An Introduction to the Numerical Simulation of Stochastic Differential Equations

Desmond J. Higham
Department of Mathematics
University of Strathclyde
Glasgow, G1 1XH
Scotland, U.K.
Telephone: +44 141 548 3716
Fax: +44 141 552 8657

Peter E. Kloeden
Institut für Mathematik
Johan Wolfgang Goethe-Universität
D-60054 Frankfurt am Main
Germany

December 21, 2012

Still to do

- Nearer the end we can list outstanding tasks here.
- Add a section summarizing notation in book.
- Make numbering of Theorems etc involve chapter number.
- Task 2
- Task3
- Etc.

Note: we are only using the basic LaTeX "book" style. No attention has been paid to fonts/sizing/numbering for headings, figures, captions, Examples/Computational Examples/Exercises/Quotes, etc. The look and feel of the text will change dramatically before a final version is prepared.

Contents

Preface

Our intention in this book is to provide a punchy, accessible introduction to the numerical solution of stochastic differential equations (SDEs). With the aim of making this topic available to the widest possible readership, we have kept the prerequisites to a minimum. We assume only a competence in algebra and calculus at the level reached by a typical first year undergraduate mathematics class. Some familiarity with basic concepts from numerical analysis and probability is also desirable, but not absolutely necessary.

Our intended readership includes

- undergraduate and beginning graduate students in mathematics, statistics, physics, economics, finance, business, computer science, engineering and the life sciences, who, perhaps having been exposed to SDE models, wish to learn more about how to simulate them,

- researchers in the above disciplines who routinely perform SDE simulations and would like to understand some of the underlying theory, and

- researchers from deterministic applied mathematics or numerical analysis who are seeking to broaden their interests.

A big motivation for this book has been the amount of positive feedback that one of us has received from the article [26] that appeared in the Education section of SIAM Review. Based on that feedback, and on discussions with colleagues, we believe that there is a definite niche for a self-contained, low-level text that puts across the fundamentals as succinctly as possible. Following the style of [26] we have made heavy use of computational examples and illustrative figures. There is, of course, much more material here than in [26]. Our guiding principles were to add enough background material to allow an outline proof of the key properties of weak convergence (Chapter 8) and strong convergence (Chapter 9) of the Euler–Maruyama method, and to explain the relevance of Ito's lemma in the derivation of higher order methods. Because comprehensive and rigorous numerical SDE books, such as [38, 46, 47], have already been published, we feel justified in focusing on accessibility and brevity.

This is not a rigorous text. Like Thomas Mikosch, author of the excellent

introductory SDE text [45], we are not *ashamed* of this lack of rigour[1]. However, we are somewhat *wary* that readers may be fooled into thinking that we are presenting the whole story. In an effort to avoid this, we have included many pointers to the wealth of high-level, technical literature that can be used to fill in the many gaps.

The authors have given post-graduate short courses based on various incarnations of this material at several insitutions across Europe and North America. Students at these courses typically had backgrounds in mathematics, physics, computer science or engineering and were interested in SDEs because of their relevance to physics, mathematical finance and the life sciences.

You can get a reasonable feel for the content of this book by skimming through the Outline bullet points that begin each chapter. Because this is a relatively new and rapidly expanding field, even in an introductory text like this we were able to include material on topics that, to our knowledge, do not appear elsewhere outside the domain of research articles:

- asymptotic and mean-square linear stability for numerical SDEs: Chapter 10,
- steady state behaviour: Chapter 13,
- SDEs with jumps: Chapter 14,
- mean exit times: Chapter 11.

There are a number of topics related to numerical SDEs that can be confusing, and can raise questions that are as much philosophical as mathematical. Within the limitations of accessibility, we have tried to explain clearly the issues surrounding

- strong versus weak solutions,
- Ito versus Stratonovich calculus,
- strong versus weak convergence,
- mean-square versus asymptotic stability.

The popularity and importance of SDEs is driven by their relevance as models in many application areas. To acknowledge this, and also to give the opportunity to present some fairly realistic computational scenarios, we have included descriptions of simple SDEs models from a range of scenarios.

Each chapter includes a list of exercises that are designed to reinforce the text and fill in some of the details. A three star labelling system has been used; one star

[1]See the quote at the end of this preface.

for questions with the shortest/easiest answers and three stars for questions with the longest/hardest answers. Worked solutions to the even-numbered exercises are obtainable from the website mentioned in the next paragraph. This leaves the even-numbered questions as an examining resource for teachers. We have also sprinkled a few quotes at the end of each chapter. You are not allowed to read these until you have done all the exercises.

A website for the book has been created at ?????? It includes

- worked solutions to the even-numbered exercises,
- listings of each Program of the Chapter,
- links to all websites mentioned in the book,
- bonus quotes,

and will no doubt also house

- a list of corrections.

A number of people deserve acknowledgement. DJH wishes to thank

- **Xuerong Mao** and **Andrew Stuart**, who have generously shared their SDE expertise in research collaborations.

Both authors are grateful to

- **Maxim Ullrich**, **Arnulf Jentzen** and **Andreas Neuenkirch**, who have provided critical feedback on one or more portions of drafts of this book.

MATLAB Programs

In our experience, the best way to understand a computational algorithm is to experiment with a suitable computer program. Hence, in addition to the numerous computational examples throughout the book, each chapter ends with a full listing of a *Program of the Chapter* that illustrates a key topic. These programs may be downloaded from the book's website, mentioned earlier in the preface.

The programs are written in MATLAB[2], a commercial software product produced by The Mathworks, whose homepage is at

```
http://www.mathworks.com/
```

Descriptive comments have been inserted into the codes and an accompanying "walk-through" commentary is included. The walk-throughs are intended to describe the key MATLAB commands as they are used; for a comprehensive guide to MATLAB, see [28]. The codes were written for illustrative purposes, rather than for efficiency or elegance. However, where appropriate in later chapters we have used MATLAB's *vectorization* facilities (essentially, this boils down to avoiding `for` loops); this makes the codes shorter and easier to assimilate, and also has the advantage of improving execution time.

We chose to present codes in MATLAB because it offers

- high-level random number generation and two and three dimensional graphical output,
- built-in commands for matrix computations and statistics,
- portability across a variety of platforms,

and also because it is widely used in universities and research laboratories.

If you have any experience of scientific computing in languages such as Java, C or FORTRAN then you should find it relatively easy to understand these MATLAB codes.

Let us emphasize that the MATLAB programs are an additional feature, not an essential part of the book. The rest of the book may be read without reference to them.

Quotes:
Stochastic differential equations is usually,
and justly,
regarded as a graduate level subject.

[2]Legal blurb about Mathworks

A really careful treatment assumes the students' familiarity with
probability theory, measure theory, ordinary differential equations,
and perhaps partial differential equations as well.
This is all too much to expect of undergrads.

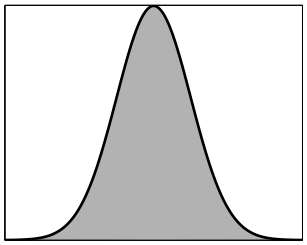
But white noise, Brownian motion and the random calculus are wonderful topics,
too good for undergraduates to miss out on.
Lawrence C. Evans, An Introduction to Stochastic Differential Equations, Version 1.2,
available at `http://math.berkeley.edu/~evans/`

Ten years ago I would not have dared to write a book like this:
a non-rigorous treatment of a mathematical theory.
I admit that I would have been ashamed,
and I am afraid that most of my colleagues in mathematics still think like this.
However, my experience with students and practitioners convinced me
that there is a strong demand for popular mathematics.
Thomas Mikosch, [45].

Example is always more efficacious than precept.
Samuel Johnson (1709–1784).

Chapter 1

Random Variables



Outline

- Discrete and continuous random variables
- Expected value and variance
- Independence
- Normal random variables
- Central Limit Theorem
- Strong Law of Large Numbers

1.1 Motivation

In order to progress to stochastic differential equations, we need an appreciation of some basic concepts from probability theory. This chapter is designed to give a very brief introduction on a need-to-know, example-driven basis. Our focus here, and throughout the book, is on developing an intuitive feeling for the main ideas, and hence we do not give rigorous definitions and proofs.

1.2 Discrete Random Variables

Suppose we roll a fair die. Since the six possible outcomes, $\{1, 2, 3, 4, 5, 6\}$, are all equally likely, we may say that the *probability* of rolling any one of those numbers is $1/6$. More formally, we may introduce a random variable X that describes the outcome of the roll, and say that

$$\mathbb{P}(X = i) = \frac{1}{6}, \quad \text{for each } i = 1, 2, \ldots, 6. \tag{1.1}$$

We read this as "the probability that $X = i$ is $1/6$".

Generalizing this concept, a *discrete random variable* X that takes possible values from the set $\{x_1, x_2, \ldots, x_K\}$ has an associated set of probabilities $\{p_1, p_2, \ldots, p_K\}$ such that

$$\mathbb{P}(X = x_i) = p_i, \quad 1 \leq i \leq K. \tag{1.2}$$

We read this as "the probability that $X = x_i$ is p_i". The set of probabilities must satisfy two reasonable conditions:

- $p_i \geq 0$ (probabilities cannot be negative), and
- $\sum_{i=1}^{K} p_i = 1$ (probabilities add up to one).

Example A *Bernoulli* random variable with parameter $p \in [0, 1]$ has $K = 2$, $x_1 = 0$, $x_2 = 1$, $p_1 = 1 - p$ and $p_2 = p$. $\diamondsuit$

We can think of a Bernoulli random variable as respresenting an experiment that has two possible outcomes. A good example is a coin toss. Suppose

- a coin is biased in such a way that the chance of `Heads` is p, and
- we win \$1 if the coin lands `Heads` and nothing if the coin lands `Tails`.

Then a Bernoulli random variable represents our winnings.

We may extend this scenario to let X represent our total winnings after n coin tosses (or, more generally, the total number of "successes" after n experiments). In this case, X takes possible values $\{0, 1, 2, \ldots, n\}$ and we have

$$\mathbb{P}(X = i) = \binom{n}{i} p^i (1 - p)^{n-i}, \quad 0 \leq i \leq n, \tag{1.3}$$

where

$$\binom{n}{i} := \frac{n!}{i!(n - i)!}$$

is a binomial coefficient. Exercise 1.2 asks you to justify (1.3).

We refer to X in (1.3) as a *Binomial* random variable with parameters n and p. A Bernoulli random variable is therefore a special case of a Binomial random variable where $n = 1$.

We may also consider the case where there are a (countably) infinite number of possible outcomes $\{x_i\}_{i=1}^{\infty}$. We then require a set of probabilities $\{p_i\}_{i=1}^{\infty}$ such that

- $p_i \geq 0$ (probabilities cannot be negative), and
- $\sum_{i=1}^{\infty} p_i = 1$ (probabilities add up to one).

Example A *Poisson* random variable X with parameter $\lambda > 0$ takes possible values $\{0, 1, 2, 3, \ldots\}$ such that

$$\mathbb{P}(X = i) = e^{-\lambda}\frac{\lambda^i}{i!}, \quad i \geq 0. \tag{1.4}$$

$\diamondsuit$

Figure 1.1 illustrates the probabilities arising for a Poisson random variable with $\lambda = 0.1$, 0.5 and 1, over the range $0 \leq i \leq 5$.

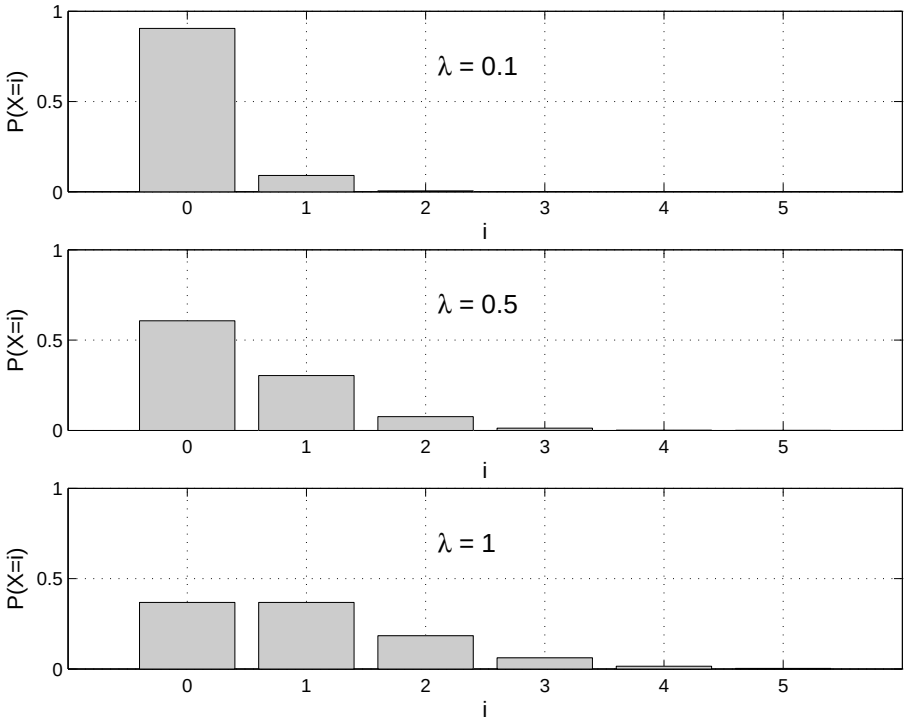


Figure 1.1: Probabilities (1.4) associated with a Poisson random variable for $\lambda = 0.1$, 0.5 and 1, over the range $0 \leq i \leq 5$.

A Poisson random variable with parameter $\lambda = np$ may be regarded as an approximation to a Binomial random variable with parameters n and p when n is large and np is of moderate size; see Exercise 1.4.

Poisson random variables play a central role in Chapter 14 on jumps.

1.3 Continuous Random Variables

A *continuous random variable* X with *density function* f is such that

$$\mathbb{P}(a \leq X \leq b) = \int_a^b f(x)\,dx, \quad \text{for any } a \leq b.$$

In words, the probability that X takes a value in the interval $[a, b]$ is given by the integral of the density function over that interval. A density function must satisfy two reasonable conditions:

- $f(x) \geq 0$ (probabilities cannot be negative), and
- $\int_{-\infty}^{\infty} f(x)\, dx = 1$ (density integrates to one).

Example If

$$f(x) = \begin{cases} \frac{1}{\beta - \alpha} & \text{for } x \in (\alpha, \beta) \\ 0 & \text{otherwise,} \end{cases} \tag{1.5}$$

then X is said to be a *uniform* random variable over (α, β). Such a random variable only takes values in the range (α, β) and, within this range, is not biased towards any particular values. $\diamondsuit$

Example A continuous random variable with density function

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{for } x \geq 0 \\ 0 & \text{otherwise,} \end{cases} \tag{1.6}$$

for some $\lambda > 0$, is said to be an *exponential* random variable with parameter λ. $\diamondsuit$

Figure 1.2 plots the exponential density function in (1.6) for the cases $\lambda = \frac{1}{2}$, 1 and 2.

So far we have glossed over the meaning of the concept of a "probability". In everyday language the word has a well-understood connotation, but pinning this down mathematically has proved surprisingly contentious. From the viewpoint of classical statistics, if we regard a discrete random variable X as representing the outcome of an experiment, for which repeated, independent runs can be made, then $\mathbb{P}(X = x_i)$ accords with the *long-run relative frequency* with which the outcome x_i will be observed. So $\mathbb{P}(X = x_i) = 0.9$ suggests that the outcome x_i should arise in about 90% of the runs. Similarly, if X is a continuous random variable then $\mathbb{P}(a \leq X \leq b)$ can be thought of as the long-run relative frequency for observations in $[a, b]$. This viewpoint is standard, and is natural for the computer simulation tasks that we cover in this book. However, we mention that the alternatve *Bayesian* viewpoint, where probability represents *level of belief*, is perhaps more systematic and widely applicable. References for the Bayesian approach are given in Section 1.10.

1.4 Expectation and Variance

A key concept in probability theory is the *expectation* of a random variable. If X is a discrete random variable, as characterized by (1.2), then the *expected value*

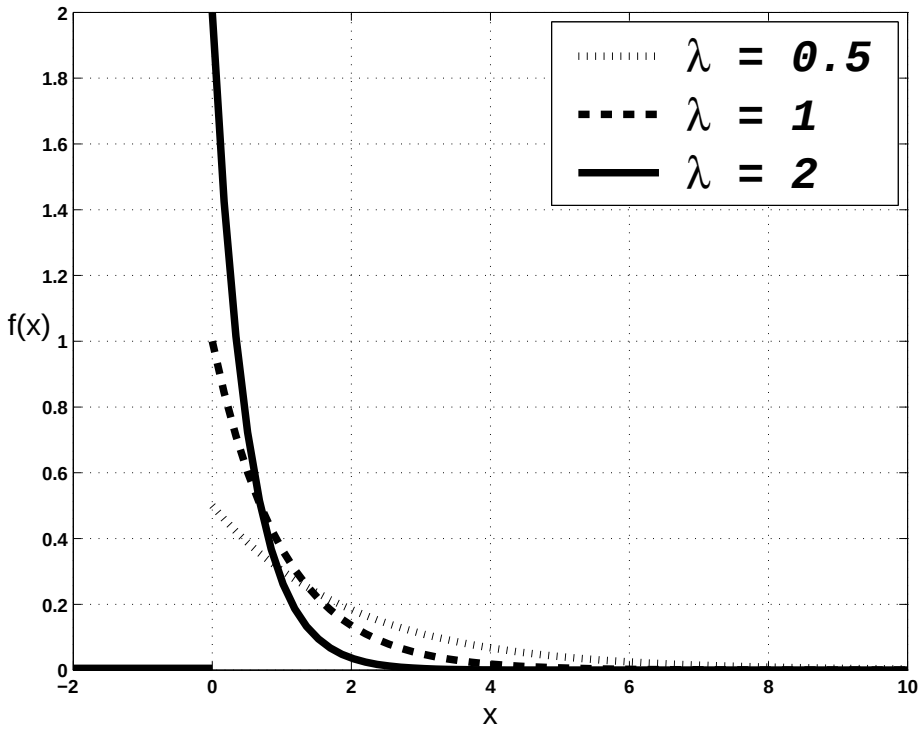


Figure 1.2: Exponential density function $f(x)$ in (1.6) for $\lambda = \frac{1}{2}$, 1 and 2.

of X, denoted $\mathbb{E}[X]$, or, when more convenient, by $\mathbb{E}X$, is defined by

$$\mathbb{E}[X] := \sum_{i=1}^{K} x_i p_i. \tag{1.7}$$

We also refer to $\mathbb{E}[X]$ as the *mean* of X. In words, the expected value is a weighted average of the possible outcomes, with the weights given by the corresponding probabilities.

Example When X represents the roll of a fair die, (1.7) gives

$$\mathbb{E}[X] = 1\frac{1}{6} + 2\frac{1}{6} + 3\frac{1}{6} + 4\frac{1}{6} + 5\frac{1}{6} + 6\frac{1}{6} = \frac{7}{2}. \tag{1.8}$$

♢

Example If X is a Bernoulli random variable with parameter p then,

$$\mathbb{E}[X] = 0 \times (1-p) + 1 \times p = p. \tag{1.9}$$

♢

Exercise 1.7 asks you to show that $\mathbb{E}[X] = np$ for a Binomial random variable with parameters n and p.

We remark that the expected value should not be interpreted as the "most likely" value since, as in (1.8) and (1.9), it may be a value that X cannot possibly take. However, $\mathbb{E}[X]$ can be regarded as an "average" value—see the Strong Law of Large Numbers in Section 1.9.

In the case where X is a discrete random variable taking a (countably) infinite number of possible values, we simply use (1.7) with $K = \infty$, and in this case by writing $\mathbb{E}[X]$ we are implicitly assuming that the infinite series is finite.

Example For a Poisson random variable with parameter λ, we have, from (1.4),

$$\begin{aligned}\mathbb{E}[X] &= \sum_{i=0}^{\infty} i e^{-\lambda} \frac{\lambda^i}{i!} \\ &= \lambda e^{-\lambda} \sum_{i=1}^{\infty} \frac{\lambda^{i-1}}{(i-1)!} \\ &= \lambda e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} \\ &= \lambda e^{-\lambda} e^{\lambda} \\ &= \lambda. \end{aligned} \tag{1.10}$$

♢

The definition of expected value extends to a continuous random variable with density function $f(x)$ according to

$$\mathbb{E}[X] := \int_{-\infty}^{\infty} x f(x)\, dx, \tag{1.11}$$

where, in writing $\mathbb{E}[X]$, we are implicitly assuming that this integral is finite.

Example If X is a uniform random variable over (α, β) then, using (1.5) in (1.11),

$$\mathbb{E}[X] = \int_{\alpha}^{\beta} \frac{x}{\beta - \alpha}\, dx = \left[\frac{x^2}{2(\beta - \alpha)} \right]_{\alpha}^{\beta} = \frac{\alpha + \beta}{2}.$$

♢

Example Using (1.6) in (1.11), an exponential random variable with parameter λ has expected value

$$\mathbb{E}[X] = \int_0^{\infty} \lambda x e^{-\lambda x}\, dx.$$

Integration by parts then gives

$$\mathbb{E}[X] = \left[-xe^{-\lambda x}\right]_0^\infty + \int_0^\infty e^{-\lambda x}\,dx = 0 + \left[\frac{e^{-\lambda x}}{-\lambda}\right]_0^\infty = \frac{1}{\lambda}. \tag{1.12}$$

♢

If X is a random variable and $\alpha \in \mathbb{R}$ then αX is another random variable, and it is intuitively reasonable that

$$\mathbb{E}[\alpha X] = \alpha \mathbb{E}[X]. \tag{1.13}$$

Exercises 1.8 and 1.9 ask you to confirm this.

Similarly, if X and Y are random variables then $X + Y$ is also a random variable, and we have

$$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]. \tag{1.14}$$

In words, the mean of the sum is the sum of the means. In particular, if α is a constant then, since α may be regarded as a random variable with expected value α, we have

$$\mathbb{E}[\alpha + X] = \alpha + \mathbb{E}[X]. \tag{1.15}$$

Generally, given any function h we may apply this function to X in order to get a new random variable. In the discrete case (1.2) we see that $h(X)$ takes possible values $\{h(x_1), h(x_2), \ldots, h(x_K)\}$ and we have

$$\mathbb{P}\left(h(X) = h(x_i)\right) = p_i,$$

whence

$$\mathbb{E}[h(X)] = \sum_{i=1}^{K} h(x_i) p_i.$$

The corresponding identity for a continuous random variable is

$$\mathbb{E}[h(X)] = \int_{-\infty}^{\infty} h(x) f(x)\,dx. \tag{1.16}$$

Example If X is an exponential random variable with parameter λ, then

$$E[X^2] = \int_0^\infty x^2 \lambda e^{-\lambda x}\,dx.$$

Integrating by parts, and using (1.12), we find

$$\mathbb{E}[X^2] = \left[-x^2 e^{-\lambda x}\right]_0^\infty + \int_0^\infty 2xe^{-\lambda x}\,dx = 0 + \frac{2}{\lambda}\mathbb{E}[X] = \frac{2}{\lambda^2}. \tag{1.17}$$

♢

The expected value is a single number that captures a "weighted average" of the possible values. The *variance*, denoted $\mathsf{var}[X]$, is another informative number. It describes how much X tends to stray from its expected value. We define the variance as

$$\mathsf{var}[X] := \mathbb{E}\left[(X - \mathbb{E}[X])^2\right]. \tag{1.18}$$

Using (1.13) and (1.14), we have

$$\begin{aligned}\mathsf{var}[X] &= \mathbb{E}\left[X^2 - 2X\mathbb{E}[X] + (\mathbb{E}[X])^2\right] \\ &= \mathbb{E}[X^2] - 2\mathbb{E}[X]\mathbb{E}[X] + (\mathbb{E}[X])^2 \\ &= \mathbb{E}[X^2] - (\mathbb{E}[X])^2, \end{aligned} \tag{1.19}$$

which is often a more convenient characterization of $\mathsf{var}[X]$. Using (1.14) in (1.19) we see that

$$\mathsf{var}[\alpha X] = \alpha^2 \mathsf{var}[X], \tag{1.20}$$

for any $\alpha \in \mathbb{R}$.

The *standard deviation* is simply the square root of the variance. In this book, for consistency, we will always work in terms of the variance, rather than the standard deviation.

Example A Bernoulli random variable with parameter p has variance given by $\mathsf{var}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = p - p^2 = p(1-p)$. $\diamondsuit$

Example A Poisson random variable with parameter λ has, using (1.10),

$$\begin{aligned}\mathbb{E}\left[X^2\right] &= \sum_{i=0}^{\infty} i^2 e^{-\lambda} \frac{\lambda^i}{i!} \\ &= \lambda \sum_{i=1}^{\infty} i e^{-\lambda} \frac{\lambda^{(i-1)}}{(i-1)!} \\ &= \lambda \sum_{j=0}^{\infty} (j+1) e^{-\lambda} \frac{\lambda^j}{j!} \\ &= \lambda \left(\sum_{j=0}^{\infty} j e^{-\lambda} \frac{\lambda^j}{j!} + \sum_{j=0}^{\infty} e^{-\lambda} \frac{\lambda^j}{j!} \right) \\ &= \lambda(\mathbb{E}[X] + 1) \\ &= \lambda(\lambda + 1). \end{aligned} \tag{1.21}$$

Hence, from (1.19), we have

$$\mathsf{var}[X] = \lambda. \tag{1.22}$$

$\diamondsuit$

Example If X is an exponential random variable with parameter λ then, from (1.12) and (1.17),

$$\mathsf{var}[X] = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

$\diamondsuit$

1.5 Markov and Lyapunov Inequalities

If $\mathbb{E}[X]$ is "moderately sized" then our intuition may suggest that X is unlikely to take very large values. This is indeed the case, according to a very useful result known as the *Markov inequality*. This result concerns random variables that take only non-negative values, and hence applies to $|X|$ where X can be any random variable. The inequality is

$$\mathbb{P}\left(|X| > a\right) \leq \frac{\mathbb{E}[|X|]}{a}, \qquad \text{for any } a > 0. \tag{1.23}$$

Applying the Markov inequality to $(X - \mathbb{E}[X])^2$, we have

$$\mathbb{P}\left((X - \mathbb{E}[X])^2 \geq a^2\right) \leq \frac{\mathbb{E}[(X - \mathbb{E}[X])^2]}{a^2} = \frac{\mathsf{var}[X]}{a^2},$$

and hence

$$\mathbb{P}\left(|X - \mathbb{E}[X]| \geq a\right) \leq \frac{\mathsf{var}[X]}{a^2}, \quad \text{for all } a > 0, \tag{1.24}$$

a result that is known as *Chebyshev's inequality*. This gives one way to quantify the sense in which the variance controls how far a random variable is spread away from its mean.

Next we give a result that relates $\mathbb{E}[X]$ and $\mathbb{E}[X^2]$. To begin, we note that for any $x, \mu \in \mathbb{R}$, it is easy to check that

$$x^2 = \mu^2 + 2\mu(x - \mu) + (x - \mu)^2,$$

and so

$$x^2 \geq \mu^2 + 2\mu(x - \mu).$$

Hence, for any random variable X we may write

$$X^2 \geq \mu^2 + 2\mu(X - \mu). \tag{1.25}$$

Choosing $\mu = \mathbb{E}[X]$ and taking expectations in (1.25), using (1.13), (1.14) and (1.15), we find that

$$\mathbb{E}[X^2] \geq \mu^2 + 2\mu\mathbb{E}[X - \mu] = \mu^2 + 2\mu\left(\mathbb{E}[X] - \mu\right) = \mu^2 = \left(\mathbb{E}[X]\right)^2.$$

Hence, we have shown that

$$\mathbb{E}[X] \leq \sqrt{\mathbb{E}[X^2]}. \tag{1.26}$$

This is known as the *Lyapunov inequality*.

Because (1.23), (1.24) and (1.26) are true for all random variables (assuming, of course, that the expected values exist), they are typically not sharp—see, for example, Exercise 1.11. However, they can provide a very convenient means to derive results about probabilities and expected values, as we will see in Chapter 7 and Chapter 9.

1.6 Independence

Loosely, two random variables are independent if knowing the value of one gives us no useful information about the value of the other. To give the classical definition of pairwise independence requires concepts that we do not develop in this book, and hence we will state an equivalent condition. Random variables X and Y are *independent* if and only if

$$\mathbb{E}[g(X)h(Y)] = \mathbb{E}[g(X)]\mathbb{E}[h(Y)], \qquad \text{for all } g, h : \mathbb{R} \mapsto \mathbb{R}. \tag{1.27}$$

More generally, a sequence of random variables $X_1, X_2, X_3, \ldots$ is said to be *independent and identically distributed*, abbreviated to *i.i.d.* if

(a) they have the same distribution (that is, in the discrete case they have the same possible values and associated probabilities, and in the continuous case they have the same density function), and

(b) knowing the values of any subsequence gives no useful information about the values of the remaining random variables.

In particular, if $X_1, X_2, X_3, \ldots$ are i.i.d. then they are pairwise independent, and so, taking both f and g to be the identity functions in (1.27).

$$\mathbb{E}[X_i X_j] = \mathbb{E}[X_i]\mathbb{E}[X_j], \quad \text{for } i \neq j. \tag{1.28}$$

We emphasize that although the additivity property (1.14) holds for any pair of random variables, the multiplicativity property (1.28) does not hold, in general, unless X and Y are independent. For example, when X is a Bernoulli random variable and $X = Y$ we have, from (1.17), and (1.12) $\mathbb{E}[XY] = \mathbb{E}[X^2] = 2/\lambda^2$, whereas $\mathbb{E}[X]\mathbb{E}[Y] = (\mathbb{E}[X])^2 = 1/\lambda^2$.

1.7 Normal Random Variables

If X has density function

$$f(x) = \frac{1}{\sqrt{2\sigma^2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \tag{1.29}$$

then we say that X is a *normal random variable* with parameters μ and σ^2. We also write $X \sim N(\mu, \sigma^2)$. Exercise 1.14 asks you to confirm that $\mathbb{E}[X] = \mu$ and $\mathsf{var}[X] = \sigma^2$.

The density function (1.29) is often refered to as a "bell-shaped curve". Figure 1.3 plots examples where $\mu = 0, \sigma = 1$; $\mu = 1, \sigma = 4$ and $\mu = -1, \sigma = 0.75$. We see that $x = \mu$ gives the peak and σ controls the spread. In the special case where $X \sim N(0, 1)$, that is, where the mean is zero and the variance is one, we say that X is a *standard normal* random variable.

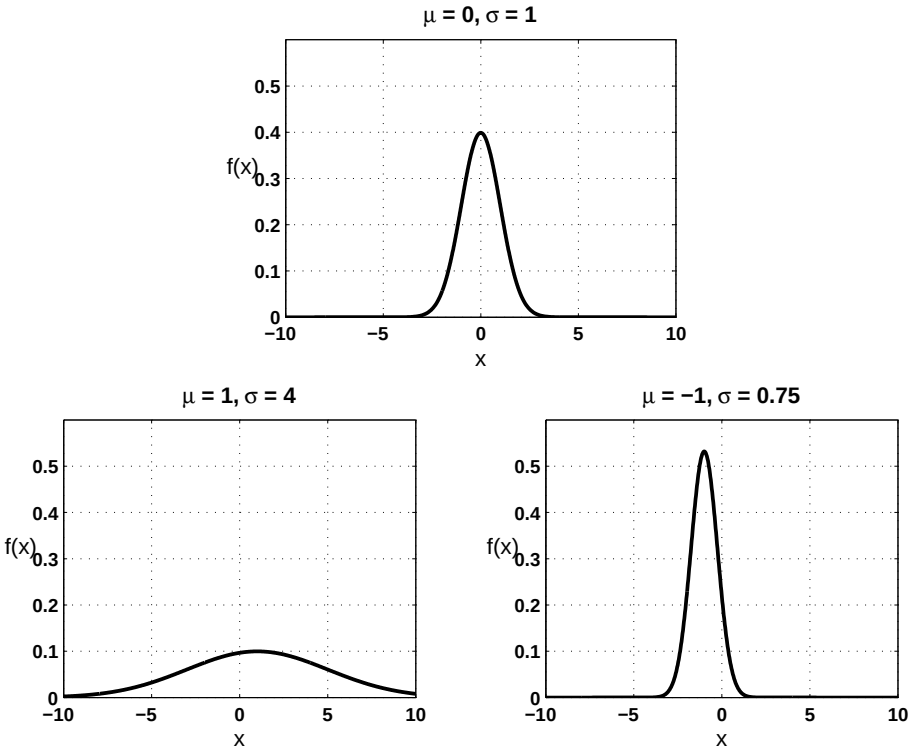


Figure 1.3: Normal density function $f(x)$ in (1.29) for three choices of (μ, σ^2).

Normal random variables play a central role in the study of stochastic differential equations, and indeed across the whole of stochastic analysis. A fundamental reason for their importance is given by the Central Limit Theorem, which is treated in the next section.

Normal random variables enjoy a number of special properties, including:

1. If $X \sim \mathsf{N}(\mu, \sigma^2)$ then $(X - \mu)/\sigma \sim \mathsf{N}(0, 1)$.

2. If $Y \sim \mathsf{N}(0, 1)$ then $\sigma Y + \mu \sim \mathsf{N}(\mu, \sigma^2)$.

3. If $X \sim \mathsf{N}(\mu_1, \sigma_1^2)$, $Y \sim \mathsf{N}(\mu_2, \sigma_2^2)$ and X and Y are independent, then $X + Y \sim \mathsf{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

4. If X and Y are normal random variables[1] then X and Y are independent if and only if $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.

We know from (1.13), (1.15) and (1.20) how shifting and scaling affects the mean and variance. Properties 1 and 2 show that these operations preserve the normality. Similarly property 3 tells us that adding two independent normal random variables produces another normal random variable. Finally, property 4 shows that the necessary condition for independence $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ is also sufficient when X and Y are normal. We emphasize that these properties are peculiar to the normal distribution, and analogous results do not hold for other types of random variable.

1.8 Central Limit Theorem

A wonderful and far-reaching result in probability theory says that the normal distribution is ubiquitous, in the sense that the sum of a large number of i.i.d. random variables behaves like a single normal random variable.

To begin, suppose we have two independent random variables, X and Y. We know from (1.14) that the means are additive. Using the independence result (1.28), we may compute the variance of the sum as

$$\begin{aligned}
\mathsf{var}[X + Y] &= \mathbb{E}[(X + Y)^2] - (\mathbb{E}[X + Y])^2 \\
&= \mathbb{E}[X^2 + 2XY + Y^2] - (\mathbb{E}[X] + \mathbb{E}[Y])^2 \\
&= \mathbb{E}[X^2] + 2\mathbb{E}[X]\mathbb{E}[Y] + \mathbb{E}[Y^2] - ((\mathbb{E}[X])^2 + 2\mathbb{E}[X]\mathbb{E}[Y] + (\mathbb{E}[Y])^2 \\
&= \mathsf{var}[X] + \mathsf{var}[Y]. \qquad (1.30)
\end{aligned}$$

Hence, for independent random variables, variances add.

Applying the same analysis to an i.i.d. sequence $\{X_i\}_{i \geq 1}$ with mean a and variance b^2, we may show that

$$\mathbb{E}\left[\sum_{i=1}^{M} X_i\right] = Ma \quad \text{and} \quad \mathsf{var}\left[\sum_{i=1}^{M} X_i\right] = Mb^2.$$

[1] More precisely, they must be *jointly* normally distributed, so that $aX + bY$ is normally distributed for any constants a and b.

hence, in view of (1.13), (1.15) and (1.20) the random variable

$$\frac{\sum_{i=1}^{M} X_i - Ma}{\sqrt{M}b} \tag{1.31}$$

has mean 0 and variance 1. The Central Limit Theorem reveals the remarkable fact that *whatever type* of random variables we use for the i.i.d. sequence $\{X_i\}_{i\geq 1}$, as $M \to \infty$ the shifted and scaled sum in (1.31) *always* behaves like a standard normal random variable. The theorem may be stated precisely as follows.

