$

Remarks

(a) The Chapman-Kolmogorov equation is written for the transitional density function of a Markov process and does not overtly arise from a stochastic differential equation. However, for the SDE

$$dx_j = \mu_j(\mathbf{x},t)\,dt + \sum_{\alpha=1}^{M} \sigma_{j\alpha}(\mathbf{x},t)\,dW_\alpha$$

it has been shown that

$$a_k(\mathbf{x},t) = \mu_k(\mathbf{x},t)\,, \qquad g_{jk}(\mathbf{x},t) = \sum_{\alpha,\beta=1}^{M} \sigma_{j\alpha}(\mathbf{x},t)\,Q_{\alpha\beta}\,\sigma_{k\beta}(\mathbf{x},t)\,, \tag{6.38}$$

and therefore in the absence of discontinuous paths, commonly called *jump processes*, the Chapman-Kolmogorov equations for the prototypical SDE takes the form

$$\frac{\partial\, f(\mathbf{z},t\,|\,\mathbf{y},T)}{\partial t} = \sum_{k=1}^{N} \frac{\partial}{\partial z_k}\Big(\frac{1}{2}\sum_{j=1}^{N} \frac{\partial\,[\,g_{jk}(\mathbf{z},t)\,f(\mathbf{z},t\,|\,\mathbf{y},T)\,]}{\partial\, z_j} - a_k(\mathbf{z},t)\,f(\mathbf{z},t\,|\,\mathbf{y},T)\Big)\,. \tag{6.39}$$

(b) Equation (6.39) is often called the "Forward Kolmogorov" equation or the "Fokker-Planck" equation because of its independent discovery by Andrey Kolmogorov (1931) and Max. Planck with his doctoral student Adriaan Fokker in (1913). The term "forward" refers to the fact that the equation is written for time and the "forward" variable $\mathbf{y}$. There is, of course, a "Backward Kolmogorov" equation written for the variable $\mathbf{x}$.

6.3.1 Intuitive derivation of the Forward Kolmogorov Equation

Let $\psi(\mathbf{x})$ be an arbitrary twice continuously differentiable function of state, then the forward Kolmogorov equation is constructed by asserting that

$$\frac{d}{dt}\int_\Omega \psi(x) f(t,x)\,dx = \int_\Omega \mathbb{E}\Big[\frac{d\psi(x)}{dt}\Big] f(t,x)\,dx\,. \tag{6.40}$$

Put simply, the rate of change of the expectation of ψ is equal to the expectation of the rate of change of ψ. Clearly the left hand side of equation (6.40) has value

$$\int_\Omega \psi(x)\frac{\partial f(t,x)}{\partial t}\,dx\,.$$

Let $\phi(x\,|\,y,t)$ denote the rate of the jump process from y at time t to x at time t^+. Ito's Lemma applied to $\psi(x)$ gives

$$\begin{aligned}\mathbb{E}[d\psi] \;=\; & \sum_{k=1}^{N} \frac{\partial\psi}{\partial x_k}\,\mu_k\,dt + \frac{1}{2}\sum_{j,k=1}^{N} \frac{\partial^2\psi}{\partial x_j \partial x_k}\,g_{jk}\,dt \\ & + dt\Big[\int \psi(x)\Big(\int \phi(x\,|\,y,t) f(t,y)\,dy - f(t,x)\int \phi(y\,|\,x,t)\,dy\Big)\,dx\Big]\end{aligned}$$

from which it follows immediately that

$$\begin{aligned}\mathbb{E}\Big[\frac{d\psi}{dt}\Big] &= \sum_{k=1}^{N}\frac{\partial\psi}{\partial x_k}\,\mu_k + \frac{1}{2}\sum_{j,k=1}^{N}\frac{\partial^2\psi}{\partial x_j\partial x_k}\,g_{jk} \\ &\quad + \int \psi(x)\Big(\int \phi(x\,|\,y,t)f(t,y)\,dy - f(t,x)\int \phi(y\,|\,x,t)\,dy\Big)\,dx\,.\end{aligned}$$

The divergence theorem is now applied to the right hand side of equation (6.40) to obtain

$$\begin{aligned}\int_\Omega \mathbb{E}\Big[\frac{d\psi(x)}{dt}\Big]f(t,x)\,dx &= \int_\Omega \sum_{k=1}^{N}\Big(\frac{\partial\psi}{\partial x_k}\,\mu_k + \frac{1}{2}\sum_{j,k=1}^{N}\frac{\partial^2\psi}{\partial x_j\partial x_k}\,g_{jk}\Big)f\,dx \\ &\quad + \int \psi(x)\Big(\int \phi(x\,|\,y,t)f(t,y)\,dy - f(t,x)\int \phi(y\,|\,x,t)\,dy\Big)\,dx \\ &= -\int_\Omega \sum_{k=1}^{N}\psi\frac{\partial(\mu_k f)}{\partial x_k}\,dx + \int_{\partial\Omega}\sum_{k=1}^{N}\psi\,\mu_k f n_k\,dS \\ &\quad + \frac{1}{2}\int_\Omega \sum_{j,k=1}^{N}\frac{\partial\psi}{\partial x_j}\,(g_{jk}f)n_k\,dS - \frac{1}{2}\int_\Omega \frac{\partial\psi}{\partial x_j}\frac{\partial(g_{jk}f)}{\partial x_k}\,dx \\ &\quad + \int \psi(x)\Big(\int \phi(x\,|\,y,t)f(t,y)\,dy - f(t,x)\int \phi(y\,|\,x,t)\,dy\Big)\,dx \\ &= -\int_{\partial\Omega}\sum_{k=1}^{N}\psi\,\Big(\mu_k f - \frac{1}{2}\frac{\partial(g_{kj}f)}{\partial x_j}\Big)n_k\,dS + \frac{1}{2}\int_\Omega \sum_{j,k=1}^{N}\frac{\partial\psi}{\partial x_j}\,(g_{jk}f)n_k\,dS \\ &\quad + \int_\Omega \psi\Big(\frac{1}{2}\frac{\partial^2(g_{jk}f)}{\partial x_j\partial x_k} - \frac{\partial(\mu_k f)}{\partial x_k}\Big)\,dx \\ &\quad + \int \psi(x)\Big(\int \phi(x\,|\,y,t)f(t,y)\,dy - f(t,x)\int \phi(y\,|\,x,t)\,dy\Big)\,dx\,.\end{aligned}$$

The surface contributions to the previous equation are now set to zero. Thus equation (6.40) leads to the identity

$$\begin{aligned}\int_\Omega \psi(x)\frac{\partial f(t,x)}{\partial t}\,dx &= \int_\Omega \psi\frac{\partial}{\partial x_k}\Big(\frac{1}{2}\frac{\partial(g_{jk}f)}{\partial x_j} - \mu_k f\Big)\,dx \\ &\quad + \int \psi(x)\Big(\int \phi(x\,|\,y,t)f(t,y)\,dy - f(t,x)\int \phi(y\,|\,x,t)\,dy\Big)\,dx\,.\end{aligned} \tag{6.41}$$

which must hold for all suitably differentiable choices of the function $\psi(x)$. In conclusion,

$$\frac{\partial f(t,x)}{\partial t} = \frac{\partial}{\partial x_k}\Big(\frac{1}{2}\frac{\partial(g_{jk}f)}{\partial x_j} - \mu_k f\Big) + \Big(\int \phi(x\,|\,y,t)f(t,y)\,dy - f(t,x)\int \phi(y\,|\,x,t)\,dy\Big)\,. \tag{6.42}$$

Of course, $\int \phi(y\,|\,x,t)\,dy$ is simply the rate at which the jump mechanism removes density from x. Typically this will be the unconditional rate of the jump process.

6.3.2 The Backward Kolmogorov equation

The construction of the backward Kolmogorov equation is a straightforward application of Ito's lemma. It follows immediately from Ito's lemma that $f(\mathbf{z}, t \,|\, \mathbf{y}, T)$ satisfies

$$df = \frac{\partial f}{\partial t}\, dt + \sum_{j=1}^{N} \frac{\partial f}{\partial y_j}\, dy_j + \frac{1}{2} \sum_{j,k=1}^{N} \frac{\partial^2 f}{\partial y_j \partial y_k}\, dy_j\, dy_k$$

in the absence of jumps, and where $\mathbf{y}$ now denote backward variables which one may regard as spot values or initial conditions. When probability density is conserved then $\mathbb{E}[\, df \,] = 0$ and therefore

$$\begin{aligned} 0 &= \frac{\partial f}{\partial t}\, dt + \sum_{j=1}^{N} \frac{\partial f}{\partial y_j}\, \mathbb{E}[\, dy_j \,] + \frac{1}{2} \sum_{j,k=1}^{N} \frac{\partial^2 f}{\partial y_j \partial y_k}\, \mathbb{E}[\, dy_j\, dy_k \,] \\ &= \frac{\partial f}{\partial t}\, dt + \sum_{j=1}^{N} \frac{\partial f}{\partial y_j}\, \mu_j(\mathbf{y})\, dt + \frac{1}{2} \sum_{j,k=1}^{N} \frac{\partial^2 f}{\partial y_j \partial y_k}\, g_{jk}\, dt \end{aligned}$$

from which it follows immediately that

$$\frac{\partial f}{\partial t} + \sum_{j=1}^{N} \mu_j(\mathbf{y})\, \frac{\partial f}{\partial y_j} + \frac{1}{2} \sum_{j,k=1}^{N} g_{jk}(\mathbf{y})\, \frac{\partial^2 f}{\partial y_j \partial y_k} = 0\,. \tag{6.43}$$

Example

Use the Backward Kolmogorov Equation to construct the characteristic function of Heston's model of stochastic volatility, namely

$$\begin{aligned} dS &= S(r\, dt + \sqrt{V}\, dW_1) \\ dV &= \kappa(\theta - V)\, dt + \sigma \sqrt{V}\, dW_2\,, \end{aligned} \tag{6.44}$$

where S and V are respectively the spot values of asset price and diffusion, dW_1 and dW_2 are increments in the correlated Wiener processes W_1 and W_2 such that $\mathbb{E}[dW_1 dW_2] = \rho\, dt$, and the parameters r (risk-free rate of interest), κ (coefficient of mean reversion), θ (mean level of the rate of diffusion) and σ (volatility of volatility) take constant values.

