An introduction to modelling and likelihood inference with stochastic differential equations

Omiros Papaspiliopoulos; revised by Alexandros Beskos

Warwick

Summary

In this short course we will approach a topic which stands at the interface of probability theory, real analysis, measure theory, numerical analysis, and statistical inference and computation. The models and methods we will consider are employed by several and diverse scientific areas such as econometrics, engineering, chemistry, statistical physics and biology. Thus, unavoidably we will have to be selective.

The general setting is the following: there is a process X which evolves continuously in time, so that $X(t)$ is the value of the process at time t. We will postulate a class of stochastic models called *diffusion processes* for describing the dynamics of X. Such processes are determined as solutions of Stochastic Differential Equations (SDEs). The models will be specified up to some unknown parameters, say θ, and we will estimate θ using likelihood methods based on an observed sample path $\{X(t); t \in [0, T]\}$ for some given time horizon T.

The exposition begins (Chapter 1) by reviewing some basic concepts in the context of Ordinary Differential Equations (ODEs). In Chapter 2 we motivate the theory by presenting some real-life applications of differential equations; the concept of the Stochastic Differential Equation will appear in this section for the first time. In Chapter 3 we explain the construction of SDEs. Then, in Chapter 4 we will show how to obtain a likelihood function under such stochastic models and how to carry out statistical inference.

Our priority throughout the notes will be to illustrate the main concepts intuitively, keeping mathematical proofs to a minimum. So, the presentation will be rather informal. The material has been selected with a view of being general enough to bring out the main challenges of the topic but at the same time introductory enough to require only a modest mathematical background.

It is worth mentioning that statistical inference for diffusion processes is

a research topic of intensive current investigation.

References

For background reading, the following references can be consulted. For ODEs, one can resort to "An introduction to Ordinary Differential Equations" by J.C. Robinson (CUP). For real analysis and theory of Lebesgue-Stieltjes integration, see "Principles of Mathematical Analysis", by W. Rudin (International Series in Pure & Applied Mathematics). For SDEs and Itô calculus, see "Stochastic Differential Equations. An introduction with applications", by B. Oksendal (Springer). For likelihood inference for diffusions based on high-frequency data see the article by G.O.Roberts and N.Polson, "Bayes factors for discrete observations from diffusion processes" (Biometrika, 1994, 81, 1, pp.11-26). For a general treatment of deterministic and stochastic dynamical systems see the review article "Deterministic and random dynamical systems: theory and numerics", by A.R. Humphries and A.M. Stuart (available from http://www.sussex.ac.uk/Units/SMS/Reports/CMAIA/CMAIA01-09.ps).

Chapter 1

Deterministic Modelling of Time-Evolving Processes

Let X be a real-valued function, $X : R \mapsto R$. We will typically denote the argument of X by s. An n-th order Ordinary Differential Equation (ODE) for X is an equation which involves X, s, and the first n derivatives of X, i.e. an equation of the form:

$$g(X^{(1)}, X^{(2)}, \ldots, X^{(n)}, X, s) = 0 \ ,$$

where $g : R^{n+2} \mapsto R$, and $X^{(i)}$ denotes the i-th derivative of X with respect to (w.r.t.) s. When a function X satisfies this equation we say that it is *a solution* of it. Differential equations play a fundamental role in pure and applied mathematics.

We will be interested in the case where s denotes *time*. In this case we will refer to $\{X(s); s \in I\}$, for some given time interval I, as the path of X. Moreover, we will concentrate on 1st order, *autonomous* (also called time-homogeneous) ODEs, i.e. equations of the form:

$$\frac{dX}{ds} = f(X(s)) \ . \tag{1.1}$$

We will refer to f as the *drift* of the process. It is helpful to think of $X(s)$ as the position of a particle at time s, in which case (1.1) specifies the velocity of the particle as a function of its location.

1.1 Solution of an ODE

Consider the ODE (1.1) with the additional restriction (*initial condition*) that $X(s_0) = x_0$ for the time instance s_0. For simplicity we may assume that $s_0 = 0$. The pair of the ODE and the initial condition is known as an initial value problem:

$$\frac{dX}{ds} = f(X(s))\,, \quad X(0) = x_0\,. \tag{1.2}$$

Notice that the initial value problem is equivalent to finding a solution to the integral equation:

$$X(t) = x_0 + \int_0^t f(X(s))ds\,. \tag{1.3}$$

Definition 1. *A solution of the initial value problem (1.2) on a open interval I that contains the origin 0, is a differentiable function X with $X(0) = x_0$ and $X^{(1)}(s) = f(X(s))$ for all $s \in I$.*

Example 1.1. Consider the initial value problem:

$$\frac{dX}{ds} = \alpha X(s), \quad X(0) = x_0 > 0\,. \tag{1.4}$$

This ODE can be used as an approximate model for the growth of a population. In this case the constant α represents the rate of growth (birthrate minus deathrate) of the population. The function:

$$X(s) = x_0 e^{\alpha s} \tag{1.5}$$

is a solution of (1.4) since one can check that: $X^{(1)}(s) = \alpha x_0 e^{\alpha s} = \alpha X(s)$, and $X(0) = x_0$. When $\alpha > 0$ the solution tends to infinity, exponentially fast as $s \to \infty$, whereas it tends to 0 exponentially fast when $\alpha < 0$. □

This example points to an interesting qualitative characteristic of (1.1). Namely, the linear drift $f(x) = \alpha x$ corresponds to functions which change exponentially with time, that is processes which change rapidly with time. Then, sub-linear drifts will provide processes that cannot evolve faster than exponentially. This property is rigorously described by *Gronwall's inequality*.

Gronwall's inequality
Consider the ODE (1.1) with $f(x) \leq \alpha x$ for some real α. Then, for any s, t with $s < t$:

$$X(t) \leq X(s) e^{\alpha(t-s)} .$$

This result provides a first step towards understanding the effect of the drift f on the dynamics of X.

Exercise 1.1. *Prove Gronwall's inequality.*

1.1.1 Existence and Uniqueness of Solutions

We consider the initial value problem (1.2). Existence and uniqueness of a solution will be implied by properties of the drift f. The following example illustrates that existence of a solution is not guaranteed.

Exercise 1.2. *i) Suppose that $f(x) = 1$, for $x \leq 0$, and $f(x) = -1$, for $x > 0$, and $x_0 = 0$. Show that the initial value problem does not have any solutions. ii) Identify a solution for the initial value problem when $f(x) = -1$ for $x \leq 0$, $f(x) = 1$ for $x > 0$ and $x_0 = 0$.*

Therefore, if f is discontinuous solutions might not exist.

Uniqueness: Suppose that X and Y are two solutions of (1.2). We consider the difference $Z(s) = X(s) - Y(s)$, thus $Z(0) = 0$. For uniqueness of a solution, it is required that $Z(s) = 0$ for all s. By assumption:

$$d|Z|^2/ds = 2Z(s)Z^{(1)}(s) = 2Z(s)(f(X(s)) - f(Y(s))) .$$

We consider the following condition.

Lipschitz condition on a set A
There exist an $0 < L < \infty$ such that for any $x, y \in A$,

$$|f(x) - f(y)| \leq L|x - y| .$$

Assuming that the Lipschitz condition holds on the real line R we have that $d|Z|^2/ds \leq 2L|Z|^2$. Gronwall's inequality now yields $|Z(s)|^2 \leq |Z(0)|^2 e^{2Ls}$. Since $Z(0) = 0$ we have that $X(s) = Y(s)$ for all s. We just showed that under the Lipschitz condition on f, there can be only one solution of (1.2).

Exercise 1.3. *i) Prove (using the Mean Value Theorem) that when f is differentiable, the Lipschitz condition on an interval $A = (a, b)$ is equivalent to $f^{(1)}$ being bounded on A. ii) Prove that if f satisfies the Lipschitz condition for all $x, y \in R$, then $f(x)$ has a linear growth bound: $|f(x)| \leq \alpha|x| + \beta$, for appropriate constants α, β.*

Notice that if f satisfies the Lipschitz condition only on $A \subset R$ with $x_0 \in A$, then the argument we used earlier can be adapted to establish that can be only one solution on a time interval $(-\delta, \delta)$ for some $\delta > 0$. To see that, consider for example that $A = (a, b)$. Since any solution X will be continuous in s, there will be a $\delta > 0$ such that $X(s) \in (a, b)$ for all $|s| < \delta$. One can now follow the arguments above to show that there can be only one solution on $(-\delta, \delta)$.

Consider the case when $f^{(1)}$ is continuous on a bounded interval which contains x_0. Due to continuity, $f^{(1)}$ is also bounded on this interval. Exercise 1.3 now implies that f will satisfy the Lipschitz condition on that interval and uniqueness on a time interval around 0 is guaranteed. For the case of discontinuous $f^{(1)}$, we present the following precautionary example.

Example 1.2. Let $f(x) = \sqrt{x}$ and $x_0 = 0$. Notice that $f^{(1)}$ is not continuous at x_0 since $f^{(1)}(x) \to \infty$, as $x \downarrow x_0$. Thus, it is impossible to establish uniqueness for any time interval appealing to the argument we have given above. In fact, this is a case when multiple solutions exist. The zero function $X(s) = 0$ is a solution, but so is the following function:

$$X(s) = \begin{cases} 0, & s \leq c, \\ (s - c)^2/4, & s > c \ , \end{cases}$$

for any $c > 0$. Notice the unpredictable behavior of the process: it waits at 0 until suddenly, at a time not determined by either its current value or its drift, it starts increasing to infinity. □

Existence: The Lipschitz condition is also sufficient for showing existence of a solution. We will not present the proof in any detail apart from mentioning its first step. The initial value problem is equivalent to finding a solution

to the integral equation

$$X(t) = x_0 + \int_0^t f(X(s))ds \ . \tag{1.6}$$

We start with an initial guess for a solution, $X_0(s) = x_0$, and we obtain X_1 via the equation above as $X_1(t) = x_0 + \int_0^t f(X_0(s))\, ds = x_0 + tf(x_0)$. We iterate this scheme to obtain a sequence of guesses X_i, $i = 1, 2, \ldots$ as follows:

$$X_n(t) = x_0 + \int_0^t f(X_{n-1}(s)), \quad n \geq 1 \ .$$

Using the Lipschitz condition we show that the sequence $X_n(t)$ converges (as $n \to \infty$) to a function $X(t)$ which solves the ODE. This approach is known as *Picard's iteration*. It is useful to remember the following pathological example.

Example 1.3. Let $f(x) = -x^3$ and $x_0 \neq 0$. Here f is not Lipschitz on R. It can be checked that

$$X(s) = x_0/\sqrt{1 + 2sx_0^2}, \quad s > -1/(2x_0^2),$$

is a solution. However, $X(s)$ explodes as $s \to -1/(2x_0^2)$, therefore the solution does not exist for all s. □

1.2 Prediction

An ODE describes the *microscopic* behavior of the process X, i.e. the motion of the process in an infinitesimal time period: for small h we have

$$X(h) = x_0 + \int_0^h f(X(s))\, ds \approx x_0 + hf(x_0) \ . \tag{1.7}$$

A reason why ODEs appear so frequently in mathematical modelling is because in many contexts we have a very good idea (from physical arguments or otherwise) about the microscopic behavior of a process on small time intervals. The *macroscopic* behavior of the process is then implied by the solution of the ODE. Note that it is usually impossible to obtain explicit solutions (even when we know that they exist) for most of the ODEs used in practice.

That is, one cannot obtain an analytical formula for $X(s)$ as a function of s and x_0. Practitioners usually resort to numerical approximations of the unavailable solutions which most times are sufficient for practical purposes. If a solution is determined (explicitly or numerically) we can then predict both the *future* ($s > 0$) and the *past* ($s < 0$) given the value of the process at present.

1.3 Qualitative Analysis of the ODE

We will now assume that a unique solution exists for the initial value problem (1.2). The structure of the drift f can provide important insight into the shape of the solution of (1.2) even if the latter is not explicitly available. The qualitative analysis of the ODE refers precisely to an investigation of the drift to obtain information for the path of X.

The first step is to identify points x^* such that $f(x^*) = 0$; these are called *stationary points*. If the process starts at x^*, it never leaves it; this is implied by the hypothesis for uniqueness a of solution. The stationary points are distinguished into *stable* and *unstable*. Stable points attract particles starting next to them, whereas unstable points force particles to move away from them. More rigorously, stable are those stationary points for which f is positive on their left and negative on their right (the other way around for unstable points). If X starts between a stable and unstable point, it will move away from the unstable point and towards the stable one without ever (this follows again from uniqueness of solution) reaching it. Thus, plotting the drift $f(x)$ against x reveals useful information about the path of X under any possible initial condition.

The above remarks are better understood through an example. In Figure 1.1 we have plotted an example drift function. The stationary points are labeled as y_0, y_1, y_2. Assume that X starts at time 0 between y_0, y_1, that is $y_0 < x_0 < y_1$. Since $X'(0) = f(x_0) < 0$, X will start decreasing and move towards y_0. In fact, $X(s) \to s_0$ as $s \to \infty$. Thinking similarly, one can understand the behavior of X for any initial position x_0.