Theorem 1 (Central Limit Thoerem) *Let $\{X_i\}_{i\geq 1}$ be i.i.d. with mean a and variance b^2. Then for all $\alpha < \beta$*

$$\lim_{M\to\infty} \mathbb{P}\left(\alpha \leq \frac{\sum_{i=1}^{M} X_i - Ma}{\sqrt{M}b} \leq \beta\right) = \frac{1}{\sqrt{2\pi}} \int_{\alpha}^{\beta} e^{-\frac{1}{2}x^2}\, dx. \tag{1.32}$$

We emphasize that the theorem applies to *any* i.i.d. sequence (with finite mean and variance). From a mathematical modeling perspective, the Central Limit Theorem allows us to argue that the combined effect of a large number of sources of "noise" can be well approximated by a single normal random variable. This helps to explain the popularity and success of the use of normal random variables to represent "random" effects, and also serves to motivate the appearance of normal increments when we define Brownian motion in Chapter 3.

Figure 2.3 in Chapter 2 illustrates the Central Limit Theorem.

1.9 Strong Law of Large Numbers

A set of i.i.d. random variables $\{X_i\}_{i=1}^{M}$ may be thought of as representing repeated, independent runs of an experiment. It is then of interest to understand how well the average $(1/M)\sum_{i=1}^{M} X_i$ approximates the mean of X_i.To get some insight into this issue, we may rearrange (1.32) to give

$$\mathbb{P}\left(\frac{\alpha b}{\sqrt{M}} \leq \frac{1}{M}\sum_{i=1}^{M} X_i - a \leq \frac{\beta b}{\sqrt{M}}\right) \approx \frac{1}{\sqrt{2\pi}} \int_{-\alpha}^{\beta} e^{-\frac{1}{2}x^2}\, dx.$$

Then taking $\alpha = -M^{\frac{1}{4}}$ and $\beta = M^{\frac{1}{4}}$ gives

$$\mathbb{P}\left(\frac{-b}{M^{\frac{1}{4}}} \leq \frac{1}{M}\sum_{i=1}^{M} X_i - a \leq \frac{b}{M^{\frac{1}{4}}}\right) \approx \frac{1}{\sqrt{2\pi}} \int_{-M^{\frac{1}{4}}}^{M^{\frac{1}{4}}} e^{-\frac{1}{2}x^2}\, dx \to 1, \quad \text{as } M \to \infty.$$

Consequently, we could argue that, for large M, $(1/M)\sum_{i=1}^{M} X_i$ will be close to the mean value a with high probability. This statement can be made precise in a very sharp manner, as the following result shows.

Theorem 2 (Strong Law of Large Numbers) *Let $\{X_i\}_{i\geq 1}$ be i.i.d. with mean a. Then*

$$\mathbb{P}\left(\lim_{M\to\infty}\frac{1}{M}\sum_{i=1}^{M}X_i = a\right) = 1. \tag{1.33}$$

Loosely, the Strong Law of Large Numbers tells us that the average of a large number of i.i.d. random variables behaves like their common mean, and the Central Limit Theorem tells us about the fluctuations around that mean. Together, these two powerful results can be used to motivate Monte Carlo simulation, a key computational tool that will be introduced in Chapter 2.

1.10 Notes and References

There are many excellent introductory books on random variables and probability that expand on the material in this chapter. Examples include [23, 53, 61]. A particularly informal, freewheeling introduction is given by [30]. *The Probability Web* at `http://www.mathcs.carleton.edu/probweb/probweb.html` is a useful on-line source for all things probabilistic, including a text [24] that the American Mathematical Society has placed in the public domain.

To study continuous random variables with complete rigour requires the tools of measure theory, which are beyond the scope of this book. Accessible introductions are [10, 57]. To learn more about the Bayesian approach to probability, we recommend [59, 61] as introductory guides and [35] as a forcefully written, encyclopaedic reference. The book [42] deals with many modern applications from a Bayesian viewpoint.

We stated simple forms of the Central Limit Theorem and the Strong Law of Large Numbers. Both results can be generalized to less restrictive circumstances, such as when the i.i.d. assumption is relaxed; see, for example, [31]. The Lyapunov inequality (1.26) is a special case of *Jensen's inequality*, which says that $\mathbb{E}[f(x)] \geq f(\mathbb{E}[X])$ whenever $f(\cdot)$ is convex.

Exercises

1.1. $\star$ Confirm that $\sum_{i=0}^{n}\mathbb{P}(X=i) = 1$ for X in (1.3).

1.2. $\star\star$ Using the "probability as long-run relative frequency" interpretation mentioned at the end of Section 1.3, show that regarding X as the sum of n Bernoulli random variables leads to (1.3).

1.3. $\star$ Confirm that $\sum_{i=0}^{n}\mathbb{P}(X=i) = 1$ for a Poisson random variable X, as defined in (1.4).

1.4. ⋆⋆ Suppose X is a Binomial random variable with parameters n and p, where n is large and $\lambda = np$ is of moderate size. For a fixed i, use the approximations

$$\frac{n(n-1)\cdots(n+i-1)}{n^i} \approx 1, \quad \left(1-\frac{\lambda}{n}\right)^n \approx e^{-\lambda} \quad \text{and} \quad \left(1-\frac{\lambda}{n}\right)^i \approx 1$$

to show that

$$\mathbb{P}\left(x=i\right) \approx e^{-\lambda}\frac{\lambda^i}{i!},$$

so that X behaves like a Possion random variable with parameter λ.

1.5. ⋆ Show that the exponential density in (1.6) integrates to 1.

1.6. ⋆⋆ A *standard Cauchy random variable* has density function

$$f(x) = \frac{1}{\pi(1+x^2)}, \qquad \text{for all } x \in \mathbb{R}.$$

Show that the integral (1.11) defining the expected value is not finite in this case.

1.7. ⋆ Show that a Binomial random variable with parameters n and p has expected value np.

1.8. ⋆ Suppose X is a discrete random variable as defined in (1.2). Given $\alpha \in \mathbb{R}$, let $Y = \alpha X$, so that Y is a random variable taking possible values $\{\alpha x_1, \alpha x_2, \ldots, \alpha x_\}$, such that $\mathbb{P}(Y = \alpha x_i) = p_i$. From the definition of expected value (1.7), show that $\mathbb{E}[Y] = \alpha\mathbb{E}[X]$.

1.9. ⋆⋆ Suppose X is a continuous random variable with density function $f_X(x)$. Given $\alpha > 0$, let $Y = \alpha X$, and denote the density function of Y by $f_Y(x)$. By comparing

$$\mathbb{P}\left(a \leq Y \leq\ b\right) = \mathbb{P}\left(\frac{a}{\alpha} \leq X \leq \frac{b}{\alpha}\right) = \int_{a/\alpha}^{b/\alpha} f_X(x)\,dx$$

with

$$\mathbb{P}\left(a \leq Y \leq b\right) = \int_a^b f_Y(x)\,dx,$$

find an expression for $f_Y(\cdot)$ in terms of $f_X(\cdot)$. Hence, show that $\mathbb{E}[Y] = \alpha\mathbb{E}[X]$. Similarly, show that this result also holds when $\alpha < 0$.

1.10. ⋆ Show that a Binomial random variable with parameters p and n has variance $np(1-p)$.

1.11. ⋆⋆ Verify that the Markov inequality (1.23) is true when X is a Bernoulli random variable with parameter p. Comment on the sharpness of the inequality in this case.

```
%CH01.M   Plot density for Normal (0,1)

clf

xvals = linspace(-10,10,1e3);
mu = 0;
sigma = 1;
denom = sqrt(2*pi)*sigma;
func = exp(-0.5*(xvals-mu).^2/(sigma^2));
yvals = (1/denom)*func;
plot(xvals,yvals)

xlim([-10,10])
ylim([0 0.5])
title('\mu = 0, \sigma = 1')
xlabel('x')
ylabel('f(x)')
grid on
```

Figure 1.4: Program of Chapter 1: `ch01.m`.

1.12. ⋆ Show that if X and Y are independent and $\alpha \in \mathbb{R}$, then αX and αY are independent.

1.13. ⋆ Show that a normal random random variable (that is, one with density given by (1.29)) has mean μ and variance σ^2.

1.14. ⋆ Taking $\alpha = -K/\sqrt{M}$ and $\beta = K/\sqrt{M}$ in (1.32), for any $K > 0$, we deduce that

$$\mathbb{P}\left(-Kb \leq \sum_{i=1}^{M} X_i - Ma \leq Kb\right) \approx \frac{1}{\sqrt{2\pi}} \int_{-K/\sqrt{M}}^{K/\sqrt{M}} e^{-\frac{1}{2}x^2}\, dx.$$

Use this to argue that "$\sum_{i=1}^{M} X_i$ fluctuates from its mean beyond any finite bound with high probability."

1.11 Program of Chapter 1 and Walkthrough

The program `ch01`, listed in Figure 1.4, produces a simple plot of the density function (1.29) for a normal random variable in the standard case where $\mu = 0$ and $\sigma = 1$, as seen in the top of Figure 1.3. MATLAB's basic two-dimensional `plot` command is used to draw the picture.

Programming Exercises

P1.1. Use MATLAB's `subplot` command to illustrate several choices of μ and σ in (1.29), along the lines of Figure 1.3.

P1.2. With MATLAB's `bar` command, produce histograms for the Poisson distribution like those in Figure 1.1.

Quotes:

A man from a village in northern France has won
the French national lottery for the second time—using the same numbers he did the first time
around 27 years ago. .....
A spokesman for the lottery organiser, Francaise des Jeux,
said the odds of the same seven digit series cropping up twice
were "virtually incalculable".
Lucky lotto numbers win again, Guardian (newspaper), August 20, 2005.

While writing my book [Stochastic Processes] I had an argument with Feller.
He asserted that everyone said "random variable"
and I asserted that everyone said "chance variable."
We obviously had to use the same name in our books,
so we decided the issue by a stochastic procedure.
That is, we tossed for it and he won.
J. Doob, quoted in Statistical Science,
source `http://www.mathcs.carleton.edu/probweb/probweb.html`

Most people have an intuitive conception of a probability,
based on their own experience.
However, a precise formulation of intuitive concepts is fraught with difficulties,
and it has been found most convenient
to axiomatise probability theory as an essentially abstract science, ...
C. W. Gardiner, [19].

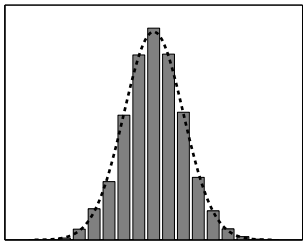
Presumably, Gauss meant 'normal' in its mathematical sense of 'perpendicular',
expressing the geometric meaning of those equations.
The minimum distance from a point (the estimate) to a plane (the constant)
is the length of the perpendicular.
But, as Pearson himself observes, the term 'normal distribution' is a bad one
because the common colloquial meaning of 'normal' is *standard* or *sane*,
implying a value judgement.
This leads many to think—consciously or subconsciously—that all other distributions
are in some way abnormal.

Actually, it is quite the other way;
it is the so-called 'normal' distribution that is abnormal
in the sense that it has many unique properties not possessed by any other.
E. T. Jaynes, [35].

This name was introduced by George Pólya (1920),
with the intention that the adjective 'central' was to modify the noun 'theorem';
i.e. it is the limit theorem which is *central to probability theory*.
Almost universally, students today think that 'central' modifies 'limit',
so that it is instead a theorem about a '*central limit*',
whatever that means.
E. T. Jaynes, [35].

Chapter 2

Computer Simulations



Outline

- Pseudo-random number generation
- Monte Carlo method
- Kernel density estimation

2.1 Motivation

In solving stochastic differential equations, we will tiptoe into the rich and rewarding field of *simulation*—the use of computational methods to study problems in probability and statistics. In this chapter we introduce some basic concepts. Pseudo-random numbers are the key ingredient for simulation algorithms, and typical tasks involve Monte Carlo approximation of expected values and reconstruction of entire probability distributions by kernel density estimation.

2.2 Pseudo-random Number Generation

It is possible to buy hardware devices that, according to established laws of physics, generate truly random numbers; that is, numbers that may be regarded as the outcome of repeated, independent runs of an experiment that can be represented by a random variable. For example, some commercial products use electrical noise or radioactive decay as the source of randomness. However, for reasons of economy, efficiency, reproducibility and personal safety, most of us must be content with *pseudo-random numbers*.

A pseudo-random number sequence is produced by a deterministic algorithm in such a way that the long-run statistics of the numbers are intended to match

those of an appropriate random variable. For example, typing `rand(10,1)` in MATLAB produces

```
>> rand(10,1)
ans =
    0.9501
    0.2311
    0.6068
    0.4860
    0.8913
    0.7621
    0.4565
    0.0185
    0.8214
    0.4447
```

Here we requested a 10 by 1 array of pseudo-random numbers consistent with the uniform distribution over $(0, 1)$. Using the function `randn` instead of `rand` specifies the N(0,1) distribution:

```
>> randn(10,1)
ans =
   -0.4326
   -1.6656
    0.1253
    0.2877
   -1.1465
    1.1909
    1.1892
   -0.0376
    0.3273
    0.1746
```

Designing, implementing and evaluating algorithms for pseudo-random number generation is an active, specialized research topic that lies beyond the scope of this book. Fortunately, high quality off-the-shelf code is available in most scientific computing languages—see Section 2.5. Our approach will be to assume that suitable pseudo-random number generators are available, and thus to focus on the central issue of how to use this technology in the numerical simulation of stochastic differential equations.

All stochastic computations in this book were performed using `rand` and `randn` to generate uniform and normal samples, respectively. In common with most software, these generators allow a "state" to be set that makes the com-

putations reproducible—if a program is re-run then using the same state setting causes the same pseudo-random numbers to be used[1].

2.3 Monte Carlo Method

Suppose that X is a random variable (a) from which we are able to sample, and (b) whose expected value $a = \mathbb{E}[X]$ we wish to approximate. Let $b^2 = \mathsf{var}[X]$ denote the variance of X. From the Strong Law of Large Numbers (Theorem 2) we know that repeated samples from a random variable can be averaged to give an asymptotically correct estimate of its mean. This is the essence of the Monte Carlo method—to estimate the mean of X, take a large number, M, of samples, $\{\xi_i\}_{i=1}^M$, and compute $(1/M)\sum_{i=1}^M \xi_i$. We also know from the Central Limit Theorem (Theorem 1) that the discrepency between this approximation and the true mean depends on the variance of X. This will allow us to assess the accuracy of our approximation.

The quantity

$$a_M := \frac{1}{M}\sum_{i=1}^{M} \xi_i \tag{2.1}$$

is known as the *sample mean*. From the definition of variance (1.18) it would be reasonable to estimate the variance as $(1/M)\sum_{i=1}^M (\xi_i - a_M)^2$. However, it is easily shown that this estimate is very slightly *biased*—see Exercise 2.1—and hence it is common practice to define the *sample variance* as the unbiased expression

$$b_M^2 := \frac{1}{M-1}\sum_{i=1}^{M} (\xi_i - a_M)^2. \tag{2.2}$$

Now suppose $Z \sim N(0,1)$. It is known that (to two decimal places)

$$\mathbb{P}\left(|Z| \leq 1.96\right) = 0.95, \tag{2.3}$$

see Exercise 2.3. In words, if we sample from a $N(0,1)$ distribution then around 95% of the time we will see values between -1.96 and $+1.96$. More generally, if $Y \sim N(\mu, \sigma^2)$ then the scaled version $(Y-\mu)/\sigma \sim N(0,1)$, so

$$\mathbb{P}\left(\left|\frac{Y-\mu}{\sigma}\right| \leq 1.96\right) = 0.95, \tag{2.4}$$

which rearranges to

$$\mathbb{P}\left(\mu - 1.96\sigma \leq Y \leq \mu + 1.96\sigma\right) = 0.95.$$

[1] We typically use `randn('state',100)` and `rand('state',100)`.

This result is often expressed in terms like "95% of normal samples lie within two standard deviations of the mean". The Central Limit Theorem (Theorem 1) tells us that, for large M, our sample mean a_M in (2.1) is approximately $N(a, b^2/M)$. So, using (2.4), it is approximately true that

$$\mathbb{P}\left(\left|\frac{a_M - a}{b/\sqrt{M}}\right| \leq 1.96\right) = 0.95.$$

This may be rearranged to

$$\mathbb{P}\left(a_M - 1.96\frac{b}{\sqrt{M}} \leq a \leq a_M + 1.96\frac{b}{\sqrt{M}}\right) = 0.95. \tag{2.5}$$

In words, with probability of approximately 0.95, the true mean will lie in the interval $[a_M - 1.96b/\sqrt{M}, a_M + 1.96b/\sqrt{M}]$. Since we do not generally know the true variance, it is typical to use the sample variance, and regard

$$\left[a_M - 1.96\frac{b_M}{\sqrt{M}}, a_M + 1.96\frac{b_M}{\sqrt{M}}\right] \tag{2.6}$$

as an approximate *95% confidence interval* for a. We may interpret this as telling us

> if we repeat the Monte Carlo simulation a large number of times, approximately 95% of the computed sample means would lie in the interval (2.6).

Looking back over how (2.6) was obtained, we can see that the confidence interval is approximate for two main reasons.

(1) The Central Limit Theorem is relevant to the limit $M \to \infty$.

(2) The exact variance b^2 has been replaced by the sample variance b_M^2.

Computational Example In Figure 2.1 we use Monte Carlo to estimate $\mathbb{E}[X]$, where X has the form $e^Z/\sqrt{e}$, with $Z \sim N(0,1)$. In this case we know that $\mathbb{E}[X] = 1$; see Exercise 2.4. We used sample sizes of $M = 10^2$, 10^3, 10^4, 10^5, 10^6 and 10^7. For each sample size, we plot the sample mean, a_M, with a cross. The vertical lines, or *error bars*, indicate the extent of the 95% confidence intervals (2.6). The error bars for $M = 10^6$ and $M = 10^7$ are virtually invisible on the scale used. Note that the horizontal axis is logarithmically scaled; this is visually convenient, but tends to hide the fact that the improvments in accuracy are coming with a huge increase in computational effort. The confidence interval with $M = 10^7$ samples had width $2 \times 1.96\frac{b_M}{\sqrt{M}}$ of 1.6×10^{-4}. It follows that to get, say, an accuracy of 10^{-6} with this approach would require a sample size of around $M = 2 \times 10^{11}$; see Exercise 2.6. $\diamondsuit$

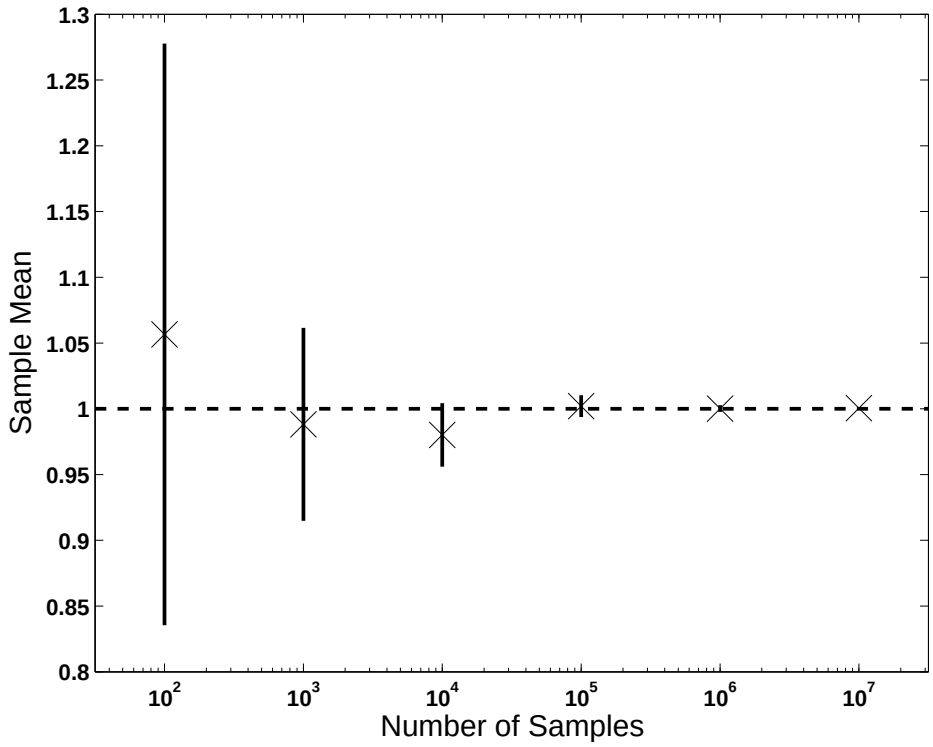


Figure 2.1: Monte Carlo approximations to $\mathbb{E}[X]$, where X has the form $e^Z/\sqrt{e}$, with $Z \sim N(0,1)$. Crosses show the sample means and vertical lines show the extent of the computed 95% confidence intervals. Horizontal dashed line is at the value of the true mean, $\mathbb{E}[X] = 1$.

Computational Example As a variation on the Monte Carlo experiment in Figure 2.1, we fixed $M = 10^4$ and, over many repetitions (with different pseudo-random numbers), counted the frequency with which the approximate 95% confidence interval (2.6) contained the true expected value $\mathbb{E}[X] = 1$. With 10^2 repetitions, the confidence interval contained the true expected value with a frequency 0.9400. With 10^3 and 10^4 repetitions the frequency rose to 0.9450 and 0.9476, respectively. This agrees well with 0.95 probability that applies as $M \to \infty$. $\diamondsuit$

It is, of course, possible to compute intervals for any desired level of confidence. Generally, if we require γ % confidence then we find α such that, for $Z \sim N(0,1)$, $\mathbb{P}(|Z| \leq \alpha) = \gamma/100$. The arguments above then lead to the interval

$$\left[a_M - \alpha \frac{b_M}{\sqrt{M}}, a_M + \alpha \frac{b_M}{\sqrt{M}} \right]. \tag{2.7}$$

Exercise 2.3 shows how α can be found in terms of the inverse cumulative distribution function for Z. For a 99% confidence interval, $\alpha = 2.58$.

Regarding a confidence interval as a measure of accuracy, there are two take-home messages from this section.

(a) The accuracy of the sample mean is proportional to the inverse square root of the number of samples. For example, to get one more digit of accuracy requires 100 times as many samples. For this reason, Monte Carlo simulation is typically only capable of producing fairly low accuracy solutions—one or two digits being common for realistic problems.

(b) The accuracy is proportional to the square root of the variance. With fixed M, we will get less accuracy for a random variable of high variance than for a random variable of low variance. This is perfectly reasonable. In some cases it is possible to ameliorate this effect by finding a second random variable Y such that $\mathbb{E}[Y] = \mathbb{E}[X]$, but $\mathsf{var}[Y] < \mathsf{var}[X]$. In this case, applying Monte Carlo to Y provides a more effective means to estimate $\mathbb{E}[X]$. This approach is known as *variance reduction*. Its usefulness varies dramatically between problems, but it can be very impressive when information about the nature of X permits the construction of an appropriate Y.

2.4 Kernel Density Estimation

We may use the "frequency" interpretation of probability mentioned in Section 1.3 as a means to reconstruct the distribution of X from a collection of samples. In the case where X is a discrete random variable taking possible values $\{x_i\}$, we simply use

$$\frac{\text{number of samples taking the value } x_i}{\text{total number of samples}}$$

as an approximation to $p_i := \mathbb{P}(X = x_i)$. For a continuous random variable, we choose a sufficiently large interval along the x axis and divide it into subintervals, or *bins*. In the simple case where we use S equally sized bins of width δx, we may write the overall interval as $[\alpha_1, \alpha_{S+1}]$ and let the ith bin cover $[\alpha_i, \alpha_{i+1}]$, for $1 \leq i \leq S$. Here, $\alpha_i = \alpha_1 + (i-1)\delta x$. We then approximate the density function $f(x)$ by a function that is piecewise constant within each bin. Approximating the area under a curve by the area of the rectangle based on the left-hand function value, as illustrated in Figure 2.2, we have

$$\mathbb{P}\left(X \in [\alpha_i, \alpha_{i+1}]\right) = \int_{\alpha_i}^{\alpha_{i+1}} f(x)\,dx \approx \delta x f(\alpha_i). \tag{2.8}$$

Then using sample frequency to approximate probability we obtain

$$f(\alpha_i) \approx \frac{1}{\delta x} \frac{\text{number of samples in } [\alpha_i, \alpha_{i+1}]}{\text{total number of samples}}. \tag{2.9}$$

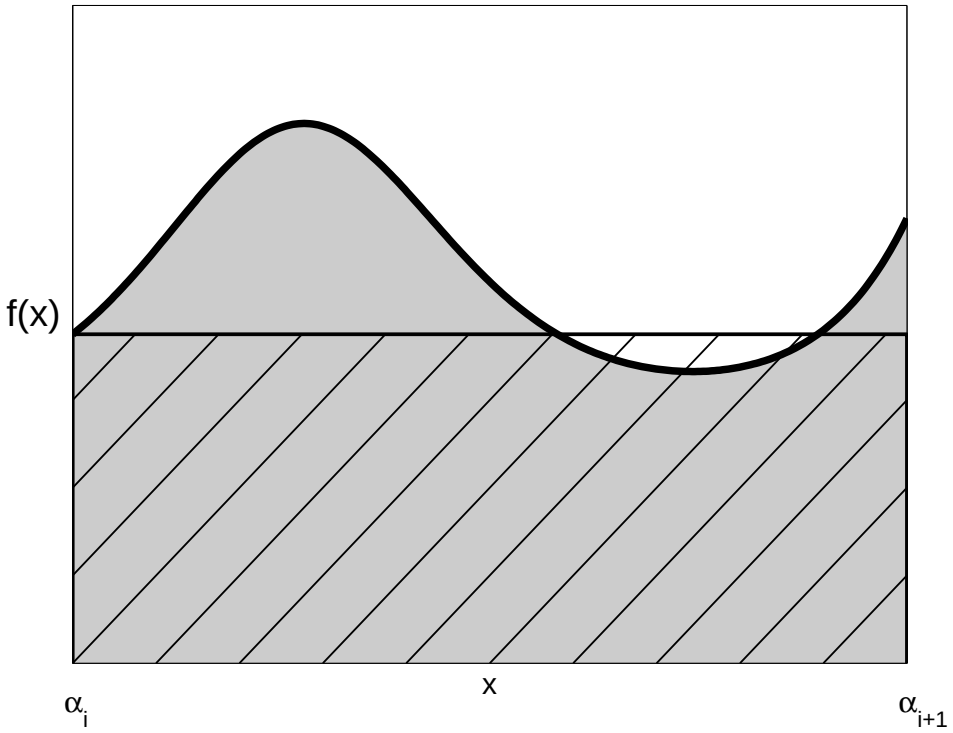


Figure 2.2: The area under the curve (shaded) can be approximated by the area of a rectangle whose height is given by the left-endpoint function value (hashed). This illustrates (2.8).

Samples that lie outside the interval $[\alpha_1, \alpha_{S+1}]$ could be dealt with by, for example, putting them into their nearest bin.

Computational Example Figure 2.3 applies the frequency approximation (2.9) in the case where X has the form

$$\frac{\sum_{i=1}^{M} Z_i - Ma}{\sqrt{M}b}, \tag{2.10}$$

where the $Z_i = e^{Y_i}$ and the Y_i are i.i.d. Bernoulli random variables with parameter $p = 0.25$. Here $a = 1 + (e-1)p$ and $b = \sqrt{p(1-p)(e-1)^2}$, so that a and b^2 are the mean and variance of Z_i, respectively; see Exercise 2.8. We generated 10^4 samples from the distribution of X and used bins $[-4, -3.5]$, $[-3.5, -3]$, ..., $[3.5, 4]$. The Central Limit Theorem (Theorem 1) shows that, for large M, X should appear to be $N(0, 1)$. To confirm this we have superimposed the $N(0, 1)$ density as a dashed line. $\diamondsuit$

The task of recovering the density function from a set of samples comes under the general heading of *kernel density estimation*.

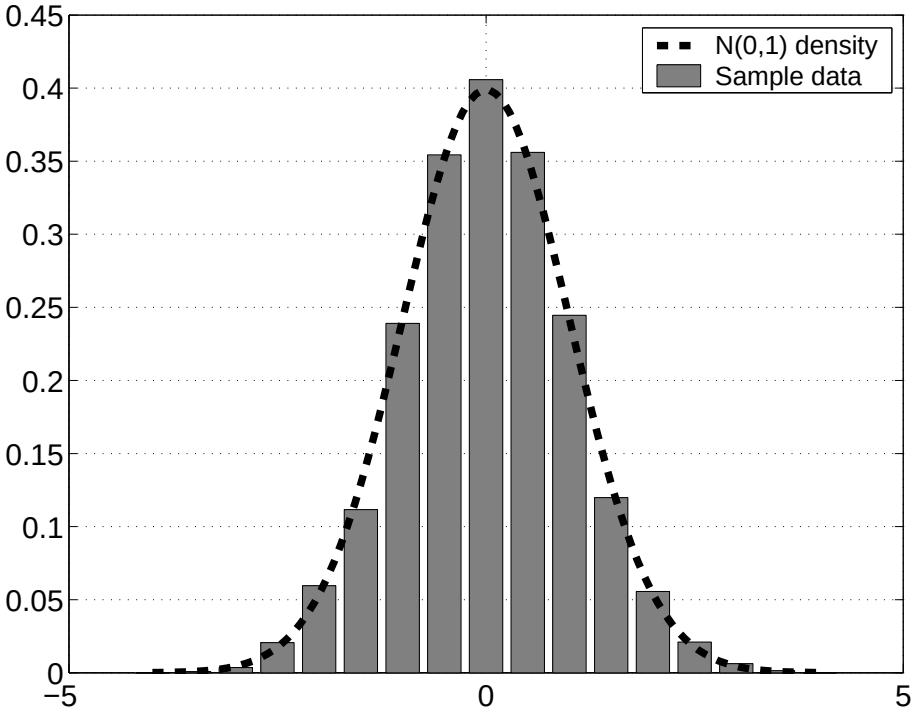


Figure 2.3: Histogram giving density function estimates (2.9) based on samples of (2.10). Dashed line superimposes the $N(0,1)$ density function.

2.5 Notes and References

A great deal more can be said about the theory and practice of designing and implementing computer simulation experiments. Informative texts are [49] and [55]. In particular, those references give recommendations for choosing the bin width as a function of the sample size in kernel density estimation. An accessible, computationally-oriented introduction to probability theory is provided by [16]. The comprehensive reference [21] focusses on issues in mathematical finance.

The pLab website at

`http://random.mat.sbg.ac.at/`

gives information on random number generation, and has links to software in a variety of computing languages.

Cleve's Corner articles "Normal Behavior," Spring 2001, and "Random Thoughts," Fall 1995, which are downloadable from

`http://www.mathworks.com/company/newsletter/clevescorner/cleve_toc.shtml`

look at the design and implementation of MATLAB's pseudo-random number generators, as does Chapter 9 of that author's lively text [48].

A readable article that discusses random number generators based on resistor or semiconductor noise is *Hardware random number generators*, by Robert Davies, which can be found at `http://www.robertnz.net/hwrng.htm`.

The Monte Carlo method dates back to the 1940s when John von Neumann, Stanislaw Ulam and Nicholas Metropolis worked on nuclear weapon projects in the Los Alamos National Laboratory. They chose the name in homage to the location of famous gambling casinos.

Exercises

2.1. ⋆ We say that a random variable Y is *unbiased* as an estimator for a random variable X if $\mathbb{E}[Y] = \mathbb{E}[X]$. Otherwise, we say Y is *biased*. Show that a_M in (2.1) is an unbiased estimator of X.

2.2. ⋆⋆ Following on from Exercise 2.1 show that $(1/M)\sum_{i=1}^{M}(\xi_i - a_M)^2$ is a biased estimator of $(X - \mathbb{E}[X])^2$, and also show that changing to b_M^2 in (2.2) produces an unbiased estimator.

2.3. ⋆⋆ The *N(0,1) distribution function*, $N(x)$, is defined as $N(x) := \mathbb{P}(Z \leq x)$ for $Z \sim N(0,1)$; that is,

$$N(x) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}s^2}\, ds.$$

The *error function*, $\mathrm{erf}(x)$, is defined as

$$\mathrm{erf}(x) := \frac{2}{\sqrt{\pi}} \int_{0}^{x} e^{-t^2}\, dt.$$

Show that $N(x)$ and $\mathrm{erf}(x)$ are related by

$$N(x) = \frac{1 + \mathrm{erf}\left(\frac{x}{\sqrt{2}}\right)}{2}.$$

Now let erfinv denote the inverse error function; so $\mathrm{erfinv}(x) = y$ means $\mathrm{erf}(y) = x$. Show that α in (2.7) can be computed via erfinv. (Note that erf and erfinv are available in MATLAB; type `help erf` or `help erfinv`.)

2.4. ⋆⋆ Using (1.16), show that $\mathbb{E}[e^Z] = \sqrt{e}$ for $Z \sim N(0,1)$.

2.5. ⋆ Given access to MATLAB's `rand`, which computes pseudo-random samples from the uniform distribution over $(0,1)$, explain how you would compute samples from a Bernouli distribution with parameter p.

2.6. ⋆⋆ In Figure 2.1, with $M = 10^6$ samples the confidence interval has width $2 \times 1.96 \frac{b_M}{\sqrt{M}} = 1.6 \times 10^{-4}$. Show that to get an accuracy of 10^{-6} with this approach would require a sample size of around $M = 2 \times 10^{11}$.

```
% CH02    Frequency with which conf interval contains exact mean.
%         Exact mean is known to equal 1.

randn('state',100);
se = sqrt(exp(1));

M =10^4;
repeats = [10^2;10^3;10^4];
len = length(repeats);
m = zeros(len,1);
s_mean = zeros(len,1);
s_std = zeros(len,1);

for i = 1:len
  reps = repeats(i);
  hits = 0;
  for k = 1:reps
      x = exp(randn(M,1))/se;   % samples
      s_mean = mean(x);         % sample mean
      s_std = std(x);           % sample standard deviation
      ebar = 1.96*s_std/sqrt(M);
      clower = s_mean-ebar;     % left endpoint of C.I.
      cupper = s_mean+ebar;     % right endpoint of C.I.
      if (s_mean-ebar<1 & s_mean+ebar>1)
        hits = hits+1;          % increment if successful
      end
  end
  hitfreq(i) = hits/reps;  % required frequency
end

disp('frequencies = '), disp(hitfreq)
```

Figure 2.4: Program of Chapter 2: `ch02.m`.

2.7. $\star$ Extend the arguments used to derive (2.6) in order to derive (2.7).

2.8. $\star$ Let $Z = e^Y$, where Y is a Bernoulli random variable with parameter p. Show that $\mathbb{E}[Z] = 1 + (e-1)p$ and $\mathsf{var}[Z] = p(1-p)(e-1)^2$.

2.6 Program of Chapter 2 and Walkthrough

The program `ch02`, listed in Figure 2.4, performs the computational experiment described on page 23. Here, the line `x = exp(randn(M,1))/se` computes the required samples, with `clower` and `cupper` giving the left and right endpoints of the corresponding approximate 95% confidence interval. The output from the pogram is

```
>> ch02
frequencies =
    0.9400    0.9450    0.9476
```

Programming Exercises

P2.1. Adapt `ch02.m` in order to produce a picture of the type shown in Figure 2.1.