Solution

Start by setting $Y = \log S$ and applying Ito's lemma to the first of equations (6.44) gives

$$dY = \frac{1}{S}(Sr\, dt + S\sqrt{V}\, dW_1) + \frac{1}{2}\Big(- \frac{1}{S^2}\Big) S^2 V\, dt \quad \to \quad dY = (r - V/2)\, dt + \sqrt{V}\, dW_1\,.$$

Let the transitional density of the Heston process be $f = f(Y, V, t \,|\, y, v, t = T)$ in which (Y, V) are spot values at time $t \le T$ and (y, v) is the terminal state at time T. Thus (Y, V) are the backward variables and (y, v) are the forward variables. The Backward Kolmogorov Equation asserts that

$$\frac{\partial f}{\partial t} + (r - V/2)\frac{\partial f}{\partial Y} + \kappa(\theta - V)\frac{\partial f}{\partial V} + \frac{1}{2}\Big(V \frac{\partial^2 f}{\partial Y^2} + 2\rho\sigma V \frac{\partial^2 f}{\partial Y \partial V} + \sigma^2 V \frac{\partial^2 f}{\partial V^2} \Big) = 0\,.$$

The characteristic function of the transitional density is the Fourier Transform of this density with respect to the forward variables, that is,

$$\phi(Y,R,t) = \int_{\mathbb{R}^2} f(Y,V,t\,|\,y,v,t=T)\,e^{i(\omega_y\,y+\omega_v\,v)}\,dy\,dv\,.$$

Taking the Fourier transform of the Backward Kolmogorov Equation with respect to the forward state therefore gives

$$\frac{\partial\phi}{\partial t} + (r-V/2)\frac{\partial\phi}{\partial Y} + \kappa(\theta-V)\frac{\partial\phi}{\partial V} + \frac{1}{2}\Big(V\frac{\partial^2\phi}{\partial Y^2} + 2\rho\sigma V\frac{\partial^2\phi}{\partial Y\partial V} + \sigma^2 V\frac{\partial^2\phi}{\partial V^2}\Big) = 0\,. \tag{6.45}$$

The terminal condition for the Backward Kolmogorov Equation is clearly $f(Y,V,T\,|\,y,v,t=T) = \delta(y-Y)\delta(v-V)$ which in turn means that the terminal condition for the characteristic function is $\phi(Y,R,T) = e^{i(\omega_y\,Y+\omega_v\,V)}$. The key idea is that equation (6.45) has a solution in the form

$$\phi(Y,R,t) = \exp\left[\beta_0(T-t) + \beta_1(T-t)Y + \beta_2(T-t)V\right] \tag{6.46}$$

where $\beta_0(0)=0$, $\beta_1(0)=i\omega_y$ and $\beta_2(0)=i\omega_v$. To appreciate why this claim is true, let $\tau = T-t$ and note that

$$\frac{\partial\phi}{\partial t} = -\Big(\frac{d\beta_0}{d\tau} + Y\frac{d\beta_1}{d\tau} + V\frac{d\beta_2}{d\tau}\Big)\phi\,, \qquad \frac{\partial\phi}{\partial Y} = \beta_1\phi\,, \qquad \frac{\partial\phi}{\partial V} = \beta_2\phi\,,$$
$$\frac{\partial^2\phi}{\partial Y^2} = \beta_1^2\phi\,, \qquad \frac{\partial^2\phi}{\partial Y\partial V} = \beta_1\beta_2\phi\,, \qquad \frac{\partial^2\phi}{\partial V^2} = \beta_2^2\phi\,.$$

Each expression is now substituted into equation (6.45) to get

$$-\Big(\frac{d\beta_0}{d\tau} + Y\frac{d\beta_1}{d\tau} + V\frac{d\beta_2}{d\tau}\Big) + (r-V/2)\beta_1 + \kappa(\theta-V)\beta_2 + \frac{1}{2}\Big(V\beta_1^2 + 2\rho\sigma V\beta_1\beta_2 + \sigma^2 V\beta_2^2\Big) = 0\,.$$

Ordinary differential equations satisfied by β_0, β_1 and β_2 are obtained by equating to zero the constant term and the coefficients of Y and V. Thus

$$\begin{aligned}
\frac{d\beta_0}{d\tau} &= r\beta_1 + \kappa\theta\beta_2\,,\\
\frac{d\beta_1}{d\tau} &= 0\,,\\
\frac{d\beta_2}{d\tau} &= -\frac{\beta_1}{2} - \kappa\beta_2 + \frac{1}{2}(\beta_1^2 + 2\rho\sigma\beta_1\beta_2 + \sigma^2\beta_2^2)
\end{aligned} \tag{6.47}$$

with initial conditions $\beta_0(0)=0$, $\beta_1(0)=i\omega_y$ and $\beta_2(0)=i\omega_v$.

6.4 Alternative view of the forward Kolmogorov equation

The derivation of the forward Kolmogorov equation in Section 6.3 was largely of a technical nature and as such the derivation provided relatively little insight into either the meaning of a stochastic differential equation or the structure of the forward Kolmogorov equation. In this section the forward Kolmogorov equation is derived via a two-stage physical argument.

Suppose that $\mathbf{X}$ is an N-dimensional random variable with sample space $\mathcal{S}$ and let $\mathbb{V}$ be an arbitrary but fixed N-dimensional volume of the sample with outer boundary $\partial\mathbb{V}$ and unit **outward** normal direction $\mathbf{n}$. The mass of probability density within $\mathbb{V}$ is

$$\int_{\mathbb{V}} f(t,\mathbf{x})\, dV\,.$$

The dependence of a probability density function on time necessarily means that probability must "flow" through the sample space $\mathcal{S}$, *i.e.* there is the notion of a **probability flux vector**, say $\mathbf{p}$, which transports probability around the sample space in order to maintain the proposed time dependence of the probability density function. Conservation of probability requires that the time rate of change of the probability mass within the volume $\mathbb{V}$ must equate to the **negative** of the total flux of probability crossing the boundary $\partial\mathbb{V}$ of the volume $\mathbb{V}$ in the outward direction, *i.e.*

$$\frac{d}{dt}\int_{\mathbb{V}} f(t,\mathbf{x})\, dV = -\int_{\partial\mathbb{V}} \sum_{k=1}^{N} p_k n_k\, dA\,. \tag{6.48}$$

The divergence theorem is applied to the surface integral on the right hand side of equation (6.48) and the time derivative is taken inside the left hand side of equation (6.48) to obtain

$$\int_{\mathbb{V}} \frac{\partial f(t,\mathbf{x})}{\partial t}\, dV = -\int_{\partial\mathbb{V}} \sum_{k=1}^{n} \frac{\partial p_k}{\partial x_k}\, dV\,.$$

However, $\mathbb{V}$ is an arbitrary volume within the sample space and therefore

$$\frac{\partial f(t,\mathbf{x})}{\partial t} = -\sum_{k=1}^{N} \frac{\partial p_k}{\partial x_k}\,. \tag{6.49}$$

Equation (6.49) is called a "conservation law" because it embodies the fact that probability is not destroyed but rather is redistributed in time.

6.4.1 Jumps

Jump processes, whenever present, act to redistribute probability density by a mechanism different from diffusion and therefore contribute to the right hand side of equation (6.49). Recall that

$$\lim_{h\to 0^+} \frac{f(\mathbf{x}, t+h\,|\,\mathbf{y}, t)}{h} = \phi(\mathbf{x}\,|\,\mathbf{y}, t\,)\,,$$

then we may think of $\phi(\mathbf{x}\,|\,\mathbf{y}, t\,)$ as the rate of jump transitions from state y at time t to state x at t^+. Evidently the rate of supply of probability density to state $\mathbf{x}$ from all other states is simply

$$\int_{\mathcal{S}} f(t,\mathbf{y}\,)\,\phi(\mathbf{x},\mathbf{y})\, d\mathbf{y}$$

whereas the loss of probability density at state $\mathbf{x}$ due to transitions to other states is simply

$$\int_{\mathcal{S}} f(t,\mathbf{x}\,)\,\phi(\mathbf{y},\mathbf{x})\, d\mathbf{y} = f(t,\mathbf{x}\,)\int_{\mathcal{S}} \phi(\mathbf{y},\mathbf{x})\, d\mathbf{y}$$

and consequently the net supply of probability to state $\mathbf{x}$ from all other states is

$$\int_{\mathcal{S}} \left(f(t, \mathbf{y}\,)\, \phi(\mathbf{x}, \mathbf{y}) - f(t, \mathbf{x}\,)\, \phi(\mathbf{y}, \mathbf{x}) \right) d\mathbf{y}\,. \tag{6.50}$$

Thus the final form of the forward Kolmogorov equation taking account of jump processes is

$$\frac{\partial f(t, \mathbf{x})}{\partial t} = -\sum_{k=1}^{n} \frac{\partial p_k}{\partial x_k} + \int_{\mathcal{S}} \left(f(t, \mathbf{y})\, \phi(\mathbf{x}, \mathbf{y}) - f(t, \mathbf{x})\, \phi(\mathbf{y}, \mathbf{x}) \right) d\mathbf{y}\,, \tag{6.51}$$

where $\phi(\mathbf{x}, \mathbf{y})$ is the rate of transition from state $\mathbf{y}$ to state $\mathbf{x}$. The fact that jump mechanisms simply redistribute probability density is captured by the identity

$$\int_{\mathbb{R}} \Big(\int_{\mathcal{S}} \Big(f(t, \mathbf{y})\, \phi(\mathbf{x}, \mathbf{y}) - f(t, \mathbf{x})\, \phi(\mathbf{y}, \mathbf{x}) \Big)\, d\mathbf{y} \Big)\, d\mathbf{x} = 0\,.$$

It is important to recognise that the analysis so far has used no information from the stochastic differential equation. Clearly the role of the stochastic differential equation is to provide the constitutive equation for the probability flux.

Special case

An important special case occurs when the intensity of the jump process, say $\lambda(t)$, is a function of t alone and is independent of the state $\mathbf{x}$. In this case $\phi(\mathbf{x}, \mathbf{y}) = \lambda(t)\psi(\mathbf{x}, \mathbf{y})$ leading to the result that

$$\frac{\partial f(t, \mathbf{x})}{\partial t} = -\sum_{k=1}^{n} \frac{\partial p_k}{\partial x_k} + \lambda(t) \Big(\int_{\mathcal{S}} f(t, \mathbf{y})\, \psi(\mathbf{x}, \mathbf{y})\, d\mathbf{y} - f(t, \mathbf{x}) \Big)\,, \tag{6.52}$$

where we have made the usual assumption that

$$\int_{\mathcal{S}} \psi(\mathbf{y}, \mathbf{x})\, d\mathbf{y} = 1\,.$$

6.4.2 Determining the probability flux from an SDE - One dimension

The probability flux vector is constructed by investigating how probability density at time t is distributed around the sample space by the mechanism of diffusion. Consider now the stochastic differential equation

$$dx = \mu(t, x)\, dt + \sqrt{g(t, x)}\, dW\,.$$

This equation asserts that during the interval $[t, t + h]$, "point-density" $f(t, x)\, \Delta x$ contained within an interval of length Δx centred on x is redistributed according to a Gaussian distribution with mean value $x + \mu(t, x)\, h + o(h)$ and variance $g(t, x)\, h + o(h)\,]$. Thus the probability density function describing the redistribution of probability density $f(t, x)\, \Delta x$ is

$$\phi(z\ :\ h, x) = \frac{\exp\Big[-\dfrac{(z - x - \mu(t, x)\, h - o(h))^2}{2g(t, x)\, h + o(h)} \Big]}{\sqrt{2\pi g(t, x)\, h + o(h)}} \tag{6.53}$$

in which z is the variables of the density. In the analysis to follow the $o(h)$ terms will be ignored in order to maintain the transparency of the calculation, but it will be clear throughout that $o(h)$ terms make no contribution to the calculation. The idea is to compute the rate at which probability flows across the boundary $x = \psi$ where ψ is an arbitrary constant parameter. Figure 6.1 illustrates the idea of an imbalance between the contribution to the region $x < \psi$ from the region $x > \psi$ and the contribution to the region $x > \psi$ from the region $x < \psi$. The rate at which this imbalance develops determines p.