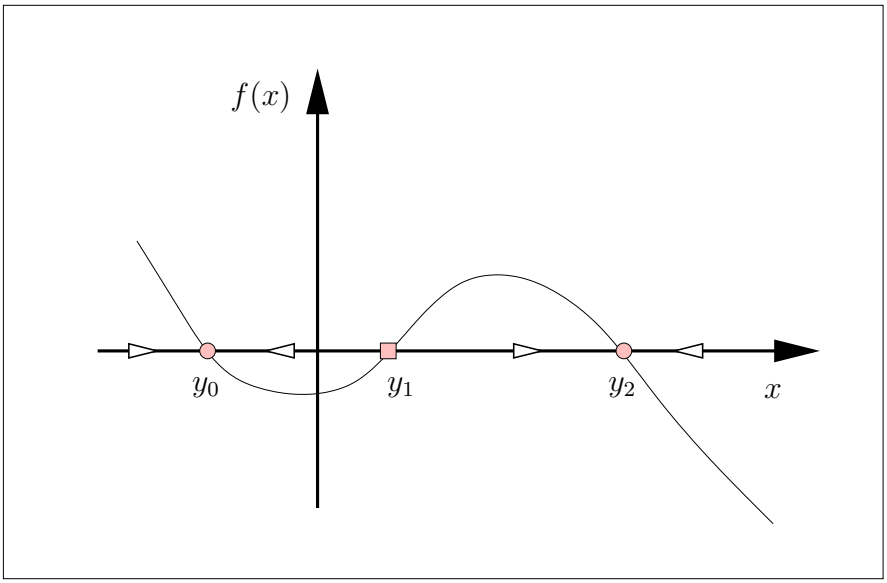

Figure 1.1: An example of a drift function. Circles indicate the stable stationary points and squares the unstable ones.

1.4 Numerical Approximation of ODEs

Suppose that we wish to calculate $X(t)$ for some time $t > 0$ (the case $t < 0$ is handled similarly). This will involve solving the initial value problem. As we have already stated though, in many cases even if a solution exists and is unique, it can be analytically intractable; that is we cannot express $X(s)$ in terms of s and x_0. In such cases, numerical methods can be employed to approximate the value of $X(t)$. Such problems fall in area of mathematics known as numerical analysis.

The simpler approximation method, known in the literature as the *Euler-Maruyama* scheme, build on an discretisation of the infinitesimal dynamics of the ODE. In particular, for small time increment h one can have:

$$X(t+h) - X(t) = \int_t^{t+h} f(X(s))ds \approx f(X(t))h \ , \tag{1.8}$$

after assuming that the function $s \mapsto f(X(s))$ is approximately constant on $[t, t+h]$. A helpful picture to have in mind is given in Figure 1.2. To approximate the complete function $\{X(t); t \in [0, T]\}$ for some given time

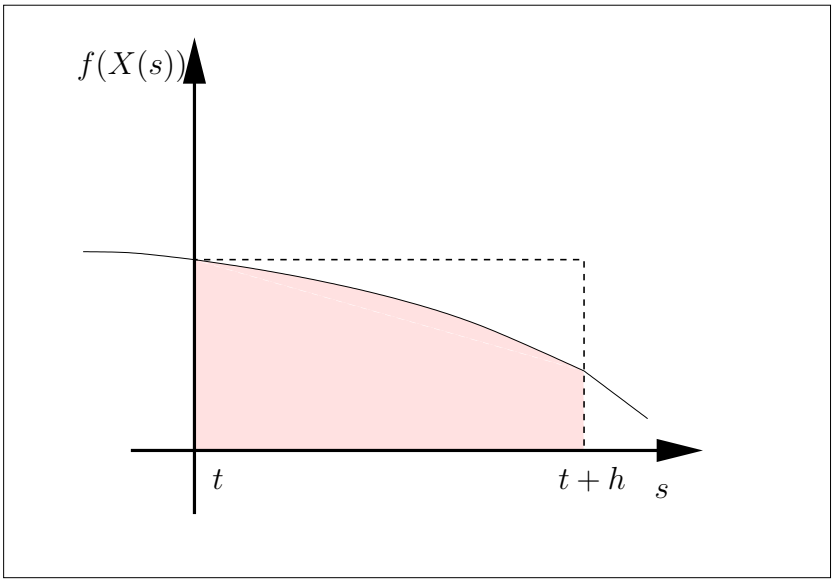

Figure 1.2: Exact value and approximation of $X(t+h) - X(t)$. The exact value is given by the shaded area. For small h, the Euler scheme approximates the shaded area with the area of a rectangle of height $f(X(t))$ and width h.

horizon T, one can divide the interval $[0, T]$ into n pieces, each of length $h = T/n$, and apply (1.8) on each of them separately. That is, if $\tilde{X}$ denotes the approximation of X:

$$\tilde{X}(ih) = \tilde{X}((i-1)h) + f(\tilde{X}((i-1)h))\, h, \quad i \geq 1 \ ,$$

with starting point $\tilde{X}(0) = x$. Note that the Euler scheme will not deliver a continuous-time path, but only a finite collection of points. Typically, successive points are connected with straight line to give the impression of a continuous curve.

The Euler scheme is one of the many available for the approximation of ODEs. Under conditions, it's error diminishes linearly with h. Alternative schemes can provide better approximations. The Euler method is nonetheless the simplest one to comprehend and implement.

Chapter 2

Differential Equations in Practice

2.1 Epidemics

In 1927, Kermack and McKendrick presented a model for describing the spread of a disease in a population. The model is simplistic and in the years that followed formed the base for the construction of much more sophisticated and accurate models. However, it can still provide an initial means for the exploration of a disease. The model divides the population into three categories: Susceptible, Infectious, Recovered. Susceptibles are those who have so far avoided the disease and can potentially get it in the future. Infectious are those that carry the disease and can transmit it to others. Recovered are those who carried the disease in the past but now recovered and can no longer transmit it. The model is called SIR after the initials of the three groups.

Let $S(t), I(t), R(t)$ denote the number of people in each of the three groups at time t. The model postulates that the size of the three groups evolve in time as follows:

$$\begin{aligned}
\frac{dS}{dt} &= B - v\,S(t)I(t) - d\,S(t) \\
\frac{dI}{dt} &= v\,S(t)I(t) - d\,I(t) \\
\frac{dR}{dt} &= g\,I(t) - d\,R(t)\ .
\end{aligned}$$

Disease	R_0
SARS	2.2 to 3.6
AIDS	2 to 5
Smallpox	3 to 5
Measles	16 to 18
Malaria	> 100

Table 2.1: The value of the basic reproductive ration R_0 for various diseases.

Here, B denotes the birth rate, d the death rate, v the contact rate of the disease and $1/g$ it's infectious period. We will provide some reasoning for this model.

First, note that it is determined in terms of differential equations: the derivatives $dS/dt, dI/dt, dR/dt$, that is the rate of change for the size of each group at time t, are given as a function of the current group sizes $S(t), I(t), R(t)$. Consider for instance the differential equation for dS/dt. The number of susceptibles $S(t)$ changes in time. It increases because of births; this is incorporated in the equation in terms of the parameter B. Then, $S(t)$ decreases due to the number of objects that get the disease and pass to the group of infectious; to incorporate that, the equation uses the term $-vS(t)I(t)$, with the minus sign indicating decrease, $S(t)I(t)$ corresponding to the interactions between the two groups $S(t), I(t)$ and v referring to the 'easyness' with which the disease can pass from one object to another. One can similarly understand the logic behind the other two equations.

Epidemiologists use a single parameter to provide an immediate index for the strength of a disease. This parameter, called the 'basic reproductive ratio', is defined as follows:

$$R_0 = v/g \ . \tag{2.1}$$

Intuitively, it could be thought of as the average number of susceptible people infected by a single infectious individual in a lifetime. Table 2.1 shows the estimated value of R_0 for various diseases. When $R_0 < 1$, the disease will eventually disappear, whereas if $R_0 > 1$ it will spread. Note that R_0 can change dynamically over time depending on measures taken by authorities to control the spread of the disease or other considerations. In the case of the Foot-and-Mouth virus spread in 2001 in the UK, the initial value for R_0 was 3.3; after the authorities imposed restrictions on the movement of

livestock and forced culling on infected farms, R_0 fell to 0.65 and the disease gradually disappeared[1]. The values in Table 2.1 correspond to a free spread of a disease, before any prophylactic measures are introduced.

The basic reproductive ratio is also related with vaccination policies. It can be shown that the percentage of the population that must be vaccinated to eliminate a disease is:

$$1 - 1/R_0 \ .$$

So, this index offers an indication of whether a vaccination policy is economically feasible given the provided means. Otherwise, alternative policies must be followed (culling, quarantine etc.)

2.2 Finance: Option Pricing

We will now consider a context where the introduction of "chance" or randomness in the evolution of a system seems inevitable: stock market. In 1997 Robert Merton and Myron Scholes received the Nobel prize in Economics for developing a mathematical framework for the fair pricing of *options*. Their work was presented several years earlier in the articles "The Pricing of Options and Corporate Liabilities" by Black and Scholes (1973), and "Theory of Rational Option Pricing", Merton (1973). (Fischer Black died before the award of the Nobel). By the time they received the Nobel, it was already realized that their work would be of dramatic impact on the way that complex financial products were priced. Fundamental in the analysis presented in the above papers is the notion of the Stochastic Differential Equation (SDE).

An option is a financial object providing the right (but not the obligation) to execute, in the future, a financial agreement on some underlying security (e.g. stock, currency, interest rate) arranged at present. The agreement involves a *buyer* and a *seller* of the option. Assume for instance that the underlying security is some stock and that the option gives the right to the buyer to acquire the stock at some future time $t = T$ at the *execution price* K. This is the description of the *European call option*. The problem is the identification of the *fair* price of the option now ($t = 0$). Black and Scholes answered this question in the context of a simple and (as understood later) rather unrealistic model. Yet, they provided a framework of mathematical

[1]See the very interesting (and easy to read) report 'Foot-and-mouth disease under control in the UK' published in *Nature* (2001)

modelling and ideas which were subsequently applied in much more robust contexts.

In the classical Black and Scholes model, the market is comprised of a *safe* and a *risky* asset. The safe asset (e.g. bank investment), denoted by $X_0(t)$ is assumed to evolve deterministically in time according to an ODE:

$$dX_0(t) = \rho X_0(t)dt, \quad X_0(0) = 1 \ , \tag{2.2}$$

for some interest rate ρ. The risky asset (e.g. a stock) will necessarily evolve randomly in time; Black and Scholes used an SDE to model it's fluctuations. They assumed that relative changes of the stock price over small time increments are described via a combination of a deterministic and a random component. That is, for small h:

$$\frac{X(t+h) - X(t)}{X_t} = a\, dt + \sigma \times \text{“random noise”}$$

for some parameter α providing a general trend for the stock price, and σ corresponding to the magnitude of the random fluctuations. For $h \to 0$, the above equation can be rigorously recognised as the following SDE:

$$dX(t) = \alpha\, X(t)dt + \sigma\, X(t)\, dB(t), \quad X(0) = x \ , \tag{2.3}$$

known in the literature as the *geometric Brownian motion*; the randomness emanates from the presence of the Brownian motion $B(t)$ in the equation. We will make sense of this type of equations in the following section. For the price of the European call option, with the stock $X(t)$ as it's underlying security, to be fair neither the buyer nor the seller of the option should be making certain profit from exchanging it. Based on this principle, and using the explicit mathematical modelling (2.2)-(2.3) for the market, Black and Scholes found that the price of the option should be the following:

$$\mathrm{E}_Q\, [\, e^{-\rho T}\, (X(T) - K)^+ \,] \ . \tag{2.4}$$

We will avoid stating details about the meaning of this formula. Note only that $\mathrm{E}_Q\, [\, \cdot \,]$ refers to expectation w.r.t. to the random quantity $X(T)$. We have only written the above expression to indicate it's *explicitness*, which came as a surprise for the option pricing people. Carrying out some calculations, one can obtain the following alternative expression for the option price:

$$\frac{e^{-\rho T}}{\sigma\sqrt{2\pi T}} \int_\gamma^\infty (x \exp\{y + (\rho - \sigma^2)T\} - K) \exp\{-y^2/(2\sigma^2 T)\} dy \ ,$$

where $\gamma = \log(K/x) - \rho T + \sigma^2 T/2$. For more details on the Black and Scholes theory, see Chapter 12 of "Stochastic Differential Equations. An introduction with applications" of Oksendal.

Chapter 3

Modelling with Stochastic Differential Equations

To realize the need for a *stochastic* analogue to (1.1), consider the case when one investigates the evolution of the number of cells in an organism infected by a virus. A candidate model could be the population growth model given in (1.4). Assume now that, even if the rate of growth of the infected cells is "on average" α, it is in fact not constant with time but fluctuates around α due to changing *unpredictable* biological conditions inside the organism. Therefore, we are in a situation where we have insufficient understanding of the effect of the biological conditions to the growth rate to specify a deterministic mathematical model describing it's dynamical evolution. Or indeed, even if we can speculate such a model it might be too complicated to analyze and our analysis may become very sensitive to model misspecification.

On the other hand, a more parsimonious approach is to try to construct a stochastic model according to which the growth rate at time t is $\alpha+$"error":

$$\frac{dX}{ds} = (\alpha + \text{“error”})X(s)\,, \quad X(0) = x_0\,. \tag{3.1}$$

We would like a specification of the "error" terms so that their mean is 0, therefore "on average" the growth rate is α. Notice that if the variance of the "error" is 0 our model reduces to an ODE. The picture we have in mind is given in Figure 3.1. There, we plot the solution of the ODE (1.4) with $x_0 = 2$, and $\alpha = 0.8$. The solution is given by the *smooth* curve which grows exponentially at a constant rate α. We have superimposed three *rough* paths with a stochastic behaviour, which imitate the dynamics prescribed in (3.1).