P2.2. Illustrate the Central Limit Theorem in the style of Figure 2.3 when Z_i in (2.10) has the form $Z_i = \sin(2\pi Y_i)$, where the Y_i are i.i.d. with uniform distribution over $(0, 1)$. (Note that Y_i can be sampled with MATLAB's `rand`.)

Quotes:

Simulation is best regarded as mathematical experimentation ...
Brian D. Ripley, [55].

There is a frequent and apparently paradoxical need
to use a computer to generate random numbers.
In modeling a stochastic system it is necessary,
to include a source of noise,
but computers are (hopefully) not noisy.
Neil Gershenfeld, in The Nature of Mathematical Modeling, Cambridge University Press, 1999.

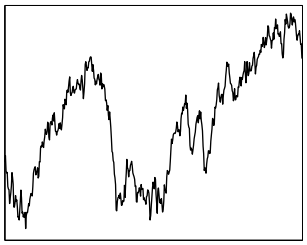
Life is a school of probability.
Walter Bagehot, in The World of Mathematics, J.R. Newman (ed.), Simon and Schuster, New York, 1956.

The generation of random numbers is too important to be left to chance.
Robert R. Coveyou,
source `http://math.furman.edu/~mwoodard/mquot.html`

Isn't it extraordinary how difficult it is to get a sample really random?
Every possible precaution, as it may seem, sometimes fails to protect one.
I remember Greenwood telling me that,
in some experiments done by drawing different coloured counters from a bag,
there seemed to be a bias against one particular colour.
On testing, they concluded that this colour had given the
counters a slightly greasy surface,
so that it tended to escape the sampler's fingers.
Yule, in a letter to Maurice Kendall
source `http://www.mathcs.carleton.edu/probweb/probweb.html`

Chapter 3

Brownian Motion



Outline

- The three defining properties
- Simulating a Brownian path
- Scaling property
- Non-differentiability and unbounded variation
- Fourier series viewpoint
- White noise interpretation

3.1 Motivation

Brownian motion is at the heart of the theory of stochastic differential equations. Its name comes from the Scottish botanist Robert Brown who, in around 1827, reported experimental observations involving the erratic behaviour of pollen grains that were bombarded by (relatively small and effectively invisible) water molecules. A mathematical theory for Brownian motion has since been developed, with famous names such as Albert Einstein and Norbert Wiener making significant contributions. Here we will introduce some of the basic concepts and illustrate them with computational examples.

3.2 Defining Brownian Motion

We begin with the classical definition that sets up Brownian motion as a *stochastic process*, that is, a random variable that changes with time.

Definition A stochastic process $W(t)$ for $t \in [0,T]$ is called *Brownian motion*, or a *standard Wiener process*, over $[0,T]$ if it satisfies the following three conditions.

C1 $W(0) = 0$.

C2 For $0 \le s < t \le T$ the random variable given by the increment $W(t) - W(s)$ is $N(0, t-s)$; equivalently, $W(t) - W(s) \sim \sqrt{t-s}\, N(0,1)$.

C3 For $0 \le s \le t \le u \le v \le T$ the increments $W(t) - W(s)$ and $W(v) - W(u)$ are independent.

$\diamondsuit$

Let's examine these conditions. The definition tells us that, at any time t, $W(t)$ is a random variable. Condition C1 says that initially, at time zero, W is always equal to zero. (More precisely, we should say that $W(0)$ takes the constant value zero with probability one.) Condition C2 shows how W evolves between time s and some later time t. The random variable describing this evolution depends only upon the length, $t-s$, of the time increment; we say that W has *stationary increments*. More precisely, the increment $W(t) - W(s)$ has a normal distribution with mean zero and variance $t - s$. Finally, condition C3 tells us that the evolution of W over $[s,t]$ is independent of its evolution over any other non-overlapping time interval $[u,v]$; we say that W has *independent increments*.

Taking $s = 0$ in C2 and using C1, we note that $W(t) = W(t) - W(0) \sim N(0,t)$, and hence, in particular,

$$\mathbb{E}\left[W(t)\right] = 0 \qquad \text{and} \qquad \mathbb{E}\left[W(t)^2\right] = t. \tag{3.1}$$

Also, for any $t \ge s \ge 0$, we have, using additivity of expected values (1.14)

$$\begin{aligned} \mathbb{E}\left[W(t)W(s)\right] &= \mathbb{E}\left[(W(s) + W(t) - W(s))W(s)\right] \\ &= \mathbb{E}\left[W(s)^2\right] + \mathbb{E}\left[(W(t) - W(s))W(s)\right]. \end{aligned} \tag{3.2}$$

By C3, $W(t) - W(s)$ and $W(s)$ are independent, so, using (1.28) and C2, we have $\mathbb{E}[(W(t) - W(s))W(s)] = \mathbb{E}[W(t) - W(s)]\mathbb{E}[W(s)] = 0 \times 0 = 0$. So, in (3.2) it follows that the *autocorellation function* for Brownian motion satisfies

$$\mathbb{E}\left[W(t)W(s)\right] = \min\left(s,t\right), \quad \text{for all } s,t \ge 0. \tag{3.3}$$

3.3 Discretized Brownian Motion

For computer simulations, it is useful to consider *discretized Brownian motion*, where $W(t)$ is specified at discrete t values. The idea behind *discretization*, which will be central to our treatment of numerical methods, is to simplify from

a continuous time interval $[0, T]$ to a discrete set of times. We will let $\delta t = T/L$ for some positive integer L, and let W_i denote $W(t_i)$ with $t_i = i\delta t$. Thus we consider $W(t)$ ony at the equally spaced points $\{t_i\}$. Condition C1 says $W_0 = 0$ and conditions C2 and C3 tell us that

$$W_{i+1} = W_i + \delta W_i, \quad i = 0, 1, 2, \ldots, L-1, \tag{3.4}$$

where each δW_i is an independent random variable of the form $\sqrt{\delta t}N(0, 1)$.

A simple pseudo-code description of (3.4) is

δt = T/L
W(0) = 0
for i = 0 to L-1
    compute a N(0,1) sample ξ_i
    W(i+1) = W(i) + $\sqrt{\delta t}\xi_i$
end

We refer to a sequence computed in this way as a *discretized Brownian path*. We emphasize that this is not an approximate path; the word "discretized" refers to the fact that an exact path is sampled discrete points in time

Figure 3.1 shows an example of a discretized Brownian path computed over $[0, 1]$ with $L = 500$. MATLAB's random number generator `randn` was used to generate $N(0, 1)$ samples. In the figure we have used a solid line to join the discrete data points W_j. This "join-the-dots" technique is helpful visually, but it is important to keep in mind that the discretized Brownian path takes values only at a discrete sequence of points in time.

The "join-the-dots" picture in (3.1) suggests that a discretized Brownian path will converge to a limit curve as $\delta t \to 0$ that is a continuous, but not very smooth. Both these properties can be made precise and rigorous. In fact, continuity of Brownian paths should actually accompany conditions C1, C2 and C3 as a technical addendum in the definition of Brownian motion, in order to make $W(t)$ unique. We will investigate the lack of smoothness issue in section 3.6, where we consider differentiability and variation.

3.4 Filling in a Brownian path

In some circumstances, we compute a discretized Brownian path $\{t_i, W_i\}_{i=0}^L$ and then wish to *refine* the discretization, that is, to compute values for the path at one or more times in between the set $\{t_i\}_{i=0}^L$. To be specific, suppose we need a new value $W_{i+\frac{1}{2}}$, to represent the path at time $t_{i+\frac{1}{2}} := \frac{1}{2}(t_i + t_{i+1})$. This raises the following question:

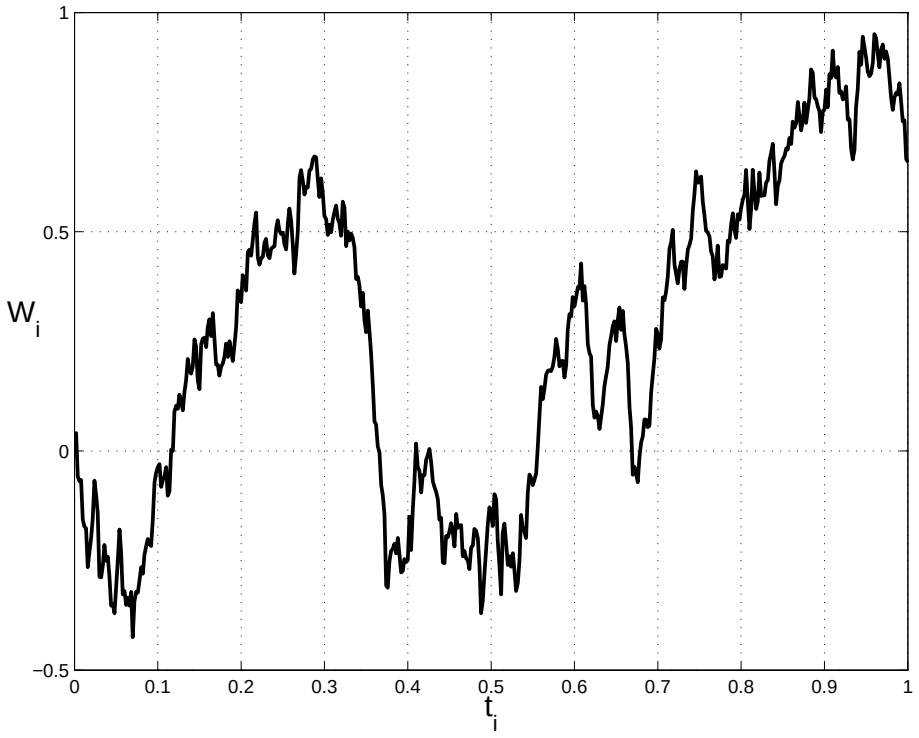


Figure 3.1: Discretized Brownian path.

> given $W(t_i)$ and $W(t_{i+1})$, how do we define a random variable $W(t_{i+\frac{1}{2}})$ in such a way that the three defining properties C1, C2 and C3 are respected?

Condition C1 is not an issue here, as we are not altering the initial value, but we must deal carefully with C2 and C3.

An obvious first try is to take the average of the neighboring values; so $W(t_{i+\frac{1}{2}}) = \frac{1}{2}(W(t_i) + W(t_{i+1}))$. However, by C2, this leads to increments

$$W(t_{i+1}) - W(t_{i+\frac{1}{2}}) = \tfrac{1}{2}\left(W(t_{i+1}) - W(t_i)\right) \sim \tfrac{1}{2}N(0, \delta t) \sim \mathsf{N}(0, \tfrac{1}{4}\delta t),$$

and similarly $W(t_{i+\frac{1}{2}}) - W(t_i) \sim \mathsf{N}(0, \frac{1}{4}\delta t)$, whereas in order to preserve C2 on the refined mesh we require these increments to be $N(0, \frac{1}{2}\delta t)$. We know from section 1.7 that the normal distribution is preserved under addition, and that variances add. This suggests taking

$$W(t_{i+\frac{1}{2}}) = \tfrac{1}{2}\left(W(t_i) + W(t_{i+1})\right) + Y_{i+\frac{1}{2}}, \quad \text{where } Y_{i+\frac{1}{2}} \sim \mathsf{N}(0, \tfrac{1}{4}\delta t), \tag{3.5}$$

with $Y_{i+\frac{1}{2}}$ independent of all other random variables used to create the path. This gives

$$W(t_{i+1}) - W(t_{i+\frac{1}{2}}) \sim \mathsf{N}(0, \tfrac{1}{2}\delta t) \quad \text{and} \quad W(t_{i+\frac{1}{2}}) - W(t_i) \sim \mathsf{N}(0, \tfrac{1}{2}\delta t),$$

as required for C2. To respect C3, we must ensure that the new increments $W(t_{i+1})-W(t_{i+\frac{1}{2}})$ and $W(t_{i+\frac{1}{2}})-W(t_i)$ are independent. Since both are normally distributed, from section 1.7 this reduces to showing that the expected value of their product is the product of their expected values. Using (1.14), $\mathbb{E}[(W(t_{i+1}) - W(t_{i+\frac{1}{2}}))(W(t_{i+\frac{1}{2}}) - W(t_i))]$ has the form

$$\mathbb{E}\left[\left(\frac{W(t_{i+1}) - W(t_i)}{2} - Y_{i+\frac{1}{2}}\right)\left(\frac{W(t_{i+1}) - W(t_i)}{2} + Y_{i+\frac{1}{2}}\right)\right].$$

This simplifies to

$$\begin{aligned}\mathbb{E}\left[\left(\frac{W(t_{i+1}) - W(t_i)}{2}\right)^2 - Y_{i+\frac{1}{2}}^2\right] &= \mathbb{E}\left[\left(\frac{W(t_{i+1}) - W(t_i)}{2}\right)^2\right] - \mathbb{E}\left[Y_{i+\frac{1}{2}}^2\right] \\ &= \frac{\delta t}{4} - \frac{\delta t}{4} \\ &= 0,\end{aligned}$$

as required. It follows that (3.5) generates a $W(t_{i+\frac{1}{2}})$ that preserves the three defining properties of Brownian motion.

Computationally, this implies that setting

$$W_{i+\frac{1}{2}} = \tfrac{1}{2}(W_i + W_{i+1}) + \tfrac{1}{2}\sqrt{\delta t}\,\xi_i, \quad \text{where } \xi_i \text{ is an independent } N(0,1) \text{ sample,} \tag{3.6}$$

allows us to "fill in" a discretized Brownian path from resolution δt to resolution $\frac{1}{2}\delta t$. Figure 3.2 illustrates this procedure.

Exercise 3.2 asks you to generalize this procedure so as to get a value for $W(\alpha t_i + (1-\alpha)t_{i+1})$ for some fixed $\alpha \in (0,1)$.

3.5 A Scaling Property

We now illustrate a remarkable scaling property. If $W(t)$ is an example of Brownian motion over $[0,T]$ then, for any fixed $c > 0$,

$$V(t) := \frac{1}{c}W(c^2 t) \tag{3.7}$$

is an example of Brownian motion over $[0, c^2T]$. To confirm this we just need to check the three defining conditions. We have $V(0) = W(0) = 0$, so C1 is satisfied. Next, for $s < t$,

$$V(t) - V(s) = \frac{1}{c}\left(W(c^2t) - W(c^2s)\right) \sim \frac{1}{c}N(0, c^2(t-s)) \sim N(0, t-s),$$

confirming C2. Finally, given $s \le t \le u \le v$,

$$\begin{aligned}V(t) - V(s) &= \frac{1}{c}\left(W(c^2t) - W(c^2s)\right), \\ V(v) - V(u) &= \frac{1}{c}\left(W(c^2v) - W(c^2u)\right).\end{aligned}$$

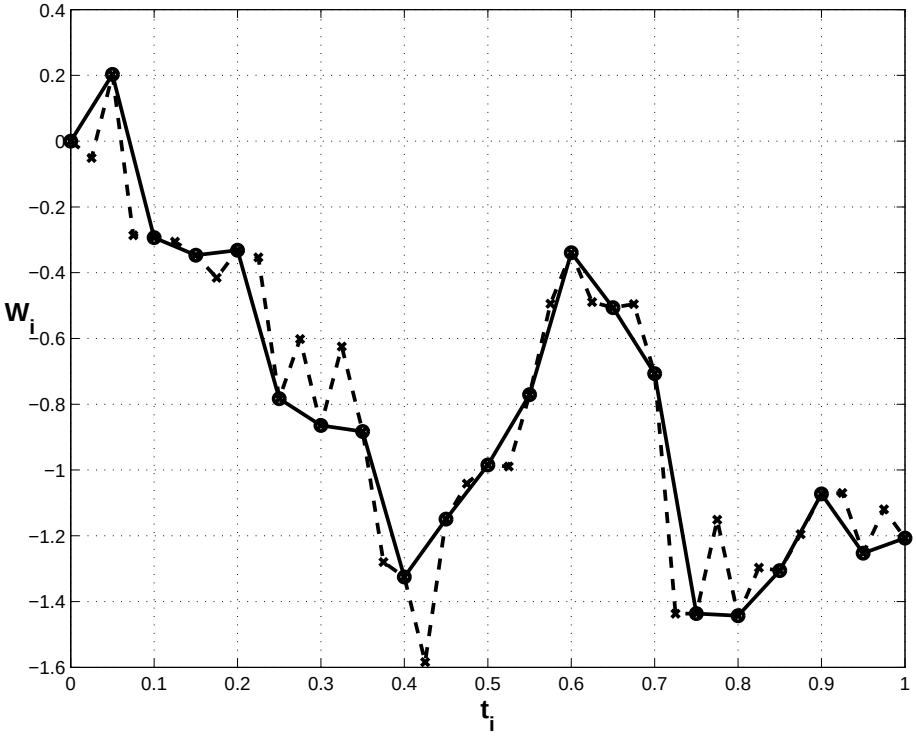


Figure 3.2: Solid lines join the points along a discrete Brownian path. Dashed lines join the points along a refined version, based on (3.6).

Since $W(c^2t) - W(c^2s)$ and $W(c^2v) - W(c^2u)$ are independent, it follows from Exercise 1.12 that $V(t) - V(s)$ and $V(v) - V(u)$ are independent, giving C3.

This result shows that Brownian motion has a fractal-like 'self-similarity'. If you zoom in with appropriate scaling in the x and y directions, you will see the same type of behavior.

To illustrate this effect, the upper plot in Figure 3.3 shows a discrete Brownian path, computed with a suitably small δt, and the lower plot zooms in on the 'first 100th' of the same path. More precisely, the lower plot shows the path over $[0, T/100]$ multiplied by a factor 10—this is consistent with taking $c = 1/10$ in (3.7). Visually, the two paths seem to be displaying the same type of 'jagged' behavior.

3.6 Non-differentiability and Unboundedness of Variation

Without worrying too much about what it means to differentiate a stochastic process, we could argue loosely that differentiability of $W(t)$ should be related to

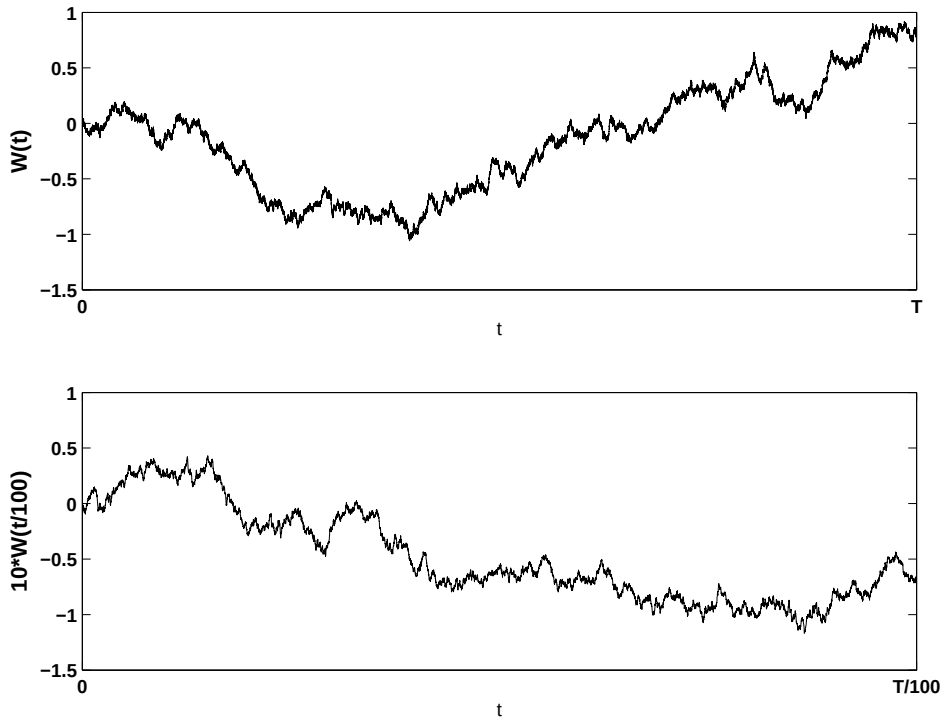


Figure 3.3: Upper plot: a discrete Brownian path, $W(t_i)$. Lower plot: a 'zoomed-in' plot of $10\,W(t_i/100)$.

the existence of a limit for $(W(t+h)-W(t))/h$ as $h \to 0$. To be specific we will take $t = 0$. Since $W(0) = 0$, boundedness of a derivative is related to the ratio $|W(h)|/h$ in the limit $h \to 0$. Now, let $h = 1/n^4$, where n is large. Since $W(t)$ has the same distribution as $V(t)$ in (3.7), we have

$$\mathbb{P}\left(\frac{|W(1/n^4)|}{1/n^4} > n\right) = \mathbb{P}\left(\frac{|V(1/n^4)|}{1/n^4} > n\right).$$

Setting $c = n^2$ in (3.7) we find

$$\mathbb{P}\left(\frac{|V(1/n^4)|}{1/n^4} > n\right) = \mathbb{P}\left(\frac{|W(1)|}{1/n^2} > n\right) = \mathbb{P}\left(|W(1)| > \frac{1}{n}\right).$$

Putting this together, we have shown that

$$\mathbb{P}\left(\frac{|W(1/n^4)|}{1/n^4} > n\right) = \mathbb{P}\left(|W(1)| > \frac{1}{n}\right).$$

Since $W(1)$ is $N(0,1)$ the term on the right-hand side is the probability that a $N(0,1)$ random variable takes a value bigger than $1/n$ in modulus. This probability clearly tends to 1 as $n \to \infty$. By considering the term on the left-hand

side, we conclude that, with probability 1, $W(t)$ is not differentiable at $t = 0$. A similar argument can be used to show that $W(t)$ is nowhere differentiable, with probability 1.

Another way of examining the roughness of Brownian motion is to consider its variation. Suppose, as before, that we divide $[0, T]$ into subintervals $[t_i, t_{i+1}]$, with $t_i = i\delta t$ and $\delta t = T/L$. For a continuously differentiable function, $f \in C^1[0, T]$, the Mean Value Theorem says that

$$f(t_{i+1}) - f(t_i) = \delta t f'(\theta_i), \qquad \text{for some } \theta_i \in (t_i, t_{i+1}).$$

It follows that the *variation* of f satisfies

$$\sum_{i=0}^{L-1} |f(t_{i+1}) - f(t_i)| = \delta t \sum_{i=0}^{L-1} |f'(\theta_i)| \leq L\delta t \max_{[0,T]} |f'(x)| = T \max_{[0,T]} |f'(x)|.$$

This shows that any $f \in C^1[0, T]$ has *finite variation*. To see whether Brownian motion has a similar property we use the inequality

$$\sum_{i=0}^{L-1} (W(t_{i+1}) - W(t_i))^2 = \sum_{i=0}^{L-1} \delta W_i^2 \leq \left(\max_{0 \leq i \leq L-1} |\delta W_i| \right) \sum_{i=0}^{L-1} |\delta W_i|. \tag{3.8}$$

First we examine $\sum_{i=0}^{L-1} \delta W_i^2$. We have

$$\mathbb{E}\left[\sum_{i=0}^{L-1} \delta W_i^2 \right] = \sum_{i=0}^{L-1} \mathbb{E}\left[\delta W_i^2 \right] = L\delta t = T \tag{3.9}$$

and

$$\mathbb{E}\left[\left(\sum_{i=0}^{L-1} \delta W_i^2 \right)^2 \right] = \mathbb{E}\left[\sum_{i=0}^{L-1} \sum_{j=0}^{L-1} \delta W_i^2 \delta W_j^2 \right] = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} \mathbb{E}\left[\delta W_i^2 \delta W_j^2 \right].$$

Now, when $i \neq j$, δW_i^2 and δW_j^2 are independent, so we may write

$$\begin{aligned} \mathbb{E}\left[\left(\sum_{i=0}^{L-1} \delta W_i^2 \right)^2 \right] &= \sum_{i=0, i \neq j}^{L-1} \sum_{j=0}^{L-1} \mathbb{E}\left[\delta W_i^2 \delta W_j^2 \right] + \sum_{i=0}^{L-1} \mathbb{E}\left[\delta W_i^4 \right] \\ &= \sum_{i=0, i \neq j}^{L-1} \sum_{j=0}^{L-1} \mathbb{E}\left[\delta W_i^2 \right] \mathbb{E}\left[\delta W_j^2 \right] + \sum_{i=0}^{L-1} \mathbb{E}\left[\delta W_i^4 \right] \end{aligned}$$

Using $\mathbb{E}[\delta W_i^2] = \delta t$ and $\mathbb{E}[\delta W_i^4] = 3\delta t^2$ (see Exercise 3.3), we obtain

$$\mathbb{E}\left[\left(\sum_{i=0}^{L-1} \delta W_i^2 \right)^2 \right] = L(L-1)\delta t^2 + 3L\delta t^2 = L^2 \delta t^2 - 2L\delta t^2 = T^2 + O(\delta t). \tag{3.10}$$

Combining (3.9) and (3.10), we find that

$$\mathsf{var}\left[\sum_{i=0}^{L-1}\delta W_i^2\right] := \mathbb{E}\left[\left(\sum_{i=0}^{L-1}\delta W_i^2\right)^2\right] - \left(\mathbb{E}\left[\sum_{i=0}^{L-1}\delta W_i^2\right]\right)^2 = T^2 + O(\delta t) - T^2 = O(\delta t). \tag{3.11}$$

So $\sum_{i=0}^{L-1}\delta W_i^2$ has mean T and variance of $O(\delta t)$. Hence, as $\delta t \to 0$ we would expect this random variable to converge to the constant value T[1]. On the other hand, each δW_i has mean zero and variance δt, so as $\delta t \to 0$ we would expect $\max_{1\le 0\le L-1}|\delta W_i|$ to converge to the constant value 0[2]. In order for the inequality (3.8) to hold it must therefore be true that $\sum_{i=0}^{L-1}|\delta W_i|$ is unbounded, with probability 1, as $\delta t \to 0$ and hence $L = T/\delta t \to \infty$. We thus say that Brownian motion has *infinite variation*.

Although Brownian motion is rough, it behaves smoothly in expectation. In Figure 3.4 we evaluate the function $u(W(t)) = e^{t+\frac{1}{4}W(t)}$ along 5 discretized Brownian paths. The sample average of $u(W(t))$ over 10^5 such paths is also plotted with a dashed linetype. Here, for each t_i we compute the Monte Carlo sample mean for each path. (All 95% confidence intervals had neglible widths to visual accuracy.) We see that although $u(W(t))$ is non-smooth over individual paths, its sample average appears to be smooth. In fact, direct computation shows that $\mathbb{E}[u(W(t))] = e^{(33/32)t}$; see Exercise 3.4. Alternatively, we will see in Chapter 6 that $u(W(t))$ is the solution to a certain stochastic differential equation, and Exercise 6.6 gives the expected value.

3.7 Fourier Series Representation

We know that *Fourier series* can be used to express deterministic functions as infinite series of trigonometric functions. A corresponding series expansion applies for Brownian motion, where the coefficients are random variables. The *Paley-Wiener representation* has the form

$$W(t) = Z_0\frac{t}{\sqrt{2\pi}} + \frac{2}{\sqrt{\pi}}\sum_{n=1}^{\infty} Z_n\frac{\sin(\frac{1}{2}nt)}{n}, \quad 0 \le t \le 2\pi, \tag{3.12}$$

where the $\{Z_i\}_{i\ge 0}$ are i.i.d. and $N(0,1)$. It can be shown that this series converges in a well-defined sense to Brownian motion over $[0, 2\pi]$.

In Figure 3.5 we draw pseudo-random samples for the Z_i and plot the curves arising when the infinite series in (3.12) is truncated to $\sum_{n=1}^{M}$, for $M = 1$, 2, 5, 10, 50 and 200. We see that the early terms in the series affect the overall shape, while the later terms add fine detail. The curves are, of course, smooth, but the

[1]This can be made rigorous: $\lim_{\delta t\to 0}\sum_{i=0}^{L-1}(W(t_{i+1}) - W(t_i))^2 = T$ with probability 1.
[2]This can be made rigorous: $\lim_{\delta t\to 0}\max_{0\le i\le L-1}|\delta W_i| = 0$ with probability 1.

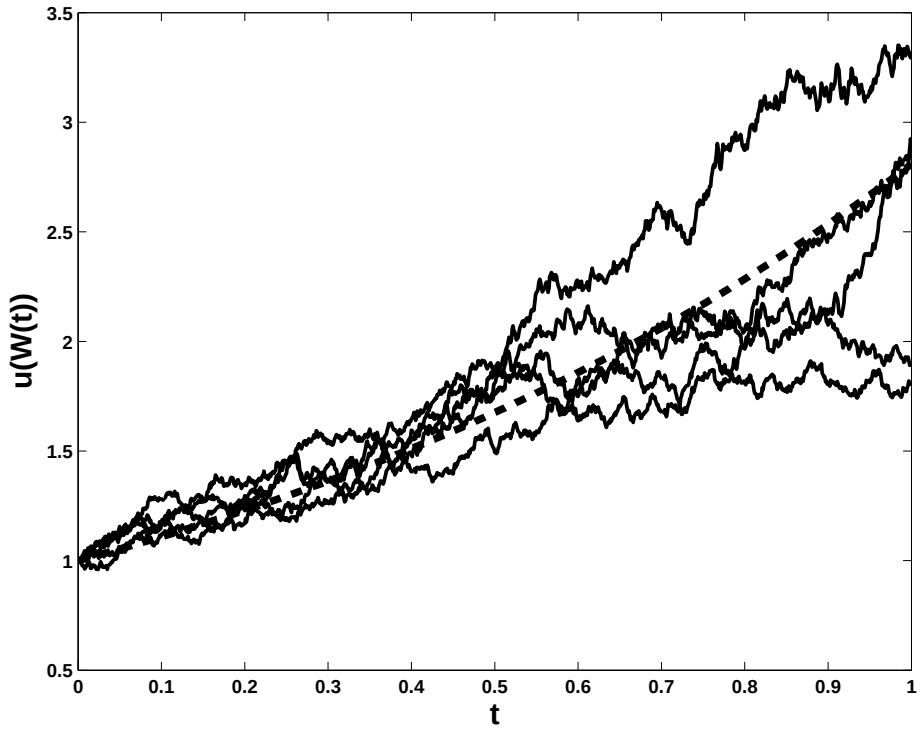


Figure 3.4: The function $e^{t+\frac{1}{4}W(t)}$ along 5 individual discretized Brownian paths (solid), and averaged over 10^5 paths (dashed).

plots are visually consistent with the fact that as $M \to \infty$ the limiting path is non-differentiable.

3.8 White Noise Interpretation

We saw in Section 3.6 that Brownian paths are not differentiable. However, many authors "pretend" that the derivative

$$\frac{dW(t)}{dt} =: \xi(t) \tag{3.13}$$

exists, and refer to it as *white noise*. Further, you may come across the "formula"

$$\mathbb{E}\left[\xi(t)\xi(s)\right] = \delta_0(s-t) \tag{3.14}$$

for the autocorrelation function, where $\delta_0(\cdot)$ is the Dirac-delta function; that is, the unit mass at 0. Here, we give a brief, heuristic discussion of this viewpoint.

For a fixed time t and a fixed time increment, h, define

$$d_h(s) := \mathbb{E}\left[\left(\frac{W(t+h)-W(t)}{h}\right)\left(\frac{W(s+h)-W(s)}{h}\right)\right]. \tag{3.15}$$

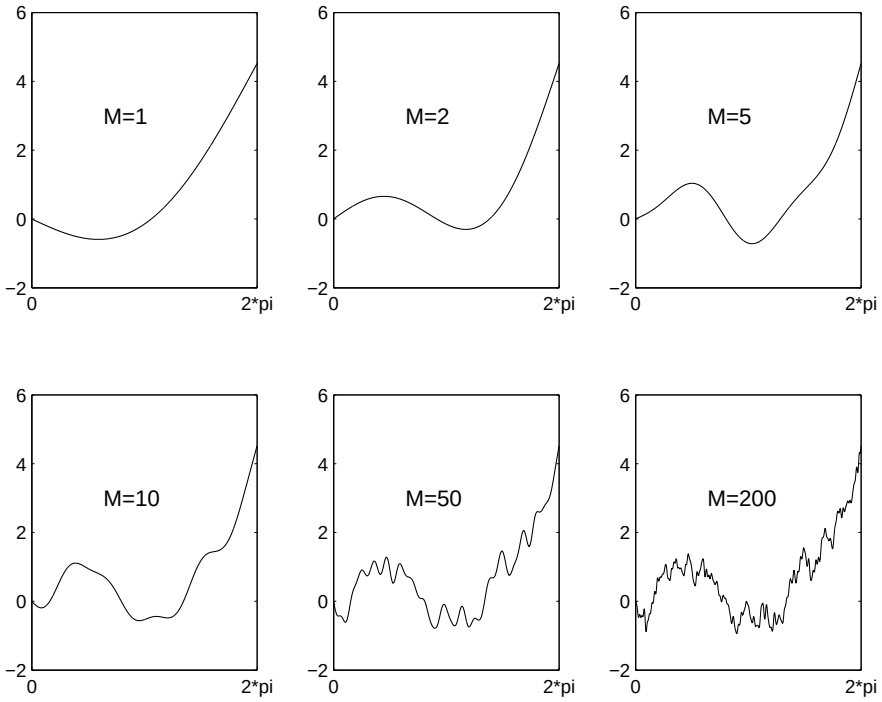


Figure 3.5: Paths based on the random Fourier series (3.12). Here the Z_i are a sequence of i.i.d. $N(0,1)$ pseudo-random numbers, and the six plots show the sum truncated after $M = 1$, 2, 5, 10, 50 and 200 sine terms.

If the expression (3.14) makes any sense, then for small h, $d_h(s)$ in (3.15) should look like $\delta_0(s-t)$. Well, using our expression (3.3) for the autocorrelation function of Brownian motion, we may expand $d_h(s)$ as

$$\begin{aligned} d_h(s) &= \frac{1}{h^2}\left(\mathbb{E}\left[W(t+h)W(s+h)\right] - \mathbb{E}\left[W(t+h)W(s)\right]\right. \\ &\quad \left. -\mathbb{E}\left[W(t)W(s+h)\right] + \mathbb{E}\left[W(t)W(s)\right]\right) \\ &= \frac{1}{h^2}\left(\min(t+h, s+h) - \min(t+h, s) - \min(t, s+h) + \min(t, s)\right) \end{aligned} \tag{3.16}$$

We see that $d_h(s)$ is piecewise linear with possible "corners" at the points $s = t-h, t, t+h$. A quick check shows that

$$d_h(t-h) = d_h(t+h) = 0 \quad \text{and} \quad d_h(t) = \frac{1}{h}.$$

A picture of $d_h(s)$ is given in Figure 3.6, and we see that $d_h(s) \to 0$ as $h \to 0$ whenever $s \neq t$, but $\int_{-\infty}^{\infty} d_h(s)ds = 1$. Hence, in some sense, $d_h(s)$ behaves like $\delta_0(s-t)$ for small h.