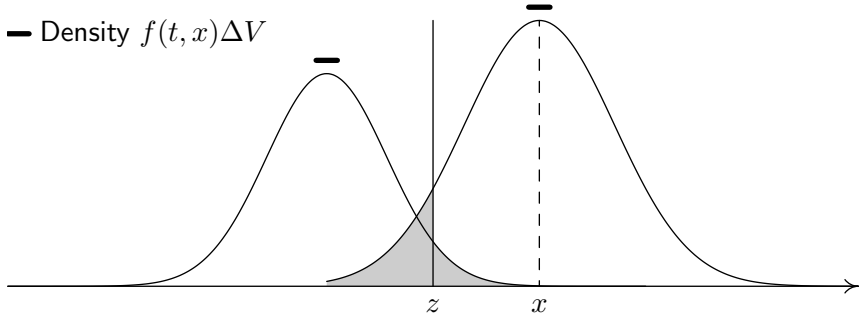

Figure 6.1: Diffusion of probability

When $x \geq \psi$, that is, the state x lies to the right of the boundary $x = \psi$, the fraction of the density $f(t,x)\Delta x$ at time t that diffuses into the region $x < \psi$ is illustrated by the shaded region of Figure 6.1 in the region $x < \psi$ and has value

$$\int_{x\geq\psi} f(t,x) \Big(\int_{z<\psi} \phi(z \,:\, h, x)\, dz \Big)\, dx\,.$$

Similarly, when $x < \psi$, the fraction of the density $f(t,x)\Delta x$ that diffuses into the region $x \geq \psi$ is illustrated by the shaded region $x \geq \psi$ of Figure 6.1 and has value

$$\int_{x<\psi} f(t,x) \Big(\int_{z\geq\psi} \phi(z : h, x)\, dz \Big)\, dx\,.$$

This rearrangement of density has taken place over an interval of duration h, and therefore the flux of probability flowing from the region $x < \psi$ into the region $x \geq \psi$ is

$$\lim_{h\to 0^+} \frac{1}{h} \Big[\int_{x<\psi} f(t,x) \Big(\int_{z\geq\psi} \phi(z : h, x)\, dz \Big)\, dx - \int_{x\geq\psi} f(t,x) \Big(\int_{z<\psi} \phi(z : h, x)\, dz \Big)\, dx \Big]\,. \tag{6.54}$$

The value of this limit is computed by L'Hopital's rule. The key component of this calculation requires the differentiation of $\phi(z \,:\, h, x)$ with respect to h. This calculation is most efficiently achieved by logarithmic differentiation from the starting equation

$$\log \phi(z : h, x) = -\frac{\log 2\pi}{2} - \frac{1}{2}\log h - \frac{1}{2}\log g(t,x) - \frac{(z - x - \mu(t,x)\, h)^2}{2g(t,x)\, h}.$$

Differentiation with respect to h and with respect to z_j gives

$$\begin{aligned}\frac{1}{\phi}\frac{\partial \phi}{\partial h} &= -\frac{1}{2h} + \frac{\mu(t,x)(z_k - x_k - \mu_k\, h)}{g(t,x)\, h} + \frac{(z - x - \mu(t,x)\, h)^2}{2g(t,x)h^2}\,. \\ \frac{1}{\phi}\frac{\partial \phi}{\partial z} &= -\frac{(z - x - \mu(t,x)\, h)}{g(t,x)h}\,.\end{aligned} \tag{6.55}$$

Evidently the previous results can be combine to show that $\phi(z : h, x)$ satisfies the identity

$$\begin{aligned}\frac{\partial \phi}{\partial h} &= -\frac{1}{2h}\,\phi - \mu(t,x)\,\frac{\partial \phi}{\partial z} - \frac{1}{2h}\,(z - x - \mu(t,x)\, h)\,\frac{\partial \phi}{\partial z} \\ &= -\frac{1}{2h}\,\frac{\partial[(z - x + \mu(t,x)\, h)\,\phi]}{\partial z}\,.\end{aligned} \tag{6.56}$$

In view of identity (6.59), L'Hopital's rule applied to equation (6.54) gives

$$\begin{aligned}p &= \lim_{h\to 0^+}\Big[\,\frac{1}{2h}\int_{x<\psi} f(t,x)\Big[-\int_{z\geq\psi}\frac{\partial[(z - x + \mu(t,x)\, h)\,\phi]}{\partial z}\,dz\Big]\,dx \\ &\quad + \frac{1}{2h}\int_{x\geq\psi} f(t,x)\Big[\int_{z<\psi}\frac{\partial[(z - x + \mu(t,x)\, h)\,\phi]}{\partial z}\,dz\Big]\,dx\Big]\,, \\ &= \lim_{h\to 0^+}\frac{1}{2h}\Big[-\int_{x<\psi} f(t,x)\Big(-(\psi - x + \mu(t,x)\, h)\,\phi\Big)\,dx \\ &\quad + \int_{x\geq\psi} f(t,x)\Big((\psi - x + \mu(t,x)\, h)\,\phi\Big)\,dx\Big]\,.\end{aligned} \tag{6.57}$$

Further simplification gives

$$p = \lim_{h\to 0^+}\frac{1}{2h}\int_{\mathbb{R}} f(t,x)\,(\psi - x + \mu(t,x)\, h)\,\phi(\psi; x, h)\,dx\,. \tag{6.58}$$

One way to make further progress is to compute the integral of p with respect to ψ over the $[a, b]$ such that $a < \psi < b$, that is, the point ψ is guaranteed to lie within the interval $[a, b]$. The computation of this integral will require the evaluation of

$$I = \int_a^b \Big(\frac{\psi - x + \mu(t,x)\, h}{2h\sqrt{g(t,x)\, h}}\Big)\,\exp\Big(-\frac{(\psi - x - \mu(t,x)\, h)^2}{2g(t,x)\, h}\Big)\,d\psi\,.$$

Under the change of variable $y = (\psi - x - \mu(t,x))/\sqrt{g(t,x)\, h}$, the integral for I becomes

$$I = \int_{(a-x-\mu(t,x)h)/\sqrt{g(t,x)\, h}}^{(b-x-\mu(t,x)h)/\sqrt{g(t,x)\, h}} \Big(\mu(t,x) + \frac{y\sqrt{g(t,x)h}}{2h}\Big)\,e^{-y^2/2}\,dy\,,$$

and therefore

$$\begin{aligned}I &= \mu(t,x)\Big[\Phi\Big(\frac{b - x - \mu(t,x)h}{\sqrt{g(t,x)\, h}}\Big) - \Phi\Big(\frac{a - x - \mu(t,x)h}{\sqrt{g(t,x)\, h}}\Big)\Big] \\ &\quad + \frac{\sqrt{g(t,x)h}}{2h}\int_{a-x-\mu(t,x)h)/\sqrt{g(t,x)\, h}}^{b-x-\mu(t,x)h)/\sqrt{g(t,x)\, h}} y e^{-y^2/2}\,dy \\ &= \mu(t,x)\Big[\Phi\Big(\frac{b - x - \mu(t,x)h}{\sqrt{g(t,x)\, h}}\Big) + \Phi\Big(\frac{a - x - \mu(t,x)h}{\sqrt{g(t,x)\, h}}\Big)\Big] \\ &\quad + \frac{\sqrt{g(t,x)h}}{2h}\Big[\exp\Big(-\frac{(b - x - \mu(t,x)h)^2}{2g(t,x)\, h}\Big) - \exp\Big(-\frac{(a - x - \mu(t,x)h)^2}{2g(t,x)\, h}\Big)\Big]\,.\end{aligned}$$

The conclusion of this calculation is that

$$\int_a^b p(t,\psi)\,d\psi = \lim_{h\to 0^+}\int_{\mathbb{R}} f(x)\mu(t,x)\Big[\Phi\Big(\frac{b-x-\mu(t,x)h}{\sqrt{g(t,x)\,h}}\Big)-\Phi\Big(\frac{a-x-\mu(t,x)h}{\sqrt{g(t,x)\,h}}\Big)\Big]\,dx$$
$$-\lim_{h\to 0^+}\int_{\mathbb{R}} f(x)\frac{\sqrt{g(t,x)h}}{2h}\Big[\,\exp\Big(-\frac{(b-x-\mu(t,x)h)^2}{2g(t,x)\,h}\Big)-\exp\Big(-\frac{(a-x-\mu(t,x)h)^2}{2g(t,x)\,h}\Big)\,\Big]\,dx\,.$$

Now consider the value of the limit under each integral. First, it is clear that

$$\lim_{h\to 0^+}\Big[\Phi\Big(\frac{b-x-\mu(t,x)h}{\sqrt{g(t,x)\,h}}\Big)-\Phi\Big(\frac{a-x-\mu(t,x)h}{\sqrt{g(t,x)\,h}}\Big)\Big]=\left[\begin{array}{ll}0 & x\in(-\infty,a)\cup(b,\infty)\\ 1 & x\in[a,b\,]\end{array}\right.$$

Second,

$$\begin{aligned}\lim_{h\to 0^+}\frac{1}{\sqrt{g(t,x)\,h}}\exp\Big(-\frac{(b-x-\mu(t,x)h)^2}{2g(t,x)\,h}\Big) &= \delta(x-b)\,,\\ \lim_{h\to 0^+}\frac{1}{\sqrt{g(t,x)\,h}}\exp\Big(-\frac{(a-x-\mu(t,x)h)^2}{2g(t,x)\,h}\Big) &= \delta(x-a)\,.\end{aligned}$$

These limiting values now give that

$$\begin{aligned}\int_a^b p(t,\psi)\,d\psi &= \int_a^b f(t,x)\mu(t,x)\,dx-\int_{\mathbb{R}} f(t,x)\Big[g(t,x)\delta(x-b)-g(t,x)\delta(x-a)\Big]\,dx\\ &= \int_a^b f(t,x)\mu(t,x)\,dx-[\,f(t,b)g(t,b)-f(t,a)g(t,a)\,]\,.\end{aligned}\tag{6.59}$$

Equation (6.59) is now divided by $(b-a)$ and limits taken such that $(b-a)\to 0^+$ while retaining the point ψ within the interval $[a.b\,]$. The result of this calculation is that

$$p(t,\psi)=\mu(t,\psi)f(t,\psi)-\frac{\partial\,[\,g(t,\psi)f(t,\psi)\,]}{\partial\psi}\,.\tag{6.60}$$

Chapter 7

Numerical Integration of SDE

The dearth of closed form solutions to stochastic differential equations means that numerical procedures enjoy a prominent role in the development of the subject.

7.1 Issues of convergence

The definition of convergence for a stochastic process is more subtle than the equivalent definition in a deterministic environment, primarily because stochastic processes, by their very nature, cannot be constrained to lie within ε of the limiting value for all $n > N$. Several types of convergence are possible in a stochastic environment.

The first measure is based on the observation that although individual paths of a stochastic process behave randomly, the expected value of the process, normally estimated numerically as the average value over a large number of realisations of the process under identical conditions, behaves deterministically. This observation can be used to measure convergence in the usual way. The central limit theorem indicates that if the variance of the process is finite, then convergence to the expected state will follow a square-root N law where N is the number of repetitions.

7.1.1 Strong convergence

Suppose that $X(t)$ is the actual solution of an SDE and $X_n(t)$ is the solution based on an iterative scheme with step size Δt, we say that X_n exhibits strong convergence to X of order α if

$$\mathrm{E}\left[\, |X_n - X\,|\,\right] < C\,(\Delta t)^{\alpha} \tag{7.1}$$

Given a numerical procedure to integrate an SDE, we may identify the order of strong convergence by the following numerical procedure.

1. Choose Δt and integrate the SDE over an interval of fixed length, say T, many times.

2. Compute the left hand side of inequality (7.1) and repeat this process for different values of Δt.

3. Regress the logarithm of the left hand side of (7.1) against $\log(\Delta t)$.

If strong convergence of order α is present, the gradient of this regression line will be significant and its value will be α, the order of strong convergence.

Although it might appear superficially that strong convergence is about expected values and not about individual solutions, the Markov inequality

$$\text{Prob}\,(\,|X| \geq A) \leq \frac{\text{E}\,[\,|X|\,]}{A} \tag{7.2}$$

indicates that the presence of strong convergence of order α offers information about individual solutions. Take $A = (\Delta t)^\beta$ then clearly

$$\text{Prob}\,(\,|X_n - X\,| \geq (\Delta t)^\beta) \leq \frac{\text{E}\,[\,|X_n - X\,|\,]}{(\Delta t)^\beta} \leq C\,(\Delta t)^{\alpha-\beta} \tag{7.3}$$

or equivalently

$$\text{Prob}\,(\,|X_n - X\,| < (\Delta t)^\beta\,) > 1 - C\,(\Delta t)^{\alpha-\beta}\,. \tag{7.4}$$

For example, take $\beta = \alpha - \varepsilon$ where $\varepsilon > 0$, then the implication of the Markov inequality in combination with strong convergence is that the solution of the numerical scheme converges to that of the SDE with probability one as $\Delta t \to 0$ - hence the term *strong convergence*.