The paths fluctuate around the deterministic solution, in a way that the mean of the "error" is 0. Later in this section we will see that the rough paths have been *simulated* according to a *Stochastic Differential Equation* (SDE).

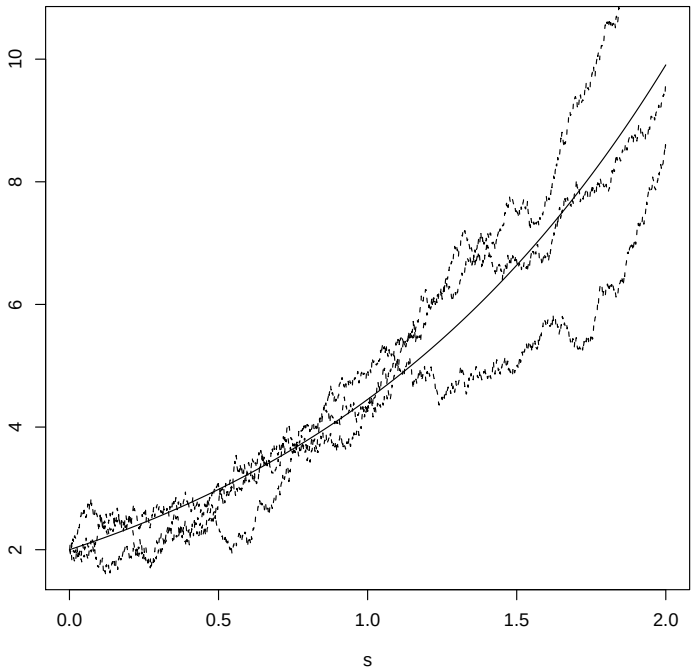

Figure 3.1: The solution (with the smooth line) of the linear ODE (1.4) with $x_0 = 2$, $\alpha = 0.8$ and three stochastic paths moving around it.

As a second motivating example, consider the data set of Figure 3.2. This is a series of Eurodollar spot deposit rates recorded every ten days in the period from 1973 to 1995 (see http://en.wikipedia.org/wiki/Eurodollar for a definition of the Eurodollar rate). Notice the roughness and stochastic behavior of the process.

Subsequently, one could imagine a general approach where uncertainty is introduced in the microscopic dynamics of a time-evolving process. Can we formally construct such models whose paths look like those in Figures 3.1 and 3.2? The purpose of this chapter is two-fold. Firstly, to demonstrate that it is highly non-trivial to give a precise mathematical description of the stochastic model we have intuitively postulated. Secondly, to give some clues

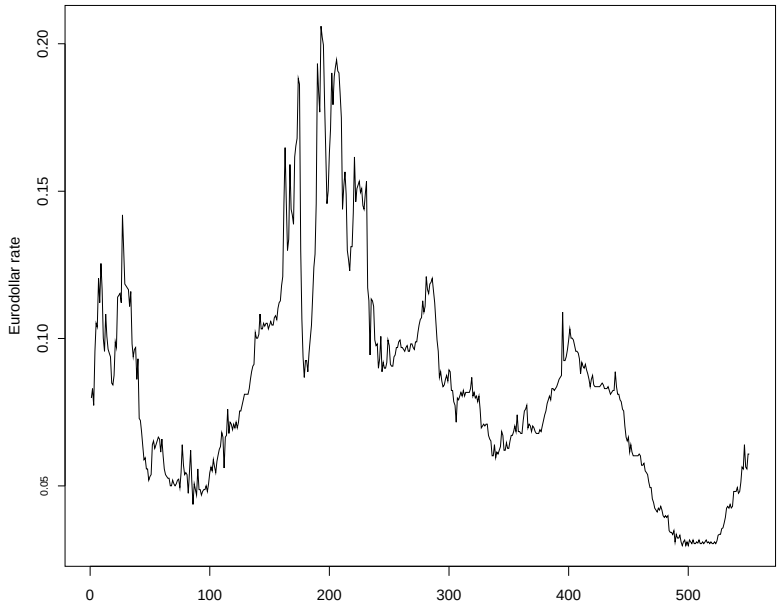

Figure 3.2: Eurodollar deposit rates in the period 1973-1995 recorded every ten days.

and intuition as to how this is done.

A Note On Randomness:
To incorporate random fluctuations in the model, mathematics use the concept of the *stochastic process*. This object is defined as a collection of *random variables*, say $\{X(t)\,;\, t \in [0,T]\}$. So, X evolves in time in a random fashion. To indicate randomness, we introduce a second argument ω and write $X(t) = X(\omega, t)$, with ω a given element of a sample space Ω. Intuitively, Ω corresponds to all possible outcomes of an experiment, and $\omega \in \Omega$ to a particular realisation of the experiment. Repeating the experiment several times results to various outcomes $(\omega_1, \omega_2, \ldots)$ and different values for the process X related with it. Each outcome is given a probability according to some distribution P. In Figure 3.3, we show three paths corresponding to three different realisations of the same random process X.

More rigorously, an experiment with random outcome corresponds to a probability space $(\Omega, \mathcal{F}, \mathrm{P})$ with Ω the sample space, $\mathcal{F}$ a σ-algebra comprised of collections of events and P a probability law. A random process is then an appropriately defined (measurable) mapping $X : (\Omega, R_+) \mapsto R$. Sometimes it is useful to think of the time instance t as being fixed and ω varying. Then

one can calculate probabilities of the form:

$$\mathrm{P}\,[\,X(t) \in (0,1)\,] = \mathrm{P}\,[\,\{\omega : X(\omega,t) \in (0,1)\}\,]$$

or $\mathrm{P}\,[\,\{X(t_1) \geq 0\} \cap \{X(t_2) < 0\}\,]$, etc. Other times, it is of interest to examine path properties of the process, for example $\mathrm{P}[\,t \mapsto X(t)$ is continuous$\,]$ or $\mathrm{P}[\,\sup_{t\in[0,1]} X(t) < 1\,]$. In this case one should consider the complete paths $t \mapsto X(\omega,t)$ for any given $\omega \in \Omega$. We will sometimes add the argument ω and write $X(t,\omega)$ to emphasize the stochastic nature of the process. Most times, though, we avoid it's use for notational simplicity.

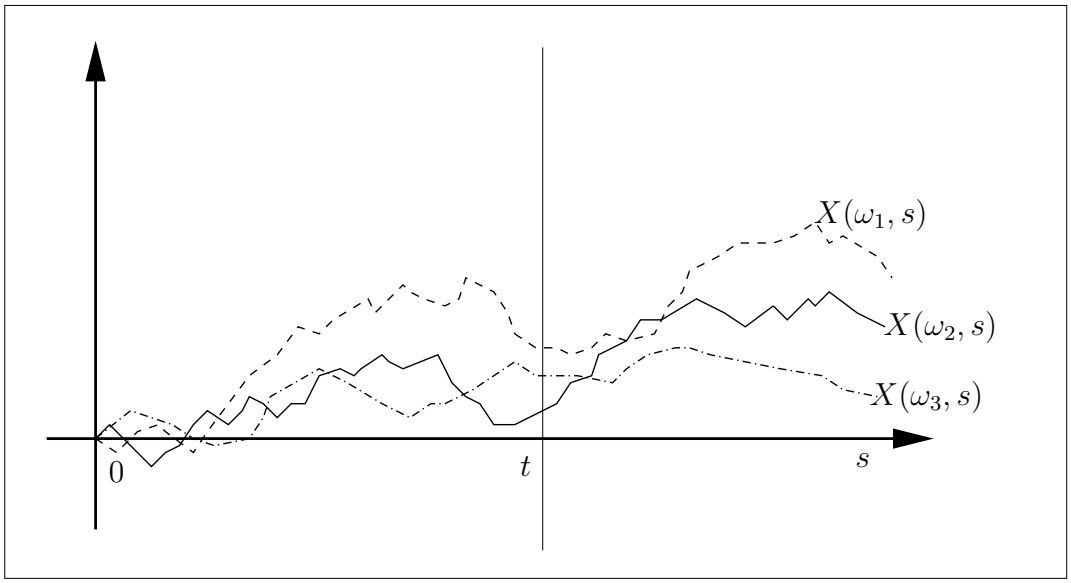

Figure 3.3: A stochastic process X. Three sample paths corresponding to three different randomly chosen (according to P) ω's.

3.1 A Stochastic Differential Equation

We will begin by looking at a stochastic analogue of the ordinary integral equation (1.3). We will give meaning to an equation of the following form:

$$X(t) = x_0 + \int_0^t f(X(s))ds + \sigma B(t), \quad t \geq 0\ . \tag{3.2}$$

There are three terms on the right-hand-side: the first two are the same as in the deterministic integral equation (1.3). We want to construct appropriately the third term to allow for randomness into the system. In terms of Figure 3.1, we want $B(t)$ to capture the discrepancy between the *smooth* and the *rough* path. At any time t, $B(t)$ has to be a random variable; it's variability will introduce variability into the path of X. Hence $B = \{B(t)\,;\, t \geq 0\}$ will be a stochastic process. In the light of the discussion for randomness above we will sometimes write $B(\omega, t)$ instead of $B(t)$ to emphasize the random nature of B. In this setting, σ is a positive scaling constant used to tune the effect of B on X: choosing $\sigma = 0$ reduces (3.2) to the ordinary integral equation (1.3).

Let us return now to (3.2) and add ω to all functions which will depend on it:

$$X(\omega, t) = x_0 + \int_0^t f(X(\omega, s))ds + \sigma B(\omega, t)\,. \tag{3.3}$$

Notice that X will be constructed via B, the former will also be a function of two variables, $(\omega, s) \mapsto X(\omega, s)$. That is, any solution of (3.3) will also be a stochastic process. The question now becomes: what properties should B have? A trivial first requirement is that $B(\omega, 0) = 0$ for (almost) all ω so that x_0 is the initial value of X. Working with the increments of X we have:

$$\begin{aligned} X(\omega, t+h) - X(\omega, t) &= \int_t^{t+h} f(X(\omega, s))ds + \sigma(B(\omega, t+h) - B(\omega, t)) \\ &\approx hf(X(\omega, t)) + \sigma(B(\omega, t+h) - B(\omega, t))\,. \end{aligned} \tag{3.4}$$

The motion at time t should be *centred* around $f(X(t))$, so we require that the expected value of the jump $B(t+h) - B(t)$ is zero:

Property 1. $\mathrm{E}\,[\,B(t+h) - B(t)\,] = 0$ *for all* t, h.

We would also like that the disturbances in different time points be independent and identically distributed, that is:

Property 2. $B(t_2) - B(t_1)$ *is independent of* $B(s_2) - B(s_1)$ *for all times* $s_1 < s_2 < t_1 < t_2$.

Property 3. *The distribution of* $B(t+h) - B(t)$ *depends only on* $|h|$ *(and not on* t*).*

Properties 2 and 3 above translate mathematically our requirement that the microscopic dynamics of X are perturbed by "pure error" which has no predictable structure. Ideally, we would also require the property that 'B has (almost everywhere) differentiable sample paths'. That is, for any fixed ω, $s \mapsto B(\omega, s)$ is differentiable in s. Were there such a process possessing Properties 1-3 and having differentiable paths, then one could easily make sense of the model (3.3) as the limit of (3.4) when divided by h and $h \to 0$. Unfortunately, it appears that there does not exist a stochastic process having Properties 1-3 and, additionally, differentiable paths.

However, not all is lost. There does exist a stochastic process B having Properties 1-3 and *continuous sample paths*. By continuous sample paths we mean that the process can be constructed so that $s \mapsto B(\omega, s)$ is continuous in s for all ω. Indeed, there exists *only one* such stochastic process, which is the *Brownian motion*. We will review some basic properties of the Brownian motion in the section that follows.

Notice that we have now made sense of the integral equation (3.2): the noise $B(t)$ appearing in the equation has been corresponded to a Brownian motion. It remains to investigate conditions under which such an equation has a solution. We have already discovered that a candidate solution cannot have differentiable paths. It appears however that, under conditions on the drift function f, there is a stochastic process X with continuous sample paths that solves (3.2). The topic of the existence and uniqueness of a solution to an SDE will be briefly addressed in Section 3.3. In retrospect, you can appreciate the flexibility in working with an integral equation instead of a differential equation: (3.3) still makes sense even when X is not differentiable but only continuous. Still, many times we equivalently express (3.3) in terms of an equation involving differentials:

$$dX(\omega, s) = f(X(\omega, s))ds + \sigma dB(\omega, s), \quad X(\omega, 0) = x_0 \ . \tag{3.5}$$

Since both X and B are not differentiable, (3.5) makes sense only when integrated w.r.t. s in which case it gives back (3.3). To derive (3.3) from (3.5) we have used the intuitive rule:

$$\int_0^t dB(\omega, s) = B(\omega, t) - B(\omega, 0) = B(\omega, t) \ .$$

3.2 Brownian Motion: An Excerpt

The building block of our model is the Brownian motion, so it worth reviewing some of it's fundamental properties. This is not a course on Brownian motion: we will only present what is completely necessary in our context. For more details we refer the reader to the several textbooks available devoted to the study of the numerous and fascinating properties of the Brownian motion.

Properties of B can be investigated in two directions. The first has to do with the finite-dimensional distributions of B. That is, the distribution of vectors of the form $\{B(s_0), B(s_1), \ldots, B(s_n)\}$ for given collections of time instances $0 < s_0 < s_1 < \cdots < s_n$. Some notation: $N(\mu, \Sigma)$ denotes a Gaussian distribution with mean μ and variance Σ. We can now state the following results:

1. $\mathrm{P}\,[\,B(0) = 0\,] = 1$.