Generally, given any stochastic process $Y(t)$, if the autocorrelation function

$$r(s,t) := \mathbb{E}\left[X(t)X(s)\right]$$

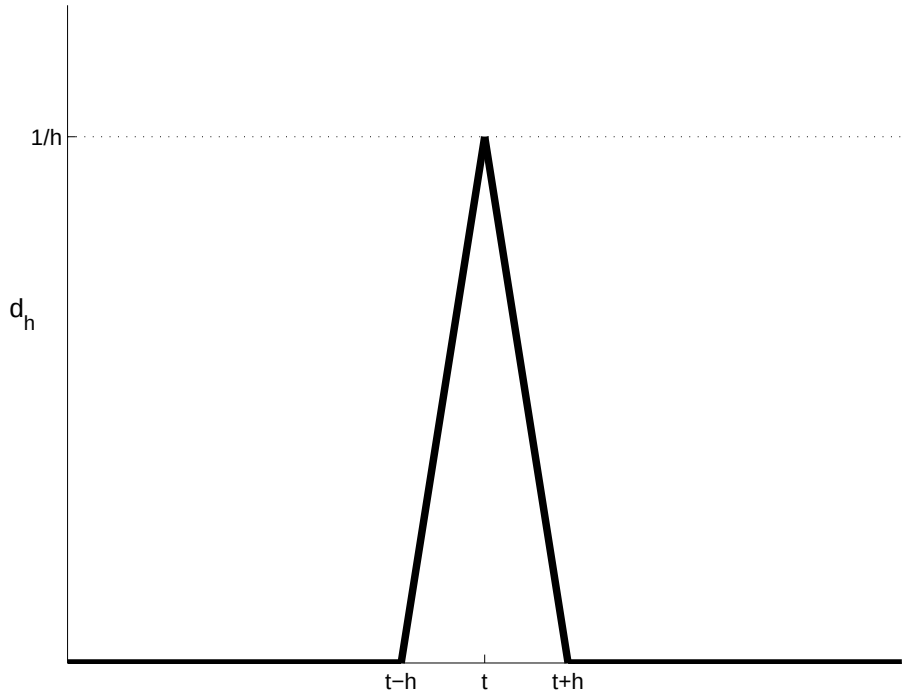


Figure 3.6: A plot of $d_h(s)$ in (3.16).

has the form $r(t,s) = c(t-s)$ for some function c, then for any frequency $\lambda \in \mathbb{R}$,

$$f(\lambda) := \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-\mathbf{i}\lambda t} c(t)\, dt,$$

is called the *spectral density* of $X(t)$. When $c(\cdot) = \delta_0(\cdot)$, we may argue that

$$f(\lambda) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-\mathbf{i}\lambda t} \delta_0(t)\, dt = \frac{1}{2\pi}.$$

Thus, for the white noise process $\xi(t)$, all frequencies have equal strength in the autocorrelation function, just as white light contains an equal mix of all colors.

3.9 Notes and References

The study of Brownian motion and related processes is a vast topic. Brownian motion is, of course, treated in the texts on stochastic differential equations that are refered to throughout this book. More specialised texts include [11, 36]. Viewpoints from the physical sciences can be found in [5, 19, 20] and a brief historical perspective appears in [25].

Norris's book [51] on discrete and continuous time Markov chains gives an alternative and very accessible means to understand Brownian motion.

The in-fill formula (3.6) is a special case of what is known as the Lévy construction of Brownian motion.

Our treatment of white noise in section 3.8 was motivated by the on-line notes of Lawrence C. Evans at

`http://math.berkeley.edu/~evans/`

There are many web-based simulators available to illustrate Brownian motion in one, two and three dimensions.

Exercises

3.1. ⋆⋆ Show that Brownian motion satisfies $\mathbb{E}|W(t) - W(s)|^2 = |t - s|$.

3.2. ⋆⋆ Generalize the formula (3.6) to the case where, given $W(t_i)$ and $W(t_{i+1})$, a value is needed for $W(t_i + \alpha\delta t)$ for some $\alpha \in (0, 1)$.

3.3. ⋆⋆ Suppose that $X \sim N(0, 1)$. We know from Exercise 1.13 that $\mathbb{E}[X] = 0$ and $\mathbb{E}[X^2] = 1$. From (1.16), the general pth moment satisfies

$$\mathbb{E}\left[X^p\right] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^p e^{-x^2/2}\, dx.$$

Show that $\mathbb{E}[X^3] = 0$ and $\mathbb{E}[X^4] = 1$. Deduce that a Brownian increment, $\delta W_i := W(t_{i+1}) - W(t_i)$, satisfies $\mathbb{E}[\delta W_i^3] = 0$ and $\mathbb{E}[\delta W_i^4] = 3\delta t^2$. Find a general expression for $\mathbb{E}[X^p]$ and hence for $\mathbb{E}[\delta W_i^p] = 0$. [You may use without proof the fact that $\int_{-\infty}^{\infty} e^{-x^2/2}\, dx = \sqrt{2\pi}$.]

3.4. ⋆⋆ Suppose that $X \sim N(0, 1)$. Show that, for $a, b \in \mathbb{R}$, $\mathbb{E}[e^{a+bX}] = e^{a+\frac{1}{2}b^2}$. Hence, deduce that $\mathbb{E}[e^{t+\frac{1}{4}W(t)}] = e^{(33/32)t}$, as observed in Figure 3.4.

3.10 Program of Chapter 3 and Walkthrough

The program `ch03`, listed in Figure 3.7, plots a discretized Brownian path in the manner of Figure 3.1. This program operates on vectors. The command `dW = sqrt(dt)*randn(1,N)` produces a one-dimensional array of independent $N(0, \delta t)$ samples, and the cumulative sum `W = cumsum(dW)` has ith component given by the sum of the first i components of `dW`.

Programming Exercises

P3.1. Alter `ch03.m` in order to illustrate the in-fill algorithm of section 3.4, as shown in Figure 3.2.

```
% CH03.M    Brownian path simulation: vectorized

clf
randn('state',100)
L = 500; T = 1; dt = T/L;

dW = sqrt(dt)*randn(1,L);   % Brownian increments
W = cumsum(dW);             % cumulative sum

plot([0:dt:T],[0,W],'r-','LineWidth',2)
xlabel('t_i','FontWeight','Bold')
ylabel('W_i','FontWeight','Bold')
grid on
```

Figure 3.7: Program of Chapter 3: `ch03.m`.

P3.2. In a similar manner to Figure 3.3, illustrate the scaling property of Brownian motion for the case $c = 1/5$.

Quotes:

The agreement of these considerations on Brownian motion with experience,
together with Planck's determination of the true molecular size from the law of radiation
convinced the sceptics,
who were quite numerous at that time (Ostwald, Mach)
of the reality of atoms.
Albert Einstein, biographical notes: in The Theory of Brownian Motion—100 Years Old, W. Gehard Pohl, The History of Science and the Cultural Integration of Europe, M. Kokowski (ed.), 2006.

For practical purposes,
Einsteins's explanation of the nature of Brownian motion
must be regarded as the beginning of
stochastic modelling of natural phenomena.
C. W. Gardiner, [19].

That a typical Brownian path has tangents nowhere is very interesting
in light of the mathematics of the eighteenth and nineteenth centuries.
Most mathematicians of that era believed
all continuous curves had to have well-defined tangents,
and there were repeated attempts to prove this.
Finally, in the middle of the nineteenth century, the German mathematician Karl Weierstrass gave an example of a continuous curve without tangents anywhere.
This example must have appeared as an aberration to most of his contempories,

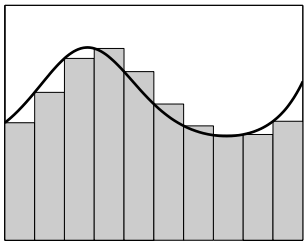
the type of example that mathematicians often call *pathological*
to indicate its unusual and unexpected nature.
But from the point of view of Brownian paths
(which, of course, were unknown at the time)
the situation is just the opposite:
the typical Brownian path has no tangents,
and the totality of paths *with* tangents are the anomalies, having probability 0.
Richard Isaac, [30].

Robert Brown was an English botanist ...
Richard Isaac, [30].

Read about Brownian motion here
[http://xanadu.math.utah.edu/java/brownianmotion/1/]
but do not believe that Brown was English (he was from Scotland).
Ray Streater, source
http://www.mth.kcl.ac.uk/~streater/Brownianmotion.html

Chapter 4

Stochastic Integrals



Outline

- Riemann sums
- Itô and Stratonovich integrals
- Properties of the Itô integral
- "$dW^2 = dt$"

4.1 Motivation

We solve differential equations by integrating. Hence, we need to know how to integrate stochastically; that is, "with respect to Brownian motion". We will do this by analogy with the Riemann sum concept that is taught for deterministic integration. Then we will explain some key properties of the resulting Itô integral.

4.2 Itô and Stratonovich Integrals

Given a suitable deterministic function, $h : \mathbb{R} \to \mathbb{R}$, the integral $\int_0^T h(t)\,dt$ can be approximated by the Riemann sum

$$\sum_{i=0}^{L-1} h(t_i)(t_{i+1} - t_i), \tag{4.1}$$

where, as in the previous chapter, $t_i = i\delta t$ with $\delta t = T/L$. Indeed, the integral $\int_0^T h(t)\,dt$ may be *defined* by taking the $\delta t \to 0$ limit in (4.1). Figure 4.1 illustrates the concept. We are approximating the area under the curve $h(t)$ by the area

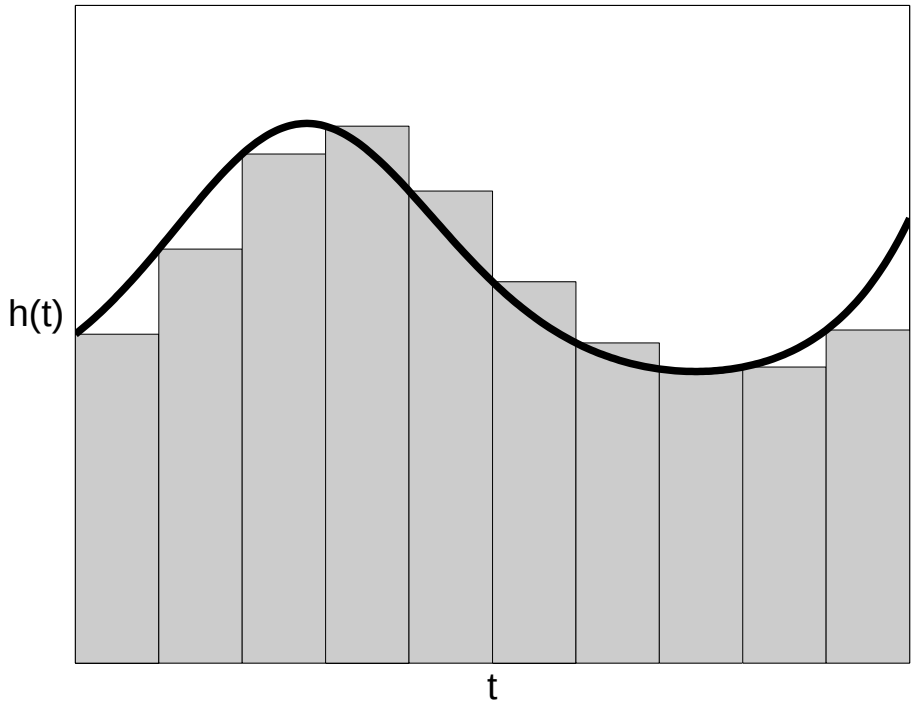


Figure 4.1: Riemann sum integral appromixation based on left-hand endpoints.

under a piecewise constant function whose height over $[t_i, t_{i+1}]$ is given by the height at the left-hand endpoint, $h(t_i)$.

In this deterministic setting, the heights used in the Riemann sum could equally well be chosen by taking the function value at any other point in the interval. For example, the midpoint version

$$\sum_{i=0}^{L-1} h(\tfrac{1}{2}(t_i + t_{i+1}))(t_{i+1} - t_i), \tag{4.2}$$

converges to the same value, $\int_0^T h(t)\,dt$, as $\delta t \to 0$. Figure 4.2 shows this alternative.

By analogy with (4.1), we may regard the sum

$$\sum_{i=0}^{L-1} h(t_i)\,(W(t_{i+1}) - W(t_i)) \tag{4.3}$$

as an approximation to a stochastic integral, $\int_0^T h(t)\,dW(t)$, with this integral being defined via the $\delta t \to 0$ limit. Here, we are thinking of integrating h with respect to Brownian motion. We note that the increments $W(t_{i+1}) - W(t_i)$ are random variables, and hence the expression (4.3) is a random variable. The

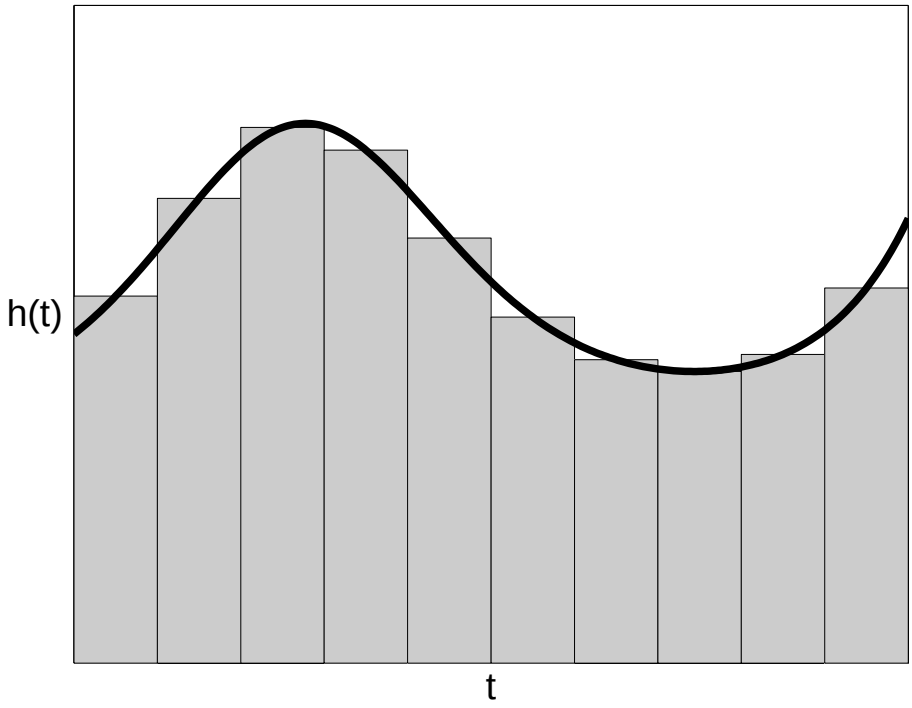


Figure 4.2: Riemann sum integral appromixation based on midpoints.

integrand, $h(t)$, may itself be a stochastic process—a random variable for each $t \in [0, T]$. In general, we must place restrictions on $h(t)$ in order for the integral to make sense. This issue will be discussed briefly in section 4.3.

Analogously to (4.2) we may use a midpoint version of the stochastic Riemann sum

$$\sum_{i=0}^{L-1} h(\tfrac{1}{2}(t_i + t_{i+1}))\,(W(t_{i+1}) - W(t_i))\,. \tag{4.4}$$

A big surprise, and the first indication that stochastic calculus is not simply a straightforward extension of deterministic calculus, is that the two Riemann sum forms (4.3) and (4.4) do not converge in any sense to the same limit. To illustrate this, we will consider the case where $h(t) \equiv W(t)$. Recalling that $\delta W_i := W(t_{i+1}) - W(t_i)$, the left-hand endpoint sum (4.3) may be written

$$\begin{aligned}\sum_{i=0}^{L-1} W(t_i)\,(W(t_{i+1}) - W(t_i)) &= \tfrac{1}{2}\sum_{i=0}^{L-1}\left(W(t_{i+1})^2 - W(t_i)^2 - (W(t_{i+1}) - W(t_i))^2\right)\\ &= \tfrac{1}{2}\sum_{i=0}^{L-1}\left(W(t_{i+1})^2 - W(t_i)^2\right) - \tfrac{1}{2}\sum_{i=0}^{L-1}\delta W_i^2.\end{aligned}$$

The first sum on the right-hand side *telescopes*—successive terms cancel leaving

only the first and last—to give $\frac{1}{2}(W(T)^2 - W(0)^2)$, which is $\frac{1}{2}W(T)^2$. We argued in section 3.6 that, as $\delta t \to 0$, $\sum_{i=0}^{L-1} \delta W_i^2$ behaves like the constant value T. It follows that if we regard $\int_0^T h(t)\, dW(t)$ as the $\delta t \to 0$ limit of (4.3) then

$$\int_0^T W(t)\, dW(t) = \tfrac{1}{2}W(T)^2 - \tfrac{1}{2}T. \tag{4.5}$$

To analyze the midpoint version with $h(t) \equiv W(t)$ in a similar way, we recall from section 3.4 that $W(\frac{1}{2}(t_i + t_{i+1}))$ can be regarded as $\frac{1}{2}(W(t_i) + W(t_{i+1})) + Z_i$, where $Z_i \sim N(0, \delta t/4)$ are i.i.d. and independent of all $W(t_i)$. Then in (4.4),

$$\sum_{i=0}^{L-1} \left(\frac{W(t_i) + W(t_{i+1})}{2} + Z_i \right) (W(t_{i+1}) - W(t_i)) = \tfrac{1}{2} \sum_{i=0}^{L-1} W(t_{i+1})^2 - W(t_i)^2 + \sum_{i=0}^{L-1} Z_i \left(W(t_{i+1}) - W(t_i) \right). \tag{4.6}$$

The first sum on the right-hand side telescopes to $\frac{1}{2}W(T)^2$. The second sum has expected value zero and variance of $O(\delta t)$; see Exercise 4.1. Hence, using the midpoint version of the Riemann sum, we conclude that

$$\int_0^T W(t)\, dW(t) = \tfrac{1}{2}W(T)^2, \quad \text{(midpoint version)}. \tag{4.7}$$

Computational Example Setting $T = 1$ and $L = 500$, so $\delta t = 1/500$, we computed a discretized Brownian path and recorded the Riemann sum (4.3) with $h(t) \equiv W(t)$. The result, `-0.2674`, differs from the limiting value $\frac{1}{2}W(1)^2 - \frac{1}{2}$ in (4.5) by an amount `0.0158`. We then filled-in this discretized Brownian path, using the technique discussed in section 3.4, and recorded the midpoint version (4.4). This gave `0.2354`, differing from the limiting value $\frac{1}{2}W(1)^2$ in (4.5) by an amount `0.0186`. ♢

We conclude from the discussion above that a Riemann sum based definition of a stochastic integral is sensitive to the point used to define the heights of the rectangles. Throughout this book we will use the notation $\int_0^T W(t)\, dW(t)$ to mean the integral defined by the $\delta t \to 0$ limit of the left-hand endpoint version (4.3). This is known as the *Itô integral*. We will use $\int_0^T W(t) \circ dW(t)$ to denote the integral defined by the $\delta t \to 0$ limit of the midpoint version (4.4). This is known as the *Stratonovich integral*. We focus almost exclusively on the Itô integral in this book. Section 5.5 gives a brief discussion of Itô versus Stratonovich in the context of SDEs.

As another example, consider the Itô integral $\int_0^T t\, dW(t)$. The approximation (4.3) may be written

$$\sum_{i=0}^{L-1} t_i \left(W(t_{i+1}) - W(t_i) \right) = \sum_{i=0}^{L-1} (t_{i+1} - \delta t) W(t_{i+1}) - t_i W(t_i)$$

$$= \sum_{i=0}^{L-1} (t_{i+1}W(t_{i+1}) - t_i W(t_i)) - \delta t \sum_{i=0}^{L-1} W(t_{i+1}).$$

The first sum telescopes to $TW(T)$ and the second sum may be regarded as a standard Riemann integral (albeit with a stochastic integrand), so we may write

$$\int_0^T t\, dW(t) = TW(T) - \int_0^T W(t)\, dt. \tag{4.8}$$

Exercise 4.3 asks you to use a similar approach to deduce that

$$\int_0^T W(t)^2\, dW(t) = \tfrac{1}{3}W(T)^3 - \int_0^T W(t)\, dt. \tag{4.9}$$

Although the simple examples $\int_0^T W(t)\, dW(t)$, $\int_0^T t\, dW(t)$ and $\int_0^T W(t)^2\, dW(t)$ are useful illustrations, they are misleading in the sense that, to an even greater extent than in the deterministic case, there are relatively few integrands for which analytical formulas can be derived for the stochastic integral.

4.3 Properties of the Itô Integral

Although we are not attempting to give a rigorous definition of the class of integrands to which Itô integration may be applied, we mention here one important restriction

> at time t, the integrand $h(t)$ must be independent of the later values $\{W(s)\}_{s>t}$ of the Brownian path.

We say that such a function h is *non-anticipative*.

Given a suitable integrand $h(t)$, and recalling the notation $\delta W_i = W(t_{i+1}) - W(t_i)$, the Riemann sum approximation $\sum_{i=0}^{L-1} h(t_i)\delta W_i$ involves the product of independent random variables $h(t_i)$ and δW_i, and hence, using (1.14) and (1.27) and recalling that $\mathbb{E}[\delta W_i] = 0$,

$$\mathbb{E}\left[\sum_{i=0}^{L-1} h(t_i)\delta W_i\right] = \sum_{i=0}^{L-1} \mathbb{E}\left[h(t_i)\delta W_i\right] = \sum_{i=0}^{L-1} \mathbb{E}[h(t_i)]\mathbb{E}[\delta W_i] = 0.$$

Since this holds for all $\delta t > 0$, we conclude that it holds for the limiting integral, and hence

$$\mathbb{E}\left[\int_0^T h(t)\, dW(t)\right] = 0. \tag{4.10}$$

The condition (4.10) is known as the *martingale property* of the Itô integral. Generally, the word martingale applies to stochastic processes whose expected

future value does not change with time. In the case of an Itô integral, for any applicable integrand the expected value is always zero. This martingale property makes the Itô integral amenable to a vast range of stochastic analysis tools.

Now we consider the expression $\mathbb{E}[(\int_0^T h(t)\,dW(t))^2]$. Using the Riemann sum approximation (4.3),

$$\mathbb{E}\left[\left(\sum_{i=0}^{L-1} h(t_i)\delta W_i\right)^2\right] = 2\sum_{i<j}^{L-1} \mathbb{E}\left[h(t_i)h(t_j)\delta W_i \delta W_j\right] + \sum_{i=0}^{L-1} \mathbb{E}\left[h(t_i)^2 \delta W_i^2\right] \tag{4.11}$$

We have specified $i < j$ in the first sum on the right-hand side of (4.11) so that we can be sure that $t_i < t_j$. Since h is non-anticipating, this guarantees that the later increment, δW_j, is independent of $h(t_i)$ and $h(t_j)$. Of course, δW_j is also independent of δW_i by definition. (Note that $h(t_j)$ is **not** independent of δW_i, in general. Consider, for example, the case $h(t) \equiv W(t)$. An earlier increment, δW_i, clearly affects the later value, $W(t_j)$, of the Brownian path.) It follows that

$$\sum_{i<j}^{L-1} \mathbb{E}\left[h(t_i)h(t_j)\delta W_i \delta W_j\right] = \sum_{i<j}^{L-1} \mathbb{E}\left[h(t_i)h(t_j)\delta W_i\right]\mathbb{E}\left[\delta W_j\right] = 0. \tag{4.12}$$

We also have

$$\sum_{i=0}^{L-1} \mathbb{E}\left[h(t_i)^2 \delta W_i^2\right] = \sum_{i=0}^{L-1} \mathbb{E}\left[h(t_i)^2\right]\mathbb{E}\left[\delta W_i^2\right] = \delta t \sum_{i=0}^{L-1} \mathbb{E}\left[h(t_i)^2\right]. \tag{4.13}$$

Using (4.12) and (4.13) in (4.11) we find that

$$\mathbb{E}\left[\left(\sum_{i=0}^{L-1} h(t_i)\delta W_i\right)^2\right] = \delta t \sum_{i=0}^{L-1} \mathbb{E}\left[h(t_i)^2\right].$$

Our final expression looks like a Riemann sum approximation to $\int_0^T \mathbb{E}[h(t)^2]\,dt$. In appropriate circumstances it is valid to take the expectation outside of the integral (effectively, this means changing the order of two integration operations) to obtain $\mathbb{E}\left[\int_0^T h(t)^2\,dt\right]$. This has given us a heuristic derivation of the *Itô isometry*

$$\mathbb{E}\left[\left(\int_0^T h(t)\,dW(t)\right)^2\right] = \mathbb{E}\left[\int_0^T h(t)^2\,dt\right]. \tag{4.14}$$

This identity, which allows us to convert from "stochastic" to "standard Riemann" integrals, will play an invaluable role in Chapter 9, where we analyse the strong convergence of a numerical method.

4.4 "$dW^2 = dt$"

In this section we go through a calculation that is illuminating on two counts.

(1) It gives an intuitive explanation of the relation "$dW^2 = dt$" that is discussed in section 6.2 as a way to remember the Itô formula.

(2) Closely related results (see Exercises 5.6 and 5.7) will be used in section 5.5 to illustrate how Itô and Stratonovich integrals are related by a simple transformation.

We will study the Riemann sum difference

$$\sum_{i=0}^{L-1} h(W(t_i))\delta W_i^2 - \sum_{i=0}^{L-1} h(W(t_i))\delta t, \tag{4.15}$$

where h is a given deterministic function. First we note that

$$\begin{aligned}\mathbb{E}\left[\sum_i h(W(t_i))\delta W_i^2 - \sum_i h(W(t_i))\delta t\right] &= \sum_i \mathbb{E}\left[h(W(t_i))\delta W_i^2\right] - \sum_i \mathbb{E}\left[h(W(t_i))\right]\delta t \\ &= \delta t \sum_i \mathbb{E}\left[h(W(t_i))\right] - \delta t \sum_i \mathbb{E}\left[h(W(t_i))\right] \\ &= 0,\end{aligned}$$

where we exploited the facts that $h(W(t_i)$ and δW_i^2 are independent and $\mathbb{E}[\delta W_i^2] = \delta t$. Also, for simplicity, we are now omitting the lower and upper limits of 0 and $L-1$ in the sums.

Having established that the mean of (4.15) is zero, we now look at the variance. Well,

$$\begin{aligned}\mathbb{E}\left[\left(\sum_i h(W(t_i))\delta W_i^2 - \sum_i h(W(t_i))\delta t\right)^2\right] &= \mathbb{E}\left[\left(\sum_i h(W(t_i))\delta W_i^2\right)\left(\sum_j h(W(t_j))\delta W_j^2\right)\right] \\ &\quad - 2\mathbb{E}\left[\sum_i \sum_j h(W(t_i))h(W(t_j))\delta W_i^2 \delta t\right] \\ &\quad + \mathbb{E}\left[\left(\sum_i h(W(t_i))\delta t\right)\left(\sum_j h(W(t_j))\delta t\right)\right] \\ &= A + B + C,\end{aligned} \tag{4.16}$$

where we have labelled the three terms in the obvious manner.

Then

$$A = 2\sum_{i<j}\sum_j \mathbb{E}\left[h(W(t_i))h(W(t_j))\delta W_i^2 \delta W_j^2\right] + \sum_i \mathbb{E}\left[h(W(t_i))^2 \delta W_i^4\right].$$

We have specified $i < j$ in the double summation so that, as explained after (4.11), we can be sure that δW_j is independent of $h(W(t_i))$, $h(W(t_j))$ and δW_i. It follows that for $i < j$

$$\begin{aligned}\mathbb{E}\left[h(W(t_i))h(W(t_j))\delta W_i^2\delta W_j^2\right] &= \mathbb{E}\left[h(W(t_i))h(W(t_j))\delta W_i^2\right]\mathbb{E}\left[\delta W_j^2\right] \\ &= \delta t\mathbb{E}\left[h(W(t_i))h(W(t_j))\delta W_i^2\right].\end{aligned}$$

Since $\mathbb{E}[\delta W_i^4] = 3\delta t^2$ (see Exercise 3.3) we find that

$$\begin{aligned}A = 2\delta t\sum_{i<j}\sum_{j}\mathbb{E}\left[h(W(t_i))h(W(t_j))\delta W_i^2\right] \\ + 3\delta t^2\sum_{i}\mathbb{E}\left[h(W(t_i))^2\right].\end{aligned} \tag{4.17}$$

In a similar way we may write

$$\begin{aligned}B = -2\delta t\sum_{i<j}\sum_{j}\mathbb{E}\left[h(W(t_i))h(W(t_j))\delta W_i^2\right] \\ - 2\delta t\sum_{i>j}\sum_{j}\mathbb{E}\left[h(W(t_i))h(W(t_j))\delta W_i^2\right] \\ - 2\delta t\sum_{i}\mathbb{E}\left[h(W(t_i))^2\delta W_i^2\right].\end{aligned}$$

When $i > j$, the increment δW_i^2 is independent of $W(t_i)$ and $W(t_j)$, so that $\mathbb{E}[h(W(t_i))h(W(t_j))\delta W_i^2] = \delta t\mathbb{E}[h(W(t_i))]h(W(t_j))]$. Hence,

$$\begin{aligned}B = -2\delta t\sum_{i<j}\sum_{j}\mathbb{E}\left[h(W(t_i))h(W(t_j))\delta W_i^2\right] \\ - 2\delta t^2\sum_{i>j}\sum_{j}\mathbb{E}\left[h(W(t_i))h(W(t_j))\right] \\ - 2\delta t^2\sum_{i}\mathbb{E}\left[h(W(t_i))^2\right].\end{aligned} \tag{4.18}$$

Finally,

$$C = 2\delta t^2\sum_{i>j}\sum_{j}\mathbb{E}\left[h(W(t_i))h(W(t_j))\right] + \delta t^2\sum_{i}\mathbb{E}\left[h(W(t_i))^2\right]. \tag{4.19}$$

Using (4.17), (4.18) and (4.19) in (4.16), we find that

$$A + B + C = 2\delta t^2\sum_{i}\mathbb{E}\left[h(W(t_i))^2\right].$$

For small δt, this quantity approximates $2\delta t\int_0^T \mathbb{E}[h(W(t))^2]\,dt$, and hence is $O(\delta t)$.

Since the difference between the two Riemann sums in (4.15) has zero mean and $O(\delta t)$ variance, it is reasonable to infer that in the $\delta t \to 0$ limit, the two sums are equal. We could write "$\int_0^T h(W(t))\,dW^2(t) = \int_0^T h(W(t))\,dt$". Although this has no real meaning, it does help to explain why the rule "$dW^2 = dt$" successfully reproduces Itô's formula in section 6.2.

4.5 Notes and References

The Itô integral is named after the Japanese mathematician Kiyosi Itô [1915–2008], who won many honours for his outstanding work. Some brief bibiliographical details can be found at

`http://www-groups.dcs.st-and.ac.uk/~history/Biographies/Ito.html`

and more technical overviews are given in [18, 34].

Our approach throughout this book is to give accessible, heuristic explanations of important results. But we aim to do so in a manner that parallels more rigorous treatments. The illustrations in this chapter based on Riemann sum style approximations could more formally be interpreted as applying to an appropriate class of *step processes*, which are also known as *simple processes*, that is, stochastic processes $Y(t)$ such that $Y(t) = Y_i$ over $[t_i, t_{i+1}]$, where each Y_i is a fixed random variable. The Itô integral, and its properties, may then be studied by

- making definitions (and deriving results) for integrands that are step processes,
- showing that a more general integrand can be approximated to arbitrary accuracy by a sequence of step processes in the $\delta t \to 0$ limit,
- defining (and deriving results for) the Itô integral of the more general integrand via the $\delta t \to 0$ limit of the Itô integral for this sequence of step processes.

We recommend [8, 37, 45] for further details at an introductory level.

Exercises

4.1. ⋆⋆ Show that the second sum on the right-hand side of (4.6) has expected value zero and variance of $O(\delta t)$.

4.2. ⋆⋆⋆ Find the mean and variance of Riemann sums of the form

$$\sum_{i=0}^{L-1} h(\alpha t_i + (1-\alpha)t_{i+1}))\,(W(t_{i+1}) - W(t_i))\,,$$

where $0 \le \alpha \le 1$.

```
%CH04     Approximate Ito and Stratonovich integrals of W dW

randn('state',100)
T = 1; L = 500; dt = T/L;

dW = sqrt(dt)*randn(1,L);          % Brownian increments
W = cumsum(dW);                    % cumulative sum

ito = sum([0,W(1:end-1)].*dW)
itoerr = abs(ito - 0.5*(W(end)^2-T))
```

Figure 4.3: Program of Chapter 4: `ch04.m`.

4.3. ⋆⋆⋆ Using the left-hand Riemann sum approximation, deduce the result (4.9). [Hint: note that $a^2(b-a) = \frac{1}{3}(b^3-a^3) - a(b-a)^2 - \frac{1}{3}(b-a)^3$.

4.4. ⋆ Verify that the Itô isometry (4.14) holds in the case $h(t) \equiv 1$.

4.5. ⋆⋆ Using the Riemann sum approximation (4.3), argue that

$$\mathbb{E}\left[\left(\int_0^T t\,dW(t)\right)^2\right] = \frac{T^3}{3}.$$

Hence, verify that the Itô isometry (4.14) holds in the case $h(t) \equiv t$.

4.6. ⋆ Using the Riemann sum approximation (4.3), argue that the linearity condition

$$\int_0^T \alpha f(t) + \beta g(t)\,dW(t) = \alpha \int_0^T f(t)\,dW(t) + \beta \int_0^T g(t)\,dW(t)$$

holds for all $\alpha, \beta \in \mathbb{R}$.

4.7. ⋆⋆⋆ Using the Riemann sum approximation (4.3), argue that the property

$$\mathbb{E}\left[\left(\int_0^T f(t)\,dW(t)\right)\left(\int_0^T g(t)\,dW(t)\right)\right] = \int_0^T \mathbb{E}[f(t)g(t)]\,dt$$

holds. [See [44, Corollary 5.19] for a more formal treatment.]

4.6 Program of Chapter 4 and Walkthrough

The MATLAB code `ch04.m` listed in Figure 4.3 performs a computational experiment from section 4.2. We evaluate the Riemann sum (4.3) in the case where $h(t) = W(t)$ for a discretized Brownian path. Here, $T = 1$, $L = 500$ and $\delta t = T/L$. The path is computed as in `ch03.m` and the main computation, `sum([0,W(1:end-1)].*dW)`, makes use of componentwise multiplication and the built-in `sum` function. The result, `ito = -0.2674`, is compared with the limiting value in (4.5) and we find a discrepancy of `itoerr = 0.0158`.

Programming Exercises

P4.1. Alter `ch04.m` in order to approximate the corresponding Stratonovich integral, and evaluate the error.

P4.2. Choose an integrand $h(t)$ and computationally confirm the Itô isometry (4.14).

Quotes:

"Everyone who is likely to pick up this book has at least heard that
there is a subject called the theory of stochastic integration
and that K. Itô is the Lebesgue of this branch of integration theory
(Paley and Weiner were its Reimann).
D. Strook and S. R. S. Varadhan; source [18].

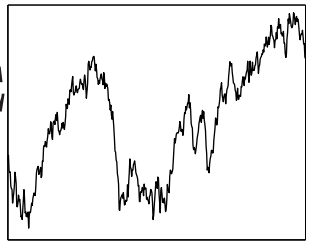
There are in this world optimists
who feel that any symbol that starts off with an integral sign
must necessarily denote something that will have every property
that they should like an integral to possess.
This of course is quite annoying to us rigorous mathematicians;
what is even more annoying
is that by doing so they often come up with the right answer.
E. J. McShane, Bulletin of the American Mathematical Society, Volume 69,
page 611, 1963.
Source `http://math.furman.edu/~mwoodard/mquot.html`

The more "obvious" definition in terms of
$G(x[\frac{1}{2}(t_i + t_{i-1}], \frac{1}{2}(t_i + t_{i-1}))$
was not used by Stratonovich in his original definition,
although the view that this provides the definition of of the Stratonovich integral
is widespread in the literature
(including the first edition of this book).
C. W. Gardiner, [19].

Nature laughs at the difficulties of integration.
Pierre-Simon de Laplace (1749 - 1827).

Chapter 5

Stochastic Differential E



Outline

- Stochastic differential equations
- Examples from application areas
- Ito versus Stratonovich

5.1 Motivation

With the Ito integral in place we are all set to define a stochastic differential equation. After explaining what it means for $X(t)$ to solve a stochastic differential equation, we will give some examples arising from a range of application areas.