7.1.2 Weak convergence

Strong convergence imposes restrictions on the rate of decay of the mean-of-the-error to zero as $\Delta t \to 0$. By contrast, weak convergence involves the error-of-the-mean. Suppose that $X(t)$ is the actual solution of an SDE and $X_n(t)$ is the solution based on an iterative scheme with step size Δt, X_n is said to converge weakly to X with order α if

$$|\,\text{E}\,[\,f(X_n)\,] - \text{E}\,[\,f(X)\,]\,| < C\,(\Delta t)^\alpha \tag{7.5}$$

where the condition is required to be satisfied for all function f belonging to a prescribed class $\mathcal{C}$ of functions. The identify function $f(x) = x$ is one obvious choice of f. Another choice for $\mathcal{C}$ is the class of functions satisfying prescribed properties of smoothness and polynomial growth.

Clearly the numerical experiment that was described in the section concerning strong convergence can be repeated, except that the objective of the calculation is now to find the difference in expected values for each Δt and not the expected value of differences.

Theorem Let $X(t)$ be an n-dimensional random variable which evolves in accordance with the stochastic differential equation

$$d\mathbf{x} = \mathbf{a}(t, \mathbf{x})\,dt + B(t, \mathbf{x})\,d\mathbf{W}\,, \qquad \mathbf{x}(t_0) = \mathbf{X}_0 \tag{7.6}$$

where $B(t,\mathbf{x})$ is an $n \times m$ array and $d\mathbf{W}$ is an m-dimensional vector of Wiener increments, not necessarily uncorrelated. If $\mathbf{a}(t,\mathbf{x})$ and $B(t,\mathbf{x})$ satisfy the conditions

$$||\phi(t,\mathbf{x}) - \phi(t,\mathbf{y})|| \leq c||\mathbf{x}-\mathbf{y}||\,, \qquad ||\phi(t,\mathbf{x})|| \leq c(1+||\mathbf{x}||)$$

for constant c, then the initial value problem (7.6) has a solution in the vicinity of $t = t_0$, and this solution is unique.

7.2 Deterministic Taylor Expansions

The Taylor series expansion of a function of a deterministic variable is well known. Since the existence of the Taylor series is a property of the function and not the variable in which the function is expanded, then the Taylor series of a function of a stochastic variable is possible. As a preamble to the computation of these stochastic expansions, the procedure is first described in the context of an expansion in a deterministic function.

Suppose $x_i(t)$ satisfies the differential equation $dx_i/dt = a_i(t,\mathbf{x})$ and let $f : \mathbb{R} \to \mathbb{R}$ be a continuously differentiable function of t and $\mathbf{x}$ then the chain rule gives

$$\frac{df}{dt} = \frac{\partial f}{\partial t} + \sum_{k=1}^{n} \frac{\partial f}{\partial x_k}\frac{dx_k}{dt} = \frac{\partial f}{\partial t} + \sum_{k=1}^{n} a_k \frac{\partial f}{\partial x_k} \qquad \longrightarrow \qquad \frac{df}{dt} = L\,f \tag{7.7}$$

where the linear operator L is defined in the obvious way. By construction,

$$\begin{aligned} f(t,\mathbf{x}_t) &= f(t_0,\mathbf{x}_0) + \int_{t_0}^{t} Lf(s,\mathbf{x}_s)\,ds\,, \\ \mathbf{x}_t &= \mathbf{x}_0 + \int_{t_0}^{t} \mathbf{a}(s,\mathbf{x}_s)\,ds\,. \end{aligned} \tag{7.8}$$

Now apply result (7.8) to $\mathbf{a}(t,\mathbf{x}_t)$ to obtain

$$\mathbf{a}(s,\mathbf{x}_s) = \mathbf{a}(t_0,\mathbf{x}_0) + \int_{t_0}^{s} L\mathbf{a}(u,\mathbf{x}_u)\,du\,, \tag{7.9}$$

and this result, when incorporated into (7.8) gives

$$\begin{aligned} \mathbf{x}_t &= \mathbf{x}_0 + \int_{t_0}^{t} \left(\mathbf{a}(t_0,\mathbf{x}_0) + \int_{t_0}^{s} L\mathbf{a}(u,\mathbf{x}_u)\,du \right) ds \\ &= \mathbf{x}_0 + \mathbf{a}(t_0,\mathbf{x}_0)\int_{t_0}^{t} ds + \int_{t_0}^{t}\int_{t_0}^{s} L\mathbf{a}(u,\mathbf{x}_u)\,du\,ds\,. \end{aligned} \tag{7.10}$$

The process is now repeated with f in equation (7.8) replaced with $L\mathbf{a}(u,x_u)$. The outcome of this operation is

$$\begin{aligned} \mathbf{x}_t &= \mathbf{x}_0 + \mathbf{a}(t_0,\mathbf{x}_0)\int_{t_0}^{t} ds + \int_{t_0}^{t}\int_{t_0}^{s} L\mathbf{a}(u,\mathbf{x}_u)\,du\,ds \\ &= \mathbf{x}_0 + \mathbf{a}(t_0,\mathbf{x}_0)\int_{t_0}^{t} ds + \int_{t_0}^{t}\int_{t_0}^{s} \left(L\mathbf{a}(t_0,\mathbf{x}_0) + \int_{t_0}^{u} L^2\mathbf{a}(w,\mathbf{x}_w)\,dw \right) du\,ds \\ &= \mathbf{x}_0 + \mathbf{a}(t_0,\mathbf{x}_0)\int_{t_0}^{t} ds + L\mathbf{a}(t_0,\mathbf{x}_0)\int_{t_0}^{t}\int_{t_0}^{s} du\,ds + \int_{t_0}^{u}\int_{t_0}^{t}\int_{t_0}^{s} L^2\mathbf{a}(w,\mathbf{x}_w)\,dw\,du\,ds \end{aligned} \tag{7.11}$$

The procedure may be repeated provided the requisite derivatives exist. By recognising that

$$\int_{t_0}^{t}\int_{t_0}^{u}\cdots\int_{t_0}^{s} dw\ \cdots du = \frac{(t-t_0)^k}{k!}$$

it follows directly from (7.11) that

$$\mathbf{x}_t = \mathbf{x}_0 + \mathbf{a}(t_0,\mathbf{x}_0)\,t + L\mathbf{a}(t_0,\mathbf{x}_0)\frac{(t-t_0)^2}{2} + \mathbf{R}_3\,, \qquad \mathbf{R}_3 = \int_{t_0}^{u}\int_{t_0}^{t}\int_{t_0}^{s} L^2\mathbf{a}(w,\mathbf{x}_w)\,dw\,du\,ds\,. \quad (7.12)$$

7.3 Stochastic Ito-Taylor Expansion

The idea developed in the previous section for deterministic Taylor expansions will now be applied to the calculation of stochastic Taylor expansions. Suppose that $\mathbf{x}(t)$ satisfies the initial value problem

$$dx_k = a_k(t,\mathbf{x})\,dt + b_{k\,\alpha}(t,\mathbf{x})\,dW_t^\alpha\,, \qquad \mathbf{x}(0) = \mathbf{x}_0\,. \quad (7.13)$$

The first task is to differentiate $f(t,\mathbf{x})$ with respect to t using Ito's Lemma. The general Ito result is derived from the Taylor expansion

$$df = \frac{\partial f}{\partial t}\,dt + \sum_{k=1}^{n}\frac{\partial f}{\partial x_k}\,dx_k + \frac{1}{2}\Big[\,\frac{\partial^2 f}{\partial t^2}\,(dt)^2 + 2\sum_{k=1}^{n}\frac{\partial^2 f}{\partial t\partial x_k}\,dt\,dx_k + \sum_{j,k=1}^{n}\frac{\partial^2 f}{\partial x_j\partial x_k}\,dx_k\,dx_j\,\Big] + \cdots\,.$$

The idea is to expand this expression to order dt, ignoring all terms of order higher than dt since these will vanish in the limiting process. The first step in the derivation of Ito's lemma is to replace dx_k from the stochastic differential equation (7.13) in the previous equation. All terms which are overtly $o(dt)$ are now discarded to obtain the reduced equation

$$\begin{aligned} df &= \frac{\partial f}{\partial t}\,dt + \sum_{k=1}^{n}\frac{\partial f}{\partial x_k}\Big[\,a_k(t,\mathbf{x})\,dt + b_{k\,\alpha}(t,\mathbf{x})\,dW_t^\alpha\,\Big] \\ &\quad + \frac{1}{2}\Big[\sum_{j,k=1}^{n}\frac{\partial^2 f}{\partial x_j\partial x_k}\,b_{k\,\alpha}(t,\mathbf{x})\,dW_t^\alpha\,b_{j\,\beta}(t,\mathbf{x})\,dW_t^\beta\,\Big] + \cdots \end{aligned}$$

For simplicity, now suppose that dW^α are uncorrelated Wiener increments so that

$$df = \Big[\,\frac{\partial f}{\partial t} + \sum_{k=1}^{n}\frac{\partial f}{\partial x_k}\,a_k(t,\mathbf{x}) + \frac{1}{2}\sum_{j,k=1}^{n}\frac{\partial^2 f}{\partial x_j\partial x_k}b_{k\,\alpha}(t,\mathbf{x})\,b_{j\,\alpha}(t,\mathbf{x})\,\Big]\,dt + \sum_{k=1}^{n}\frac{\partial f}{\partial x_k}\,b_{k\,\alpha}(t,\mathbf{x})\,dW_t^\alpha$$

which in turn leads to the integral form

$$f(t,x_t) = f(t_0,x_0) + \int_{t_0}^{t} L^0 f(s,\mathbf{x}_s)\,ds + \int_{t_0}^{t} L^\alpha f(s,\mathbf{x}_s)\,dW_s^\alpha \quad (7.14)$$

in which the linear operators L^0 and L^α are defined by the formulae

$$\begin{aligned} L^0 f &= \frac{\partial f}{\partial t} + \sum_{k=1}^{n}\frac{\partial f}{\partial x_k}\,a_k(t,\mathbf{x}) + \frac{1}{2}\sum_{j,k=1}^{n}\frac{\partial^2 f}{\partial x_j\partial x_k}b_{k\,\alpha}(t,\mathbf{x})\,b_{j\,\alpha}(t,\mathbf{x})\,, \\ L^\alpha f &= \sum_{k=1}^{n}\frac{\partial f}{\partial x_k}\,b_{k\,\alpha}(t,\mathbf{x})\,. \end{aligned} \quad (7.15)$$

In terms of these operators, it now follows that

$$f(t,\mathbf{x}_t) = f(t_0,\mathbf{x}_0) + \int_{t_0}^{t} L^0 f(s,\mathbf{x}_s)\,ds + \int_{t_0}^{t} L^\alpha f(s,\mathbf{x}_s)\,dW_s^\alpha \tag{7.16}$$

$$x_k(t) = x_k(0) + \int_{t_0}^{t} a_k(s,\mathbf{x}_s)\,ds + \int_{t_0}^{t} b_{k\,\alpha}(s,\mathbf{x}_s)\,dW_s^\alpha\,. \tag{7.17}$$