2. The random vector $(B(s_0), B(s_1), \ldots, B(s_n))$ follows a multivariate Gaussian distribution with mean μ and variance Σ determined as follows:

$$\mu = \Big(\mathrm{E}\,[\,B(s_i)\,]\Big)_i = 0, \quad \Sigma = \Big(\mathrm{Cov}(B(s_i), B(s_j))\Big)_{i,j} = \big(\min\{s_i, s_j\}\big)\ .$$

3. From result 2. we obtain that, for $t, s > 0$:

$$B(t) - B(s) \sim N(0, |t - s|)\ .$$

   Thus, Brownian motion has stationary increments.

4. Result 2. also implies that for any $s_0 < s_1 < s_2$:

$$B(s_0),\ B(s_1) - B(s_0),\ B(s_2) - B(s_1) \text{ are independent}\ .$$

   Therefore, Brownian motion has independent increments, and the step $B(s_2) - B(s_1)$ is independent of all the past $\{B(s)\,;\, s \leq s_1\}$, for any $0 < s_1 < s_2$.

The second direction of investigation is to consider properties of the paths $s \mapsto B(\omega, s)$ for fixed ω. In that respect a Brownian motion can be constructed in a way that all it's sample paths are continuous, i.e. the map $s \mapsto B(\omega, s)$ is continuous for any ω. At the same time, all sample paths of B are nowhere differentiable.

Simulation of Brownian Paths:
Can we *simulate* Brownian paths on our computer? Note first that a complete path $\{B(t)\,;\, t \in [0,T]\}$ entails infinitely many points, so one cannot expect to generate such a continuous-time object. However, what is definitely feasible is to simulate a *skeleton* of B, that is a finite subset of the complete continuous path. So, one can instead simulate the vector $\{B(s_0), B(s_1), \ldots, B(s_n)\}$ for some pre-specified time instances $s_0 < s_0 < \cdots < s_n$, based on the finite-dimensional properties of a Brownian motion described in results 1.-4. above. Figure 3.4 shows a Brownian path simulated on a computer. We simulated the locations $B(i/10^2)$, for $i = 1, 2, \ldots 10^4$ and connected the consecutive points with straight lines to resemble a continuous path.

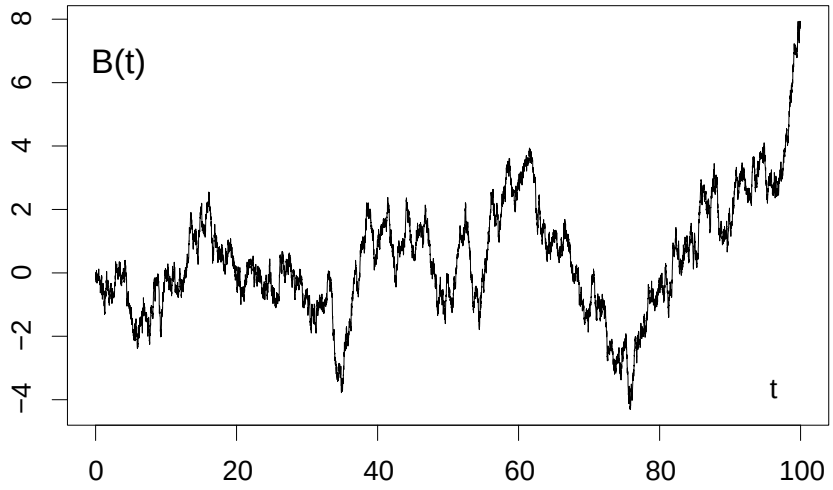

Figure 3.4: A Brownian path $t \mapsto B(\omega, t)$ on a computer.

The simulation is performed by exploiting the independency of the increments of the Brownian motion. For the case of the simulation of the vector $\{B(s_0), B(s_1), \ldots, B(s_n)\}$ we proceed as follows. We know the marginal distribution of the first point, $B(s_0) \sim N(0, s_1)$. Then, for the fist increment, $B(s_1) - B(s_0) \sim N(0, s_1 - s_0)$ independently of $B(s)$ for any $s \leq s_0$. Simulating this two Gaussian variables will give us the Brownian path at s_1 since $B(s_1) = B(s_0) + (B(s_1) - B(s_0))$. Thinking iteratively, there is a general algorithm (which could for example be implemented in R):

1. Simulate $Z_0 \sim N(0,1)$. Set $B(\omega, s_0) = \sqrt{s_0} Z_0$.

2. For $i = 1 : n$, simulate $Z_i \sim N(0,1)$.
   Set $B(\omega, s_i) = B(\omega, s_{i-1}) + \sqrt{s_i - s_{i-1}}\, Z_i$.

So, a Brownian path (meaning, a skeleton of it) is simulated using only standard Gaussian variates. Changing the Gaussian variates Z_i essentially corresponds to simulating skeletons of paths which correspond to different ω's. Therefore, if we want to simulate m different paths we run the algorithm m times, changing the Z_i's we use at each run. If the Gaussian variates are independent across the runs, we will obtain independent Brownian paths. Figure 3.5 shows three independent Brownian paths on $[0,1]$. The corresponding skeletons have been simulated at very fine time-increments, $s_i - s_{i-1} = 1/10^4$ to provide a satisfactory approximation to the continuous underlying paths. As in the case of Figure 3.4 we have joined the locations $B(\omega, s_i)$ over the consecutive instances s_i by interpolating straight lines. In Figure 3.6 we plot

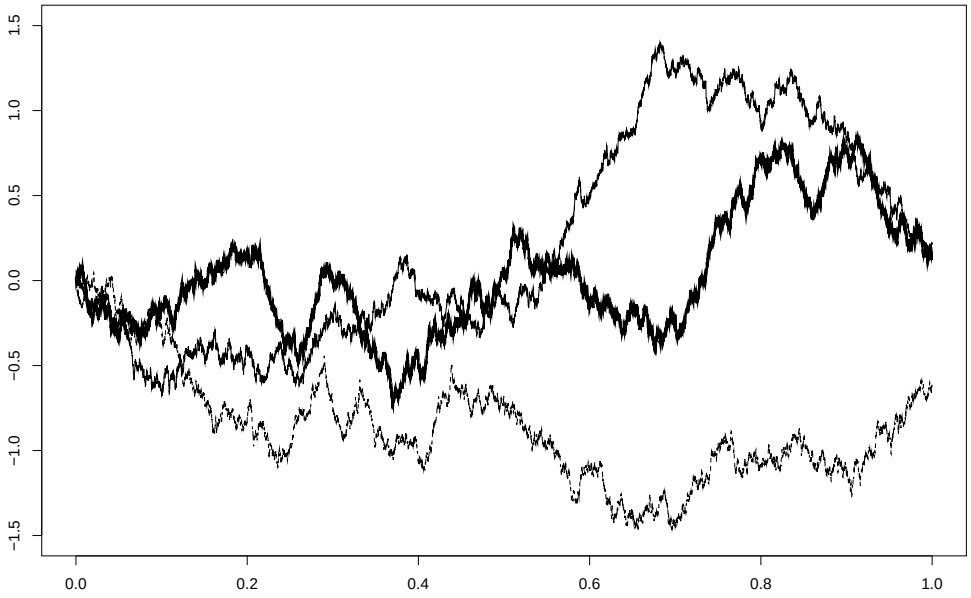

Figure 3.5: Three independent Brownian paths on $[0,1]$.

two *dependent* Brownian paths. Notice that the one path is the reflected version of the other along the $y = 0$ horizontal line. We simulated these paths by choosing two sets of Gaussian seeds, $(Z_0, \ldots, Z_n)$ for the first skeleton and $(Z'_0, \ldots, Z'_n)$ for the second skeleton with $Z'_i = -Z_i$.

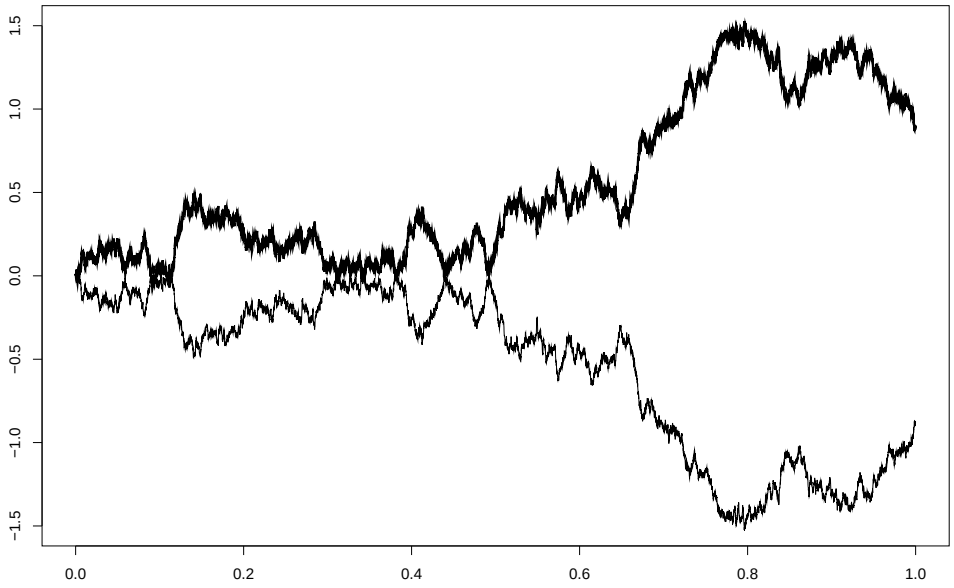

Figure 3.6: Two dependent Brownian paths on $[0, 1]$.

3.3 Solutions of SDEs

As a solution to the two equivalent expressions (3.3), (3.5), one should understand a stochastic process X, whose value at time t, $X(t)$, is expressed in terms of x_0, t and the Brownian path B, and whose time-evolution is described by (3.3). So now the solution will also depend on the Brownian motion, whereas in the ODE setting it only depended on the initial condition and the time argument. From the expression (3.3), one should expect that $X(t)$ will only depend on the past of B, $\{B(s); s \leq t\}$, i.e. the path of B up to any time t will fully specify the path of X up to t. In probabilistic terms, this is formally stated as 'X being adapted to the filtration generated by B'.

Example 3.1 (Constant drift)**.** Consider the SDE:

$$dX(s) = \mu ds + \sigma dB(s), \quad X(0) = x_0 \ , \tag{3.6}$$

i.e. the drift is constant, $f(x) = \mu$. The solution of this SDE is the stochastic process:

$$X(t) = x_0 + \mu t + \sigma B(t) \, . \tag{3.7}$$

This is trivially verified by writing the equation in it's formal integral form.

Notice that in this case we get an exact expression for the increments of X:

$$X(t+h) - X(t) = h\mu + \sigma(B(t+h) - B(t)) \ .$$

In Figure 3.7 we plot on the same graph a path of the constant drift SDE with $\mu = 3$, $\sigma = 4$ (path X_1) and another path for $\mu = 3$, $\sigma = 1$ (path X_2); on the same graph we have plotted the Brownian motion B that drives the two paths.

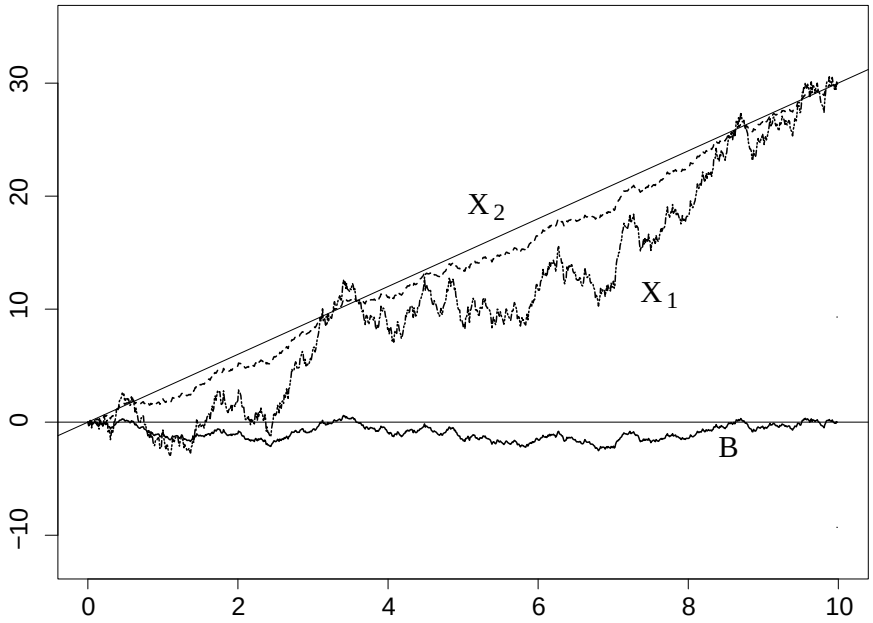

Figure 3.7: Paths of solutions of (3.6). The first, X_1, corresponds to the parameter choice $\mu = 1$, $\sigma = 4$; the second, X_2, to $\mu = 1$, $\sigma = 1$. Both paths are driven by the Brownian motion B. We have also included the mean value of the solutions, $\mathrm{E}[X(t)] = \mu t$ (straight line).