5.2 Stochastic Differential Equations

In the deterministic setting, given a function $f : \mathbb{R} \to \mathbb{R}$ we may consider the corresponding ordinary differential equation (ODE)

$$\frac{dx(t)}{dt} = f(x(t)). \tag{5.1}$$

Here, a solution $x(t)$ is a function whose derivative satisfies (5.1). Typically the solution is required over some time interval $0 \leq t \leq T$, and the *initial condition* $x(0)$ is given. The Fundamental Theorem of Calculus tells us that differentiating is the opposite of integrating, so that $x(t)$ may also be regarded as a function for which

$$x(t) - x(0) = \int_0^t f(x(s))\, ds.$$

It is this means of relating $x(t)$ to an integral equation that extends naturally to a workable definition of a stochastic differential equation (SDE).

Given functions f and g, we say that the stochastic process $X(t)$ is a solution to the SDE

$$dX(t) = f(X(t))dt + g(X(t))dW(t) \tag{5.2}$$

if $X(t)$ solves the integral equation

$$X(t) - X(0) = \int_0^t f(X(s))\,ds + \int_0^t g(X(s))\,dW(s). \tag{5.3}$$

Here, the second term on the right-hand side is an Ito stochastic integral, as introduced in Chapter 4, and a solution $X(t)$ is a random variable at each time t. The notation used in (5.2) to describe the SDE is merely a shorthand for the integral equation (5.3). It is as close as we can get to the deterministic notation (5.1), given the restriction that $W(t)$ is not differentiable and hence we are not allowed to write "$dW(t)/dt$". Let us re-emphasize that

> the symbols $dX(t)$, dt and $dW(t)$ in (5.2) have no meaning to us, other than to play a role in the shorthand notation that we are using to represent the integral formulation (5.3).

It is common to refer to $f(x)$ in (5.2) as the *drift coefficient* and $g(x)$ as the *diffusion coefficient*. As described for the deterministic case, we will consider the setting where a solution $X(t)$ is required over a finite time interval $0 \le t \le T$, and a given random variable $X(0)$ specifies the initial condition.

A simple, but very widely occuring SDE is the multiplicative noise linear case of $f(x) = \mu x$ and $g(x) = \sigma x$ in (5.2), so that

$$dX(t) = \mu X(t)dt + \sigma X(t)dW(t). \tag{5.4}$$

Here μ and σ are real constants. This SDE is used to model the evolution of a stock price in the Black–Scholes theory for financial option valuation [21, 27, 29]. In the case where $\sigma = 0$ and $X(0)$ is constant, (5.4) reduces to the ODE $dx(t) = \mu x(t)dt$, which has solution $x(t) = x(0)e^{\mu t}$. This corresponds to a non-risky deposit in a bank with interest rate μ. The stochastic term $\sigma X(t)dW(t)$ models random fluctuations, and σ is known as the *volatility*. The solution of (5.4) is

$$X(t) = X(0)e^{(\mu - \frac{1}{2}\sigma^2)t + \sigma W(t)}. \tag{5.5}$$

We will confirm this in Chapter 6. The stochastic process $X(t)$ in (5.5) is sometimes called *geometric Brownian motion*. Exercise 5.1 asks you to find an explicit formula for the density function of $X(t)$ when $X(0)$ is constant. Figure 5.1 shows the density at times $t = 1$ and $t = 5$ with volatilities $\sigma = 0.3$ and $\sigma = 0.5$, when $X(0) = 1$ and $\mu = 0.05$.

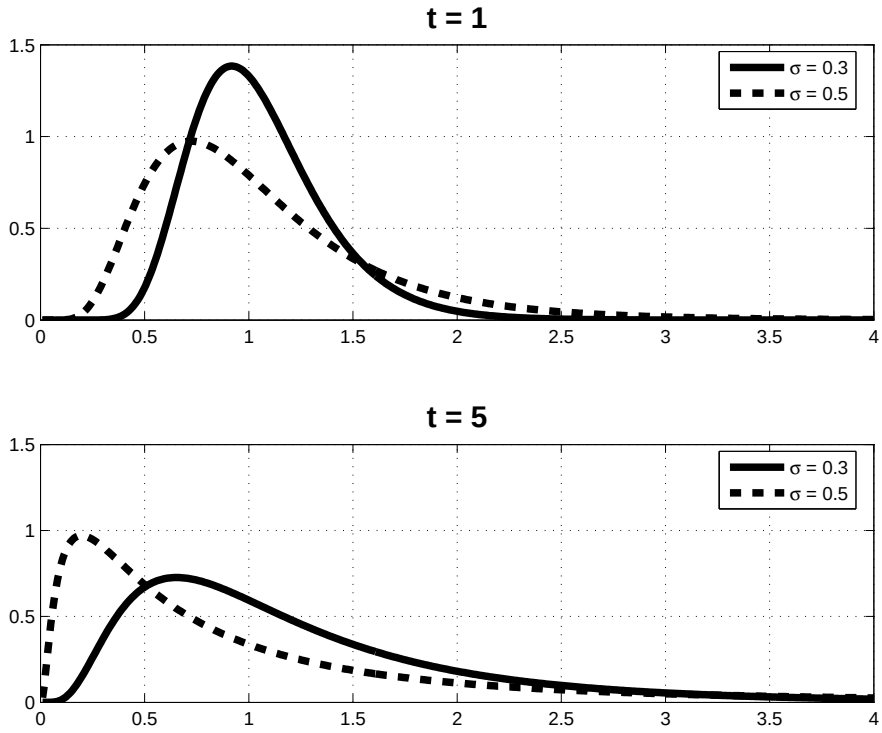


Figure 5.1: Density for the SDE solution $X(t)$ in (5.5) when $X(0) = 1$, $\mu = 0.05$ and $\sigma = 0.3$ (solid) and $\sigma = 0.5$ (dashed), as derived in Exercise 5.1. Upper picture $t = 1$. Lower picture $t = 5$.

Using Exercise 3.4, we may take expected values in (5.5) to give

$$\mathbb{E}\left[X(t)\right] = \mathbb{E}\left[X(0)\right] e^{\mu t}, \tag{5.6}$$

which explains why μ is known as the *expected growth rate* in the stock price model. We could also obtain (5.6) without directly using the expression for the solution. Taking expected values in the integral formulation (5.3), with $f(x) = \mu x$ and $g(x) = \sigma x$, and using the martingale property (4.10), we find that

$$\mathbb{E}\left[X(t)\right] - \mathbb{E}\left[X(0)\right] = \mu \int_0^t \mathbb{E}\left[X(s)\right] \, ds. \tag{5.7}$$

This is a deterministic integral equation for the function $\mathbb{E}\left[X(t)\right]$. It is readily solved to give (5.6). This trick works on any SDE that has a linear drift coefficient; see Exercise 5.2.

Computational Example In Figure 5.2 we show paths of the solution (5.5). More precisely, we show "join-the-dots" curves for $\{t_i, X(0)e^{(\mu^2 - \frac{1}{2}\sigma)t_i + \sigma W_i}\}$ where the W_i come from a discretized Brownian path. In each of the four

pictures we use a different value for the volatility, σ, with $\mu = 1$ and $X(0) = 1$ fixed. The same 10 discretized Brownian paths were used for each picture. Along with the 10 solution paths, we also plot the mean solution value as a thick dashed line. We see from the pictures that increasing σ produces more "jagged" paths that tend to spread further from the mean. (Note that $\mathsf{var}[X(t)]$, which can be found via Exercise 6.6, increases as a function of t.) $\diamondsuit$

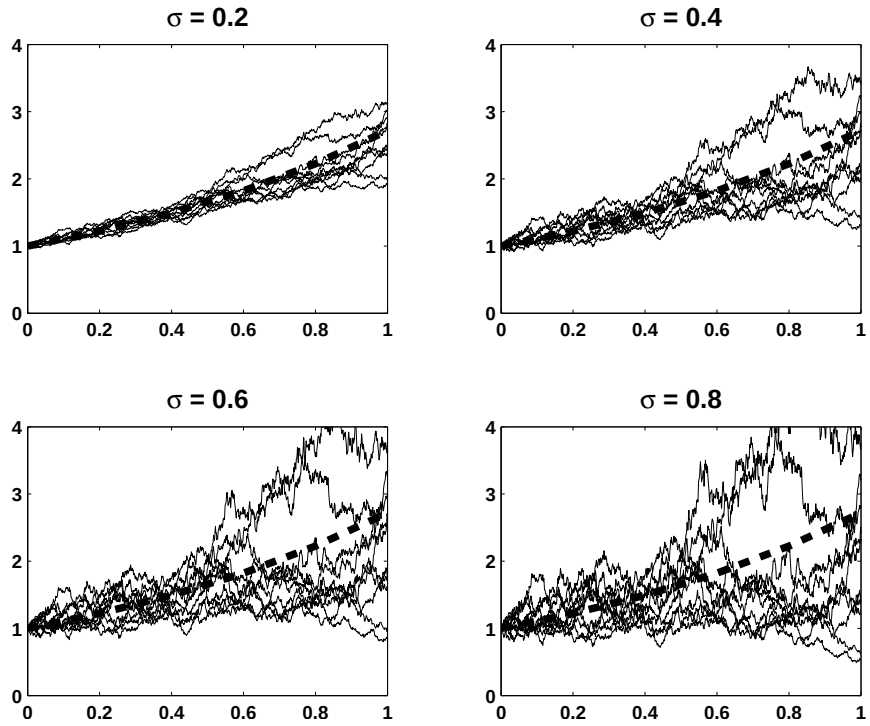


Figure 5.2: Each picture shows solution paths (5.5) for the same 10 discretized Brownian paths. The four pictures correspond to four different volatility values, $\sigma = 0.2$, 0.4, 0.6 and 0.8. Here, $\mu = 1$ and the mean over all paths, (5.6), is shown as a thick dashed line.

5.3 Examples of Stochastic Differential Equations

EXAMPLES TO INCLUDE:

- Finance thing in Alan Bryden's thesis (weak only?)

5.3.1 Asset Prices and Interest Rates

The *mean-reverting square root process* takes the form

$$dX(t) = \lambda\left(\mu - X(t)\right)dt + \sigma\sqrt{X(t)}dW(t). \tag{5.8}$$

Here λ, μ and σ are positive constants. Exercise 5.3 asks you to confirm that

$$\mathbb{E}[X(t) - \mu] = e^{-\lambda t}(\mathbb{E}[X(0)] - \mu) \tag{5.9}$$

and Exercise 6.7 in Chapter 6 asks you to confirm that

$$\mathbb{E}\left[X(t)^2 - \mu^2 - \frac{\sigma^2\mu}{2\lambda}\right] = (2\mu + \frac{\sigma^2}{\lambda})(\mathbb{E}[X(0)] - \mu)e^{-\lambda t} + \left(\mathbb{E}[X(0)^2] + (\mu + \frac{\sigma^2}{2\lambda})(\mu - 2\mathbb{E}[X(0)])\right)e^{-2\lambda t}. \tag{5.10}$$

It follows that

$$\lim_{t\to\infty}\mathbb{E}\left[X(t)\right] = \mu \quad \text{and} \quad \lim_{t\to\infty}\mathbb{E}\left[X(t)^2\right] = \mu^2 + \frac{\sigma^2\mu}{2\lambda}.$$

Hence, μ is the long-term mean, λ is the rate at which the mean converges, and the noise coefficient σ affects the variance. It can be shown that if $X(0) \geq 0$ with probability one, then this non-negativity holds for all $t \geq 0$. Further, the solution may attain the value zero if and only if $\sigma^2 > 2\lambda\mu$; see [44, Section 9.2] or [39, Section 7.1.5] for more details.

The mean-reverting square root process (5.8) is widely used in mathematical finance as an alternative to geometric Brownian motion (5.5). The idea of taking the diffusion coefficient to be proportional to the square root of the solution can be traced as least as far back as [15], and it is often argued to be more realistic than the linear proportionality used for geometric Brownian motion—the square root dampens the influence of the noise when $X(t)$ is large. With $\mu = 0$ and $\lambda \in \mathbb{R}$ the SDE (5.8) is also known as Feller's branching diffusion [37], which approximates a branching process in population dynamics.

As an interest rate model, the SDE is often named after Cox, Ingersoll and Ross from [14]. The process is also used to model asset prices and volatilities; see Chapter 16.

No explicit solution for (5.8) is available, but the *transition density* (that is the density of the random variable that describes $X(s)$ given the value of X at some earlier time) can be written down and simulated with; see, [7].

Predating the Cox, Ingersoll and Ross model is the additive noise version

$$dX(t) = \lambda\left(\mu - X(t)\right)dt + \sigma dW(t) \tag{5.11}$$

from [60], which is known as the Vasicek interest rate model. This SDE is also widely known as the *mean-reverting Ornstein-Uhlenbeck process*. Exercise 6.4 gives the solution to (5.11). This model does not guarantee to maintain a positive interest rate.

5.3.2 Opinion Polls

In [40] the SDE

$$dX(t) = -\mu \left(\frac{X(t)}{1 - X(t)^2} \right) dt + \sigma dW(t), \tag{5.12}$$

is proposed to model an opinion poll conducted in an environment where two political parties, A and B, are dominant. Here, $X(t) = P_A(t) - P_B(t)$, where $P_A(t)$ and $P_B(t)$ denote the proportions of the population that intend to vote for parties A and B, respectively, at time t. The parameters $\mu > 0$ and σ are constant. Since the proportions $P_A(t)$ and $P_B(t)$ must lie between 0 and 1, we may assume that $-1 < X(0) < 1$ with probability one. Any sensible model for $X(t)$ must then maintain the constraint $-1 < X(t) < 1$, and it can be shown that (5.12) does this when $\mu > \frac{1}{2}\sigma^2$.

5.3.3 Population Dynamics

The SDE

$$dX(t) = rX(t)\left(K - X(t)\right) dt + \beta X(t) dW(t), \tag{5.13}$$

models the evolution of a population in a crowded environment, with $X(t)$ denoting the population density at time t. With $\beta = 0$ this is a standard logistic ODE model for population growth. The constant $K > 0$ is the *carrying capacity* of the environment and the constant $r > 0$ is such that $1/r$ is a characteristic *timescale*; see, for example, [50]. The constant coefficient β governs the strength of the multiplicative noise term. It can be shown, see, for example, [38, Section 4.4] or [52, Exercise 5.15] that the SDE (5.13) has solution

$$X(t) = \frac{X(0)\, e^{(rK - \frac{1}{2}\beta^2)t + \beta W(t)}}{1 + X(0)\, r \int_0^t e^{(rK - \frac{1}{2}\beta^2)s + \beta W(s)}\, ds}. \tag{5.14}$$

Any sensible population model should produce non-negative solutions. From (5.14), we see that the model (5.13) behaves well in this respect—if $X(0) \geq 0$ with probability one, then, for any $t > 0$, $X(t) \geq 0$ with probability one.

5.3.4 Epidemic Model

In [56] a stochastic model for disease transmission in a population takes the form

$$dX(t) = F(X(t))dt + G(X(t))dW(t), \tag{5.15}$$

where

$$\begin{aligned} F(x) &= (p-1)Bx + (\beta_0 C - \alpha)(1-x)x, \\ G(x) &= \rho C(1-x)x. \end{aligned}$$

Here, $X(t)$ denotes the proportion of infected animals. The constants in this model are p, the probability of vertical disease transmission (mother to offspring); B, the birth rate; β_0, the mean disease transmission rate; ρ, the strength of the fluctuations in the disease transmission rate; α, the increase in mortality rate due to disease; and C, the contact rate between individuals.

5.3.5 Political Opinions

Cobb [12] proposes the SDE

$$dX(t) = r\,(G - X(t))\,dt + \sqrt{\epsilon X(t)\,(1 - X(t))}dW(t), \tag{5.16}$$

as a model for the motion of an individual through the liberal–conservative political spectrum. Here $X(t) = 0$ denotes an extreme liberal and $X(t) = 1$ denotes an extreme conservative. The positive parameters G and r give the long-term mean value and the rate at which this mean is approached (in the same way that we saw for the mean-reverting square root process (5.8)). The positive parameter ϵ controls the strength of the noise term. Generally, the drift term reflects an assumption that a person has a tendency to move toward the average persuasion of the whole population. The diffusion term has been constructed so that people who hold extreme views are less likely to undergo random fluctuations than people nearer the centre of the political spectrum. Cobb shows that by varying the parameters in the model it is possible to go from a broad political consensus to a polarization of opinion.

The same SDE is also widely used in population dynamics to measure the frequency of genes or alleles, where is it known as the Wright-Fisher diffusion model [37].

5.3.6 Double-Well Potential

If a deterministic ODE has the form $dx/dt = -V'(x(t))$, where the function V is bounded below, then V is called a *potential*. In this case, by the deterministic chain rule we have

$$\frac{d}{dt}V(x(t)) = V'(x(t))\frac{dx(t)}{dt} = -\left(V'(x(t))\right)^2.$$

We see that along any solution curve, $x(t)$, the potential $V(\cdot)$ is non-increasing, and is strictly decreasing unless the solution reaches a stationary point. In Figure 5.3 we plot a *double-well potential*

$$V(x) = x^2(x-2)^2. \tag{5.17}$$

Here, ODE solutions starting with $X_0 < 0$ follow solution paths that monotonically decay to zero—they slide down the left-hand branch of the left-hand well.

Similarly, ODE solutions starting with $0 < X_0 < 1$ slide down the right-hand branch of the left-hand well, ODE solutions starting with $1 < X_0 < 2$ slide down the left-hand branch of the right-hand well, and ODE solutions starting with $2 < X_0$ slide down the right-hand branch of the right-hand well. If we add noise to the problem, we make it possible for solution paths to overcome the potential barrier, crossing over the central hump and thus moving from one well to the other. The additive noise version of the potential problem is

$$dX(t) = -V'(X(t))dt + \sigma dW(t),$$

where σ is a constant. For the double-well case (5.17) this becomes

$$dX(t) = \left(-8X(t) + 12X(t)^2 - 4X(t)^3\right) dt + \sigma dW(t). \tag{5.18}$$

We use this SDE in Chapter 11, where we study mean exit times.

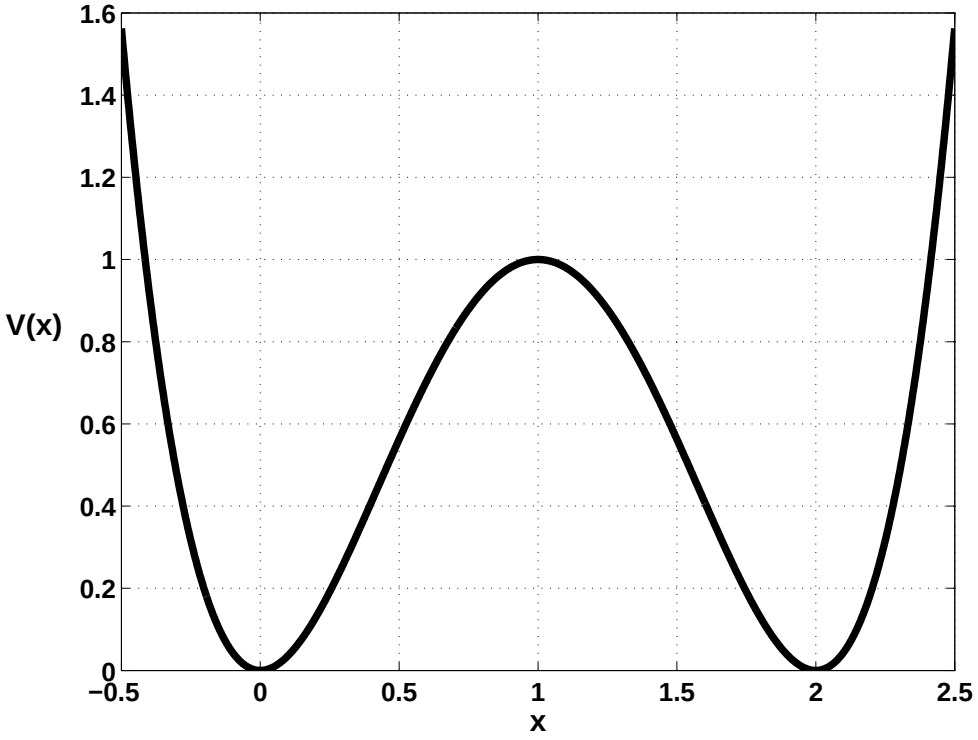


Figure 5.3: The double-well potential function $V(x)$ in (5.17).

5.4 Existence and Weak versus Strong Solutions

Although we are omitting most technical details, it is appropriate to mention that the mathematical theory for the existence and uniqueness of solutions to

SDEs typically places quite severe restrictions on the types of function f and g in (5.2). The standard conditions on which much theory is based are globally Lipschitz drift and diffusion; that is, there exists a constant L such that f and g in (5.2) satisfy

$$|f(x)-f(y)| \le L|x-y| \quad \text{and} \quad |g(x)-g(y)| \le L|x-y| \quad \text{for all } x, y \in \mathbb{R}. \tag{5.19}$$

We will make use of these conditions when we analyze convergence of a numerical method, while noting that many of the examples discussed above do not satisfy them; see Exercises 5.4 and 5.5.

When given an SDE to solve, a reasonable question to ask is "Do you have a particular Brownian path in mind, or must the solution be valid for any Brownian path?" This question is essentially distinguishing between weak and strong solutions. A *strong solution*, such as (5.5) for (5.4), is a stochastic process that is determined for any Brownian process. In this book, we will always use the word solution in this strong sense. By contrast, a *weak solution* needs to be given only for a certain Brownian process, which is then regarded as part of the solution. There are examples of SDEs for which only weak solutions exist; see section 5.6 for a reference.

We remark that these "weak and strong" concepts are not the same as those arising in weak and strong convergence of numerical methods in Chapter 7 and beyond.

5.5 Ito versus Stratonovich

BOX THE FOLLOWING:

****

This section can be skipped by readers who wish to focus only on Ito SDEs. To minimize confusion, we have restricted discussion of Stratonovich integrals and SDEs to this section (apart from the introductory material in section 4.2) Hence, elsewhere in the book all integrals and SDEs can be assumed to be in Ito form. In order to confine Stratonovich calculus to this section, we found it necessary to make a forward reference to the Ito formula, which is discussed in Chapter 6.

****

In section 4.2 we pointed out that the use of the left-hand endpoint in the Riemann sum construction (4.3) for the Ito stochastic integral is a vital element. Other choices do not lead to the same $\delta t \to 0$ limit in general. In particular, the midpoint value based sum (4.4) leads to what is known as the Stratonovich stochastic integral. In this section we pursue some of the connections and differences between Ito and Stratonovich integrals, and the resulting SDEs.

We recall that $\int_0^T h(t)\,dW(t)$ and $\int_0^T h(t)\circ dW(t)$ denote the Ito and Stratonovich integrals, respectively.

We begin with a heuristic analysis of the Riemann sum (4.4) when the integrand is a function of Brownian motion; that is, has the form $h(W(t))$ for a given function h. We let $t_{i+\frac{1}{2}} := (t_i + t_{i+1})/2$. Using a Taylor approximation,

$$h(W(t_{i+\frac{1}{2}})) \approx h(W(t_i)) + h'(W(t_i))\left(W(t_{i+\frac{1}{2}})) - W(t_i)\right).$$

Hence,

$$\begin{aligned}
\sum_{i=0}^{L-1} h(W(t_{i+\frac{1}{2}}))\delta W_i &\approx \sum_{i=0}^{L-1} h(W(t_i))\delta W_i + \sum_{i=0}^{L-1} h'(W(t_i))\left(W(t_{i+\frac{1}{2}}) - W(t_i)\right)\delta W_i \\
&= \sum_{i=0}^{L-1} h(W(t_i))\delta W_i + \sum_{i=0}^{L-1} h'(W(t_i))\left(W(t_{i+\frac{1}{2}}) - W(t_i)\right)^2 \\
&\quad + \sum_{i=0}^{L-1} h'(W(t_i))\left(W(t_{i+\frac{1}{2}}) - W(t_i)\right)\left(W(t_{i+1}) - W(t_{i+\frac{1}{2}})\right) \\
&= S_1 + S_2 + S_3, \qquad (5.20)
\end{aligned}$$

where S_1, S_2 and S_3 denote the three sums. We see that S_1 approximates $\int_0^T h(W(t))\,dt$. Exercise 5.6 with $q = h'$ shows that S_2 approximates $\frac{1}{2}\int_0^T h'(W(t))\,dt$. Exercise 5.7 with $q = h'$ shows that S_3 is approximately zero. Overall, it is reasonable to conclude from (5.20) that

$$\int_0^T h(W(t)) \circ dW(t) = \int_0^T h(W(t))dW(t) + \tfrac{1}{2}\int_0^T h'(W(t))dt. \qquad (5.21)$$

Equation (5.21) shows a relation between the Ito and Stratonovich integrals. In section 4.2 we deduced that $\int_0^T W(t)dW(t) = W(T)^2 - T/2$, whereas $\int_0^T W(t)\circ dW(t) = W(T)^2$, and is simple to check that this is consistent with the case $h(x) = x$ in (5.21).

Computational Example In Figure 5.4 we illustrate the result (5.21) for the case where $h(x) = x^3$. We generated a finely spaced discrete Brownian path at points t_i and then, for each t_i, we computed approximations to $\int_0^{t_i} h(W(s)) \circ dW(s)$ and $\int_0^{t_i} h(W(s))\,dW(s)$, using the corresponding Riemann sums. The upper left picture shows the Stratonovich integrals, the upper right picture shows the Ito integrals, and the lower left picture shows the remainder when Ito is subtracted from Stratonovich. From (5.21), this difference should be $\frac{1}{2}\int_0^{t_i} h'(W(s))\,ds$. The lower right picture shows the Riemann sum approximation to this integral, and we see that it agrees to visual accuracy. $\diamondsuit$

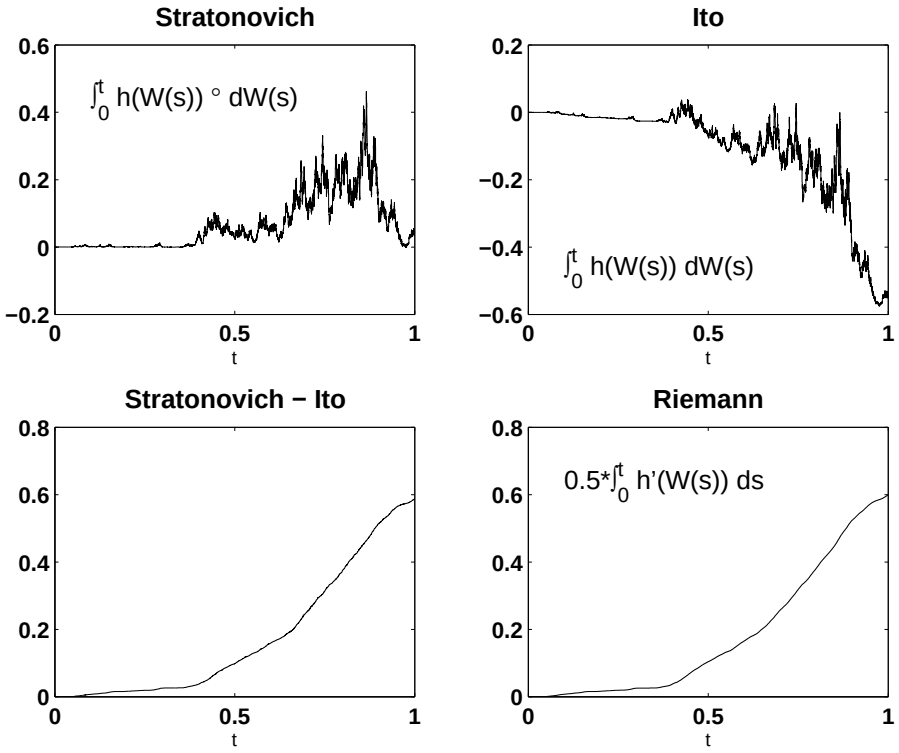


Figure 5.4: Illustration of (5.21) for the case where $h(x) = x^3$, using Riemann sum approximations from a finely-spaced discrete Brownian path. Upper left: Stratonovich integral $\int_0^{t_i} h(W(s)) \circ dW(s)$. Upper right: Ito integral $\int_0^{t_i} h(W(s))dW(s)$. Lower left: The difference $\int_0^{t_i} h(W(s)) \circ dW(s) - \int_0^{t_i} h(W(s))dW(s)$. Lower right: Riemann integral $\frac{1}{2}\int_0^{t_i} h'(W(s))\,ds$.

In Chapter 6 we discuss the Ito formula (6.5). As a slight digression we note here that applying the Ito formula with $A(t) = 0$, $B(t) = 1$, so that $X(t) = W(t)$, and $u(t, x) = r(x)$, gives

$$dr(W(t)) = \tfrac{1}{2}r''(W(t))dt + r'(W(t))dW(t).$$

In (Ito) integral form this becomes

$$r(W(t)) - r(W(0)) = \tfrac{1}{2}\int_0^T r''(W(t))dt + \int_0^T r'(W(t))dW(t). \tag{5.22}$$

We see from (5.21) that the right-hand side of (5.22) is precisely $\int_0^T h(W(t)) \circ dW(t)$ for $h(x) = r'(x)$. Hence, the Stratonvich integral satisfies

$$\int_0^T r'(W(t)) \circ dW(t) = r(W(T)) - r(W(0)). \tag{5.23}$$

This is analogous to the deterministic integral property $\int_0^T r'(x)dx = r(T) - r(0)$, and hence it is said that the Stratonovich integral satisfies the *classical chain rule*. For example, taking $r(t) = e^t$ gives

$$\int_0^T e^{W(t)} \circ dW(t) = e^{W(T)} - 1. \tag{5.24}$$

This example also shows, that the Stratonovich integral does not satisfy the martingale property (4.10) of the Ito integral; see Exercise 5.8.

The relation (5.21) between Ito and Stratonovich integrals can be generalized to the case where the integrand has the form $h(X(t))$ and $X(t)$ is an *Ito process* satisfying (6.3). In this circumstance we have

$$\int_0^T h(X(t)) \circ dW(t) = \int_0^T h(X(t))dW(t) + \tfrac{1}{2}\int_0^T B(X(t))h'(X(t))dt. \tag{5.25}$$

Just as in the Ito case, we may use the concept of the Stratonovich integral in order to define a Stratonovich stochastic differential equation. We write

$$dX(t) = \widetilde{f}(X(t))dt + \widetilde{g}(X(t)) \circ dW(t) \tag{5.26}$$

to mean that $X(t)$ satisfies the integral equation

$$X(t) = X(0) + \int_0^t \widetilde{f}(X(s))\, ds + \int_0^t \widetilde{g}(X(s)) \circ dW(s).$$

Now starting with an Ito SDE

$$X(t) = X(0) + \int_0^t f(X(s))\, ds + \int_0^t g(X(s))dW(s), \tag{5.27}$$

we may use (5.25) with $A(t) \equiv f(X(t))$, $B(t) \equiv g(X(t))$ and $h(x) = g(x)$ to re-write the Ito integral as

$$\int_0^t g(X(s))dW(s) = \int_0^t g(X(s)) \circ dW(s) - \tfrac{1}{2}\int_0^t g(X(s))g'(X(s))ds.$$

We conclude that the Ito SDE is equivalent to the Stratonovich SDE

$$dX(t) = (f(X(t)) - \tfrac{1}{2}g(X(t))g'(X(t)))dt + g(X(t)) \circ dW(t), \tag{5.28}$$

in the sense that if $X(t)$ solves one then it also solves the other. It is thus possible to transfer between the Ito and Stratonovich SDE forms.

Both Ito and Stratonovich SDE formulations are widely used. It is natural, but ultimately not very enlightening, to ask which version is "better". The Ito integral enjoys the martingale property (4.10) that can be very useful for

analyzing the behavior of solutions. On the other hand, the Stratonovich integral has a natural chain rule (5.23) that can be helpful for determining analytical expressions for SDE solutions. When SDE models are developed, modelers must decide whether the diffusion coefficient, $g(X(t))$, should be interpreted in the Ito or Stratonovich sense. There seems to be no consensus about which version is more relevant in general. In the field of mathematical finance, however, the Ito form is almost ubiquitous thanks to the *efficient market hypothesis*—if we assume that current prices reflect all known information and any change in price is due to random, instantaneous, developments, then the left endpoint Riemann sum (4.3) follows naturally. In summary, whether an SDE should arise in Ito or Stratonovich form is a *mathematical modeling* issue. Once the form is specified, it is a simple matter to convert from one form to the other.

To avoid confusion, in this book, apart from in section 4.2 and this section, we deal exclusively with the Ito SDE form.

5.6 Notes and References

Add something about Global Lipschitz being stringent.

Weak solution reference is [37].

Give refs for all results. Maybe mention Gikhod as an early reference (Michael Tretyakov).

A revised and extended version of [12] is available at `www.aetheling.com/docs/SDE.pdf`

Potential problems such as (5.18) arise as rigorous limits of high-dimensional determistic ODE models for interacting particles, see, for example, the references in [2].

Exercises

5.1. ⋆⋆⋆ When $X(0)$ is constant, the solution $X(t)$ in (5.5) is said to have a *lognormal* distribution. (This is because its log has a normal distribution.) Show that the density function for $X(t)$ is then given by

$$f(x) = \frac{\exp\left(\frac{-(\log(x/X(0))-(\mu-\sigma^2/2)t)^2}{2\sigma^2 t}\right)}{x\sigma\sqrt{2\pi t}}, \quad \text{for } x > 0, \tag{5.29}$$

with $f(x) = 0$ for $x \leq 0$.

5.2. ⋆⋆ Suppose an SDE has a linear drift coefficient; that is, $f(x) = a_1 x + a_2$ in (5.2), for some constants $a_1 \neq 0$ and a_2. Letting $m(t) := \mathbb{E}[X(t)]$, by taking expectations in the integral equation (5.3) and using the martingale property (4.10), show that

$$m(t) - m(0) = a_1 \int_0^s m(s)\, ds + a_2 t,$$

and hence $m(t)$ satisfies the ODE

$$\frac{dm(t)}{dt} = a_1 m(t) + a_2, \qquad m(0) = \mathbb{E}[X(0)].$$

Solve this ODE to deduce that

$$\mathbb{E}[X(t)] = -\frac{a_2}{a_1} + \left(\mathbb{E}\left[X(0)\right] + \frac{a_2}{a_1}\right) e^{a_1 t}.$$

5.3. $\star$ Use Exercise 5.2 to verify (5.9) for the mean-reverting square root process (5.8).

5.4. $\star\star$ For which of the SDEs in section 5.3 are the global Lipschitz conditions (5.19) satisfied?

5.5. $\star$ Use the Mean Value Theorem to show that f and g satisfy a global Lipschitz condition if they have a uniformly bounded first derivative.

5.6. $\star\star\star$ Given a function q, show that the Riemann sum difference

$$\sum_{i=0}^{L-1} q(W(t_i))(W(t_{i+\frac{1}{2}}) - W(t_i))^2 - \tfrac{1}{2}\sum_{i=0}^{L-1} q(W(t_i))\delta t$$

has mean zero and variance of $O(\delta t)$. Here, $t_{i+\frac{1}{2}} := (t_i + t_{i+1})/2$. [Hint: follow the style of analysis in section 4.4.]

5.7. $\star\star\star$ Given a function q, show that the Riemann sum

$$\sum_{i=0}^{L-1} q(W(t_i))(W(t_{i+\frac{1}{2}}) - W(t_i))(W(t_{i+1}) - W(t_{i+\frac{1}{2}}))$$

has mean zero and variance of $O(\delta t)$. ???Check??? [Hint: follow the style of analysis in section 4.4.]

5.8. $\star$ Using (5.24) and Exercise 3.4, or otherwise, show that the Stratonovich integral does not share the martingale property (4.10) of the Ito integral.