By analogy with the treatment of the deterministic case, the Ito formula (7.16) is now applied to $a_k(t,\mathbf{x})$ and $b_{k\,\alpha}(t,\mathbf{x})$ to obtain

$$\begin{aligned} a_k(t,\mathbf{x}_t) &= a_k(t_0,\mathbf{x}_0) + \int_{t_0}^{t} L^0 a_k(s,\mathbf{x}_s)\,ds + \int_{t_0}^{t} L^\alpha a_k(s,\mathbf{x}_s)\,dW_s^\alpha \\ b_{k\,\alpha}(t,\mathbf{x}_t) &= b_{k\,\alpha}(t_0,\mathbf{x}_0) + \int_{t_0}^{t} L^0 b_{k\,\alpha}(s,\mathbf{x}_s)\,ds + \int_{t_0}^{t} L^\beta b_{k\,\alpha}(s,\mathbf{x}_s)\,dW_s^\beta \end{aligned} \tag{7.18}$$

Formulae (7.18) are now applied to the right hand side of equation (7.17) to obtain initially

$$\begin{aligned} x_k(t) = x_k(0) &+ \int_{t_0}^{t} \Big(a_k(t_0,\mathbf{x}_0) + \int_{t_0}^{s} L^0 a_k(u,\mathbf{x}_u)\,du + \int_{t_0}^{s} L^\alpha a_k(u,\mathbf{x}_u)\,dW_u^\alpha \Big)\,ds \\ &+ \int_{t_0}^{t} \Big(b_{k\,\alpha}(t_0,\mathbf{x}_0) + \int_{t_0}^{s} L^0 b_{k\,\alpha}(u,\mathbf{x}_u)\,du + \int_{t_0}^{s} L^\beta b_{k\,\alpha}(u,\mathbf{x}_u)\,dW_u^\beta \Big)\,dW_s^\alpha \end{aligned}$$

which simplifies to the stochastic Taylor expansion

$$x_k(t) = x_k(0) + a_k(t_0,\mathbf{x}_0)\int_{t_0}^{t} ds + b_{k\,\alpha}(t_0,\mathbf{x}_0)\int_{t_0}^{t} dW_s^\alpha + R_k^{(2)} \tag{7.19}$$

where $R_k^{(2)}$ denotes the remainder term

$$\begin{aligned} R_k^{(2)} &= \int_{t_0}^{t}\int_{t_0}^{s} L^0 a_k(u,\mathbf{x}_u)\,du\,ds + \int_{t_0}^{t}\int_{t_0}^{s} L^\alpha a_k(u,\mathbf{x}_u)\,dW_u^\alpha\,ds \\ &\quad + \int_{t_0}^{t}\int_{t_0}^{s} L^0 b_{k\,\alpha}(u,\mathbf{x}_u)\,du\,dW_s^\alpha + \int_{t_0}^{t}\int_{t_0}^{s} L^\beta b_{k\,\alpha}(u,\mathbf{x}_u)\,dW_u^\beta\,dW_s^\alpha\,. \end{aligned} \tag{7.20}$$

Each term in the remainder R_2 may now be expanded by applying the Ito expansion procedure (7.16) in the appropriate way. Clearly,

$$\begin{aligned} \int_{t_0}^{t}\int_{t_0}^{s} L^0 a_k(u,\mathbf{x}_u)\,du\,ds &= L^0 a_k(t_0,\mathbf{x}_0)\int_{t_0}^{t}\int_{t_0}^{s} du\,ds \\ &+ \int_{t_0}^{t}\int_{t_0}^{s}\int_{t_0}^{u} L^0 L^0 a_k(w,\mathbf{x}_w)\,dw\,du\,ds + \int_{t_0}^{t}\int_{t_0}^{s}\int_{t_0}^{u} L^\gamma L^0 a_k(w,\mathbf{x}_w)\,dW_w^\gamma\,du\,ds \end{aligned}$$

$$\begin{aligned} \int_{t_0}^{t}\int_{t_0}^{s} L^\alpha a_k(u,\mathbf{x}_u)\,dW_u^\alpha\,ds &= L^\alpha a_k(t_0,\mathbf{x}_0)\int_{t_0}^{t}\int_{t_0}^{s} dW_u^\alpha\,ds \\ &+ \int_{t_0}^{t}\int_{t_0}^{s}\int_{t_0}^{u} L^0 L^\alpha a_k(w,\mathbf{x}_w)\,dw\,dW_u^\alpha\,ds + \int_{t_0}^{t}\int_{t_0}^{s}\int_{t_0}^{u} L^\gamma L^\alpha a_k(w,\mathbf{x}_w)\,dW_w^\gamma\,dW_u^\alpha\,ds \end{aligned}$$

$$\int_{t_0}^{t}\int_{t_0}^{s} L^0 b_{k\,\alpha}(u,\mathbf{x}_u)\,du\,dW_s^\alpha = L^0 b_{k\,\alpha}(t_0,\mathbf{x}_0)\int_{t_0}^{t}\int_{t_0}^{s} du\,dW_s^\alpha$$

$$+\int_{t_0}^{t}\int_{t_0}^{s}\int_{t_0}^{u} L^0 L^0\, b_{k\,\alpha}(w,\mathbf{x}_w)\,dw\,du\,dW_s^\alpha + \int_{t_0}^{t}\int_{t_0}^{s}\int_{t_0}^{u} L^\gamma\, L^0 b_{k\,\alpha}(w,\mathrm{x}_w)\,dW_w^\gamma\,du\,dW_s^\alpha$$

$$\int_{t_0}^{t}\int_{t_0}^{s} L^\beta\, b_{k\,\alpha}(u,\mathbf{x}_u)\,dW_u^\beta\,dW_s^\alpha \;=\; L^\beta\, b_{k\,\alpha}(t_0,\mathbf{x}_0)\int_{t_0}^{t}\int_{t_0}^{s} dW_u^\beta\,dW_s^\alpha$$

$$+\int_{t_0}^{t}\int_{t_0}^{s}\int_{t_0}^{u} L^0\, L^\beta\, b_{k\,\alpha}(w,\mathrm{x}_w)\,dw\,dW_u^\beta\,dW_s^\alpha + \int_{t_0}^{t}\int_{t_0}^{s}\int_{t_0}^{u} L^\gamma\, L^\beta\, b_{k\,\alpha}(w,\mathrm{x}_w)\,dW_w^\gamma\,dW_u^\beta\,dW_s^\alpha\,.$$

The general principle is apparent in this expansion. One must evaluate stochastic integrals involving combinations of deterministic (du) and stochastic (dW) differentials. For our purposes, it is enough to observe that

$$\begin{aligned} x_k(t) \;=\; & x_k(0) + a_k(t_0,\mathbf{x}_0)\int_{t_0}^{t} ds + b_{k\,\alpha}(t_0,\mathbf{x}_0)\int_{t_0}^{t} dW_s^\alpha + L^0\,a_k(t_0,\mathbf{x}_0)\int_{t_0}^{t}\int_{t_0}^{s} du\,ds \\ & + L^\alpha\,a_k(t_0,\mathbf{x}_0)\int_{t_0}^{t}\int_{t_0}^{s} dW_u^\alpha\,ds + L^0\,b_{k\,\alpha}(t_0,\mathbf{x}_0)\int_{t_0}^{t}\int_{t_0}^{s} du\,dW_s^\alpha \\ & + L^\beta\,b_{k\,\alpha}(t_0,\mathbf{x}_0)\int_{t_0}^{t}\int_{t_0}^{s} dW_u^\beta\,dW_s^\alpha + R_k^{(3)}\,. \end{aligned} \tag{7.21}$$

This solution can be expressed more compactly in terms of the multiple Ito integrals

$$I_{ijk\cdots} = \int\int\cdots\int dW^i\,dW^j\cdots dW^k$$

where $dW^k = ds$ if $k = 0$ and $dW^k = dW_s^\alpha$ if $k = \alpha > 0$. Using this representation, the solution (7.21) takes the short hand form

$$\begin{aligned} x_k(t) \;=\; & x_k(0) + a_k(t_0,\mathbf{x}_0)\,I_0 + b_{k\,\alpha}(t_0,\mathbf{x}_0)\,I_\alpha + L^0\,a_k(t_0,\mathbf{x}_0)\,I_{00} + L^\alpha\,a_k(t_0,\mathbf{x}_0)\,I_{\alpha\,0} \\ & + L^0\,b_{k\,\alpha}(t_0,\mathbf{x}_0)\,I_{0\,\alpha} + L^\beta\,b_{k\,\alpha}(t_0,\mathbf{x}_0)\,I_{\beta\,\alpha} + R_k^{(3)}\,. \end{aligned} \tag{7.22}$$

7.4 Stochastic Stratonovich-Taylor Expansion

The analysis developed in the previous section for the Ito stochastic differential equation

$$dx_k = a_k(t,\mathbf{x})\,dt + \sum_\alpha b_{k\,\alpha}(t,\mathbf{x})\,dW_t^\alpha\,, \qquad \mathbf{x}(0) = \mathbf{x}_0\,. \tag{7.23}$$

can be repeated for the Stratonovich representation of equation (7.23), namely,

$$dx_k = \bar{a}_k(t,\mathbf{x})\,dt + \sum_\alpha b_{k\,\alpha}(t,\mathbf{x}) \circ dW_t^\alpha\,, \qquad \mathbf{x}(0) = \mathbf{x}_0\,. \tag{7.24}$$

in which

$$\bar{a}_i = a_i - \frac{1}{2}\sum_{k,\alpha} b_{k\,\alpha}\,\frac{\partial b_{i\,\alpha}}{\partial x_k}\,.$$

The first task is to differentiate $f(t,\mathbf{x})$ with respect to t using the Stratonovich form of Ito's Lemma. It can be demonstrated for uncorrelated Wiener increments dW^α that

$$f(t,x_t) = f(t_0,x_0) + \int_{t_0}^{t} \bar{L}^0\,f(s,\mathbf{x}_s)\,ds + \int_{t_0}^{t} \bar{L}^\alpha\,f(s,\mathbf{x}_s) \circ dW_s^\alpha \tag{7.25}$$

in which the linear operators $\bar{L}^0$ and $\bar{L}^\alpha$ are defined by the formulae

$$\bar{L}^0 f = \frac{\partial f}{\partial t} + \sum_k \frac{\partial f}{\partial x_k} \bar{a}_k(t, \mathbf{x}) \,, \qquad \bar{L}^\alpha f = \sum_k \frac{\partial f}{\partial x_k} b_{k\,\alpha}(t, \mathbf{x}) \,. \tag{7.26}$$

In terms of these operators, it now follows that

$$f(t, \mathbf{x}_t) = f(t_0, \mathbf{x}_0) + \int_{t_0}^t \bar{L}^0 f(s, \mathbf{x}_s)\, ds + \int_{t_0}^t \bar{L}^\alpha f(s, \mathbf{x}_s) \circ dW_s^\alpha \tag{7.27}$$

$$x_k(t) = x_k(0) + \int_{t_0}^t \bar{a}_k(s, \mathbf{x}_s)\, ds + \int_{t_0}^t b_{k\,\alpha}(s, \mathbf{x}_s) \circ dW_s^\alpha \,. \tag{7.28}$$

The analysis of the previous section is now repeated to yield

$$\begin{aligned} x_k(t) = {} & x_k(0) + \bar{a}_k(t_0, \mathbf{x}_0) \int_{t_0}^t ds + b_{k\,\alpha}(t_0, \mathbf{x}_0) \int_{t_0}^t \circ dW_s^\alpha + \bar{L}^0 \bar{a}_k(t_0, \mathbf{x}_0) \int_{t_0}^t \int_{t_0}^s du\, ds \\ & + \bar{L}^\alpha \bar{a}_k(t_0, \mathbf{x}_0) \int_{t_0}^t \int_{t_0}^s \circ dW_u^\alpha\, ds + \bar{L}^0 b_{k\,\alpha}(t_0, \mathbf{x}_0) \int_{t_0}^t \int_{t_0}^s du \circ dW_s^\alpha \\ & + \bar{L}^\beta b_{k\,\alpha}(t_0, \mathbf{x}_0) \int_{t_0}^t \int_{t_0}^s \circ dW_u^\beta \circ dW_s^\alpha + R_k^{(3)} \,. \end{aligned} \tag{7.29}$$