Notice that the fluctuations of the Brownian path imply the corresponding fluctuations for X. More generally, the distribution of B implies a distribution for X. In the context of this example SDE the relationship between the two processes is apparent: X is a linear transformation of B. In particular, we readily get that the finite-dimensional distributions of X are also Gaussian. In more general SDEs the distribution of X can be more involved.

□

Notice that, whereas the SDE describes the microscopic dynamics of X, the solution of the SDE describes the characteristics of the process at longer time horizons. Given the solution we can make probabilistic assessments of the state of the process at any time t. For instance, in the case of the constant drift SDE (3.6) we know that $X(t) \sim N(\mu t, \sigma^2 t)$. We can use this information to make predictions.

The SDE (3.6) is one of the few SDEs admitting an explicit solution. In Section 3.7 we will see that we need an appropriate calculus for solving such SDEs, in the same way that we use ordinary calculus (integration by parts, chain rule) to solve ODEs. On the other hand, we can numerically approximate the solution after appropriately adjusting the Euler scheme of Section 1.4 in the stochastic context at hand. Before that one needs to investigate existence and uniqueness of solution. Such considerations are approached in a similar manner as for ODEs. The Lipschitz condition again naturally arises as a condition which ensures uniqueness, and Picard's iteration is used to establish existence. A major difference is that since we now work with random variables we will need to use expectations and a different limit theory. We refer to Oksendal's book for details (see p.70).

3.4 Numerical Approximation of SDEs

We henceforth take for granted that there is a unique solution X of (3.3). We aim to develop a reasonable discrete-time approximation of X which can be readily used for simulation purposes on a computer. In Example 3.1, sample paths of X were easy to obtain since the solution of the SDE was simply a linear transformation of a Brownian motion. We will now be interested in cases where such solutions are not analytically available.

As with the Euler scheme for the approximation of ODEs, the basis of our method will be the approximation of the infinitesimal dynamics of the SDE. In particular, for a small time increment h we have that:

$$\begin{aligned} X(t+h) - X(t) &= \int_t^{t+h} f(X(s))ds + \sigma(B(t+h) - B(t)) \\ &\approx hf(X(t)) + \sigma(B(t+h) - B(t)) \,. \end{aligned}$$

So, assuming that we are given the location $X(t)$, we have the approximation:

$$X(t+h) - X(t) \overset{approx.}{\sim} N(hf(X(t)), \sigma^2 h) \,. \tag{3.8}$$

One can now proceed similarly as in the case of an ODE. Assume that the objective is the approximation of $X(T)$ for some time instance T. For some large enough integer n, we divide the interval $[0, T]$ into n pieces of equal length $h = T/n$. We then apply iteratively the approximation (3.8) on each of the subintervals $[(i-1)T/n, iT/n]$, for $i = 1, 2, \ldots n$. Analytically, we set:

$$\tilde{X}(ih) = \tilde{X}((i-1)h) + f(\tilde{X}((i-1)h))\, h + \sigma\sqrt{h}\, Z_i \ , \tag{3.9}$$

with starting point $\tilde{X}(0) = x$; also, $\{Z_i\}$ is a collection of independent standard Gaussian variables. One can now use the discrete-time process $\tilde{X}$ to approximate sample paths of the continuous-time process X on a computer. The above numerical method is called the *stochastic Euler scheme*.

Example 3.2 (Ornstein-Uhlenbeck process)**.** Consider the SDE:

$$dX(s) = \alpha(\mu - X(s))ds + \sigma dB(s), \quad X(0) = x_0 \ , \tag{3.10}$$

for constants $\alpha, \mu \in R$, $\sigma > 0$. As in the case of an ODE, one can infer about the behavior of the solution of the SDE by examining the drift function. In fact, the Ornstein-Uhlenbeck process belongs to the family of *mean-reverting* models: when $\alpha > 0$, sample paths of X will revert towards the mean μ, with α modelling the speed of reversion. When $\alpha < 0$, sample paths will eventually escape to (plus or minus) infinity. In terms of the discussion about stationary points in Section 1.3, μ is a stable stationary point when $\alpha > 0$ and unstable otherwise. In Figure 3.1 we plotted sample paths of the Ornstein-Uhlenbeck process started at $x_0 = 2$, for parameter values $(\alpha, \mu, \sigma) = (0.8, 0, 0.04)$. Also, Figure 3.8 presents sample paths for various parameter values (always with $\alpha > 0$) and starting values on the time horizon $[0, 20]$. The Ornstein-Uhlenbeck SDE can be analytically solved; the mean of the solution is:

$$\mathrm{E}\,[\,X(t)\,] = e^{at}(x_0 - \mu) + \mu \tag{3.11}$$

which in it's turn solves the linear ODE corresponding to $\sigma = 0$. The smooth lines in Figure 3.8 show the solutions of this ODE for the various selections of μ, α and x_0.

Notice the contrast between ODE and SDE illustrated in Figure 3.8. When $x_0 \neq \mu$ the drift pushes X to move towards μ. However, due to the stochasticity introduced by B, the path of X does not move monotonically towards μ, as in the deterministic case, but rather fluctuates around it. In

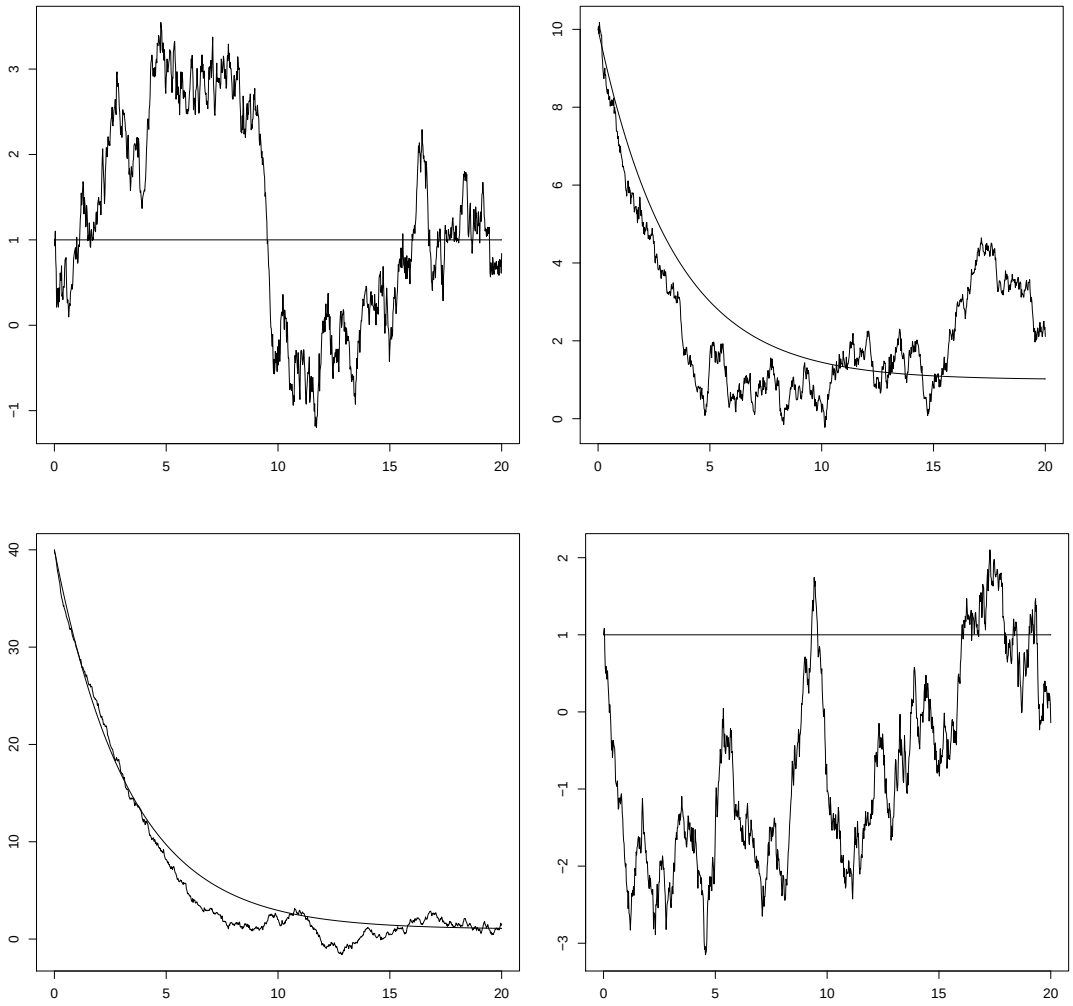

Figure 3.8: Sample paths of the Ornstein-Uhlenbeck process on $[0, 20]$. The parameter values (α, μ, σ) are $(0.3, 1, 1)$ for the first three graphs and $(0.03, 1, 1)$ for the last one. With solid lines we have superimposed the corresponding ODE solutions for $\sigma \equiv 0$.

general, when X is far from μ the drift is strong and dictates the dynamics of the process; then the effect of stochasticity is small. On the contrary, when X is close to μ, the random perturbations dominate the dynamics. □

Example 3.3 (Double-well)**.** Consider the SDE whose drift is given by the

following function:

$$f(x) = x\left(\frac{8}{(1+x^2)^2} - 2\right). \tag{3.12}$$

This drift defines two stable stationary points, +1 and -1, and one unstable stationary point, 0. Figure 3.9 shows the graph of f. You can check that f satisfies the Lipschitz condition on R. In the deterministic case ($\sigma \equiv 0$) the

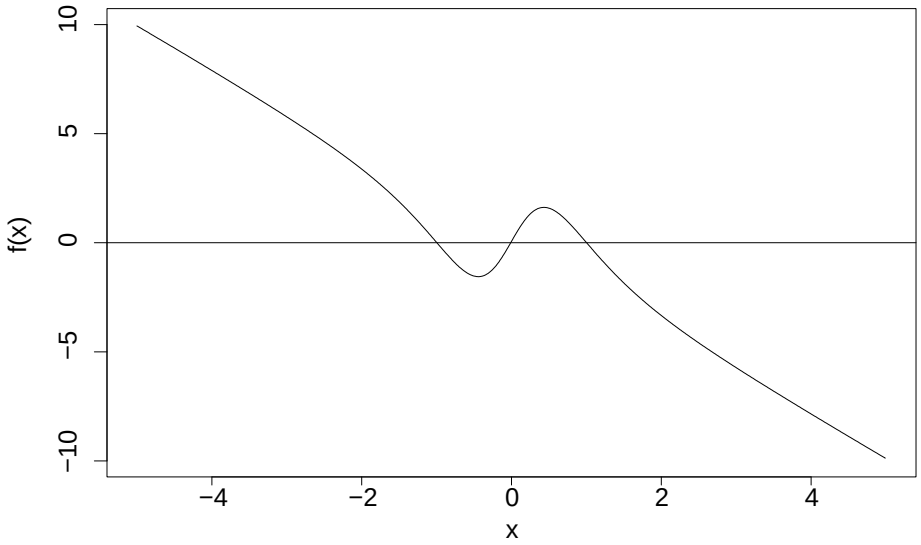

Figure 3.9: The drift (3.12) of the double-well process.

path will stabilize around the stable point closer to x_0, i.e. to $+1$ if $x_0 > 0$ or to -1 if $x_0 < 0$. When stochasticity is introduced ($\sigma \neq 0$) the behavior of the paths becomes more complex. There will be periods where X moves around one of the stable points, but now there will also be transitions from the one stationary point to the other. The frequency of these transitions will be determined by the value of σ. Figure 3.10 (top panel) shows simulated sample paths of the double-well process on the time horizon $[0, 200]$ for three different values of σ, all paths having started at the stable point $+1$. Notice that when $\sigma = 1$, X moves around both stable points making a few transitions between them. These transitions become less frequent if $\sigma = 0.8$ whence the path X is more concentrated around the stable solutions. When $\sigma = 0.5$ we observe no transitions on $[0, 200]$ and X moves only around $+1$. The bottom panel of Figure 3.10 plots histograms of the locations of the sample paths (more precisely, of the locations of the discrete-time Euler approximation used to draw the graphs). □