5.9. $\star\star$ In Figure 5.4, the integral $\int_0^t h'(W(s))\,ds$ appears be smoother, as a function of t, than the stochastic integrals $\int_0^t h(W(s))dW(s)$ and $\int_0^t h(W(s)) \circ dW(s)$. Give a heuristic explanation for this by considering Riemann sum approximations of the differences

$$\frac{\int_0^{t+\delta t} h'(W(s))\,ds - \int_0^t h'(W(s))\,ds}{\delta t} \quad \text{and} \quad \frac{\int_0^{t+\delta t} h(W(s))dW(s) - \int_0^t h(W(s))dW(s)}{\delta t}.$$

5.10. $\star$ Write the Stratonovich SDE (5.26) in Ito form.

Listing 5.7.1: Program of Chapter 5: `ch05.m`.

5.11. ⋆ An SDE of the form

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

where α is a constant, is said to have *additive noise*. Show that in this case the Stratonovich form is the same as the Ito form.

5.7 Program of Chapter 5 and Walkthrough

The program `ch05`, listed in Figure 5.7.1, plots density functions for the linear SDE solution $X(T)$ in (5.5) when $X(0) = 1$, $\mu = 0.05$ and $T = 1$. Plots for two volatilities, $\sigma = 0.3$ and $\sigma = 0.5$, are superimposed. This density function is given as (5.29) in Exercise 5.1. The picture is similar to those in Figure ref???. The array `y1` stores the density function values at equally spaced points in `x` for the first volatility value, `sigma = 0.3`. The curve is plotted with a solid red linetype. A similar computation for `sigma = 0.5` produces a blue dotted curve.

Programming Exercises

P5.1. Adapt `ch05` to give a `waterfall` plot illustrating how the lognormal density function varies with σ for fixed $T = 1$. [Type `help waterfall` to learn about this 3D plotting facility.]

P5.2. Pick a function $h(\cdot)$ and, for a particular discretized Brownian path, plot $\int_0^t h(W(s)) \circ dW(s)$, $\int_0^t h(W(s))\, dW(s)$ and $\frac{1}{2}\int_0^t h'(W(s))\, ds$, as in Figure 5.4.

Quotes:

. . . I finally devised stochastic differential equations,
after painstaking solitary endeavours.
My first paper was thus developed; . . .
Kiyosi Ito,
source
`http://www-groups.dcs.st-and.ac.uk/~history/Mathematicians/Ito.htm`

In applications one has to make a decision
as to which kind of integral, Ito or Stratonovich,
is more appropriate. There is no clear answer
as to which type of differential equation should be used.

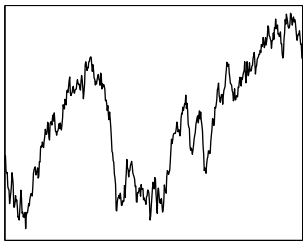
Thomas Mikosch, [45].

The Stratonovich integral was developed simultaneously by D. Fisk
in the United States, as part of his PhD thesis.
However, it was rejected for publication as being too trivial.
In the second half of his thesis he invents quasimartingales,
and that half was indeed piublished.
Robert Jarrow and Philip Protter, [34].

It is true that there is nothing in a stochastic differential equation
that is not in a Fokker–Planck equation,
but the stochastic differential equation is so much easier
to write down and manipulate that only an excessively zealous purist
would try to eschew the technique.
C. W. Gardiner, [19].

Chapter 6

The Ito Formula



Outline

- Ito formula
- Solutions of stochastic differential equatons

6.1 Motivation

We saw in Chapter 4 that stochastic calculus is not a straightforward extension of the deterministic case. In this chapter we give further evidence of this principle by outlining the Ito formula, which is a stochastic version of the chain rule. The Ito formula is a fundamental result that has many applications. We will use it in this chapter to obtain solution solutions to some simple SDEs. In later chapters we will use the Ito formula to analyze convergence of numerical methods and to derive higher order methods.

6.2 The Ito Formula

If $x(t)$ solves the deterministic ODE $dx/dt = f(x(t))$, then for any smooth function v the chain rule says that

$$\frac{d}{dt}v(x(t)) = \frac{dv(x(t))}{dx} \times \frac{dx}{dt} = v'(x(t))f(x(t)). \tag{6.1}$$

Now suppose that $X(t)$ solves the Ito SDE (5.2). What is the appropriate analogue of (6.1)? A reasonable guess is $dv = (dv/dX)dX$, so that

$$dv(X(t)) = v'(X(t))(f(X(t))dt + g(X(t))dW(t)).$$

However, this turns out not to be correct[1], an extra term arises, so that

$$dv(X(t)) = v'(X(t))dX + \tfrac{1}{2}g(X(t))^2 v''(X(t))dt,$$

Using (5.2) this becomes

$$dv(X(t)) = \left(f(X(t))v'(X(t)) + \tfrac{1}{2}g(X(t))^2 v''(X(t))\right)dt + g(X(t))v'(X(t))dW(t). \tag{6.2}$$

The result (6.2) is a special case of the Ito formula. We will continue by stating the general result and then give an intuitive explanation that also acts as a memory aid.

To state the Ito formula, we first need the concept of an *Ito process*, which is simply a process $S(t)$ of the form

$$dS(t) = A(t)dt + B(t)dW(t). \tag{6.3}$$

This is, of course, just a shorthand notation for

$$S(t) = S(0) + \int_0^t A(s)\,ds + \int_0^t B(s)\,dW(s). \tag{6.4}$$

Here, $A(s)$ and $B(s)$ can themselves be stochastic processes, and we are glossing over the issue of what conditions they must satisfy. Given a function $u : \mathbb{R}\times\mathbb{R} \to \mathbb{R}$, we may then let

$$Y(t) := u(t, S(t)).$$

The *Ito formula* says that

$$dY(t) = u_t(t, S(t))dt + u_x(t, S(t))dS(t) + \tfrac{1}{2}u_{xx}(t, S(t))B(t)^2 dt. \tag{6.5}$$

Here, subscripts denote partial derivatives, so that

$$u_t \equiv \frac{\partial u}{\partial t}, \quad u_x \equiv \frac{\partial u}{\partial x}, \quad u_{xx} \equiv \frac{\partial^2 u}{\partial x^2}.$$

Inserting (6.3), we may write (6.5) as

$$dY(t) = (u_t(t, S(t)) + u_x(t, S(t))A(t) + \tfrac{1}{2}u_{xx}(t, S(t))B(t)^2)dt + u_x(t, S(t))B(t)dW(t). \tag{6.6}$$

We emphasize that (6.6) is merely a shorthand notation for the integral version

$$Y(t) - Y(0) = \int_0^t u_t(r, S(r)) + u_x(r, S(r))A(r) + \tfrac{1}{2}u_{xx}(r, S(r))B(r)^2 dr + \int_0^t u_x(r, S(r))B(r)dW(r). \tag{6.7}$$

It is the u_{xx} term in (6.7) that could not be predicted from deterministic calculus, and for this reason that term is sometimes referred to as the *Ito correction*. The

[1] It *is* correct in the Stratonovich framework, but here we are using Ito calculus.

Ito formula is also known as the *Ito lemma*. The simpler version (6.2) corresponds to the case where u is independent of t, so $u_t \equiv 0$, with $A(s) \equiv f(X(s))$ and $B(s) \equiv g(X(s))$.

We can get an intuitive understanding of the Ito formula by using the "multiplicaton table"

$$\begin{array}{cc} dtdt = 0 & dtdW(t) = 0 \\ dW(t)dt = 0 & dW(t)dW(t) = dt \end{array}$$

to arrive at the same result. If δt and δS are small, deterministic perturbations then, using a standard Taylor series expansion in the two variables,

$$u(t+\delta t, S+\delta S) = u(t,s) + u_t(t,S)\delta t + u_x(t,S)\delta S + \tfrac{1}{2}u_{tt}(t,S)\delta t^2 + u_{tx}(t,S)\delta t\delta S + \tfrac{1}{2}u_{xx}(t,S)\delta S^2 + \text{higher order terms.}$$

This suggests that we may regard du as

$$du = u_t dt + u_x dS + \tfrac{1}{2}u_{tt}dt^2 + u_{xt}dtdS + \tfrac{1}{2}u_{xx}dS^2.$$

Substituting $Adt + BdW$ for dS and using the multiplication table gives

$$du = \left(u_t + u_x A + \tfrac{1}{2}u_{xx}B^2\right)dt + u_x B dW,$$

in agreement with (6.6).

Although the multiplication table has no mathematical meaning, it provides a useful device for those who have difficulty remembering the Ito formula and prefer to "derive" it from a Taylor expansion. Also, the analysis in section 4.4 gives some feeling for why "$dW^2 = dt$" might lead to a valid conclusion.

6.3 Using the Ito Formula

We will now show how the Ito formula can be used to find solutions of simple SDEs. More precisely, we use it to confirm that expressions "pulled out of a hat" are SDE solutions.

First we consider linear SDEs with multiplicative noise. Letting $A(t) \equiv 0$ and $B(t) \equiv 1$ in (6.3), so that $S(t) \equiv W(t)$, and letting $u(t,x) = y_0 e^{\mu t + \sigma x}$, the Ito formula (6.6) says that

$$dY(t) = y_0 e^{\mu t + \sigma W(t)}\left((\mu + \tfrac{1}{2}\sigma^2)dt + \sigma dW(t)\right),$$

which has the form

$$dY(t) = (\mu + \tfrac{1}{2}\sigma^2)Y(t)dt + \sigma Y(t)dW(t). \tag{6.8}$$

We have thus shown that

$$Y(t) = Y(0)e^{\mu t + \sigma W(t)} \tag{6.9}$$

solves the SDE (6.8)[2].

Equivalently, the solution to the SDE

$$dY(t) = \mu Y(t)dt + \sigma Y(t)dW(t). \tag{6.10}$$

has the form

$$Y(t) = Y(0)e^{(\mu - \frac{1}{2}\sigma^2 t) + \sigma W(t)}. \tag{6.11}$$

Exercise 6.1 asks you to confirm this.

Next, again letting $A(t) \equiv 0$ and $B(t) \equiv 1$ in (6.3), so that $S(t) \equiv W(t)$, this time we let $u(t,x) = \sinh(C + t + x)$, where C is a constant. From (6.6)

$$\begin{aligned} dY(t) &= (\cosh(C + t + W(t)) + \tfrac{1}{2}\sinh(C + t + W(t)))dt + \cosh(C + t + W(t))dW(t) \\ &= \left(\sqrt{1 + \sinh^2(C + t + W(t))} + \tfrac{1}{2}\sinh(C + t + W(t))\right)dt \\ &\quad + \sqrt{1 + \sinh^2(C + t + W(t))}dW(t) \\ &= \left(\sqrt{1 + Y(t)^2} + \tfrac{1}{2}Y(t)\right)dt + \sqrt{1 + Y(t)^2}dW(t). \end{aligned}$$

So $Y(t) = \sinh(C + t + W(t))$ solves the SDE

$$dY(t) = \left(\sqrt{1 + Y(t)^2} + \tfrac{1}{2}Y(t)\right)dt + \sqrt{1 + Y(t)^2}dW(t)$$

with initial condition $Y(0) = \sinh(C)$.

Taking $A(t) \equiv 0$ and $B(t) \equiv \sigma e^{\lambda t}$ in (6.3) we have

$$S(t) = \sigma \int_0^t e^{\lambda s} dW(s).$$

Choosing $u(t,x) = e^{-\lambda t}x$ in (6.6) then gives

$$\begin{aligned} dY(t) &= -\lambda S(t)e^{-\lambda t}dt + e^{-\lambda t}\sigma e^{\lambda t}dW(t) \\ &= -\lambda Y(t)dt + \sigma dW(t). \end{aligned}$$

Hence, $Y(t) = \sigma e^{-\lambda t}\int_0^t e^{\lambda s}dW(s)$ solves the SDE

$$dY(t) = -\lambda Y(t)dt + \sigma dW(t), \tag{6.12}$$

which is known as the *Ornstein–Uhlenbeck process*.

As another example, we consider the mean-reverting square root process (5.8) with $\lambda = 1$; that is,

$$dX(t) = (\mu - X(t))\,dt + \sigma\sqrt{X(t)}dW(t). \tag{6.13}$$

[2]Our argument only works for $Y(0) \in \mathbb{R}$, but the result remains true when $Y(0)$ is a random variable

Regarding this as an Ito process with $A(t) = \mu - X(t)$ and $B(t) = \sigma\sqrt{X(t)}$, we may let $u(x) = \sqrt{x}$ to get, from (6.6),

$$\begin{aligned} dY &= \left(\frac{1}{2\sqrt{X(t)}}(\mu - X(t)) - \tfrac{1}{2}\frac{\sigma^2 X(t)}{4X(t)^{3/2}} \right) dt + \frac{1}{2\sqrt{X(t)}}\sigma\sqrt{X(t)}dW(t) \\ &= \left(\frac{4\mu - \sigma^2}{8\sqrt{X(t)}} - \frac{\sqrt{X(t)}}{2} \right) dt + \tfrac{1}{2}\sigma dW(t). \end{aligned}$$

We conclude that the square root of the solution to the mean-reverting square root process, $V(t) = \sqrt{X(t)}$, satisfies the SDE

$$dV = \left(\frac{4\mu - \sigma^2}{8V(t)} - \tfrac{1}{2}V(t) \right) dt + \tfrac{1}{2}\sigma dW(t). \tag{6.14}$$

As a final example, we take $A(t) \equiv 0$ and $B(t) \equiv 1$ in (6.3), so that $S(t) \equiv W(t)$, and let $u(x) = x^m$. Then the Ito formula (6.6) gives

$$d(W(t)^m) = \tfrac{1}{2}m(m-1)W(t)^{m-2}dt + mW(t)^{m-1}dW(t).$$

In the case where $m = 2$, we have

$$d(W(t)^2) = dt + 2W(t)dW(t),$$

or, in integral form,

$$W(t)^2 - W(0)^2 = t + 2\int_0^t W(s)\, dW(s).$$

This agrees with the expression (4.5) that we derived from first principles.

6.4 Notes and References

Eemphasize: SDe solutions found by 'trick's' few general techniques Hence num methods important.

Exercises

6.1. $\star$ Give that (6.9) solves (6.8), show that (6.11) solves (6.10).

6.2. $\star\star$ Use the Ito formula to show that $d(tW(t)) = Wdt + tdW(t)$, and hence verify the expression (4.8) that was derived from first principles. [Hint: take $A \equiv 0$ and $B \equiv 1$.]

6.3. ⋆⋆ Consider the linear SDE

$$dS(t) = \mu S(t)dt + \sigma S(t)dW(t), \quad S(0) = S_0,$$

where S_0 is constant. Taking $A(t) = \mu S(t)$ and $B(t) = \sigma S(t)$ in (6.3) and letting $u(t, x) = \log x/S_0$ in (6.6), show that $Y(t) = \log(S(t)/S_0)$ satisfies

$$dY(t) = (\mu - \tfrac{1}{2}\sigma^2)dt + \sigma dW(t).$$

Integrate this to give

$$Y(t) = (\mu - \tfrac{1}{2}\sigma^2)t + \sigma W(t).$$

[This is another way to show that (6.10) has solution (6.11).]

6.4. ⋆ Consider the Vasicek model, (5.11), also known as the mean-reverting Ornstein-Uhlenbeck process, with $X(0)$ constant. From Exercise 5.2 it follows that $m(t) := \mathbb{E}[X(t)]$ satisfies $dm(t) = \lambda(\mu - m(t))dt$. Subtracting this from (5.11), show that $Y(t) := X(t) - m(t)$ satisfies the Ornstein-Uhlenbeck process (6.12) with initial condition $Y(0) = 0$. Deduce that $X(t) = \mu + e^{-\mu t}(X(0) - \mu) + \sigma e^{-\lambda t} \int_0^t e^{\lambda s} dW(s)$.

6.5. ⋆⋆ Let $A(t) \equiv 0$ and $B(t) \equiv 1$ in (6.3), so that $S(t) = W(t)$, and let $u(t, x) = 1/(1 - x)$. Use the Ito formula (6.6) to show that

$$Y(t) = \frac{1}{1 - W(t)}$$

solves the SDE

$$dY(t) = Y(t)^3 dt + Y(t)^2 dW(t).$$

[This solution exists only until the first time that $W(t) = 1$; see, for example, [B and Z, Example 7.8].]

6.6. ⋆⋆⋆ For the linear SDE (5.4), apply the Ito formula to $u(x) = x^2$ and then take expectations in order to get an expression for $\mathbb{E}[X(t)^2]$. Check your answer by computing $\mathbb{E}[X(t)^2]$ directly via (5.5) and Exercise 3.3.

6.7. ⋆⋆⋆ Use the Ito formula and the expression (5.9) to verfiy the result (5.10) for the mean-reverting square root process (5.8).

Prog of chapter: plots of (6.11) paths.

Programming Exercises

1. Plot OU process.
2. Do KDE histogram of OU process at some time T.

Quotes:

The Itô Lemma is the most important tool in Itô calculus.

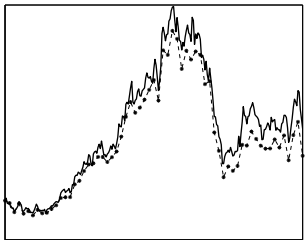
Thomas Mikosch, [45].

Over and over, Björk shows that the secret of success in Financial Engineering
is "RAIL"
which stands for the "Relentless Application of Ito's Lemma".
Bruce Graham, reviewing the book *Arbitrage Theory in Continuous Time*, by Tomas Björk, see
`http://www.amazon.com`

Obviously, this key theorem of Itô is much more than the status the lowly nomenclature "lemma" affords it,
and we prefer Itô's own description: "formula".
Robert Jarrow and Philip Protter, [34].

Chapter 7

Euler–Maruyama



Outline

- Euler–Maruyama
- Weak convergence
- Weak Euler–Maruyama method
- Strong convergence

7.1 Motivation

We are now in a position to begin the main topic of this book—the numerical simulation of SDEs. We will look at the simplest, most widely used method, and illustrate its basic convergence properties.

7.2 Euler–Maruyama

Suppose we have to solve numerically the SDE

$$dX(t) = f(X(t))dt + g(X(t))dW(t), \quad 0 \le t \le T, \quad X(0) \text{ given.} \tag{7.1}$$

Following the standard discretization approach used for deterministic ODEs, we define a *stepsize* $\Delta t = T/N$ for some integer N and compute approximate solutions at times $t_i = i\Delta t$. By definition, the exact solution of (7.1) satisfies

$$X(t_{n+1}) = X(t_n) + \int_{t_n}^{t_{n+1}} f(X(s))\,ds + \int_{t_n}^{t_{n+1}} g(X(s))\,dW(s). \tag{7.2}$$

Letting X_n denote our approximation to $X(t_n)$, the *Euler–Maruyama* (EM) method proceeds according to

$$X_{n+1} = X_n + \Delta t f(X_n) + \Delta W_n g(X_n), \tag{7.3}$$

where $\Delta W_n := W(t_{n+1}) - W(t_n)$ is a Brownian path increment. Comparing (7.2) and (7.3) we see that over $[t_n, t_{n+1}]$ the EM method is using simple left-endpoint Riemann sums to approximate the two integrals. This is consistent with the definition of the Ito integral in Chapter 4.

Just as we computed discrete Brownian paths in section 3.3, so we can compute discrete, approximate solution paths using (7.3). Since $\Delta W_n \sim N(0, \Delta t)$, a pseudo-code for such a simulation is

Δt = T/N
compute X(0) from the initial data distribution
for n = 0 to N-1
    compute a N(0,1) sample ξ_n
    X(n+1) = X(n) + Δt f(X(n)) + $\sqrt{\Delta t}\xi_n$ g(X(n))
end

Computational Example In Figure 7.1 we applied EM to the linear SDE (5.4) with $\mu = 2$, $\sigma = 0.1$ and $X(0) = 1$ for $0 \le t \le 1$. To begin, we computed a discretized Brownian path, as described in section 3.3, with resolution $\delta t = 2^{-8}$. We used a stepsize of $\Delta t = 4\delta t$ for the numerical method. The solid line shows the exact solution (5.5). Here, we evaluated the solution at the discrete points on the path, and joined the dots for clarity. Asterisks show the discrete numerical solution, joined with with a dashed line. We see that the numerical method gives a reasonable approximation to the true solution path. The discrepancy at the endpoint, $t = 1$, is 0.69. Reducing the stepsize to $\Delta t = 2\delta t$ and then $\Delta t = \delta t$ produced endpoint discrepancies of 0.16 and 0.08, respectively. $\diamondsuit$

It seems reasonable to suggest that the EM approximation $X_n \approx X(t_n)$ will generally improve as N inceases; that is, as Δt is reduced. There are two standard approaches for measuring the error $X_n - X(t_n)$, leading to the concepts of weak and strong convergence. We address them in turn in the next two sections.

7.3 Weak Convergence

In many applications, we are only concerned with the average solution behavior. Given a function Φ, the *weak error*

$$e_{\Delta t}^{\text{weak}} := \sup_{0 \le t_n \le T} |\mathbb{E}[\Phi(X_n)] - \mathbb{E}[\Phi(X(t_n))]| \tag{7.4}$$

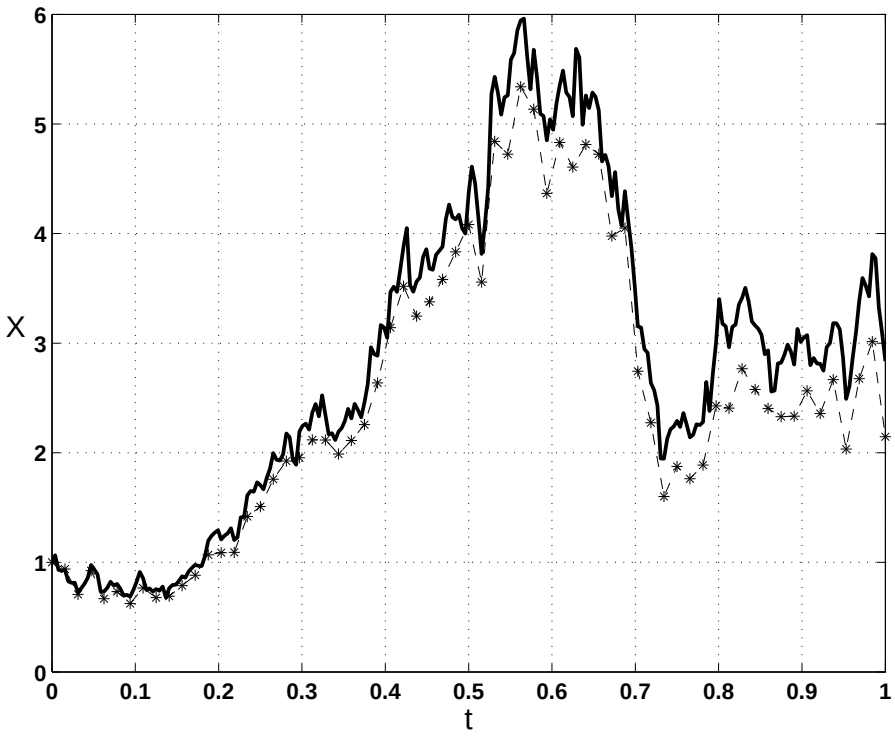


Figure 7.1: True solution (solid) and EM approximation (asterisks).

measures how well the method can approximate the mean of $\Phi(X(t))$. It is typical to restrict Φ to be a member of a class of functions, such as polynomials of degree at most k. We say that the method *converges weakly* if, for any Φ in this class,

$$e_{\Delta t}^{\text{weak}} \to 0, \quad \text{as } \Delta t \to 0.$$

Further, we say that a method has *weak order of convergence* p if there exist a constant K and a stepsize level $\Delta t^\star$ (both depending upon Φ) such that[1]

$$e_{\Delta t}^{\text{weak}} \leq K \Delta t^p, \quad \text{for all } 0 < \Delta t \leq \Delta t^\star.$$

In practice we must estimate $\mathbb{E}[\Phi(X(t_n))]$ by Monte Carlo simulation over many paths. The error definition (7.4) involves the exact expected value $\mathbb{E}[\Phi(X_n)]$ of the numerical solution, but of course in reality there is an inevitable "$1/\sqrt{M}$" size error from the sample mean approximation—see section 2.3.

Computational Example The left-hand picture in Figure 7.2 illustrates weak convergence of the EM method on the linear SDE (5.4) with $\mu = 2$, $\sigma = 0.1$ and $X(0) = 1$. Note from (5.6) that the solution has expected

[1]More precisely, we take p to be the *largest* value for which this holds.

value $\mathbb{E}[X(t)] = e^{\mu t}$. First, for a stepsize $\Delta t = 2^{-5}$ we computed $|a_M - e^{\mu T}|$, where a_M is the sample average of the EM endpoint approximation, X_N, for 10^5 discrete Brownian paths over $[0, 1]$. Thus, we are recording the weak endpoint error in the EM sample average. We then repeated this computation, with the same number of Brownian paths, for different stepsizes $\Delta t = 2^{-6}, 2^{-7}, 2^{-8}$ and 2^{-9}. In the picture, we use logarithmic scaling on both axes, so we are actually plotting the log of these errors on the vertical axis against $\log(\Delta t)$ on the horizontal axis. These points appear as asterisks, and they are joined by straight lines for clarity. If we had a relation of the form

$$|\mathbb{E}[X_N] - \mathbb{E}[X(T)]| = C\Delta t, \tag{7.5}$$

for some constant C, then, trivially,

$$\log|\mathbb{E}[X_N] - \mathbb{E}[X(T)]| = \log(C) + \log(\Delta t).$$

For reference, we have superimposed a dashed straight line of slope 1. The two curves match well, suggesting that (7.5) is valid. We tested this further by assuming a general power law relation $\log|\mathbb{E}[X_N] - \mathbb{E}[X(T)]| = C\Delta t^q$, so that $\log|\mathbb{E}[X_N] - \mathbb{E}[X(T)]| = \log(C) + q\log(\Delta t)$. A least squares fit for $\log(C)$ and q produced a value of 1.011 for q with a least squares residual of 0.05. (See section 7.5 for more details of how to do a least squares fit.) Finally, since we are computing a sample average, rather than the exact expected value $\mathbb{E}[X_N]$, we also recorded the sample variance and the corresponding approximate 95% confidence intervals (2.6). These had maximum width of less than 10^{-2}, which is negligible compared with the overall error. Hence, we are using enough paths to make the computed quantity $|a_M - \mathbb{E}[X(T)]|$ an accurate approximation to the weak endpoint error $|\mathbb{E}[X_N] - \mathbb{E}[X(T)]|$. Of course, if smaller stepsizes were to be used, then the number of sample paths would have to be increased in order to maintain this effect; Exercise 7.3 looks at the balance between exact weak error and sample error. $\diamondsuit$

That computational example suggests that the EM method is giving weak order one convergence. It can be confirmed that this is true in general—for a wide class of problems EM converges with weak order one. An outline proof of this result is given in Chapter 8. We note that in the deterministic case, that is, when $g(x) \equiv 0$ and $X(0)$ is a constant, the classical Euler method has order one: $|x(t_n) - x_n| = O(\Delta t)$. Hence, in moving from ODEs to SDEs, the Euler method is able to maintain the same accuracy when error is viewed in a weak sense.

In fact, it turns out that to obtain weak order equal to one it is not even necessary to compute exact discrete Brownian paths. It is sufficient for the increments ΔW_n in (7.3) to have the correct low order moments. Hence, we may

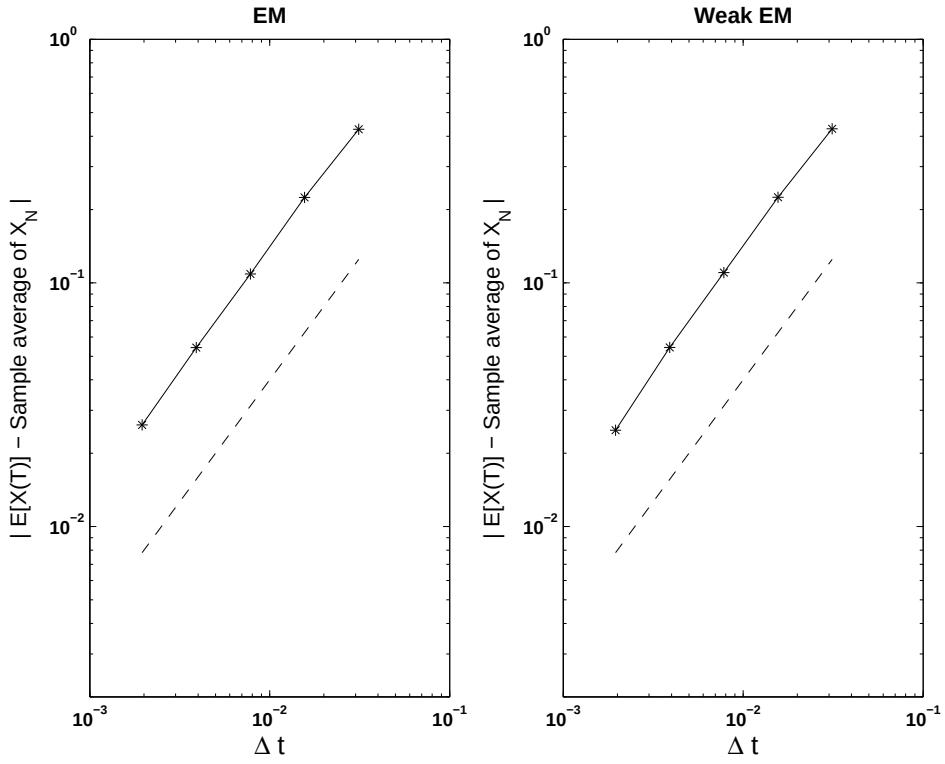


Figure 7.2: Weak endpoint error plots. Dashed line is a reference slope. Left: Euler–Maruyama. Right: Weak Euler–Maruyama.

use the *weak Euler–Maruyama* method

$$X_{n+1} = X_n + \Delta t f(X_n) + \widehat{\Delta W}_n g(X_n), \tag{7.6}$$

where $\widehat{\Delta W}_n$ are i.i.d. random variables taking the two possible values $\pm\sqrt{\Delta t}$ with equal probability; that is,

$$\mathbb{P}\left(\widehat{\Delta W}_n = \sqrt{\Delta t}\right) = \tfrac{1}{2} = \mathbb{P}\left(\widehat{\Delta W}_n = -\sqrt{\Delta t}\right). \tag{7.7}$$

Computational Example The right-hand picture in Figure 7.2 repeats the experiment from the left-hand side, using the weak EM method (7.6). We observe very similar behaviour; a least square fit gave a power of $q = 1.03$ with a residual of 0.07. This is consistent with weak convergence of order 1. $\diamondsuit$

Exercise 7.1 asks you to check that $\widehat{\Delta W}_n$ in (7.7) has $\mathbb{E}[\widehat{\Delta W}_n^r] = \mathbb{E}[\Delta W_n^r]$ for $r = 1, 2, 3$, but not for $r = 4$.

The motivation for using the weak method (7.6) is that samples from the distribution of $\widehat{\Delta W}_n$ can typically be computed more cheaply than samples from

the distribution of ΔW_n. Hence, switching to weak EM may reduce computation time on a large simulation.

7.4 Strong Convergence

Whereas weak convergence measures the "error of the means", strong convergence measures the "mean of the errors". At each time t_n, the absolute value of the error, $|X_n - X(t_n)|$, is a random variable. The *strong error*

$$e_{\Delta t}^{\text{strong}} := \sup_{0 \le t_n \le T} \mathbb{E}\left[|X_n - X(t_n)|\right] \tag{7.8}$$

is found by taking the expectation of this error at each time point. From a simulation viewpoint, we can think of the strong error as the pathwise error averaged over all paths.

We say that a method *converges strongly* if

$$e_{\Delta t}^{\text{strong}} \to 0, \quad \text{as } \Delta t \to 0,$$

and we say that a method has *strong order* p on a class of SDEs if there exist K and $\Delta t^\star$ such that[2]

$$e_{\Delta t}^{\text{strong}} \le K \Delta t^p, \quad \text{for all } 0 < \Delta t \le \Delta t^\star.$$

Computational Example We illustrate strong convergence in Figure 7.3. Here we applied EM to the linear SDE (5.4) with $\mu = 2$, $\sigma = 1$ and $X(0) = 1$. We know from from (5.5) that that the solution is $X(t) = e^{(\mu - \frac{1}{2}\sigma^2)t + \sigma W(t)}$. We computed $5,000$ discretized Brownian paths over $[0, 1]$ with $\delta t = 2^{-11}$ and, for each path, we applied EM with stepsizes δt, $2\delta t$, $4\delta t$, $16\delta t$, $32\delta t$ and $64\delta t$. We then computed the sample mean approximation to the strong endpoint error, $\mathbb{E}[|X_N - X(T)|]$ for each stepsize. These values are plotted with asterisks. We also computed the sample variance for $|X_N - X(T)|$ at each stepsize level, and found that the 95% confidence intervals (2.6) had widths that were at least an order of maginutide smaller than the sample means. So, to visual accuracy, the asterisks may be regarded as the exact strong endpoint errors. For reference, a dashed line with slope of $\frac{1}{2}$ is added to the picture. As in Figure 7.2, the axes are logarithmically scaled, so a straight line corresponds to a power-law behaviour, and in this case the behaviour is consistent with strong order of $\frac{1}{2}$. As a further check, a least-squares fit of $\log \mathbb{E}|X_N - X(T)|] = \log(C) + q \log(\Delta t)$ for $\log(C)$ and q produced a value $q = 0.51$ with a least squares residual of 0.03. $\diamondsuit$

[2]More precisely, we take p to be the *largest* value for which this holds.

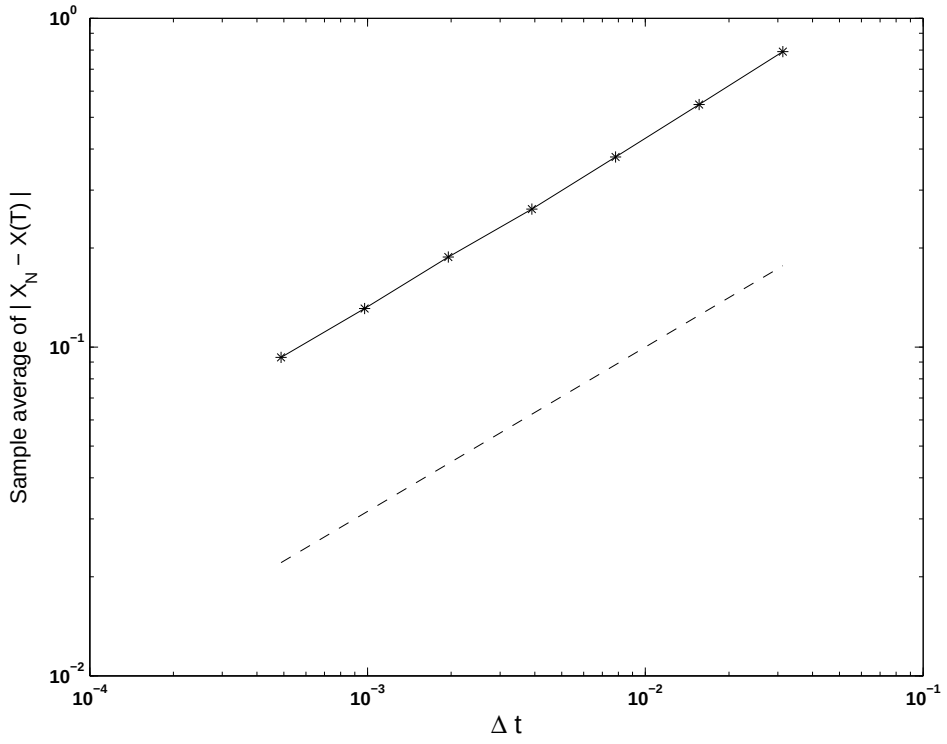


Figure 7.3: Strong endpoint error plots. Dashed line is a reference slope.