This solution can be expressed more compactly in terms of the multiple Stratonovich integrals

$$J_{ijk\cdots} = \int\int \cdots \int \circ dW^i \circ dW^j \cdots \circ dW^k$$

where $dW^k = ds$ if $k = 0$ and $dW^k = \circ\, dW_s^\alpha$ if $k = \alpha > 0$. Using this representation, the solution (7.29) takes the short hand form

$$\begin{aligned} x_k(t) = {} & x_k(0) + \bar{a}_k(t_0, \mathbf{x}_0)\, J_0 + \sum_\alpha b_{k\,\alpha}(t_0, \mathbf{x}_0)\, J_\alpha + \bar{L}^0 \bar{a}_k(t_0, \mathbf{x}_0)\, J_{00} \\ & + \sum_\alpha \bar{L}^\alpha \bar{a}_k(t_0, \mathbf{x}_0)\, J_{\alpha\,0} \bar{L}^0 b_{k\,\alpha}(t_0, \mathbf{x}_0)\, J_{0\,\alpha} + \sum_{\alpha,\beta} \bar{L}^\beta b_{k\,\alpha}(t_0, \mathbf{x}_0)\, J_{\beta\,\alpha} + R_k^{(3)} \,. \end{aligned} \tag{7.30}$$

7.5 Euler-Maruyama algorithm

The Euler-Maruyama algorithm is the simplest numerical scheme for the solution of SDE's, being the stochastic equivalent of the deterministic Euler scheme. The algorithm is

$$x_k(t) = x_k(0) + a_k(t_0, \mathbf{x}_0)\, I_0 + \sum_\alpha b_{k\,\alpha}(t_0, \mathbf{x}_0)\, I_\alpha \,. \tag{7.31}$$

It follows directly from the definitions of $I_{ijk\cdots}$ that

$$I_0 = (t - t_0) \,, \qquad I^\alpha = W^\alpha(t) - W^\alpha(t_0)$$

and consequently

$$x_k(t) = x_k(0) + a_k(t_0, \mathbf{x}_0)\, (t - t_0) + \sum_\alpha b_{k\,\alpha}(t_0, \mathbf{x}_0)\, (W^\alpha(t) - W^\alpha(t_0)) \,. \tag{7.32}$$

In practice, the interval $[a, b]$ is divided into n subintervals of length h where $h = (b-a)/n$. The Euler-Maruyama scheme now becomes

$$x_k^{j+1} = x_k^j + a_k(t_j, \mathbf{x}_j)\, h + \sum_\alpha b_{k\,\alpha}(t_j, \mathbf{x}_j)\, \Delta W_j^\alpha\,, \qquad \Delta W_j^\alpha \sim \mathrm{N}\,(0, \sqrt{h}) \tag{7.33}$$

where x_k^j denotes the value of the k-th component of $\mathbf{x}$ at time $t_j = t_0 + jh$. It may be shown that the Euler-Maruyama has strong order of convergence $\gamma = 1/2$ and weak order of convergence $\gamma = 1$.

7.6 Milstein scheme

The principal determinant of the accuracy of the Euler-Maruyama is the stochastic term which is in practice $\sqrt{h}$ accurate. Therefore, to improve the quality of the Euler-Maruyama scheme, one should first seek to improve the accuracy of the stochastic term. Equation (7.22) indicates that the accuracy of the Euler-Maruyama scheme can be improved by including the term

$$\sum_{\alpha,\beta} L^\beta\, b_{k\,\alpha}(t_0, \mathbf{x}_0)\, I_{\beta\,\alpha}$$

in the Euler-Maruyama algorithm (7.33). This now requires the computation of

$$I_{\beta\,\alpha} = \int_{t_0}^{t}\int_{t_0}^{s} dW_u^\beta\, dW_s^\alpha = \int_{t_0}^{t} (W^\beta(s) - W^\beta(t_0))\, dW_s^\alpha\,.$$

When $\alpha = \beta$, it has been seen in previous analysis that

$$\begin{aligned} I_{\alpha\,\alpha} &= \int_{t_0}^{t} (W^\alpha(s) - W^\alpha(t_0))\, dW_s^\alpha \\ &= \frac{1}{2}\Big[\, (W^\alpha(t))^2 - (W^\alpha(t_0))^2 - (t - t_0)\,\Big] - W^\alpha(t_0)\,(W^\alpha(t) - W^\alpha(t_0)) \\ &= \frac{1}{2}\Big[\, \Big(\, W^\alpha(t) - W^\alpha(t_0)\,\Big)^2 - (t - t_0)\,\Big]\,. \end{aligned}$$

When $\alpha \neq \beta$, however, $I_{\alpha\,\beta}$ is not just a combination of $(W^\alpha(s) - W^\alpha(t_0))$ and $(W^\beta(s) - W^\beta(t_0))$. The value must be approximated even in this apparently straightforward case. Therefore, the numerical solution of systems of SDEs in which the evolution of the solution is characterised by more than one independent stochastic process is likely to be numerically intensive by any procedure other than the Euler-Maruyama algorithm. However, one exception to this working rule occurs whenever symmetry is present in the respect that

$$L^\beta\, b_{k\,\alpha}(t_0, \mathbf{x}_0) = L^\alpha\, b_{k\,\beta}(t_0, \mathbf{x}_0)\,.$$

In this case,

$$L^\beta\, b_{k\,\alpha}(t_0, \mathbf{x}_0)\, I_{\beta\,\alpha} = \frac{1}{2}\, L^\beta\, b_{k\,\alpha}(t_0, \mathbf{x}_0)\, (\, I_{\beta\,\alpha} + I_{\alpha\,\beta}\,)$$

which can now be computed from the identity

$$I_{\beta\,\alpha} + I_{\alpha\,\beta} = (\, W^\beta(t) - W^\beta(t_0)\,)(\, W^\alpha(t) - W^\alpha(t_0)\,)\,.$$

For systems of SDEs governed by a single Wiener process, it is relatively easy to develop higher order schemes. For the system

$$dx_k = a_k(t, \mathbf{x})\, dt + b_k(t, \mathbf{x})\, dW_t\,, \qquad \mathbf{x}(0) = \mathbf{x}_0 \tag{7.34}$$

the Euler-Maruyama algorithm is

$$x_k^{j+1} = x_k^j + a_k(t_j, \mathbf{x}_j)\, h + b_k(t_j, \mathbf{x}_j)\, \Delta W_j\,, \qquad \Delta W_j \sim \mathrm{N}\,(0, \sqrt{h}) \tag{7.35}$$

and the Milstein improvement to this algorithm is

$$x_k^{j+1} = x_k^j + a_k(t_j, \mathbf{x}_j)\, h + b_k(t_j, \mathbf{x}_j)\, \Delta W_j + \frac{1}{2}\, L^1\, b_k(t_j, \mathbf{x}_j)\, [\,(\Delta W_j)^2 - h\,]\,, \quad \Delta W_j \sim \mathrm{N}\,(0, \sqrt{h})\,. \tag{7.36}$$

The operator L^1 is defined by the rule $L^1\, f = \sum_k \frac{\partial f}{\partial x_k}\, b_k(t, \mathbf{x})$ and this in turn yields the Milstein scheme

$$x_k^{j+1} = x_k^j + a_k(t_j, \mathbf{x}_j)\, h + b_k(t_j, \mathbf{x}_j)\, \Delta W_j + \frac{1}{2} \sum_n b_n\, \frac{\partial b_k(t_j, \mathbf{x}_j)}{\partial x_n}\, [\,(\Delta W_j)^2 - h\,]\,, \tag{7.37}$$

where $\Delta W_j \sim \mathrm{N}\,(0, \sqrt{h})$. For example, in one dimension, the SDE $dx = a\, dt + b\, dW_t$ gives rise to the Milstein scheme

$$x^{j+1} = x^j + a(t_j, x^j)\, h + b(t_j, x^j)\, \Delta W_j + \frac{b}{2}\, \frac{db(t_j, x^j)}{dx}\, [\,(\Delta W_j)^2 - h\,]\,, \quad \Delta W_j \sim \mathrm{N}\,(0, \sqrt{h})\,. \tag{7.38}$$

7.6.1 Higher order schemes

For general systems of SDEs its already clear that higher order schemes are difficult unless the equations possess a high degree of symmetry. Whenever the SDEs are dependent on a single stochastic process, some progress may be possible. To improve upon the Milstein scheme, it is necessary to incorporate all terms which behave like $h^{3/2}$ into the numerical scheme. Expression (7.22) overtly contains two such terms in

$$L^\alpha\, a_k(t_0, \mathbf{x}_0)\, I_{\alpha\,0} + L^0\, b_{\,k\,\alpha}(t_0, \mathbf{x}_0)\, I_{0\,\alpha}$$

but there is one further term hidden in $R_k^{(3)}$, namely

$$L^\gamma\, L^\beta\, b_{\,k\,\alpha}(t_0, \mathrm{x}_0) \int_{t_0}^{t} \int_{t_0}^{s} \int_{t_0}^{u} dW_w^\gamma\, dW_u^\beta\, dW_s^\alpha = L^\gamma\, L^\beta\, b_{\,k\,\alpha}(t_0, \mathrm{x}_0)\, I_{\gamma\,\beta\,\alpha}\,.$$

Restricting ourselves immediately to the situation of a single Wiener process, the additional terms to be added to Milstein's scheme to improve the accuracy to a strong Taylor scheme of order 3/2 are

$$L^1\, a_k(t_0, \mathbf{x}_0)\, I_{10} + L^0\, b_k(t_0, \mathbf{x}_0)\, I_{01} + L^1\, L^1\, b_k(t_0, \mathrm{x}_0)\, I_{111} \tag{7.39}$$

in which the operators L^0 and L^1 are defined by

$$L^0\, f = \frac{\partial f}{\partial t} + \sum_k \frac{\partial f}{\partial x_k}\, a_k(t, \mathbf{x}) + \frac{1}{2} \sum_{j,k} \frac{\partial^2 f}{\partial x_j \partial x_k} b_k(t, \mathbf{x})\, b_j(t, \mathbf{x})\,, \qquad L^1\, f = \sum_k \frac{\partial f}{\partial x_k}\, b_k(t, \mathbf{x})\,. \tag{7.40}$$

The values of I_{01}, I_{10} and I_{111} have been computed previously in the exercises of Chapter 4 for the interval $[a, b]$. The relevant results with respect to the interval $[t_0, t]$ are

$$\begin{aligned} I_{10} + I_{01} &= (t - t_0)(W_t - W_0)\,, \\ I_{111} &= \frac{(W_t - W_0)}{6}\Big[\,(W_t - W_0)^2 - 3(t - t_0)\,\Big]\,. \end{aligned} \tag{7.41}$$