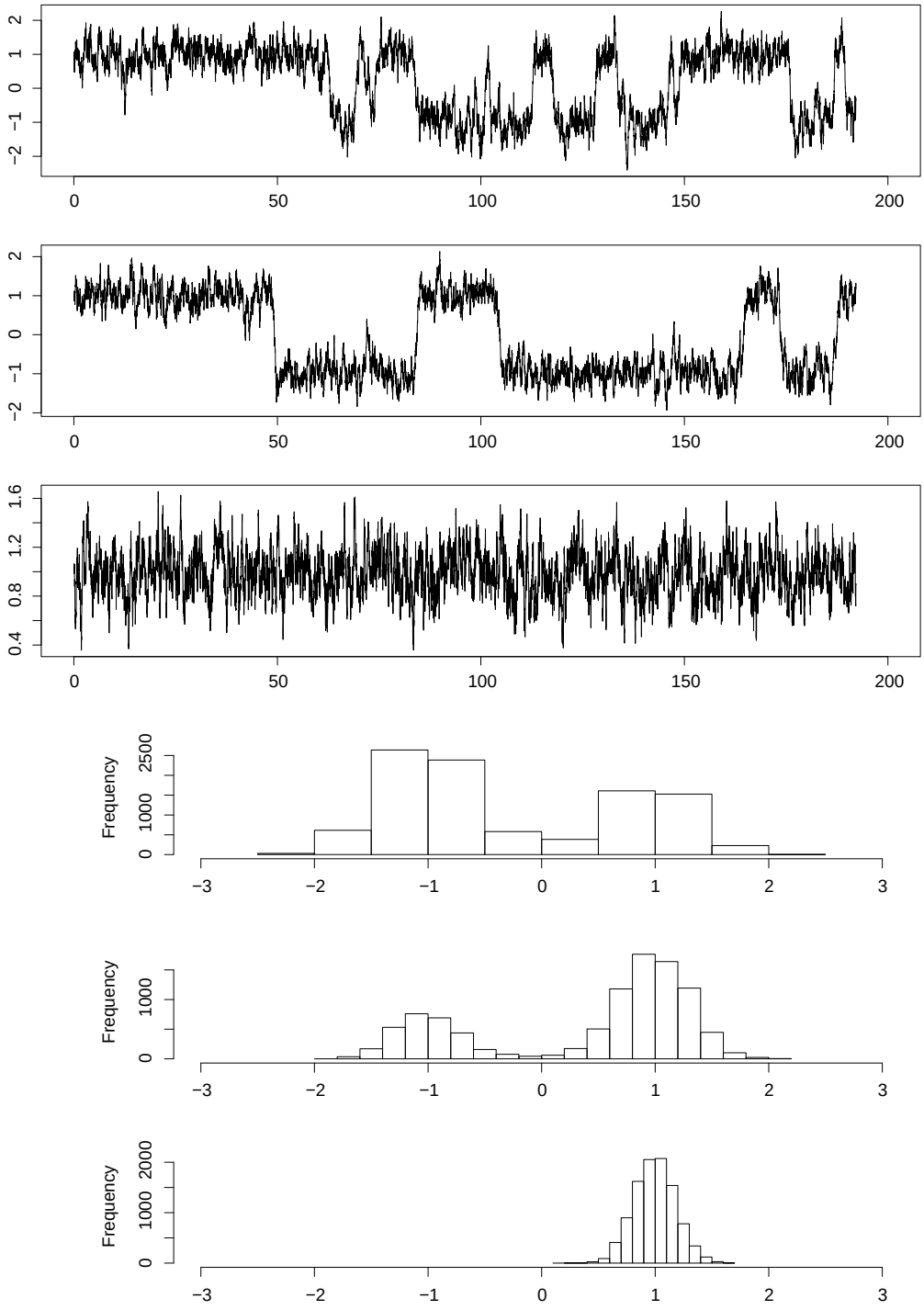

Figure 3.10: Top panel: Sample paths of the double-well process on the time interval $[0, 200]$ with $\sigma = 1$ (top), $\sigma = 0.8$ (middle) and $\sigma = 0.5$ (bottom). Bottom panel: Frequencies for locations visited by each of the three paths above.

Example 3.4 (Non-Lipschitz case)**.** It is interesting to consider a case where the drift does not satisfy a Lipschitz condition. This is to illustrate that the Lipschitz condition is sufficient but by no means necessary condition for the existence and uniqueness of a solution of an SDE. We will revisit the drift function we first discussed in Section 1.1:

$$f(x) = -x^3 \ .$$

Observe that 0 is a stable stationary point. In Section 1.1 we saw that the solution of the ODE with this drift is $X(s) = x_0/\sqrt{1 + 2sx_0^2}$, $s > -1/(2x_0^2)$. In Figure 3.11 we plot the solution for $x_0 = 1$. We have already seen that the solution explodes as $s \downarrow -1/(2x_0^2)$. An interesting interpretation (or consequence) of this exploding behavior is that, even if the initial location is large, X will move close to the origin extremely fast. For instance, if t_1 is the time that X reaches 1, then one can easily find that $t_1 = (1 - 1/x_0^2)/2$ which is not larger than $1/2$ irrespectively of how large x_0 might be.

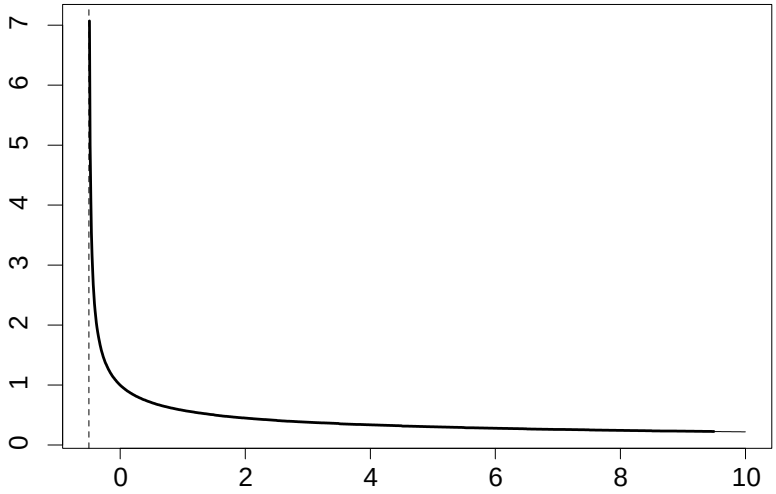

Figure 3.11: The solution of the ODE with and drift $f(x) = -x^3$, started at $x_0 = 1$.

We now examine the behavior of the corresponding SDE with the same drift. Figure 3.12 shows sample paths of the process started from three different locations, $x_0 = 1$, $x_0 = 5$ and $x_0 = 20$. As a first remark, the process moves around 0, which is not surprising given that 0 is a stable stationary point. Notice also the extremely rapid returns of X to values near 0 from any initial value. For example, contrast Figure 3.12 with Figure 3.8.

□

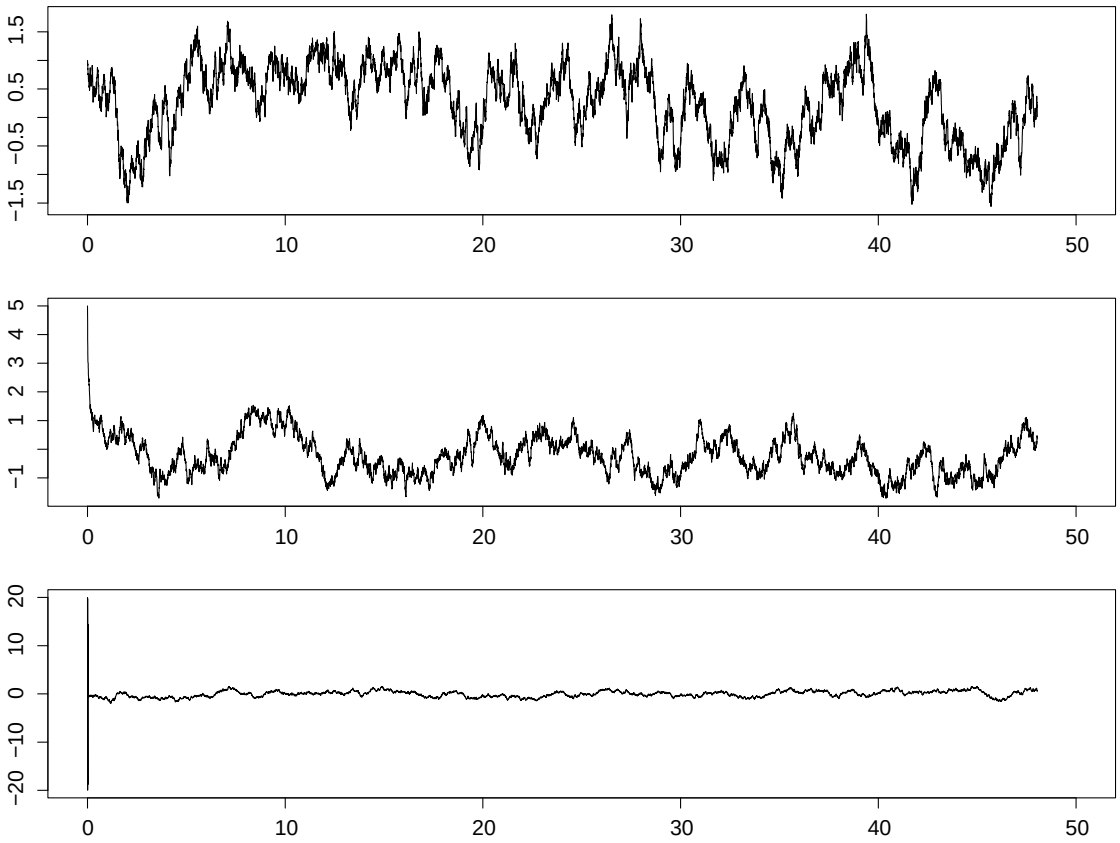

Figure 3.12: Sample paths of the SDE with drift $f(x) = -x^3$, $\sigma = 1$, and three different starting values, $x_0 = 1$ (top), $x_0 = 5$ (middle) and $x_0 = 20$ (bottom).

3.5 Towards a More Robust Model

The model we have so far constructed:

$$dX(s) = f(X(s))ds + \sigma dB(s), \quad X(0) = x_0\,, \tag{3.13}$$

though capable of describing many natural processes, needs to be appropriately generalized. Such a generalization is motivated both by mathematical and practical statistical considerations. We will now illustrate both these arguments. Recall that the model we have so far considered implies (see

section 3.4) that for small h:

$$\begin{aligned} \mathrm{E}\,[\,X(t+h) - X(t) \mid X(t)\,] &\approx f(X(t))\,h\ , \\ \mathrm{Var}\,[\,X(t+h) - X(t) \mid X(t)\,] &\approx \sigma^2 h\ . \end{aligned} \tag{3.14}$$

We begin with the mathematical considerations. Assume that X is the solution of the above SDE. Let Y be a stochastic process defined as a transformation of X, that is $Y(t) = g(X(t))$ for some sufficiently regular function $g : R \mapsto R$. Then, it is not true in general that Y also has a similar representation as X, that is Y cannot be expressed as the solution of an SDE of the type we have so far considered. Consider the following example. Let $X(t) = B(t)$, i.e. X is the solution corresponding to $f \equiv 0$, $\sigma = 1$ and $x_0 = 0$. Take $g(x) = e^x$, so that $Y(t) = e^{B(t)}$. Then, for small h:

$$\begin{aligned} Y(t+h) - Y(t) &= \exp\{B(t+h)\} - \exp\{B(t)\} \\ &= \exp\{\sqrt{h}\,Z + B(t)\} - \exp\{B(t)\} \quad [\text{for a } Z \sim N(0,1)] \\ &= \exp\{B(t)\}(\exp\{\sqrt{h}\,Z\} - 1) \\ &= Y(t)(\exp\{\sqrt{h}\,Z\} - 1) \\ &\approx Y(t)(\sqrt{h}\,Z + h\,Z^2/2) \quad [\text{by expanding the exponential}] \end{aligned}$$

So, in this case:

$$\begin{aligned} \mathrm{E}\,[\,Y(t+h) - Y(t) \mid Y(t)\,] &\approx \frac{Y(t)}{2}\,h \\ \mathrm{Var}\,[\,Y(t+h) - Y(t) \mid Y(t)\,] &\approx Y(t)^2 h\ . \end{aligned} \tag{3.15}$$

The conditional expectation is linear in $Y(t)$, as in the case of the Ornstein-Uhlenbeck model of Example 3.2. There is however a striking difference between the infinitesimal dynamics of Y in (3.15) and those in (3.14): the conditional variance of $Y(t+h) - Y(t)$ depends on $Y(t)$. Dependence of the variance of the increments on the value of the process is a feature that is not allowed in our model (3.13). The expressions in (3.15) in fact point towards a continuous-time model of the type,

$$dY(s) = \frac{Y(s)}{2} ds + Y(s) dB(s)\ . \tag{3.16}$$

Note that our current theory is insufficient to attach a well-defined mathematical meaning to this expression, since if we integrate both sides w.r.t.

time, we get:

$$Y(t) = y_0 + \int_0^t \frac{Y(s)}{2} ds + \int_0^t Y(s) dB(s) \ . \tag{3.17}$$

Therefore, to make sense of (3.16) we need first to make sense of integration w.r.t. Brownian motion. This will be the subject of the next section.

Models like (3.16) are not required only due to mathematical completeness, but arise rather naturally for dealing with practical modelling issues. Recall the Eurodollar data set of Figure 3.2. A careful examination of these data suggests that the variability of the process depends on it's current location. When the Eurodollar rate is high, it varies wildly, whereas when it is low it's variability diminishes. One should expect that a model which allows state-dependent variance would be more appropriate than (3.13) to describe such data.

Therefore, for both mathematical and statistical reasons, an extension of the model we have currently at hand is required. We wish to preserve the same structure, where the drift dictates the motion of the process, but allow for a term which controls the local variability of the motion. Hence, we advocate a model which writes as follows:

$$X(t) = x_0 + \int_0^t f(X(s)) ds + \int_0^t \sigma(X(s)) dB(s) \ , \tag{3.18}$$

where now $\sigma : R \mapsto R_+$ is a non-constant function. So, we need to understand integration w.r.t. Brownian motion. Note that up to now we managed to avoid such an issue by restricting ourselves to the special case where σ is a constant function, in which case (3.18) reduces to (3.13).

3.6 Integration and Bounded Variation

To attach a formal mathematical meaning to (3.18) we will need to appropriately define integrals of the type:

$$\int_0^t g(\omega, s) dB(\omega, s) \, , \tag{3.19}$$

where the integrator is the Brownian motion B and the integrand is a stochastic process g. Such an object is called a *stochastic integral*. A specific case of

interest is for $g(\omega, s) = \sigma(X(\omega, s))$ when one retrieves the stochastic integral in (3.18). The integration in (3.19) cannot be carried out using standard calculus, for various reasons:

1. The result of the integral will depend on ω, so it will be a random variable. Consequently, probability theory will intervene at some point to attach meaning to, and manipulate, such integrals.