That computational experiment suggests that the EM method converges with strong order $p = \frac{1}{2}$, and this can be shown to be the case in general. An outline proof is given in Chapter 9.

We remark that the $p = \frac{1}{2}$ strong order for EM marks a significant departure from the deterministic ODE case, showing that the noise term adds genuine complexity to the computational task.

Although the strong convergence definition involves an expected value, it has implications for individual paths. If

$$\mathbb{E}\left[|X_n - X(t_n)|\right] \leq K\Delta t^{\frac{1}{2}},$$

then taking $a = \Delta t^{\frac{1}{4}}$ in the Markov inequality (1.23) gives

$$\mathbb{P}\left(|X_n - X(t_n)| \geq \Delta t^{\frac{1}{4}}\right) \leq K\Delta t^{\frac{1}{4}}$$

or, equivalently,

$$\mathbb{P}\left(|X_n - X(t_n)| < \Delta t^{\frac{1}{4}}\right) \geq 1 - K\Delta t^{\frac{1}{4}}.$$

In words, along any path the error will be small with probability close to one.

As the names suggest, strong convergence is a more stringent requirement than weak convergence in the sense that for $\Phi(x) \equiv x$ in (7.4), $e_{\Delta t}^{\text{weak}} \leq e_{\Delta t}^{\text{strong}}$;

see Exercise 7.4. We note however that this inequality is clearly not sharp in the sense that $e_{\Delta t}^{\text{weak}}$ is actually an order of $\Delta t^{\frac{1}{2}}$ smaller than $e_{\Delta t}^{\text{strong}}$.

As a final point we mention that EM can be shown to have strong order $p = 1$ on SDEs with additive noise; that is, where $g(x)$ is constant. This is explained further in section???Milstein????

Computational Example EM on OU process??? or don't bother ???? $\diamondsuit$

7.5 Notes and References

Mention that weak endpoint error can also be iused–we do this in conv proof.

Mention (1) sample mean error, random number bias, roundoff error as in SIAM Review and Ref Mitsui.

Mention here how to sample $\widehat{\Delta W}_n$ cheaply?

Mention sup over time inside expectation in strong error.

Ref original Maruyama paper.

Say how to do a least squares fit?

PK note: mention that strong is sometimes defined qith square, whence 'rate' is one.

PK note: Euler not a stochastic person. Maruyama was a student of Ito, cite 1955 paper

Exercises

7.1. $\star$ For $\widehat{\Delta W}_n$ in (7.7) and ΔW_n in (7.3), show that $\mathbb{E}[\widehat{\Delta W}_n^r] = \mathbb{E}[\Delta W_n^r]$ for $r = 1, 2, 3$, but not for $r = 4$. Also, show that $\overline{\Delta W}_n$ defined by

$$\mathbb{P}\left(\overline{\Delta W}_n = \sqrt{3\Delta t}\right) = \tfrac{1}{6} = \mathbb{P}\left(\overline{\Delta W}_n = -\sqrt{3\Delta t}\right), \quad \mathbb{P}\left(\overline{\Delta W}_n = 0\right) = \tfrac{2}{3}$$

satisfies $\mathbb{E}[\overline{\Delta W}_n^{\,r}] = \mathbb{E}[\Delta W_n^r]$ for $r = 1, 2, 3, 4, 5$. (This three point distribution can be used in the construction of methods with weak order two.)

7.2. $\star$ Show how samples from $\widehat{\Delta W}_n$ in (7.7) may be computed from calls to a uniform $(0, 1)$ pseudo-random number generator.

7.3. $\star\star$ Suppose that the EM expected value, $\mathbb{E}[X_n]$, is computed as a Monte Carlo sample mean, a_M, as in section 2.3. (This approach was used in Figure 7.2.) The triangle inequality shows that

$$|a_M - \mathbb{E}[X(t_n)]| \leq |a_M - \mathbb{E}[X_n]| + |\mathbb{E}[X_n] - \mathbb{E}[X(t_n)]|.$$

Use this to argue that, from the point of view of weak convergence, to get the smallest error for a given amount of computational effort, the number of sample paths, M, and the number of timesteps, N, should be chosen so that M is proportional to N^2.

7.4. ⋆ Suppose X is a random variable. Show that $|\mathbb{E}[X]| \leq \mathbb{E}[|X|]$. Deduce that with $\Phi(x) \equiv x$ in (7.4), $e_{\Delta t}^{\text{weak}} \leq e_{\Delta t}^{\text{strong}}$.

Prog of chapter: EM path plus true path for $dX = \mu X dt + \sigma X dW(t)$.

prog ex 1: Could do chain rule example like in SIAM review.

prog ex 2: repeat prog of chapter (EM plus true) for e.g. sinh example in Ito chapter or $1/(1 - W(t))$ example in Ex 6.4.

Quotes:

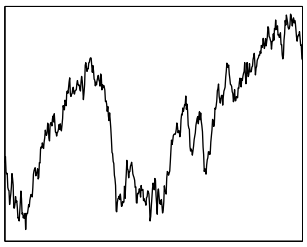
Numerical solutions are needed for different aims.
One purpose is to visualize a variety of sample paths of the solution.
...
A second objective (perhaps the most important one)
is to achieve reasonable approximations to
the distributional characteristics of the solution
to a stochastic differential equation. They include
expectations, variances, covariances and higher-order moments.
Thomas Mikosch, [45].

Since SDEs generalize ODEs, their numerical analysis must encounter
at least all the problems known for the deterministic case.
Eckhard Platen, [54].

It may very well be said that
the best way to understand SDEs is to work with their numerical solutions.
Salih N. Neftci, in An Introduction to the Mathematics of Financial
Derivatives, Academic Press, 2nd Edition, 2000.

Chapter 8

Weak Convergence



Outline

- Feynman–Kac formula
- Weak convergence proof

8.1 Motivation

We defined weak convergence in Chapter 7 and illustrated the concept numerically. We observed that the Euler–Maruyama method converges with weak order equal to one. In this chapter we give an outline proof of this result. Our aim is to get across the main ideas. In particular, we show how the Ito formula plays a role in relating SDEs to PDEs through what is known as the Feynman–Kac formula. This chapter is independent of the remainder of the book, and hence could be skipped without causing any knock-on effects.

8.2 Feynman–Kac Formula

We recall from Chapter 7 that the concept of weak convergence involves the expected value of some function of the SDE solution, $\mathbb{E}[\Phi(X(T)]$. In this section we show how this quantity arises from the solution of a related partial differential equation (PDE). This will play a crucial role in the analysis of the weak error of the EM method.

To begin, fix a time t and a point x. Now consider the SDE

$$dX(s) = f(X(s))ds + g(X(s))dW(s), \quad X(t) = x.$$

Let $u(t,x)$ be the solution of the PDE

$$u_t(t,x) + f(x)u_x(t,x) + \tfrac{1}{2}g(x)^2 u_{xx}(t,x) = 0, \qquad -\infty < x < \infty, \quad 0 < t < T, \tag{8.1}$$

$$u(T,x) = \Phi(x). \tag{8.2}$$

Here, $u(T,x) = \Phi(x)$ is a *final time* condition, specifying the solution value at time T. Applying the Ito formula (6.5) to $u(t,X(t))$, with $A(s) \equiv f(X(s))$ and $B(s) \equiv g(X(s))$ in (6.3) gives

$$\begin{aligned} du(t,X(t)) = \big(u_t(t,X(t)) + f(X(t))u_x(t,X(t)) + \tfrac{1}{2}g(X(t))^2 u_{xx}(t,X(t))\big)\,dt \\ + g(t,X(t))u_x(t,X(t))dW(t). \end{aligned}$$

From (8.1), the term in parentheses vanishes to give

$$du(t,X(t)) = g(t,X(t))u_x(t,X(t))dW(t).$$

In integral form this is

$$u(T,X(T)) - u(t,x) = \int_t^T g(t,X(t))u_x(t,X(t))dW(t).$$

On taking expectations, using the martingale property (4.10), the Ito integral disappears, leaving

$$\mathbb{E}\left[u(T,X(T))\right] = u(t,x),$$

or, from (8.2),

$$\mathbb{E}\left[\Phi(X(T))\right] = u(t,x). \tag{8.3}$$

The equation (8.1)–(8.2) is called the *Feynman–Kac* PDE and (8.3) shows that solving this PDE backwards from time T reveals the expected final-time solution functional $\mathbb{E}[\Phi(X(T))]$, given that $X(t) = x$. The relation (8.3) is known as the *Feynman–Kac* formula.

In the next section, we use the Feynman–Kac formula as part of a weak convergence proof for Euler–Maruyama. Before that, we make a short digression that relates Feynman–Kac to the classical Black–Scholes formula for financial option valuation. Options are considered further in Chapter 12.

Example At time zero (today), a *European call option* gives its holder the right to purchase a specified asset from the writer at time T at an agreed price E [27, 29, 62]. We mentioned in section 5.2 that the Black–Scholes framework models the asset price $X(t)$ at time t by the linear SDE

$$dX(t) = \mu X(t)dt + \sigma X(t)dW(t), \quad \text{with } X(0) \text{ given}, \tag{8.4}$$

where μ is the expected growth rate and σ is the volatility. Since $X(0)$ represents today's asset price, we may regard it as a known constant. Under

appropriate assumptions, it can be shown that a fair purchase price for the option at time zero may be determined from the expected payoff under the *risk-neutrality* assumption $\mu = r$, where r is the interest rate. More precisely, the purchase price is the expected value

$$\mathbb{E}\left[e^{-rT}\max\left(X(T)-E,0\right)\right]$$

when $\mu = r$ in (8.4). The Feynman–Kac formula (8.3) tells us that this fair purchase price is given by $u(0, X(0))$, where $u(t, x)$ solves the PDE (8.1)–(8.2) with

$$f(x) = rx, \quad g(x) = \sigma x, \quad \Phi(x) = e^{-rT}\max\left(x-E,0\right).$$

This PDE has solution

$$u(t,x) = xN(d_1) - Ee^{-r(T-t)}N(d_2), \tag{8.5}$$

where $N(\cdot)$ is the $\mathsf{N}(0,1)$ distribution function, defined by

$$N(y) := \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{y} e^{-\frac{s^2}{2}}\,ds, \tag{8.6}$$

and

$$d_1 = \frac{\log(x/E) + (r+\frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}}, \tag{8.7}$$

$$d_2 = d_1 - \sigma\sqrt{T-t}. \tag{8.8}$$

This is known as the *Black–Scholes formula* for the value of a European call option. Exercise 8.1 asks you to confirm that $u(t, x)$ in (8.5) solves the relevant Feynman–Kac PDE. Figure 8.1 illustrates the function $u(t, x)$ in (8.5). Here, we have superimposed as a solid white curve the value taken by u along a solution path of the SDE. This shows the how the option value varies as an asset price evolves from time $t = 0$ to $t = T$. Further details about this type of picture can be found in [27, Chapter 11]. ♢

8.3 Outline of Weak Proof

For simplicity, we will consider an SDE with constant initial data; that is

$$dX((t) = f(X(t))dt + g(X(t))dW(t), \quad \text{with } X(0) \text{ given and constant.} \tag{8.9}$$

Also, rather than $e^{\Delta t}_{\text{weak}}$ in (7.4), which involves the worst case of the weak error over all $0 \le t_n \le T$, we will focus on the weak error at the final time T. Hence our aim is to show that

$$\mathbb{E}\left[\Phi(X_N)\right] - \mathbb{E}\left[\Phi(X(T))\right] = O(\Delta t),$$

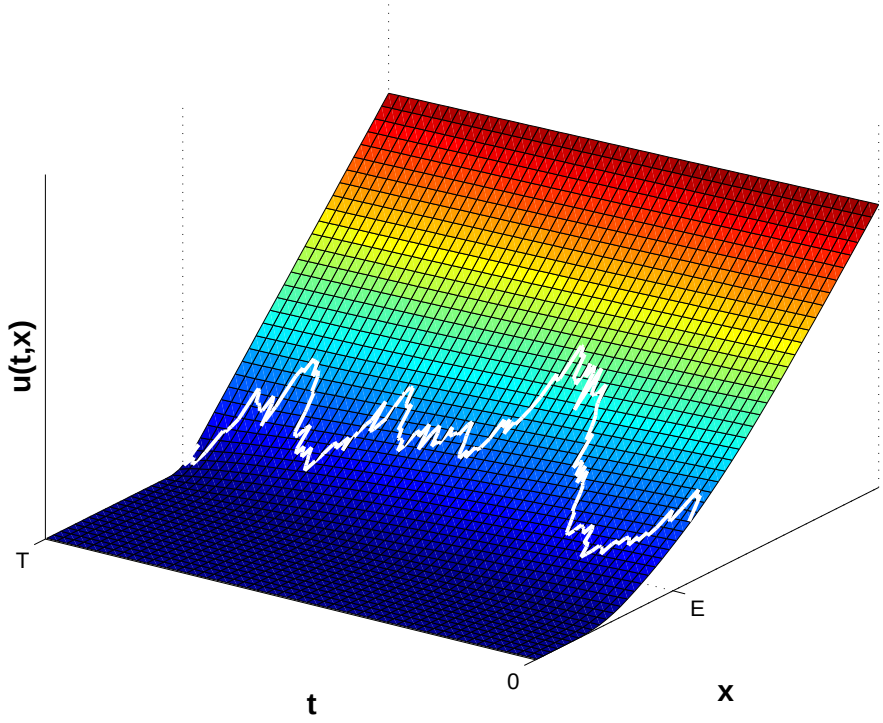


Figure 8.1: A Black–Scholes surface $u(t, x)$ from (8.5), with $u(t, X(t))$ for a single asset path superimposed.

for the EM method (7.3).

Now the EM approximation (7.3) is defined at the meshpoints $\{t_i\}_{i=0}^N$. For the purpose of analysis it is useful to extend this to an approximation that is defined for all $0 \le t \le T$. We will make use of two such interpolants. First, the "piecewise constant" or "step" process

$$\overline{X}(t) = X_n, \qquad \text{for } t_n \le t < t_{n+1}. \tag{8.10}$$

Here, our time t approximation is given by the discrete-time numerical solution from the closest earlier meshpoint. Second, the "piecewise linear" or "join-the-dots" process

$$\widehat{X}(t) = X_n + f(X_n)(t - t_n) + g(X_n)(W(t) - W(t_n)), \qquad \text{for } t_n \le t < t_{n+1}. \tag{8.11}$$

Here, the approximation at time $t_n + \gamma$, for $0 \le \gamma < \Delta t$, is given by taking an Euler–Maruyama step from X_n of length γ. We may also write $\widehat{X}(t)$ in integral form as

$$\widehat{X}(t) = X_0 + \int_0^t f(\overline{X}(s))\, ds + \int_0^t g(\overline{X}(s))\, dW(s), \tag{8.12}$$

where $\overline{X}(t)$ is defined in (8.10). Exercise 8.3 asks you to confirm this. By construction, $\widehat{X}(t_n) = X_n$ for all $0 \le n \le N$, and, in particular, $\widehat{X}(T) = X_N$. Hence,

our aim is to bound $|\mathbb{E}[\phi(\widehat{X}(T))] - \mathbb{E}[\phi(X(T))]|$. Introducing the continuous time process $\widehat{X}(t)$ satisfying (8.12) allows us to invoke the Ito formula.

To keep the analysis as straightforward as possible, we will assume that all necessary functions of the numerical solution $\overline{X}(t)$ have bounded expected value. This will cover terms such as $\mathbb{E}[u(t, \overline{X}(t))]$ and $\mathbb{E}[f(\overline{X}(t))]$ as well as more general products such as $\mathbb{E}[u_{xx}(t, \overline{X}(t))g(\overline{X}(t))^2]$. A more complete proof would require a separate stage where boundedness of these quantities is established and such analysis typically relies on the global Lipschitz conditions (5.19).

Letting $A(t) = f(\overline{X}(t))$ and $B(t) = g(\overline{X}(t))$ we may apply the Ito formula (6.5) to $u(t, \overline{X}(t))$, where $u(t, x)$ solves the Feyman–Kac PDE (8.1)–(8.2). This gives

$$\begin{aligned} du(t, \widehat{X}(t)) = \Big(u_t(t, \widehat{X}(t)) + u_x(t, \widehat{X}(t))f(\overline{X}(t)) + \tfrac{1}{2}u_{xx}(t, \widehat{X}(t))g(\overline{X}(t))^2\Big)\, dt \\ + u_x(t, \widehat{X}(t))g(\overline{X}(t))dW(t). \end{aligned} \tag{8.13}$$

At this stage, it is helpful to compactify our notation by letting $\widehat{u}$ denote $u(t, \widehat{X}(t))$, $\overline{f}$ denote $f(\overline{X}(t))$, and so on, so that (8.13) becomes

$$d\widehat{u} = \big(\widehat{u}_t + \widehat{u}_x\overline{f} + \tfrac{1}{2}\widehat{u}_{xx}\overline{g}^2\big)\, dt + \widehat{u}_x\overline{g}dW(t).$$

Now, from (8.1), we may replace $\widehat{u}_t$ by $-\widehat{u}_x\widehat{f} - \frac{1}{2}\widehat{u}_{xx}\widehat{g}^2$ to give

$$d\widehat{u} = \Big((\overline{f} - \widehat{f})\widehat{u}_x + \tfrac{1}{2}(\overline{g}^2 - \widehat{g}^2)\widehat{u}_{xx}\Big)\, dt + \widehat{u}_x\overline{g}dW(t).$$

In integral form this means

$$\begin{aligned} u(T, \widehat{X}(T)) - u(0, \widehat{X}(0)) = \int_0^T (\overline{f} - \widehat{f})\widehat{u}_x\, dt + \int_0^T \tfrac{1}{2}(\overline{g}^2 - \widehat{g}^2)\widehat{u}_{xx}\, dt \\ + \int_0^T \overline{g}\widehat{u}_x\, dW(t). \end{aligned} \tag{8.14}$$

By construction, $\widehat{X}(T) = X_N$ and $\widehat{X}(0) = X_0 = X(0)$. Also, $u(T, X_N) = \Phi(X_N)$ from (8.2), and the Feynman–Kac formula (8.3) tells us that $u(0, X(0)) = \mathbb{E}[\Phi(X(T)]$. Hence, taking expectations in (8.14) we have

$$\mathbb{E}[\Phi(X_N)] - \mathbb{E}[\Phi(X(T))] = \mathbb{E}\left[\int_0^T (\overline{f} - \widehat{f})\widehat{u}_x\, dt + \int_0^T \tfrac{1}{2}(\overline{g}^2 - \widehat{g}^2)\widehat{u}_{xx}\, dt,\right]. \tag{8.15}$$

where the Ito integral has disappeared thanks to martingale property (4.10).

From (8.15) we get the weak endpoint error bound

$$|\mathbb{E}[\Phi(X_N)] - \mathbb{E}[\Phi(X(T))]| = \left|\mathbb{E}\left[\int_0^T (\overline{f} - \widehat{f})\widehat{u}_x\, dt\right]\right| + \left|\mathbb{E}\left[\int_0^T \tfrac{1}{2}(\overline{g}^2 - \widehat{g}^2)\widehat{u}_{xx}\, dt\right]\right|. \tag{8.16}$$

This bound looks promising. The right-hand side involves the factors $\overline{f} - \widehat{f}$ and $\overline{g}^2 - \widehat{g}^2$. Here we are taking the difference between f evaluated along (i) the step function $\overline{X}(t)$ and (ii) the "join-the-dots" function $\widehat{X}(t)$, and similarly for g^2. By construction, $\overline{X}(t)$ and $\widehat{X}(t)$ match at each meshpoint, and it is reasonable to suspect that they cannot separate too much while covering the Δt distances between meshpoints. Our final step is to confirm this suspicion.

To deal with the first term on the right-hand side of (8.16), we let

$$e(t,x) = (f(X_n) - f(x))\, u_x(t,x), \quad \text{for } t \in [t_n, t_{n+1}). \tag{8.17}$$

With $A(t) \equiv f(\overline{X}(t))$ and $B(t) \equiv g(\overline{X}(t))$ we may apply the Ito formula (6.5) to give

$$e(t, \widehat{X}(t)) = \left(\widehat{e}_t + \widehat{e}_x \overline{f} + \tfrac{1}{2}\widehat{e}_{xx}\overline{g}^2\right) dt + \widehat{e}_x \overline{g} dW(t).$$

Integrating from t_n to a time $s \in [t_n, t_{n+1})$, and noting the crucial fact that $e(t_n, \widehat{X}(t_n)) = (f(X_n) - f(X_n))u_x(t_n, X_n) = 0$, we have

$$e(s, \widehat{X}(s)) = \int_{t_n}^{s} \widehat{e}_t + \widehat{e}_x \overline{f} + \tfrac{1}{2}\widehat{e}_{xx}\overline{g}^2 \, dt + \int_{t_n}^{s} \widehat{e}_x \overline{g} \, dW(t). \tag{8.18}$$

Taking expected values, and moving the expectation operation inside the integral, we find that

$$\begin{aligned}
\mathbb{E}\left[e(s, \widehat{X}(s))\right] &= \mathbb{E}\left[\int_{t_n}^{s} \widehat{e}_t + \widehat{e}_x \overline{f} + \tfrac{1}{2}\widehat{e}_{xx}\overline{g}^2 \, dt\right] \\
&= \int_{t_n}^{s} \mathbb{E}[\widehat{e}_t] + \mathbb{E}[\widehat{e}_x \overline{f}] + \tfrac{1}{2}\mathbb{E}[\widehat{e}_{xx}\overline{g}^2] \, dt.
\end{aligned}$$

Now, under our assumption that all appropriate functions of the expected value of the numerical solution are bounded, because the integral is taken over an interval $[t_n, s)$ of length less than Δt, there is a constant K_1 such that $\left|\mathbb{E}\left[e(s, \widehat{X}(s))\right]\right| \leq K_1 \Delta t$. This bound holds for any $0 \leq s \leq T$. From the definition of e in (8.17) we deduce

$$\begin{aligned}
\left|\mathbb{E}\left[\int_0^T (\overline{f} - \widehat{f})\widehat{u}_x \, dt\right]\right| &= \left|\int_0^T \mathbb{E}\left[(\overline{f} - \widehat{f})\widehat{u}_x\right] \, dt\right| \\
&\leq \left|\int_0^T K_1 \Delta t \, dt\right| \\
&\leq K_1 T \Delta t.
\end{aligned}$$

A similar argument shows that

$$\left|\mathbb{E}\left[\int_0^T \tfrac{1}{2}(\overline{g}^2 - \widehat{g}^2)\widehat{u}_{xx} \, dt\right]\right| \leq K_2 T \Delta t,$$

for some constant K_2. These are the bounds that we need to complete the analysis. Inserting them into (8.16) gives

$$|\mathbb{E}[\Phi(X_N)] - \mathbb{E}[\Phi(X(T))]| \leq (K_1 + K_2) T \Delta t.$$

8.4 Notes and References

PUt A HAT OVER THE C in KAC Mention global Lip ⇒ moment bounds.

NOTES ARE We ASSUMINg IC CONSTANT

MIGHT HAVE TO MENTION THAT WE ONLY DO FINAL TIME WEAk ERROR HERE

More goes here.

Exercises

8.1. ⋆⋆⋆ Confirm that $u(t, x)$ in (8.5) solves the Feynman–Kac PDE (8.1)–(8.2) in the case where $f(x) = rx$, $g(x) = \sigma x$ and $\Phi(x) = e^{-rT} \max(x - E, 0)$.

8.2. ⋆⋆ Could do Greeks. Ref my book for asnwers.

8.3. ⋆⋆ Confirm that $\widehat{X}(t)$ in (8.11) may be written in the form (8.12). Also, conform that $\overline{X}(t_n) = X_n$ for all $0 \leq n \leq N$.

8.4. ⋆ [On a homework exercise form the web. MATH 700, Q5?] Consider the linear SDE $dX(t) = \mu X(t)dt + \sigma X(t)\delta W(t)$. Derive formulas or the first and second moments of the solution at time T; that is, $\mathbb{E}[X(T)]$ and $\mathbb{E}[X(T)^]$. Then derive formulas for $\mathbb{E}[X_n]$ and $\mathbb{E}[X_n^2]$ from the Euler-Maruyama method applied to this problem. By letting $\Delta t \to 0$ with $T = N\delta t$ fixed, show that $\mathbb{E}[X_n]$ and $\mathbb{E}[X_n^2]$ converge to the exact solution moments $\mathbb{E}[X(T)]$ and $\mathbb{E}[X(T)^]$. Can we get order equal to 1? [You have thereby confirmed a special case of the weak convergence result proved in this Chapter.] In the SDE chapter, follow Mikosch page 156, secton 3.3.4 to derive ODE for mean, and in the Ito chapter derive ODE for second moment.

Prog of chapte: weak log-log err plot.

Prog Ex

1) Copy FK/BS figure

2) Do delta version??? Or something similar to prog of chapter.

Quotes:

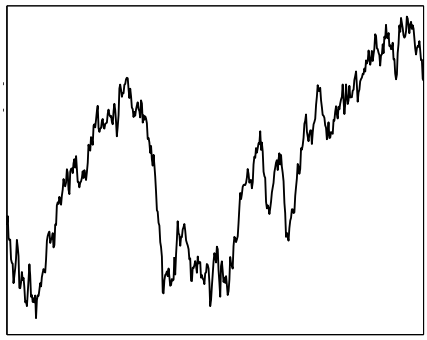
Build up your weaknesses until they become your strong points.
Knute Rockne (1888–1931), source
`http://math.furman.edu/~mwoodard/mquot.html`

The race is not [always] to the swift, nor the battle to the strong.
Ecclesiastes 9:11

Life is largely a matter of expectation.
Horace (65 BC - 8 BC), source
`http://math.furman.edu/~mwoodard/mquot.html`

Chapter 9

Strong Co



Outline

- Background assumptions
- Strong convergence proof

9.1 Motivation

Our numerical experiments in Chapter 7 indicated that EM converges with strong order $\frac{1}{2}$. In this chapter we sketch a proof of this result. Important ingredients will be the Cauchy–Schwarz inequality, the Ito Isometry and Gronwall's Lemma. Like Chapter 8 on weak convergence, this is a stand-alone chapter that you can skip without affecting your understanding of the remainder of the book.

9.2 Setting Up the Proof

Our aim is to prove that the EM method (7.3) satisfies

$$\sup_{0\leq t_n\leq T} \mathbb{E}\left[|X_n - X(t_n)|\right] \leq C\Delta t^{\frac{1}{2}}, \tag{9.1}$$

for all sufficiently small Δt, where C does not depend on Δt. As in the weak convergence proof of Chapter 8, it is convenient to move to a continuous time approximation. Recalling the definition of the "piecewise constant" process $\overline{X}(t)$ in (8.10), we will let

$$Z(t) := \sup_{0\leq s\leq t} \mathbb{E}\left[\left(\overline{X}(s) - X(s)\right)^2\right].$$

If we can show that

$$Z(t) \leq K\Delta t, \tag{9.2}$$

for all sufficiently small Δt, where K does not depend on Δt, then, by the Lyapunov inequality (1.26) we have

$$\sup_{0\leq s\leq T} \mathbb{E}\left[|\overline{X}(s) - X(s))|\right] \leq K^{\frac{1}{2}}\Delta t^{\frac{1}{2}}.$$

By construction, $\overline{X}(t) = X_n$ for all $0 \leq t_n \leq T$, and hence the required result (9.1) would follow.

In the deterministic setting, *Gronwall's Lemma* is a key tool for establishing convergence of numerical methods, and we will also find it useful in our strong SDE convergence context. Gronwall's Lemma says that if the deterministic function $\alpha(t)$ is integrable and

$$0 \leq \alpha(t) \leq A + B\int_0^t \alpha(r)\,dr, \quad 0 \leq t \leq T,$$

for some constants $A, B > 0$, then

$$\alpha(t) \leq Ae^{Bt}, \quad 0 \leq t \leq T.$$

Exercise 9.1 asks you to prove this result. It follows from Gronwall's Lemma that an inequality of the form

$$Z(t) \leq D\Delta t + E\int_0^t Z(r)\,dr, \quad 0 \leq t \leq T, \tag{9.3}$$

implies our desired result (9.2) with $K = De^{ET}$. Hence, our aim in the next section is to establish (9.3).

To obtain the result, we will assume that the drift and diffusion coefficients satisfy the global Lipschitz conditions (5.19). We note that this assumption implies the *linear growth conditions*

$$f(x)^2 \leq \widehat{L}\left(1 + x^2\right) \quad \text{and} \quad g(x)^2 \leq \widehat{L}\left(1 + x^2\right), \quad \text{for all } x \in \mathbb{R}, \tag{9.4}$$

for some constant $\widehat{L}$; see Exercise 9.2. We will also assume that the exact and numerical second moments, $\mathbb{E}[X(t)^2]$ and $\mathbb{E}[\overline{X}(t)^2]$, are bounded over $[0, T]$.

9.3 Outline of Strong Proof

Given any point $s \in [0, T]$, we let n_s denote the integer such that $s \in [t_{n_s}, t_{n_s+1})$. (In other words, $\overline{X}(s) = X_{n_s}$.) We then have

$$\overline{X}(s) - X(s) = X_{n_s} - X(s)$$

$$
\begin{aligned}
&= X_{n_s} - \left(X_0 + \int_0^s f(X(r))dr + \int_0^s g(X(r))dW(r) \right) \\
&= \sum_{i=0}^{n_s-1} (X_{i+1} - X_i) - \int_0^s f(X(r))dr - \int_0^s g(X(r))dW(r) \\
&= \sum_{i=0}^{n_s-1} f(X_i)\Delta t + \sum_{i=0}^{n_s-1} g(X_i)\Delta W_i - \int_0^s f(X(r))dr - \int_0^s g(X(r))dW(r). \qquad (9.5)
\end{aligned}
$$

Now, by construction,

$$
\int_{t_i}^{t_{i+1}} f\left(\overline{X}(t)\right)\, dt = \int_{t_i}^{t_{i+1}} f\left(X_i\right)\, dt = f\left(X_i\right) \int_{t_i}^{t_{i+1}} dt = f\left(X_i\right) \Delta t, \qquad (9.6)
$$

and similarly

$$
\int_{t_i}^{t_{i+1}} g\left(\overline{X}(t)\right)\, dW(t) = g\left(X_i\right) \Delta W_i. \qquad (9.7)
$$

Hence, in (9.5),

$$
\begin{aligned}
\overline{X}(s) - X(s) &= \int_0^{t_{n_s}} f(\overline{X}(r))\, dr + \int_0^{t_{n_s}} g(\overline{X}(r))\, dW(r) - \int_0^s f(X(r))\, dr - \int_0^s g(X(r))\, dW(r) \\
&= \int_0^{t_{n_s}} f(\overline{X}(r)) - f(X(r))\, dr + \int_0^{t_{n_s}} g(\overline{X}(r)) - g(X(r))\, dW(r) \\
&\quad - \int_{t_{n_s}}^s f(X(r))\, dr - \int_{t_{n_s}}^s g(X(r))\, dW(r). \qquad (9.8)
\end{aligned}
$$

Loosely, the first two integrals on the right-hand side of (9.8) involve the errors incurred by the EM method from time zero to time t_{n_s}, and the third and fourth integrals represent the error made by approximating $X(s)$ with X_{n_s}.

Now, for any $a, b, c, d \in \mathbb{R}$ we have

$$
(a+b+c+d)^2 \leq 4(a^2+b^2+c^2+d^2), \qquad (9.9)
$$

see Exercise 9.3. It follows that squaring both sides in (9.8) and taking expected values gives

$$
\begin{aligned}
\mathbb{E}\left[\left(\overline{X}(s) - X(s)\right)^2\right] \leq 4 \Bigg\{ & \mathbb{E}\left[\left(\int_0^{t_{n_s}} f(\overline{X}(r)) - f(X(r))dr\right)^2\right] \\
& + \mathbb{E}\left[\left(\int_0^{t_{n_s}} g(\overline{X}(r)) - g(X(r))dW(r)\right)^2\right] \\
& + \mathbb{E}\left[\left(\int_{t_{n_s}}^s f(X(r))dr\right)^2\right] \\
& + \mathbb{E}\left[\left(\int_{t_{n_s}}^s g(X(r))dW(r)\right)^2\right] \Bigg\}. \qquad (9.10)
\end{aligned}
$$

We will now bound the four terms inside the curly parentheses. First, using the Cauchy–Schwarz inequality in the form of (9.16) from Exercise 9.4, and then using the Lipschitz condition (5.19), we find that

$$\begin{aligned}\mathbb{E}\left[\left(\int_0^{t_{n_s}} f(\overline{X}(r)) - f(X(r))\,dr\right)^2\right] &\le t_{n_s}\int_0^{t_{n_s}} \mathbb{E}\left[\left(f(\overline{X}(r)) - f(X(r))\right)^2\right]\,dr \\ &\le TL^2\int_0^{t_{n_s}} \mathbb{E}\left[\left(\overline{X}(r) - X(r)\right)^2\right]\,dr \\ &\le TL^2\int_0^{s} Z(r)\,dr. \end{aligned} \tag{9.11}$$

Similarly, using (9.16) and the linear growth condition (9.4),

$$\begin{aligned}\mathbb{E}\left[\left(\int_{t_{n_s}}^{s} f(X(r))\,dr\right)^2\right] &\le (s - t_{n_s})\int_{t_{n_s}}^{s} \mathbb{E}\left[f(X(r))^2\right]\,dr \\ &\le \Delta t\widehat{L}\int_{t_{n_s}}^{s} 1 + \mathbb{E}\left[X(r)^2\right]\,dr \\ &\le C_1\Delta t^2, \end{aligned} \tag{9.12}$$

for some constant C_1, under our assumption that $\mathbb{E}[X(r)^2]$ is bounded.

Using the Ito isometry (4.14) along with the Lipschitz condition (5.19), we find that

$$\begin{aligned}\mathbb{E}\left[\left(\int_0^{t_{n_s}} g(\overline{X}(r)) - g(X(r))\,dW(r)\right)^2\right] &= \int_0^{t_{n_s}} \mathbb{E}\left[\left(g(\overline{X}(r)) - g(X(r))\right)^2\right]\,dr \\ &\le L^2\int_0^{t_{n_s}} \mathbb{E}\left[\left(\overline{X}(r) - X(r)\right)^2\right]\,dr \\ &\le L^2\int_0^{s} Z(r)\,dr. \end{aligned} \tag{9.13}$$

Similarly, using the Ito isometry and the linear growth bound,

$$\begin{aligned}\mathbb{E}\left[\left(\int_{t_{n_s}}^{s} g(X(r))\,dW(r)\right)^2\right] &= \int_{t_{n_s}}^{s} \mathbb{E}\left[g(X(r))^2\right]\,dr \\ &\le \widehat{L}\int_{t_{n_s}}^{s} \mathbb{E}\left[1 + X(r)^2\right]\,dr \\ &\le C_2\Delta t, \end{aligned} \tag{9.14}$$

for some constant C_2.

Using (9.11)–(9.14) in (9.10) leads to a bound of the form (9.3), which thereby completes our proof.

9.4 Remarks About the Proof

Looking over the analysis in Section 9.3, we note from (9.12) that the deterministic integral over $[t_{n_s}, s]$ makes an $O(\Delta t^2)$ contribution to the error bound, whereas the stochastic integral contributes an $O(\Delta t)$ term via (9.14). This highlights the essential difference between the deterministic and stochastic error analyses. Without the term (9.14), the Gronwall lemma would allow us to recover the $O(\Delta t)$ error bound that holds for Euler's method applied to an ODE.

9.5 Notes and References

Say the proof is a cut down version of K & P.

Mention refs for sup inside $\mathbb{E}$. Mention HMS non-global Lipschitz work. Need to put Cauchy–Schwarz somewhere.