Formulae (7.41) determine the coefficients of the new terms to be added to Milstein's algorithm to generate a strong Taylor approximation of order 3/2. Of course, the Riemann integral I_{10} must still be determined. This calculation is included as a problem in Chapter 4, but its outcome is that I_{10} is a Gaussian process with a mean value of zero, variance $(t - t_0)^3/3$ and has correlation $(t - t_0)^2/2$ with $(W_t - W_0)$. Thus $\Delta Z_j = I_{10}$ may be simulated as a Gaussian deviate with mean zero, variance $h^3/3$ and such that it has correlation $h^2/2$ with ΔW_j. The extra terms are

$$\begin{aligned} \sum_m b_m\, a_{k\,,m}\, \Delta Z_j + \Big[\,\frac{\partial b_k}{\partial t} + \sum_m a_m\,\frac{\partial b_k}{\partial x_m} + \frac{1}{2}\sum_{m,r} b_m\, b_r\,\frac{\partial^2 b_k}{\partial x_m \partial x_r}\,\Big]\,(h\Delta W_j - \Delta Z_j) \\ + \frac{1}{6}\sum_{r,m} b_r \frac{\partial}{\partial x_r}\Big(\,b_m\,\frac{\partial b_k}{\partial x_m}\Big)\,(\Delta W_j)[\,(\Delta W_j)^2 - 3h\,] \end{aligned} \tag{7.42}$$

where all derivatives are understood to be evaluated at time t_j and state $\mathbf{x}_j$. Consider, for example, the simplest scenario of a single equation. In this case, the extra terms (7.42) become

$$\begin{aligned} & b\,a'\,\Delta Z_j + \Big[\,b_t + a\,b' + \frac{1}{2}\,b^2\,b''\,]\;(h\Delta W_j - \Delta Z_j) + \frac{1}{6}\,b\,(\,b\,b')'\,(\Delta W_j)[\,(\Delta W_j)^2 - 3h\,] \\ = \;& b\,a'\,\Delta Z_j + \Big[\,b_t + a\,b' + \frac{1}{2}\,b^2\,b''\,]\,(h\Delta W_j - \Delta Z_j) + \frac{b}{2}\,[\,(b')^2 + b''\,]\,[\,\frac{(\Delta W_j)^2}{3} - h\,]\,\Delta W_j\,. \end{aligned} \tag{7.43}$$

Chapter 8

Exercises on SDE

Exercises on Chapter 2

Q 1. Let $W_1(t)$ and $W_2(t)$ be two Wiener processes with correlated increments ΔW_1 and ΔW_2 such that $\mathbb{E}\,[\Delta W_1 \Delta W_2] = \rho \Delta t$. Prove that $\mathbb{E}\,[W_1(t) W_2(t)] = \rho\, t$. What is the value of $\mathbb{E}\,[W_1(t) W_2(s)]$?

Q 2. Let $W(t)$ be a Wiener process and let λ be a positive constant. Show that $\lambda^{-1} W(\lambda^2 t)$ and $t\, W(1/t)$ are each Wiener processes.

Q 3. Suppose that $(\varepsilon_1\,,\,\varepsilon_2)$ is a pair of uncorrelated $\mathrm{N}\,(0,1)$ deviates.

(a) By recognising that $\xi_x = \sigma_x\, \varepsilon_1$ has mean value zero and variance σ_x^2, construct a second deviate ξ_y with mean value zero such that the Gaussian deviate $X = [\,\xi_x\,,\,\xi_y\,]^{\mathrm{T}}$ has mean value zero and correlation tensor

$$\mathbf{\Omega} = \begin{bmatrix} \sigma_x^2 & \rho\, \sigma_x \sigma_y \\ \rho\, \sigma_x \sigma_y & \sigma_y^2 \end{bmatrix}$$

where $\sigma_x > 0$, $\sigma_y > 0$ and $|\,\rho\,| < 1$.

(b) Another possible way to approach this problem is to recognise that every correlation tensor is similar to a diagonal matrix with positive entries. Let

$$\alpha = \frac{\sigma_x^2 - \sigma_y^2}{2}\,, \qquad \beta = \frac{1}{2}\sqrt{(\sigma_x^2 + \sigma_y^2)^2 - 4(1-\rho^2)\sigma_x^2\sigma_y^2} = \sqrt{\alpha^2 - \rho^2 \sigma_x^2 \sigma_y^2}\,.$$

Show that

$$Q = \frac{1}{\sqrt{2\beta}} \begin{bmatrix} \sqrt{\beta + \alpha} & -\sqrt{\beta - \alpha} \\ \sqrt{\beta - \alpha} & \sqrt{\beta + \alpha} \end{bmatrix}$$

is an orthogonal matrix which diagonalises $\mathbf{\Omega}$, and hence show how this idea may be used to find $X = [\,\xi_x\,,\,\xi_y\,]^{\mathrm{T}}$ with the correlation tensor $\mathbf{\Omega}$.

(c) Suppose that $(\varepsilon_1, \ldots, \varepsilon_n)$ is a vector of n uncorrelated Gaussian deviates drawn from the distribution $\mathrm{N}(0,1)$. Use the previous idea to construct an n-dimensional random column vector X with correlation structure $\mathbf{\Omega}$ where $\mathbf{\Omega}$ is a positive definite $n \times n$ array.

Q 4. Let X be normally distributed with mean zero and unit standard deviation, then $Y = X^2$ is said to be χ^2 distributed with one degree of freedom.

(a) Show that Y is Gamma distribution with $\lambda = \rho = 1/2$.

(b) What is now the distribution of $Z = aX^2$ if $a > 0$ and $X \sim \mathrm{N}(0, \sigma^2)$.

(b) If $X_1, \cdots, X_n$ are n independent Gaussian distributed random variables with mean zero and unit standard deviation, what is the distribution of $Y = \mathbf{X}^{\mathrm{T}}\mathbf{X}$ where $\mathbf{X}$ is the n dimensional column vector whose k-th entry is X_k.

Exercises on Chapter 3

Q 5. Calculate the bounded variation (when it exists) for the functions

(a) $f(x) = |x|$ $\quad x \in [-1, 2]$ $\qquad$ (b) $g(x) = \log x$ $\quad x \in (0, 1]$

(c) $h(x) = 2x^3 + 3x^2 - 12x + 5$ $\quad x \in [-3, 2]$ $\qquad$ (d) $k(x) = 2\sin 2x$ $\quad x \in [\pi/12, 5\pi/4]$

(e) $n(x) = H(x) - H(x-1)$ $\quad x \in \mathbb{R}$ $\qquad$ (f) $m(x) = (\sin 3x)/x$ $\quad x \in [-\pi/3, \pi/3]$.

Q 6. Use the definition of the Riemann integral to demonstrate that

$$\text{(a)} \int_0^1 x\,dx = \frac{1}{2}, \qquad \text{(b)} \int_0^1 x^2\,dx = \frac{1}{3}.$$

You will find useful the formulae $\sum_{k=1}^n k = n(n+1)/2$ and $\sum_{k=1}^n k^2 = n(n+1)(2n+1)/6$.

Q 7. The functions f, g and h are defined on $[0, 1]$ by the formulae

$$\text{(a)} \quad f(x) = \begin{cases} x\sin(\pi/x) & x > 0 \\ 0 & x = 0 \end{cases}$$

$$\text{(b)} \quad g(x) = \begin{cases} x^2\sin(\pi/x) & x > 0 \\ 0 & x = 0 \end{cases}$$

$$\text{(c)} \quad h(x) = \begin{cases} (x/\log x)\sin(\pi/x) & x > 0 \\ 0 & x = 0, 1 \end{cases}$$

Determine which functions have bounded variation on the interval $[0, 1]$.

Exercises on Chapter 4

Q 8. Prove that

$$\int_a^b (dW_t)^2\, G(t) = \int_a^b G(t)\, dt$$

Q 9. Prove that

$$\int_a^b W^n\, dW_t = \frac{1}{n+1}\Big[\, W(b)^{n+1} - W(a)^{n+1}\,\Big] - \frac{n}{2}\int_a^b W^{n-1}\, dt\,.$$

Q 10. The connection between the Stratonovich and Ito integrals relied on the claim that

$$\lim_{n\to\infty} \mathbb{E}\Big[\sum_{k=1}^{n} \frac{\partial f(t_{k-1}\, W_{k-1})}{\partial W}\Big((W_{k-1/2} - W_{k-1})(W_k - W_{k-1}) - \frac{t_k - t_{k-1}}{2}\Big)\Big]^2 = 0$$

provided $\mathbb{E}\,[\,|\partial f(t,W)/\partial W|^2\,]$ is integrable over the interval $[a,b]$. Verify this unsubstantiated claim.

Q 11. Compute the value of

$$\Phi = \int_a^b W_t\, dW_t$$

where the definition of the integral is based on the choice $\xi_k = (1-\lambda)t_{k-1} + \lambda t_k$ and $\lambda \in [0,1]$.

Q 12. Solve the stochastic differential equation

$$dx_t = a(t)\, dt + b(t)\, dW_t\,, \qquad x_0 = X_0$$

where X_0 is constant. Show that the solution x_t is a Gaussian deviate and find its mean and variance.

Q 13. The stochastic integrals I_{01} and I_{10} defined by

$$I_{10} = \int_a^b \int_0^s dW_u\, ds\,, \qquad\qquad I_{01} = \int_a^b \int_0^s du\, dW_s\,.$$

occur frequently in the numerical solution of stochastic differential equations. Show that

$$I_{10} + I_{01} = (b-a)(W_b - W_a)\,.$$

Q 14. The stochastic integrals I_{111} defined by

$$I_{111} = \int_a^t \int_a^s \int_a^u dW_w\, dW_u\, dW_s$$

occurs frequently in the numerical solution of stochastic differential equations. Show that

$$I_{111} = \frac{(W_b - W_a)}{6}\Big[\,(W_b - W_a)^2 - 3(b-a)\,\Big]\,.$$

Q 15. Show that the value of the Riemann integral

$$I_{10} = \int_a^b \int_0^s dW_u\, ds$$

may be simulated as a Gaussian deviate with mean value zero, variance $(b-a)^3/3$ and such that its correlation with $(W_b - W_a)$ is $(b-a)^2/2$.

Exercises on Chapter 5

Q 16. Solve the degenerate Ornstein-Uhlenbeck stochastic initial value problem

$$dx = -\alpha\, x + \sigma\, dW\,, \qquad x(0) = x_0$$

in which α and σ are positive constants and x_0 is a random variable. Deduce that

$$\mathrm{E}\,[\,X\,] = \mathrm{E}\,[\,x_0\,]\, e^{-\alpha\, t}\,, \qquad \mathrm{V}\,[\,X\,] = \mathrm{V}\,[\,x_0\,]\, e^{-2\alpha t} + \frac{\sigma^2}{2\alpha}\,[\,1 - e^{-2\alpha t)}\,]\,.$$

Q 17. It is given that the instantaneous rate of interest satisfies the equation

$$dr = \mu(t, r)\, dt + \sqrt{g(t, r)}\, dW\,. \tag{8.1}$$

Let $B(t, R)$ be the value of a zero-coupon bond at time t paying one dollar at maturity T. Show that $B(t, R)$ is the solution of the partial differential equation

$$\frac{\partial B}{\partial t} + \mu(t, r)\frac{\partial B}{\partial r} + \frac{g(t, r)}{2}\,\frac{\partial^2 B}{\partial r^2} = rB \tag{8.2}$$

with terminal boundary condition $B(T, r) = 1$.

Q 18. It is given that the solution of the initial value problem

$$dx = -\alpha\, x + \sigma\, x\, dW\,, \qquad x(0) = x_0$$

in which α and σ are positive constants is

$$x(t) = x(0)\, \exp\Big[\, -(\alpha + \sigma^2/2)t + \sigma\, W(t)\,\Big]$$

Show that

$$\mathrm{E}\,[\,X\,] = x_0\, e^{-\alpha\, t}\,, \qquad \mathrm{V}\,[\,X\,] = e^{-2\alpha\, t}(e^{\,\sigma^2\, t} - 1)\, x_0^2\,.$$

Q 19. Benjamin Gompertz (1840) proposed a well-known law of mortality that had the important property that financial products based on male and female mortality could be priced from a single mortality table with an age decrement in the case of females. Cell populations are also well-know to obey Gompertzian kinetics in which $N(t)$, the population of cells at time t, evolves according to the ordinary differential equation

$$\frac{dN}{dt} = \alpha N \log\Big(\frac{M}{N}\Big)\,,$$

where M and α are constants in which M represents the maximum resource-limited population of cells. Write down the stochastic form of this equation and deduce that $\psi = \log N$ satisfies an OU process. Further deduce that mean reversion takes place about a cell population that is smaller than M, and find this population.