2. In general, g and B will be stochastically dependent. Working with dependent random variables can be difficult. It turns out that, to define a stochastic integral, g must be confined within a specific family of stochastic processes.

3. A third reason why (3.19) lies outside ordinary calculus is more subtle and relates to the concept of *variation*.

We begin our discussion with the last point. Let us recall the construction of the well-known Riemann integral:

$$\text{Riemann integral:} \quad \int_a^b g(s)ds$$

for real-valued *continuous functions*, $g : R \mapsto R$. (It is not necessary to restrict to continuous functions, but it does simplify things). The integral can be obtained as the (unique) limit of approximating sums of the type:

$$\sum_{i=1}^{n} g(s_i)\Delta t_i \,. \tag{3.20}$$

We have set $\Delta t_i = t_i - t_{i-1}$ where $a = t_0 \leq t_1 \leq \cdots \leq t_{n-1} \leq t_n = b$ is a *partition* of $[a, b]$; s_i is *any* point in the interval $[t_{i-1}, t_i]$. See also Figure 3.13. More precisely, one should think of a sequence of partitions indexed by n:

$$a = t_0^{(n)} \leq t_1^{(n)} \leq \cdots \leq t_{n-1}^{(n)} \leq t_n^{(n)} = b$$

so that a limit of (3.20) for $n \to \infty$ can make sense. We adopt such a strategy for the sequel, but suppress further use of the superscript (n). We are assuming that $\sup_i\{\Delta t_i\} \to 0$ as $n \to \infty$, i.e. the partition gets finer as we increase n. We will be making this assumption for all sequences of partitions mentioned in the rest of the section without further notice.

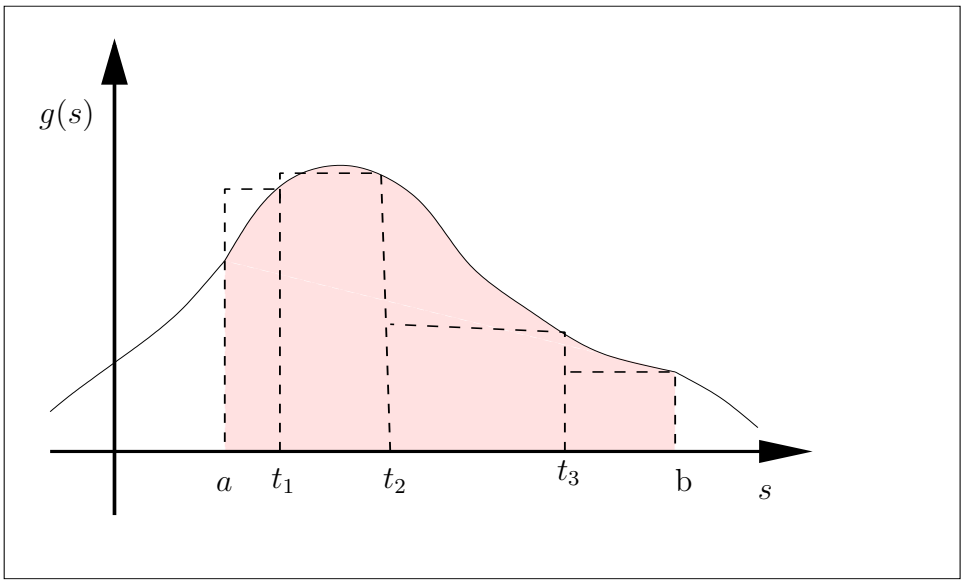

Figure 3.13: The surface of the shaded area corresponds to $\int_a^b g(s)ds$. An approximation is given by the area covered by the rectangular boxes, using a partition $0 = t_0 < t_1 < t_2 < t_3 < t_4 = b$. In this figure we have taken $s_i = t_i$.

Riemann integration can be generalised to Riemann-Stieltjes integration which can calculate the integral of g w.r.t. an arbitrary *increasing* function F:

$$\text{Riemann-Stieltjes integral:} \quad \int_a^b g(s)dF(s) \ . \tag{3.21}$$

Similarly to the Riemann integral, (3.21) is defined as the limit ($n \to \infty$) of approximating sums of the form:

$$\sum_{i=1}^{n} g(s_i)\Delta F_i \, , \tag{3.22}$$

over a partition $a = t_0 \leq \cdots \leq t_n = b$ of $[a, b]$, where we have set:

$$\Delta F_i = F(t_i) - F(t_{i-1}) \ .$$

As above, s_i is any point of the interval $[t_{i-1}, t_i]$. In this generalisation F needs not be continuous. Note that the Riemann integral is the special case of the Riemann-Stieltjes integral where F is the identity mapping.

Example 3.5. Let X be a random variable with distribution function F. Then it's expected value is given by the Riemann-Stieltjes integral expression $\int_{-\infty}^{\infty} x dF(x)$, provided the integral exists (i.e. provided $\lim_{k\to\infty} \int_{-k}^{k} x dF(x)$ exists). Notice that this compact notation caters for both continuous and discrete random variables. □

Riemann-Stieltjes integration obeys all the familiar rules of Riemann integration. Also, it can be further generalised. We can define the integral of g w.r.t. functions F of *bounded total variation*.

Definition 2. *Consider the function $F : [a, b] \mapsto R$. For any $p > 0$, the limit $V^{(p)}[F, a, b] = \lim_{n\to\infty} V_n^{(p)}[F, a, b]$, where:*

$$V_n^{(p)}[F, a, b] = \sum_{i=1}^{n} |\Delta F_i|^p$$

is defined over a partition of $[a, b]$, is called the p-th variation of F. $V^{(1)}[F, a, b]$ is called total variation and $V^{(2)}[F, a, b]$ quadratic variation. We say that a function is of bounded variation if it's total variation on any compact interval $[a, b]$ is finite.

Exercise 3.1. *Show that (trivially) if F is increasing on $[a, b]$ then it is of bounded total variation on that interval.*

Exercise 3.2. *Show that if $F : R \mapsto R$ satisfies the Lipschitz condition on $[a, b]$ then it is of bounded total variation on that interval.*

Functions of bounded variation appear naturally in the context of Riemann-Stieltjes integration because of their connection with increasing functions. It can be proved that any function F of bounded variation can be written as the difference of two monotonically increasing functions:

$$F(s) = F_1(s) - F_2(s), \;\; F_1, F_2 \text{ increasing.}$$

Then, the integral of g w.r.t. F is defined as:

$$\int_a^b g(s)dF(s) = \int_a^b g(s)dF_1(s) - \int_a^b g(s)dF_2(s) \;,$$

where the two terms on the right-hand side are standard Riemann-Stieltjes integrals. Even in this general framework, when g is continuous and F is of

bounded variation, we can make sense of $\int_a^b g(s)dF(s)$ as the (unique) limit of approximating sums $\sum_{i=1}^n g(s_i)\Delta F_i$, calculated using partitions of $[a,b]$, where s_i is *any* point inside each interval $[t_{i-1}, t_i]$.

Therefore, one may naturally think of defining the stochastic integral (3.19) as a Riemann-Stieltjes integral. Such a direction would require the Brownian paths to be of bounded variation. Unfortunately, this is not the case. From the properties of the Gaussian distribution one can find that:

$$\mathrm{E}\,[\,|B(t+h) - B(t)|\,] = \sqrt{\frac{2}{\pi}}\sqrt{h}\ .$$

So, the Brownian path in small time-increments h varies wildly: it's increments are of the order $\sqrt{h}$. Indeed, we have the following theorem.

Theorem 3.1.

1. *All sample paths of the Brownian motion on* $[0,t]$ *have the same quadratic variation* t*. That is, for all* $\omega \in \Omega$*:*

$$\lim_{n\to\infty} \sum_{i=1}^{n} |B(\omega, t_i) - B(\omega, t_{i-1})|^2 = t\ . \tag{3.23}$$

2. *The total variation of all Brownian paths on* $[0,t]$ *is infinite.* [1]

Proof.
We will use the shorthand notation:

$$\Delta B_i = B(\omega, t_i) - B(\omega, t_{i-1})\ .$$

1. We will prove a slightly different statement, namely that the following is true:

$$\lim_{n\to\infty} \mathrm{E}\left[\left(\sum_{i=1}^{n} |\Delta B_i|^2 - t\right)^2\right] \to 0\,.$$

Firstly, we expand the square inside the expectation to get three terms $(\sum(\Delta B_i)^2)^2 + t^2 - 2t\sum(\Delta B_i)^2$. We further expand the square in the first term to get $\sum_i (\Delta B_i)^4 + \sum_{i\neq j}(\Delta B_i)^2(\Delta B_j)^2$. Taking the expectation of the four

[1] For these results to hold, it is required that the sequence of partitions considered is increasing, that is the partition corresponding to $n = n_1$ is a subset of the one corresponding to $n = n_2$ when $n_1 < n_2$.

terms we get $3\sum_i(\Delta t_i)^2 + \sum_{i\neq j}\Delta t_i\Delta t_j - t^2$. Notice that $t^2 = (\sum_i \Delta t_i)^2$. Thus, we can simplify the previous expression as $2\sum_i(\Delta t_i)^2$. Notice that $\sum_i(\Delta t_i)^2 = \sum_i \Delta t_i\Delta t_i = \sum_i t_i\Delta t_i - \sum_i t_{i-1}\Delta t_i \to t^2/2 - t^2/2 = 0$, as $\Delta t_i \to 0$, i.e. as $n \to \infty$. Thus, we have proved the result.

2. We will show that continuous functions of bounded non-zero quadratic variation have necessarily infinite total variation. Assume that F is a continuous real-valued function of non-zero quadratic variation on $[a, b]$. Notice first that, under the assumption of increasing sequence of partitions (see footnote), the triangular inequality implies that $V_n^{(1)}[F, a, b]$ is increasing with n. Thus, $V_n^{(1)}[F, a, b]$ will either converge to some finite real or diverge to infinity. From the definition of the quadratic and total variation we get that:

$$V_n^{(2)}[F, a, b] \leq \sup_i |\Delta F_i| \times V_n^{(1)}[F, a, b]$$

Notice however, that every continuous function on a closed interval is also uniformly continuous on it. Using this, the fact that $\sup_i \Delta t_i \to 0$ implies that:

$$\sup_i |\Delta F_i| \to 0 \ .$$

So, if the total variation is finite, the quadratic variation will be 0. Equivalently, non-zero quadratic variation implies that the total variation is infinite. □

Now we know that our integration problem does not fit into the ordinary calculus. So, what might go wrong?

Example 3.6. Consider the stochastic integral:

$$\int_0^t B(\omega, s)dB(\omega, s)\,. \tag{3.24}$$

For a given partition of $[0, t]$, the following sums correspond to two reasonable, Riemann-Stieltjes type, approximations of (3.24):

$$I_n(\omega) = \sum_{i=1}^n B(\omega, t_{i-1})\Delta B_i\,, \quad J_n(\omega) = \sum_{i=1}^n B(\omega, t_i)\Delta B_i\,. \tag{3.25}$$

These sums have the same structure as (3.22) and correspond to two different choices of the intermediate times s_i: $I_n(\omega)$ chooses $s_i = t_{i-1}$ whereas $J_n(\omega)$

chooses $s_i = t_i$. One would expect that as $n \to \infty$, $I_n(\omega)$ and $J_n(\omega)$ would both converge (in some mode) to the *same* limit which would correspond to the value of the integral (3.24). However:

$$\begin{aligned} J_n(\omega) &= \sum_{i=1}^{n} B(\omega, t_i)\Delta B_i \\ &= \sum_{i=1}^{n} (\Delta B_i + B(\omega, t_{i-1}))\Delta B_i = \sum_{i=1}^{n} (\Delta B_i)^2 + I_n(\omega) \ . \end{aligned}$$

Thus, $I_n(\omega) - J_n(\omega) \to t$ as $n \to \infty$ using Theorem 3.1! Hence, Brownian motion varies in small time-increments so rapidly that the choice of the intermediate time in the integral approximating sum affects the value of the limit. This is remarkably different from the Riemann-Stieltjes integration where we know that such choice is irrelevant and a unique limit exists for any $s_i \in [t_{i-1}, t_i]$. One can prove that:

$$\lim_{n\to\infty} \mathrm{E}\left[\left(I_n(\omega) - \frac{B(\omega,t)^2}{2} - \frac{t}{2}\right)^2\right] \to 0 \, ,$$

which suggests the random variable $B(\omega, t)^2/2 - t/2$ as a canditate calculation of the stochastic integral (3.24). Of course, in ordinary calculus we know that $\int_0^t F(s)dF(s) = F(t)^2/2$. You can already see that standard calculus rules brake down in the context of stochastic integration. □

It should be obvious by now that a different calculus is required to define integrals of the form (3.19). One possibility is to define the integral as the limit of sums of the type $I_n(\omega)$ defined in (3.25). This leads to what is known as the *Itô's stochastic integral*. It's properties can then be studied using martingale theory. Similarly, one can define stochastic integrals w.r.t. more general stochastic processes X; for instance (3.19) can be generalised to:

$$\int_0^t g(\omega, s)dX(\omega, s) \ ,$$

where X is the solution of (3.18) by considering appropriate limits as $n \to \infty$, of sums of the type

$$\sum_{i=1}^{n} g(\omega, t_{i-1})\Delta X_i \, .$$

Unfortunately, it is beyond the scope of this course to delve into the beautiful theory of stochastic integration. We refer to the book of Oksendal for further reading.

3.7 Stochastic Calculus - Itô's Formula

Similarly to ordinary calculus, identification of stochastic integrals relates with the consideration of appropriate derivatives. Let $X = \{X(t); t \geq 0\}$ be the diffusion process determined as the solution of the general equation:

$$dX(s) = b(X(s))ds + \sigma(X(s))dB(s), \quad X(0) = x \ . \tag{3.26}$$

Let $g : [0, \infty) \times R \mapsto R$ be a function one time differentiable in the first argument, twice continuously differentiable in the second, and consider the stochastic process:

$$Y(s) = g(s, X(s)) \ .$$

The celebrated *Itô's formula* dictates that $Y(s)$ is also a diffusion process and provides the SDE solved by $Y(s)$. From a practical perspective, it gives the means for making sense of the infinitesimal differences $dY(s)$ and, then, for calculating stochastic integrals. Analytically, Itô's formula states the following:

Itô's Formula:

$$dY(s) = \partial_s g(s, X(s))ds + g^{'}(s, X(s))dX(s) + \frac{1}{2}g^{''}(s, X(s))(dX(s))^2 \ .$$

The expression is similar to the one from ordinary calculus, apart from the second-derivative term on the right-hand side usually called *Itô's correction*. One can substitute $dX(s)$ in the above equation with it's equal from (3.26). When doing so, the quadratic term $(dX(s))^2$ will give quantities of the type $(dB(s))^2$, $ds \cdot dB(s)$, $(ds)^2$. The following rule should then be applied:

$$(dB(s))^2 = ds; \quad (ds)^2 = 0; \quad dB(s) \cdot ds = 0 \ .$$

Example 3.7. We will use Itô's formula to find the integral:

$$\int_0^t B(s)dB(s) \ .$$

Consider the process $Y(s) = (B(s))^2$. Applying Itô's formula in this context, with $g(x) = x^2$, and $X(s) \equiv B(s)$, gives the following:

$$d(B(s))^2 = 2B(s)dB(s) + (dB(s))^2 \ ,$$

so, using the rule $(dB(s))^2 = ds$, and integrating, we find that:

$$\int_0^t B(s)dB(s) = \frac{1}{2}\int_0^t d(B(s))^2 - \frac{1}{2}\int_0^t ds \equiv \frac{B(t)^2}{2} - \frac{t}{2} \ . \quad \square$$

Exercise 3.3. *Use Itô's formula to prove that:*

$$\int_0^t f(s)dB(s) = f(t)B(t) - \int_0^t f^{'}(s)B(s)ds \ .$$

Exercise 3.4. *Use Itô's formula to solve the following differential equation:*

$$dX(s) = \beta X(s)ds + \sigma X(s)dB(s) \ .$$

(Hint: Divide both sides with X_s; then use Itô's formula to see that the left-hand side is equal to $d(\log X(s))$ plus a constant.)

Chapter 4

Likelihood Inference

4.1 Introduction

We will now consider carrying out statistical inference for diffusion models. We will initially confine ourselves to the following family of (scalar) SDEs:

$$dX(t) = f(X(t);\theta)dt + \sigma\, dB(t), \quad X(0) = x_0\ . \tag{4.1}$$

for unknown parameters σ and $\theta = (\theta_1, \ldots, \theta_m)$ to be estimated based on observed data; x_0 is assumed known. For instance, in the case of the Ornstein-Uhlenbeck diffusion in Example 3.2: $m = 2, \theta_1 = \alpha, \theta_2 = \mu$. Notice that θ relates with the drift and σ with the variance of the process. We will see that this distinction between drift and variance parameters is crucial for inference purposes.

We assume possession of data corresponding to an observed sample path $X = \{X(t); t \in [0, T]\}$ of the process described by the model (4.1). In practice of course we cannot have continuous-time data, only very high frequency ones. Thus, the actual data will be:

$$X^{(n)} = (X(t_0), X(t_1), \ldots, X(t_n))$$

for some given time instances $0 = t_0 < t_1 < \cdots < t_{n-1} < t_n = T$. We will sometimes write X_i instead of $X(t_i)$ to simplify the mathematical expressions. Following the notation we have already established, we set:

$$\Delta t_i = t_i - t_{i-1}\,; \quad \Delta X_i = X_i - X_{i-1}\ .$$

For simplicity, and without loss of generality, we assume that the data are equidistant, so $\Delta t_i = T/n$. We can thus omit the index i and write just Δt.

We mentioned in Section 3.3 that the joint distribution of $X^{(n)}$, thus the likelihood function, is unavailable when (4.1) does not have an explicit solution. However, we can use the Euler approximation to write down an approximate likelihood, say $L^{(n)}(\theta, \sigma)$, which will be an increasingly accurate approximation of the exact one as $\Delta t \to 0$ (or equivalently $n \to \infty$). Following the Euler scheme:

$$L^{(n)}(\theta, \sigma) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2 \Delta t}} \exp\{-\frac{1}{2\sigma^2 \Delta t} (X_i - X_{i-1} - f(X_{i-1}; \theta)\Delta t)^2\} \quad (4.2)$$

We can now consider the log-likelihood:

$$\begin{aligned} \ell^{(n)}(\theta, \sigma) &= c - n \log \sigma - \frac{1}{2\sigma^2 \Delta t} \sum_{i=1}^{n} (\Delta X_i - f(X_{i-1}; \theta)\Delta t)^2 \\ &= c - n \log \sigma - \frac{1}{2\sigma^2} \left\{ \frac{\sum_{i=1}^{n} (\Delta X_i)^2}{\Delta t} + \sum_{i=1}^{n} f(X_{i-1}; \theta)^2 \Delta t - 2 \sum_{i=1}^{n} f(X_{i-1}; \theta) \Delta X_i \right\} . \end{aligned}$$

We can now find the value of the parameters that maximise the likelihood function. The equations:

$$\frac{\partial \ell^{(n)}}{\partial \sigma} = 0; \quad \frac{\partial \ell^{(n)}}{\partial \theta_j} = 0, \ j = 1, 2, \ldots m \ ,$$

yield the following expressions:

$$\begin{aligned} &\sigma^2 = \frac{1}{T} \sum_{i=1}^{n} (\Delta X_i)^2 + \frac{1}{n} \sum_{i=1}^{n} f(X_{i-1}; \theta)^2 \Delta t - \frac{2}{n} \sum_{i=1}^{n} f(X_{i-1}; \theta) \Delta X_i \ , \\ &\sum_{i=1}^{n} f(X_{i-1}; \theta) \frac{\partial}{\partial \theta_j} f(X_{i-1}; \theta) \Delta t - \sum_{i=1}^{n} \frac{\partial}{\partial \theta_j} f(X_{i-1}; \theta) \Delta X_i = 0 \ , \end{aligned} \quad (4.3)$$

the first coming from the derivative w.r.t. σ and the rest (for $j = 1, \ldots, m$) from the derivative w.r.t. θ_j. The solution of equations (4.3) w.r.t. σ, θ is the maximum likelihood estimator (MLE) of the unknown parameters.

Notice that, as $n \to \infty$:

$$\begin{aligned}\sum_{i=1}^{n} f(X_{i-1};\theta)^2\Delta t_i &\rightarrow \int_0^T f(X(s);\theta)^2 ds \\ \sum_{i=1}^{n} f(X_{i-1};\theta)\Delta X_i &\rightarrow \int_0^T f(X(s);\theta) dX(s)\,.\end{aligned}$$

So, the terms $1/n\sum_{i=1}^{n} f(X(t_{i-1});\theta)^2\Delta t_i$ and $1/n\sum_{i=1}^{n} f(X(t_{i-1});\theta)\Delta X_i$ appearing in (4.3) converge to zero as $n \to \infty$. Therefore, for large n, we can omit these terms from the score function, and get the following, familiar in the literature, approximate MLE of σ:

$$\hat{\sigma} = \left(\frac{1}{T}\sum_{i=1}^{n}(\Delta X_i)^2\right)^{1/2}. \tag{4.4}$$

Notice that here that the term "approximate MLE" refers to two different approximations: the approximation of the true log-likelihood of $X^{(n)}$ by the log-likelihood corresponding to an Euler approximation of the dynamics, and the approximation of the score function by the simpler estimating equation where the $O(1/n)$ terms have been omitted.

Example 4.1 (Constant drift)**.** We consider the constant drift diffusion first introduced in Example 3.1:

$$dX(s) = \theta ds + \sigma dB(s), \quad s \in [0,T]\,. \tag{4.5}$$

In this case, the Euler approximation of the likelihood of a discretely observed data set coincides with the real likelihood function.

Exercise 4.1. *Show that the MLE of θ and σ obtained by solving equations (4.3) are*

$$\begin{aligned}\hat{\sigma}^2 &= \frac{\sum_{i=1}^{n}(\Delta X_i)^2}{T} - \frac{1}{n}\frac{(X(T)-x_0)^2}{T} \\ \hat{\theta} &= \frac{X(T)-x_0}{T}\,.\end{aligned} \tag{4.6}$$

Show that using (4.4) instead, yields (clearly) the same estimate for θ, but $\hat{\sigma}^2 = \sum_{i=1}^{n}(\Delta X_i)^2/T$.

A first observation is that $(x_0, X(T))$ are *sufficient* statistics for θ. It is interesting that from all the path X only two points are used for the estimation of θ, and in particular the estimate is the same regardless how large n is. Inference for σ has very different characteristics.

Exercise 4.2. *Show that if X is generated from the process (4.5) then*

$$\frac{\sum_{i=1}^{n}(\Delta X_i)^2}{T} \to \sigma^2 \text{ , almost surely.}$$

(Hint: use Theorem 3.1.)

So, $\frac{\sum_{i=1}^{n}(\Delta X_i)^2}{T}$ is a *consistent* estimator of σ^2 as $n \to \infty$. The information for σ increases as n increases, and at the limit we get infinite information for σ. On the contrary, the information about θ is finite, as the following exercise asks you to show.

Exercise 4.3. *Find the Fisher information for θ. Show that the information about θ does not increase with n, but with T.*

□

The situation described in this example illustrates a common characteristic of statistical inference procedures for SDEs. Namely, the variance parameters can be perfectly estimated from a path $X = \{X(s); s \in [0, T]\}$ (or estimated with very small sampling variance from high frequency data). On the other hand the same path X contains only finite information about the drift parameters. The information about drift parameters increases with T, the time horizon, not with n, the fineness of the partition.

Exercise 4.4. *Consider the Ornstein-Uhlenbeck diffusion:*

$$dX(s) = a(\mu - X(s))ds + \sigma dB(s) \ .$$

Solve the estimating equations (4.3) using (4.4), to provide estimates for the parameters a, μ, σ.

4.2 The General Case

Consider now the case of non-constant diffusion coefficient:

$$dX(t) = f(X(t); \theta)dt + \sigma(X(t); \psi)dB(t), \quad X(0) = x_0 \ ,$$

where the unknown parameters $\theta \in R^m$, $\psi \in R^l$, for positive integers m, l, are to be specified given the observations

$$X^{(n)} = (X(t_0), X(t_1), \ldots, X(t_n))$$

over a sample path of X.

Working as in the case of constant σ, one can obtain an approximate likelihood function similar to the one in (4.2) with the only difference that σ should now be replaced by $\sigma(X_{i-1}; \psi)$. The estimating equations (4.3), appropriately adjusted, would then provide the means for finding the MLEs of the unknown parameters. In this general context, however, such equations will be scarcely analytically solvable. Instead, the expressions look simpler if one assumes directly that the data are of high frequency, so that the following result can be used directly.

$$\sum_{i=1}^{n} (\Delta X_i)^2 \to \int_0^T \sigma^2(X(t); \psi) dt \ .$$

This limit constitutes an apparent generalization of (4.2) obtained there for the case of constant σ. From a practical point of view, an estimator of ψ can be determined via the solution (with ψ being the unknown) of the equation:

$$\sum_{i=1}^{n} (\Delta X_i)^2 = \sum_{i=1}^{n} \sigma^2(X_{i-1}; \psi) \Delta t \ . \tag{4.7}$$

Given an estimate of ψ, the estimating equations for θ are then to the ones in (4.3), though considerably more difficult to solve. Analytically, working as in the previous section, an estimator of θ will be determined by solving:

$$\sum_{i=1}^{n} \frac{f(X_{i-1}; \theta) \frac{\partial}{\partial \theta_j} f(X_{i-1}; \theta)}{\sigma^2(X_{i-1}; \psi)} \Delta t_i - \sum_{i=1}^{n} \frac{\frac{\partial}{\partial \theta_j} f(X_{i-1}; \theta)}{\sigma^2(X_{i-1}; \psi)} \Delta X_i = 0 \ . \tag{4.8}$$

Example 4.2. A popular SDE in econometrics is the following:

$$dX(t) = a(\mu - X(t)) + \sigma \sqrt{X(t)} dB(t) \ ,$$

called the *CIR model* after the initials of Cox, Ingersoll, Ross, that advocated the use of the diffusion for modelling the evolution of interest rates in a paper

in Econometrica in 1985. Solving equation (4.7) for this specific SDE where now:

$$\sigma(x;\psi) = \sigma\sqrt{x}$$

provides the following estimator of σ:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(\Delta X_i)^2}{T} \Big/ \frac{\sum_{i=1}^{n} X_{i-1}}{n} \ .$$

□