Q 20. A two-factor model of the instantaneous rate of interest, $r(t)$, proposes that $r(t) = r_1(t)+r_2(t)$ where $r_1(t)$ and $r_2(t)$ satisfy the square-root processes

$$\begin{aligned} dr_1 &= \alpha_1(\theta_1 - r_1)\,dt + \sigma_1\sqrt{r_1}\,dW_1\,, \\ dr_2 &= \alpha_2(\theta_2 - r_2)\,dt + \sigma_2\sqrt{r_2}\,dW_2\,, \end{aligned} \tag{8.3}$$

in which dW_1 and dW_2 are independent increments in the Wiener processes W_1 and W_2. Let $B(t, R_1, R_2)$ be the value of a zero-coupon bond at time t paying one dollar at maturity T.

Construct the partial differential equation satisfied by $B(t, R_1, R_2)$ and write down the terminal boundary condition for this equation.

Q 21. Solve the stochastic differential equation

$$dx_t = a(t)\,dt + b(t)\,dW_t\,, \qquad x_0 = X_0$$

where X_0 is constant. Show that the solution x_t is a Gaussian deviate and find its mean and variance.

Q 22. Solve the stochastic differential equation

$$dx_t = -\frac{x_t\,dt}{1+t} + \frac{dW_t}{1+t}\,, \qquad x_0 = 0\,.$$

Q 23. Solve the stochastic differential equation

$$dx_t = -\frac{x_t\,dt}{2} + \sqrt{1-x_t^2}\,dW_t\,, \qquad x_0 = a \in [-1, 1]\,.$$

Q 24. Solve the stochastic differential equation

$$dx_t = dt + 2\sqrt{x_t}\,dW_t\,, \qquad x(0) = x_0\,.$$

Q 25. Solve the stochastic differential equation

$$dx = [\,a(t) + b(t)\,x\,]\,dt + [\,c(t) + d(t)\,x\,]\,dW_t$$

by making the change of variable $y(t) = x(t)\,\phi(t)$ where $\phi(t)$ is to be chosen appropriately.

Q 26. In an Ornstein-Uhlenbeck process, $x(t)$, the state of a system at time t, satisfies the stochastic differential equation

$$dx = -\alpha(x - X)\,dt + \sigma\,dW_t$$

where α and σ are positive constants and X is the equilibrium state of the system in the absence of system noise. Solve this SDE. Use the solution to explain why $x(t)$ is a Gaussian process, and deduce its mean and variance.

Q 27. Let $\mathbf{x} = (x_1, \ldots, x_n)$ be the solution of the system of Ito stochastic differential equations

$$dx_k = a_k\, dt + b_{k\,\alpha}\, dW_\alpha$$

where the repeated Greek index indicates summation from $\alpha = 1$ to $\alpha = m$. Show that $\mathbf{x} = (x_1, \ldots, x_n)$ is the solution of the Stratonovich system

$$dx_k = \Big[\, a_k - \frac{1}{2}\, b_{j\,\alpha} b_{k\,\alpha,j} \,\Big]\, dt + b_{k\,\alpha} \circ dW_\alpha\,.$$

Let $\phi = \phi(t, \mathbf{x})$ be a suitably differentiable function of t and $\mathbf{x}$. Show that ϕ is the solution of the stochastic differential equation

$$d\phi = \Big[\, \frac{\partial \phi}{\partial t} + \bar{a}_k\, \frac{\partial \phi}{\partial x_k} \,\Big]\, dt + \frac{\partial \phi}{\partial x_k}\, b_{k\,\alpha} \circ dW_\alpha$$

where $\bar{a}_k = a_k - b_{k\,\alpha,j}\, b_{j\,\alpha}/2$.

Q 28. The displacement $x(t)$ of the harmonic oscillator of angular frequency ω satisfies $\ddot{x} = -\omega^2 x$. Let $z = \dot{x} + i\omega x$. Show that the equation for the oscillator may be rewritten

$$\frac{dz}{dt} = i\omega z\,.$$

The frequency of the oscillator is randomized by the addition of white noise of standard deviation σ to give random frequency $\omega + \sigma\xi(t)$ where $\xi \sim \mathrm{N}\,(0, 1)$. Determine now the stochastic differential equation satisfied by z.

Solve this SDE under the assumption that it should be interpreted as a Stratonovich equation, and use the solution to construct expressions for

$$(a) \quad \mathrm{E}\,[\, z(t)\,] \qquad\qquad (b) \quad \mathrm{E}\,[\, z(t)\, z(s)\,] \qquad\qquad (c) \quad \mathrm{E}\,[\, z(t)\, \bar{z}(s)\,]\,.$$

Exercises on Chapter 6

Q 29. It has been established in a previous example that the price $B(t, R)$ of a zero-coupon bond at time t and spot rate R paying one dollar at maturity T satisfies the Bond Equation

$$\frac{\partial B}{\partial t} + \mu(t, r)\frac{\partial B}{\partial r} + \frac{g(t, r)}{2}\, \frac{\partial^2 B}{\partial r^2} = rB$$

with terminal condition $B(T, r) = 1$ when the instantaneous rate of interest evolves in accordance with the stochastic differential equation $dr = \mu(t, r)\, dt + \sqrt{g(t, r)}\, dW$.

The popular square-root process proposed by Cox, Ingersol and Ross, commonly called the CIR process, corresponds to the choices $\mu(t, r) = \alpha(\theta - r)$ and $g(t, r) = \sigma^2 r$. It is given that the anzatz $B(t, r) = \exp[\beta_0(T - r) + \beta_1(T - t)r]$ is a solution of the Bond Equation in this case provided the coefficient functions $\beta_0(T-t)$ and $\beta_1(T-t)$ satisfy a pair of ordinary differential equations. Determine these equations with their boundary conditions.

Solve these ordinary differential equations and deduce the value of $B(t,R)$ where R is the spot rate of interest.

Q 30. In a previous exercise it has been shown that the price $B(t,R_1,R_2)$ of a zero-coupon bond at time t paying one dollar at maturity T satisfies the partial differential equation

$$\frac{\partial B}{\partial t}+\alpha_1(\theta_1-r_1)\frac{\partial B}{\partial r_1}+\alpha_2(\theta_2-r_2)\frac{\partial B}{\partial r_2}+\frac{\sigma_1^2 r_1}{2}\frac{\partial^2 B}{\partial r_1^2}+\frac{\sigma_2^2 r_2}{2}\frac{\partial^2 B}{\partial r_2^2}=(r_1+r_2)B\,,$$

when the instantaneous rate of interest, $r(t)$, is driven by a two-factor model in which $r(t)=r_1(t)+r_2(t)$ in which $r_1(t)$ and $r_2(t)$ evolve stochastically in accordance with the equations

$$\begin{aligned} dr_1 &= \alpha_1(\theta_1-r_1)\,dt+\sigma_1\sqrt{r_1}\,dW_1\,,\\ dr_2 &= \alpha_2(\theta_2-r_2)\,dt+\sigma_2\sqrt{r_2}\,dW_2\,, \end{aligned} \tag{8.4}$$

where dW_1 and dW_2 are independent increments in the Wiener processes W_1 and W_2.

Given that the required solution has generic solution $B(t,r_1,r_2)=\exp[\beta_0(T-r)+\beta_1(T-t)r_1+\beta_2(T-t)r_2]$, construct the ordinary differential equations satisfied by the coefficient functions β_0, β_1 and β_2. What are the appropriate initial conditions for these equations?

Q 31. The position $x(t)$ of a particle executing a uniform random walk is the solution of the stochastic differential equation

$$dx_t=\mu\,dt+\sigma\,dW_t\,,\qquad x(0)=X\,,$$

where μ and σ are constants. Find the density of x at time $t>0$.

Q 32. The position $x(t)$ of a particle executing a uniform random walk is the solution of the stochastic differential equation

$$dx_t=\mu(t)\,dt+\sigma(t)\,dW_t\,,\qquad x(0)=X\,,$$

where μ and σ are now prescribed functions of time. Find the density of x at time $t>0$.

Q 33. The state $\mathbf{x}(t)$ of a particle satisfies the stochastic differential equation

$$d\mathbf{x}_t=\mathbf{a}\,dt+\mathbf{b}\,d\mathbf{W}_t\,,\qquad \mathbf{x}(0)=\mathbf{X}\,,$$

where $\mathbf{a}$ is a constant vector of dimension n, $\mathbf{b}$ is a constant $n\times m$ matrix and $d\mathbf{W}$ is a vector of Wiener increments with $m\times m$ covariance matrix Q. Find the density of $\mathbf{x}$ at time $t>0$.

Q 34. The state $x(t)$ of a system evolves in accordance with the stochastic differential equation

$$dx_t=\mu\,x\,dt+\sigma\,x\,dW_t\,,\qquad x(0)=X\,,$$

where μ and σ are constants. Find the density of x at time $t>0$.

Q 35. The state $x(t)$ of a system evolves in accordance with the Ornstein-Uhlenbeck process

$$dx = -\alpha\,(x-\beta)\,dt + \sigma\,dW_t\,, \qquad x(0) = X\,,$$

where α, β and σ are constants. Find the density of x at time $t > 0$.

Q 36. If the state of a system satisfies the stochastic differential equation

$$dx = a(x,t)\,dt + b(x,t)\,dW\,, \qquad x(0) = X\,,$$

write down the initial value problem satisfied by $f(x,t)$, the probability density function of x at time $t > 0$. Determine the initial value problem satisfied by the cumulative density function of x.

Q 37. Cox, Ingersoll and Ross proposed that the instantaneous interest rate $r(t)$ should follow the stochastic differential equation

$$dr = \alpha(\theta - r)dt + \sigma\sqrt{r}\,dW\,, \qquad r(0) = r_0\,,$$

where dW is the increment of a Wiener process and α, θ and σ are constant parameters. Show that this equation has associated transitional probability density function

$$f(t,r) = c\Big(\frac{v}{u}\Big)^{q/2} e^{-(\sqrt{u}-\sqrt{v})^2}\, e^{-2\sqrt{uv}}\, I_q(2\sqrt{uv})\,,$$

where $I_q(x)$ is the modified Bessel function of the first kind of order q and the functions c, u, v and the parameter q are defined by

$$c = \frac{2\alpha}{\sigma^2(1-e^{-\alpha(t-t_0)})}\,, \qquad u = cr_0\,e^{-\alpha(t-t_0)}\,, \qquad v = cr\,, \qquad q = \frac{2\alpha\theta}{\sigma^2} - 1\,.$$

Exercises on Chapter 7

Q 38. Consider the problem of numerically integrating the stochastic differential equation

$$dx = a(t,x)\,dt + b(t,x)\,dW\,, \qquad x(0) = X_0\,.$$

Develop an iterative scheme to integrate this equation over the interval $[0,T]$ using the Euler-Maruyama algorithm.

It is well-know that the Euler-Maruyama algorithm has strong order of convergence one half and weak order of convergence one. Explain what programming strategy one would use to demonstrate these claims.

Q 39. Compute the stationary densities for the stochastic differential equations

(a) $dX = (\beta - \alpha X)\,dt + \sigma\,\sqrt{X}\,dW$

(b) $dX = -\alpha \tan X \, dt + \sigma \, dW$

(c) $dX = [(\theta_1 - \theta_2)\cosh(X/2) - (\theta_1 + \theta_2)\sinh(X/2)] \, \cosh(X/2) \, dt + 2\cosh(X/2) \, dW$

(d) $dX = \dfrac{\alpha}{X} \, dt + dW$

(e) $dX = \left(\dfrac{\alpha}{X} - X\right) dt + dW$