Citations for any theory quoted here.

Prog of chapter: from EM chapter, strong log-log plot.

Prog Ex's: 1. repeat above for another SDE 2. Pursue high prob argument on page 72 when $X(t)$ is known in terms of $W(t)$.

Exercises

9.1. $\star\star$ The scalar, linear ODE

$$\frac{du(t)}{dt} = Bu(t),$$

where $B > 0$ is a constant, has solution $u(t) = u(0)e^{Bt}$. Suppose $u(0) > 0$. Convince yourself that if $u(t)$ satisfies the *differential inequality*

$$\frac{du(t)}{dt} \leq Bu(t), \quad \text{for all } 0 \leq t \leq T,$$

then

$$u(t) \leq u(0)e^{Bt}.$$

Use this result to establish the Gronwall inequality (9.3).

9.2. $\star$ By taking $y = 0$ in (5.19), or otherwise, show that the linear growth conditions (9.4) follow from the gobal Lipschitz conditions (5.19).

9.3. $\star$ Show that $(a+b)^2 \leq 2(a^2+b^2)$ for any $a, b \in \mathbb{R}$. Hence, establish (9.9).

9.4. Given vectors $x, y \in \mathbb{R}^N$, the Cauchy–Schwarz inequality says that $\left(\sum_{i=1}^N x_i y_i\right)^2 \leq \left(\sum_{i=1}^N x_i^2\right)\left(\sum_{i=1}^N y_i^2\right)$. Hence, for any δt,

$$\left(\sum_{i=1}^N x_i y_i \delta t\right)^2 \leq \left(\sum_{i=1}^N x_i^2 \delta t\right)\left(\sum_{i=1}^N y_i^2 \delta t\right). \tag{9.15}$$

Suppose now that $a(r)$ and $b(r)$ are stochastic processes defined over a time interval $0 \leq r \leq S$. Discretizing this time interval, let $r_i = (i-1)\delta r$ with $\delta r = S/(N+1)$, and suppose that $x_i = a(r_i)$ and $y_i = b(r_i)$. We may then interpret (9.15) as an inequality involving Riemann sum type approximations to integrals. Using this approach, give a heuristic justification for the inequalities

$$\mathbb{E}\left[\left(\int_0^S a(r)b(r)dr\right)^2\right] \leq \mathbb{E}\left[\int_0^S a(r)^2 dr \int_0^S b(r)^2 dr\right]$$

and

$$\mathbb{E}\left[\left(\int_0^S b(r)dr\right)^2\right] \leq S\int_0^S \mathbb{E}\left[b(r)^2\right] dr. \tag{9.16}$$

Quotes:

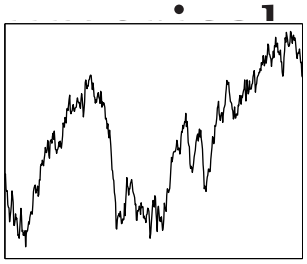
Every path serves a purpose.
Gene Oliver,
source `http://math.furman.edu/~mwoodard/mquot.html`

Don't expect to build up the weak by pulling down the strong.
Calvin Coolidge (1872–1933),
source `http://math.furman.edu/~mwoodard/mquot.html`

I dreamed a thousand new paths ...
I woke and walked my old one.
Chinese proverb,
source `http://math.furman.edu/~mwoodard/mquot.html`

Chapter 10

Implicit Methods and N... Stability



Outline

- Stochastic θ-method
- Linear test equation
- Mean-square stability
- Asymptotic stability

10.1 Motivation

Numerical analysts aim to understand the important features of computational algorithms: Are certain methods better than others? Can a method be improved? What types of problem suit this method? In general, we do not know the exact answer to the problem (if we did, we wouldn't need the numerical method!) so we cannot hope to get an exact expression for the error. For this reason, we try to gain insights by, for example,

(1) looking at asymptotic limits,

(2) looking at simple test problems.

The weak and strong convergence theory in Chapters 8 and 9 comes into the first category. There, we fix a finite time interval $[0, T]$ and look at some measure of the error in the asymptotic limit $\Delta t \to 0$, which is relevant for sufficiently small stepsizes. In this chapter we look at linear stability theory, which combines categories (1) and (2). First, we fix Δt and look at the asymptotic limit $T \to \infty$.

This is relevant to the circumstance where Δt is not particularly small and long-time behavior of the SDE solution is of interest. Second, in order to make the analysis tractable, we focus on a simple test problem.

It is well known from deterministic numerical analysis that, in terms of stability behavior, the Euler method is inferior to many implicit methods, and hence we begin by defining a useful class of implicit SDE methods on which to base our analysis.

10.2 Stochastic Theta Method

Given a parameter $\theta \in [0, 1]$, the corresponding *stochastic θ-method* for the SDE (5.2) takes the form

$$X_{n+1} = X_n + (1-\theta)\Delta t f(X_n) + \theta \Delta t f(X_{n+1}) + \Delta W_n g(X_n). \tag{10.1}$$

In the case where $\theta = 0$, this is simply the Euler–Maruyama method (7.3). For general $\theta \neq 0$, (10.1) is an implicit equation. Given X_n we must solve the equation to obtain X_{n+1}.

A pseudo-code for a θ-method simulation is thus

Δt = T/N
compute X(0) from the initial data distribution
for n = 0 to N-1
    compute a N(0,1) sample ξ_n
    solve for X(n+1) in the equation ...
    X(n+1) = X(n) + $(1-\theta)\Delta t$ f(X(n)) + $\theta \Delta t$ f(X(n+1)) + $\sqrt{\Delta t}\xi_n$ g(X(n))
end

When f is nonlinear, computing X_{n+1} means solving a nonlinear equation. This is a standard task in numerical analysis and there are many methods available. Since it is not central to this book, we simply refer the reader to the references in section 10.6. However, we do note that, in general, a nonlinear equation may have zero, one or many solutions. For (10.1), if we assume that f satisfies the global Lipschitz condition (5.19), then it can be shown that a unique solution exists for all sufficiently small Δt. Also for certain classes of nonlinear functions f, a unique solution exists for any $\Delta t > 0$.

In the deterministic case, where $g(x) \equiv 0$ and $X(0)$ is constant, (10.1) reduces to the widely studied deterministic θ-method and the particular cases $\theta = \frac{1}{2}$ and $\theta = 1$ are known as the *trapezoidal* and *backward Euler* methods, respectively.

10.3 Linear Test Equation

We focus in this chapter on the linear multiplicative noise SDE

$$dX(t) = \mu X(t)dt + \sigma X(t)dW(t), \tag{10.2}$$

where μ and σ are real constants, whose solution was found in Chapter 6 to be

$$X(t) = X(0)e^{(\mu - \frac{1}{2}\sigma^2)t + \sigma W(t)}. \tag{10.3}$$

We know from Exercise 6.6 that

$$\mathbb{E}\left[X(t)^2\right] = \mathbb{E}\left[X(0)^2\right] e^{(2\mu+\sigma^2)t}.$$

It follows that

$$\lim_{t\to\infty} \mathbb{E}\left[X(t)^2\right] = 0 \Leftrightarrow \mu + \tfrac{1}{2}\sigma^2 < 0. \tag{10.4}$$

The property on the left-hand side of (10.4) is known as *mean-square stability*, and we see that it is completely characterized by the condition $\mu + \frac{1}{2}\sigma^2 < 0$.

An alternative concept is *asymptotic stability*, which is defined as $\lim_{t\to\infty} |X(t)| = 0$, with probability one. We know that $W(t) \sim N(0,t)$ and hence $W(t)$ can be regarded as a $N(0,1)$ random variable scaled by $\sqrt{\Delta t}$. Because of this square root behavior, it is reasonable to expect the linear term, $(\mu - \frac{1}{2}\sigma^2)t$ in (10.3) to dominate as $t \to \infty$. It can be shown that this is indeed the case, so that the asymptotic stability of (10.2) is characterized through

$$\lim_{t\to\infty} |X(t)| = 0, \text{ with prob. } 1 \Leftrightarrow \mu - \tfrac{1}{2}\sigma^2 < 0. \tag{10.5}$$

Both mean-square and asymptotic stability involve the solution "tending to zero" as $t \to \infty$. However, the two definitions are different, and they have different characterizations. Loosely, mean-square stability requires that the average over all paths of the square of the solution tends to zero, whereas asymptotic stability asks for every individual solution path to tend to zero. We see from (10.4) and (10.5) that mean-square stability is stronger than asymptotic stability for the SDE (10.3), in the sense that

$$\text{mean} - \text{square stable} \Rightarrow \mu + \tfrac{1}{2}\sigma^2 < 0 \Rightarrow \mu - \tfrac{1}{2}\sigma^2 < 0 \Rightarrow \text{asymptotically stable}.$$

However, the "mean-square $\Rightarrow$ asymptotic" implication fails to hold for other classes of SDE. Generally, the two stability types should simply be regarded as useful concepts that can be studied in relation to long-time behavior. Asymptotic stability is perhaps more relevant in many applications, but mean-square stability analysis is more common as it tends to lead to more tractable mathematics.

Having defined and characterized two long-time properties that can hold for our SDE (10.2), we may now investigate whether the stochastic θ-method reproduces them. This motivates the next two sections.

10.4 Mean-Square Stability of the Theta Method

Copying the definition for the SDE, we could say that a numerical method is mean-square stable if $\lim_{n\to\infty} \mathbb{E}[X_n^2] = 0$. The θ-method applied to the test equation (10.2) gives

$$(1 - \theta\Delta t\mu)\, X_{n+1} = X_n \left(1 + (1-\theta)\Delta t\mu + \sqrt{\Delta t}\sigma V_n\right), \tag{10.6}$$

where the $V_n \sim N(0,1)$ are i.i.d. Squaring both sides and taking expectations, we find that

$$(1 - \theta\Delta t\mu)^2\, \mathbb{E}\left[X_{n+1}^2\right] = \left((1 + (1-\theta)\Delta t\mu)^2 + \Delta t\sigma^2\right) \mathbb{E}\left[X_n^2\right], \tag{10.7}$$

see Exercise 10.1. Assuming that $1 - \theta\Delta t\mu \neq 0$, we thus have

$$\mathbb{E}\left[X_{n+1}^2\right] = \frac{(1 + (1-\theta)\Delta t\mu)^2 + \Delta t\sigma^2}{(1 - \theta\Delta t\mu)^2} \mathbb{E}\left[X_n^2\right].$$

It is then clear that the mean-square stability condition for the θ-method reduces to

$$\lim_{n\to\infty} \mathbb{E}[X_n^2] = 0 \quad \Leftrightarrow \quad \frac{(1 + (1-\theta)\Delta t\mu)^2 + \Delta t\sigma^2}{(1 - \theta\Delta t\mu)^2} < 1. \tag{10.8}$$

Some simple algebra rearranges this to

$$\lim_{n\to\infty} \mathbb{E}[X_n^2] = 0 \quad \Leftrightarrow \quad \Delta t(1-2\theta)\mu^2 < -2(\mu + \tfrac{1}{2}\sigma^2), \tag{10.9}$$

see Exercise 10.2. Note that the expression $\mu + \frac{1}{2}\sigma^2$, which determines mean-square stability of the underlying SDE, conveniently appears on the right-hand side of (10.9). In comparing (10.4) and (10.9), it is useful to isolate three cases.

Case 1: $0 \leq \theta < \frac{1}{2}$**.** In this regime

$$\text{SDE stable } \Rightarrow \theta\text{−method stable for } \Delta t < \frac{2|\mu + \frac{1}{2}\sigma^2|}{(1-2\theta)\mu^2} \tag{10.10}$$

and

$$\text{SDE not stable } \Rightarrow \theta\text{−method not stable for any } \Delta t > 0. \tag{10.11}$$

Case 2: $\theta = \frac{1}{2}$**.** In this regime

$$\text{SDE stable } \Leftrightarrow \theta\text{−method stable for all } \Delta t > 0. \tag{10.12}$$

Case 3: $\frac{1}{2} < \theta \leq 1$**.** In this regime

$$\text{SDE stable } \Rightarrow \theta\text{−method stable for all } \Delta t > 0 \tag{10.13}$$

and

$$\text{SDE not stable } \Rightarrow \theta\text{−method not stable for } \Delta t < \frac{2|\mu + \frac{1}{2}\sigma^2|}{(2\theta - 1)\mu^2}. \tag{10.14}$$

Exercise 10.3 asks you to check these facts.

In summary, for $0 \leq \theta < \frac{1}{2}$ the method has a *finite region of stability*—on a stable SDE the range of Δt for stability is bounded above. For $\theta = \frac{1}{2}$ the method reproduces the stability of the SDE perfectly, for any Δt. For $\frac{1}{2} < \theta \leq 1$, the method is *overstable*—it is always stable on a stable SDE but can also be stable on an unstable SDE.

The simple characterizaton (10.9) opens up a neat visualization of the stability properties. First, we let

$$x := \Delta t\mu \qquad \text{and} \qquad y := \Delta t\sigma^2.$$

Note that, by construction, $y \geq 0$ and hence the points (x, y) are restricted to the upper half of the x-y plane. With this notation, the condition $\mu + \frac{1}{2}\sigma^2 < 0$ in (10.4) that determines stability of the SDE translates into the wedge $y < -2x$. The corresponding condition (10.9) for the θ-method becomes $y < (2\theta - 1)x^2 - x$, creating a region between a parabola and the x-axis. For $\theta < \frac{1}{2}$ we have a "sad" parabola, for $\theta = \frac{1}{2}$ the parabola collapses to a line, and for $\theta > \frac{1}{2}$ the parabola is "happy". Figure 10.1 illustrates the cases $\theta = 0$, 0.5 and 0.75 and 1, for a finite portion of the upper half x-y plane. Here, the horizontal hashing gives the wedge where the SDE is stable and the vertical hashing shows the region under the parabola where the method is stable. To interpret the pictures, first we note that, given values for μ and σ in (10.2) the point $(x, y) = (\mu, \sigma^2)$ corresponds to the stepsize $\Delta t = 1$. Using a different stepsize, so that $(x, y) = (\mu\Delta t, \sigma^2\Delta t)$, corresponds to moving along the ray that conects (μ, σ^2) with the origin. To have a stable θ-method, we must be on the ray **and** below the parabola. The pictures also illustrate the relations (10.10)–(10.13).

10.5 Asymptotic Stability of the Theta Method

To analyze asymptotic stability it is convenient to write (10.6) in the form

$$X_{n+1} = (a + bV_n)X_n, \tag{10.15}$$

where

$$a := \frac{1 + (1 - \theta)\Delta t\mu}{1 - \theta\Delta t\mu}, \qquad b := \frac{\sqrt{\Delta t}\sigma}{1 - \theta\Delta t\mu},$$

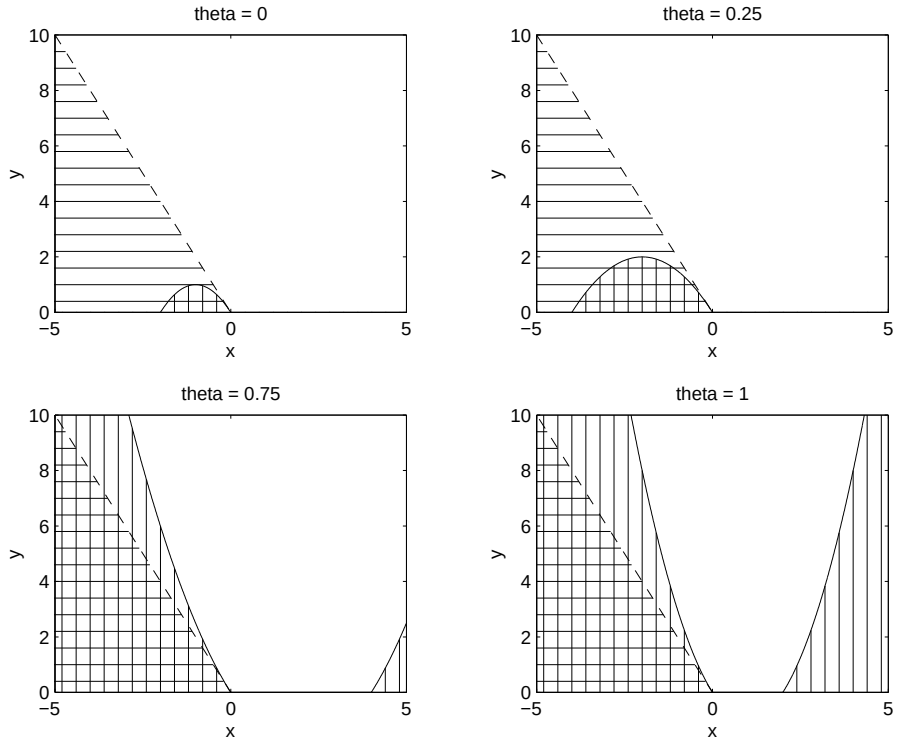


Figure 10.1: Mean-square stability regions for the stochastic θ-method (vertical hashing), and the underlying SDE (horizontal hashing). Top left: $\theta = 0$, Top right: $\theta = 0.25$, Lower left: $\theta = 0.75$, Lower right: $\theta = 1$.

and we assume $\theta \Delta t \mu \neq 1$.

From (10.15) we have

$$|X_n| = \left(\prod_{i=0}^{n-1} |a + bV_i| \right) X_0.$$

A key step is now to take logs, converting the product of i.i.d. random variables into a sum, so that

$$\log |X_n| = \log |X_0| + \sum_{i=0}^{n-1} \log |a + bV_i|. \tag{10.16}$$

Letting $\mu := \mathbb{E}[\log |a + bV_n|]$, the Strong Law of Large Numbers (Theorem 2) tells us that

$$\lim_{n \to \infty} \left(\frac{\sum_{i=0}^{n-1} \log |a + bV_i|}{n} - \mu \right) = 0, \qquad \text{with probability 1.}$$

Hence,

if $\mu < 0$: then $\lim_{n\to\infty} \sum_{i=0}^{n-1} \log|a+bV_i| = -\infty$ with probability 1, so that, from (10.16), $\lim_{n\to\infty} |X_n| = 0$, and

if $\mu > 0$: then $\lim_{n\to\infty} \sum_{i=0}^{n-1} \log|a+bV_i| = \infty$ with probability 1, so that, from (10.16), $\lim_{n\to\infty} |X_n| = \infty$.

Analysis for $\mu = 0$ is slightly more delicate, but it can be shown that $\lim_{n\to\infty} |X_n| = 0$ with probability 1 does not hold in this intermediate case; see section 10.6. Hence, overall, asymptotic stability of the stochastic θ-method is completely characterized by

$$\lim_{n\to\infty} |X_n| = 0 \text{ with probability 1} \qquad \Leftrightarrow \qquad \mathbb{E}[\log|a+bV_n|] < 0.$$

The condition on the right-hand side may be rearranged to

$$\log|a| + \gamma(c) < 0, \qquad \text{where } c := b/a \text{ and } \gamma(c) := \mathbb{E}[\log|1+cV_n|]. \tag{10.17}$$

There is no simple analytical expression for $\gamma(c)$, but given values for μ, σ, θ and Δt, we could approximate $\gamma(c) := (1/\sqrt{2\pi}) \int_{-\infty}^{\infty} \log|1+cx| e^{-s^2/2}\, ds$ using numerical quadrature and hence check whether condition (10.17) holds. In Figure 10.2, we plot $\gamma(c)$ for $c > 0$. A symmetry argument shows that the stability is unchanged if $a \mapsto -a$ or $b \mapsto -b$, so the restriction to $c > 0$ is not significant; see Exercise 10.4.

For the special case $\theta = 1$, it is possible to visualize an asympotic stability region directly from Figure 10.2. Recalling that $x := \Delta t\mu$ and $y := \Delta t\sigma^2$, when $\theta = 1$ we have $a = 1/(1-x)$ and $b = \sqrt{y}/(1-x)$, so that $c = b/a = \sqrt{y}$. it follows from (10.17) that the θ-method is stable when $x < 1 - e^{\gamma(\sqrt{y})}$ and when $x < 1 + e^{\gamma(\sqrt{y})}$. The result is shown with vertical hashing in Figure 10.3. From (10.5) the SDE is stable when $y > 2x$, and this wedge is shown with horizontal hashing.

Although the condition (10.17) is not analytically convenient, it is possible to prove that the favourable mean-square stability results that we saw in section 10.4 do not carry through to this asymptotic stability setting: even for $\frac{1}{2} \le \theta \le 1$ there are μ, σ and Δt for which the SDE is stable, but not the stochastic θ-method; see section 10.6.

NUMERICAL EXPERIMENTS First, copy from SIAM rev Fig 5 and Fig 6. Add a comment about wierdness of as but not m-s. Second, copy from SINUM Fig 7.1 and 7.2

10.6 Notes and References

Mention reasons why linear test problem is relevant. E.g. in my SciCADE talk.]

Ref for existence/uniqueness of implicit equation.

Refs for solving nonlnear eqns (sec 10.2)

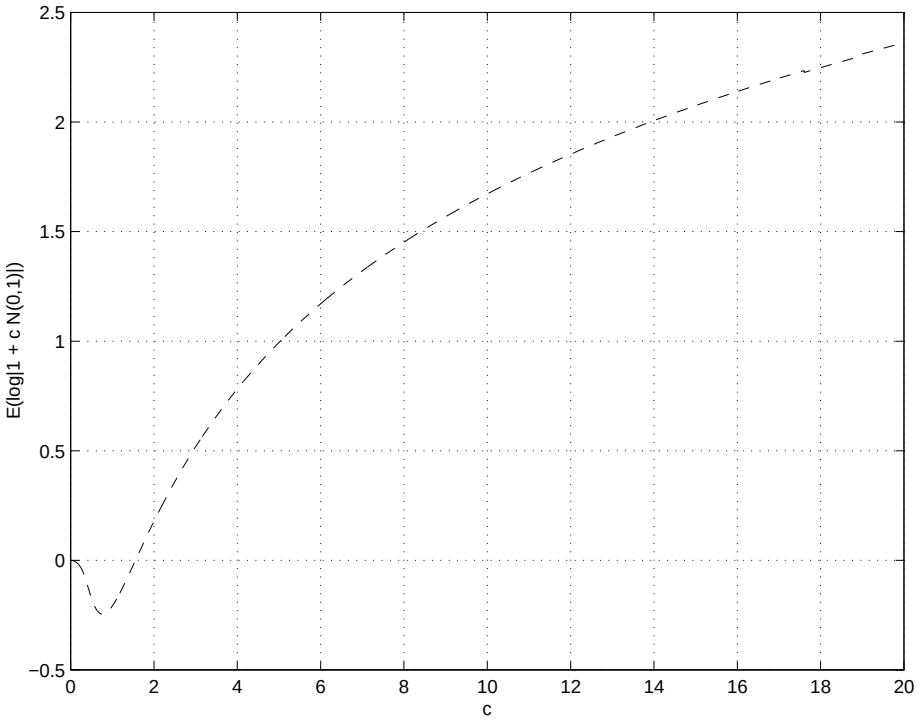


Figure 10.2: Plot of $\gamma(c) := \mathbb{E}[\log|1 + cV_n|]$ against c.

Refs for existence/uniqueness of solutions to (10.1).
Mention LIL result for dealing with $\mu = 0$
Mention balanced methods.
Ergodicity?
Ref Mao for proving (10.5?
General refs for SDE stability (e.g. Mao). mention μ, σ complex.
Add an exercise on invariant measures for θ-method on $dX = \lambda X dt + \sigma dW$ (see Alan Bryden's thesis).

Exercises

10.1. $\star$ Confirm that (10.7) follows from (10.6).

10.2. $\star$ Confirm that (10.9) follows from (10.8).

10.3. $\star\star$ Show that (10.10)–(10.14) follow from (10.4) and (10.9).

10.4. $\star$ Show that the asymptotic stability condition $\log|a| + \gamma(c) < 0$ in (10.17) is unchanged if a is replaced by $-a$ or if b is replaced by $-b$.

Quotes:

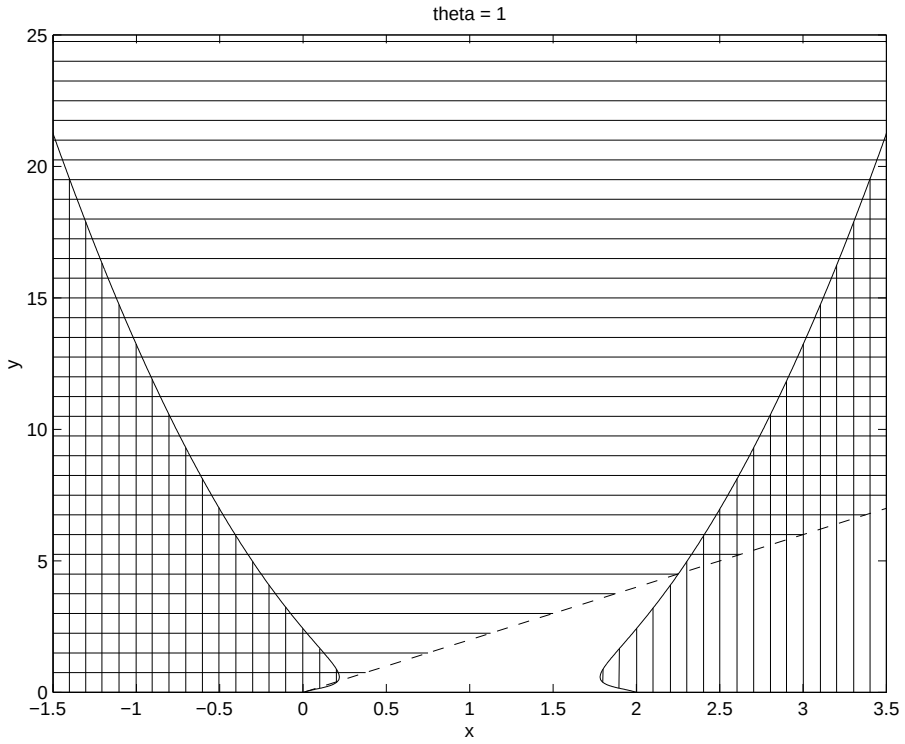


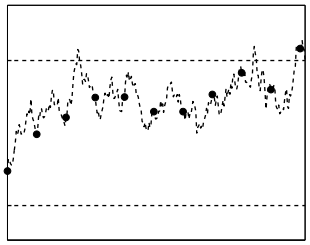
Figure 10.3: Asymptotic stability region for the stochastic θ-method with $\theta = 1$ (vertical hashing) and the underlying SDE (horizontal hashing).

The first step towards amendment is the recognition of error.
Seneca (5BC–65AD),
source `http://math.furman.edu/~mwoodard/mquot.html`

What really matters in a numerical scheme is that it should be
numerically stable, can be conveniently implemented,
and generates fast highly accurate results.
Eckhard Platen, source *An introduction to numerical methods for stochastic differential equations*, Acta Numerica, 1999, 197–246.

Chapter 11

Mean Exit Times



Outline

- Mean exit time
- Monte Carlo approach
- Computational experiments

11.1 Motivation

When does a particle first roam more that a set distance from its starting point? When does an interest rate first leave a specified band? Exit time problems like this are common in physics and finance. In this chapter we use numerical SDE approximation in a Monte Carlo setting to compute *mean exit times*.

11.2 Background

Given the scalar SDE

$$dX(t) = f(X(t))dt + g(X(t))dW(t), \quad X(0) = X_0, \tag{11.1}$$

where, for simplicity, we assume that X_0 is a constant lying in some interval (a, b), we may define the random variable T_{exit} to be

$$T_{\text{exit}} := \inf\{t : X(t) = a \text{ or } X(t) = b\}.$$

In words, T_{exit} is the first time that the solution leaves the interval (a, b). Our task is then to find the mean exit time

$$T_{\text{exit}}^{\text{mean}} := \mathbb{E}\left[T_{\text{exit}}\right].$$

11.3 Monte Carlo for Mean Exit Time

A straightforward Monte Carlo approach to approximate $T_{\text{exit}}^{\text{mean}}$ may be summarized as follows, where we use the Euler–Maruyama method (7.3) to simulate the SDE (11.1).

Choose a stepsize, Δt
Choose a number of paths, M
for s = 1 to M
    Set $t_n = 0$ and $X_n = X_0$
    While $X_n > a$ and $X_n < b$
        Compute a N(0,1) sample ξ_n
        Replace X_n by $X_n + \Delta t f(X_n) + \sqrt{\Delta t}\,\xi_n g(X_n)$
        Replace t_n by $t_n + \Delta t$
    end
    set $T_{\text{exit}}^s = t_n - \frac{1}{2}\Delta t$
end
set $a_M = \frac{1}{M}\sum_{s=1}^{M} T_{\text{exit}}^s$
set $b_M^2 = \frac{1}{M-1}\sum_{s=1}^{M}(T_{\text{exit}}^s - a_M)^2$

This produces an approximate mean exit time a_M and an approximate 95% confidence interval (2.6).

With this approach, we apply the numerical method on the discrete grid $\{t_i\}$ and, for each path, record the midpoint value $\frac{1}{2}(t_{n-1} + t_n)$, where t_n is the first gridpoint at which the numerical solution leaves the range (a, b). There are two sources of error that we have already encountered in this book.

1. The sampling error arising from the approximation of an expected value by a sample mean.

2. The discretization error arising because we use a numerical method to approximate the SDE solution paths[1].

However, there is also a third source of error, caused by the fact that we are recording solution values only at the discrete points $\{t_i\}$. Even if we follow the SDE solution paths exactly we run the risk of missing an exit point—within an interval $t_i < t < t_{i+1}$ the path may leave the range (a, b) and then return unnoticed. Figure 11.1 illustrates this possibility.

Our aim here is to examine the accuracy of the Monte Carlo approach to estimating a mean exit time.

[1]Because the numerical integration proceeds until X_n leaves the range (a, b), rather than over a fixed interval of time, the standard finite-time convergence theory, as outlined in Chapter 8 and 9, does not apply directly. We are not aware of any work that fills this gap in the literature.

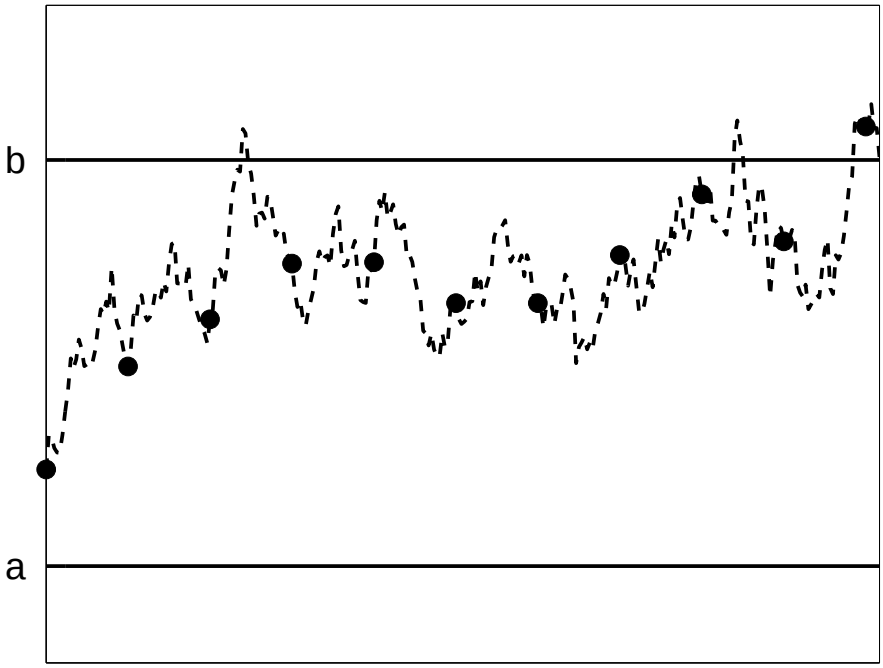


Figure 11.1: A solution path (dashed curve) and values along the path at a discrete set of times (circles). The continuous-time path sneaks above the upper bound b twice without being spotted by the discrete samples. Only the third exit leads to a discrete value outside the range.

Under appropriate conditions, the mean exit time problem may also be formulated as a deterministic ODE. Letting $u(x)$ denote the mean exit time $T_{\text{exit}}^{\text{mean}}$ when $X_0 = x$, it can be shown that

$$\tfrac{1}{2}g(x)^2\frac{d^2u}{dx^2} + f(x)\frac{du}{dx} = -1, \quad \text{for } a < x < b, \quad \text{with } u(a) = u(b) = 0. \quad (11.2)$$

This gives an alternative approach for computing mean exit times—solve the ODE (11.2) either analytically or numerically. With an exact solution at hand we will be able to investigate the accuracy of the Monte Carlo approach.

11.4 Computational Experiments

We consider first the case where $f(X) = \mu X$ and $g(X) = \sigma X$, with $0 < a < X_0 < b$. We recall from Chapter 5 that this SDE is used to model asset prices, so the mean exit time gives us the expected value of the first time that the asset either dips below the lower bound a or rises above the upper bound b. For this

SDE, (11.2) becomes

$$\tfrac{1}{2}\sigma^2 x^2 \frac{d^2u}{dx^2} + \mu x \frac{du}{dx} = -1, \quad \text{for } a < x < b, \quad \text{with } u(a) = u(b) = 0. \tag{11.3}$$

It can be shown that (11.3) has the solution

$$u(x) = \tfrac{1}{\frac{1}{2}\sigma^2 - \mu} \left(\log(x/a) - \frac{1 - (x/a)^{1-2\mu/\sigma^2}}{1 - (b/a)^{1-2\mu/\sigma^2}} \log(b/a) \right). \tag{11.4}$$

Exercise 11.1 asks you to confirm this, and Exercises 11.2 to 11.6 look at various properties of this formula.

Figure 11.2 illustrates the function $u(x)$ in the case where $\mu = 0.1$, $\sigma = 0.2$, $a = 0.5$ and $b = 2$. We see that the curve is convex with a fairly sharply defined maximum value around $x \approx 0.75$. This is intuitively reasonable. With intial data x close to a, the first exit times will typically be close to zero (because the paths start very close to a boundary). As the initial data moves further away from a, the expected time to leave the range increases. At the other extreme, as the initial data approaches the upper limit b, the mean first exit time again tends to zero. The intermediate initial value $x \approx 0.75$ where $T_{\text{exit}}^{\text{mean}}$ is largest, is closer to a than to b. This asymetry can be understood by looking at the exact solution (5.5) of the SDE and its expected value (5.6). We see that there is a general upward drift of solution paths so, to get the maximum expected mean first exit time, the initial data should be closest to the lower bound.

Keeping $\mu = 0.1$, $\sigma = 0.2$, $a = 0.5$, $b = 2$ and taking $X_0 = 1$, we applied the Monte Carlo method with $M = 5 \times 10^4$ paths and a stepsize of $\Delta t = 10^{-2}$. For this SDE we know how the exact solution evolves, so instead of using the Euler–Maruyama method we used the exact solution. That is, in the pseudocode description in section 11.3, instead of

$$\text{Replace } X_n \text{ by } X_n + \Delta t \mu X_n + \sqrt{\Delta t}\, \xi_n \sigma X_n$$

we used

$$\text{Replace } X_n \text{ by } X_n \exp\left((\mu - \tfrac{1}{2}\sigma^2)\Delta t + \sqrt{\Delta t}\, \xi_n \sigma \right)$$

Figure 11.3 gives a histogram of the M exit times that arose. The sample mean for the exit times was 7.8056, with a sample variance of 5.6. This gives an approximate 95% confidence interval of $[7.7561, 7.8552]$. Using the formula (11.4), the exact value for the mean exit time is $T_{\text{exit}}^{\text{mean}} = 7.6450$. We conclude that

(a) the exact answer is well outside the confidence interval, and

(b) the Monte Carlo method is overestimating the mean exit time.

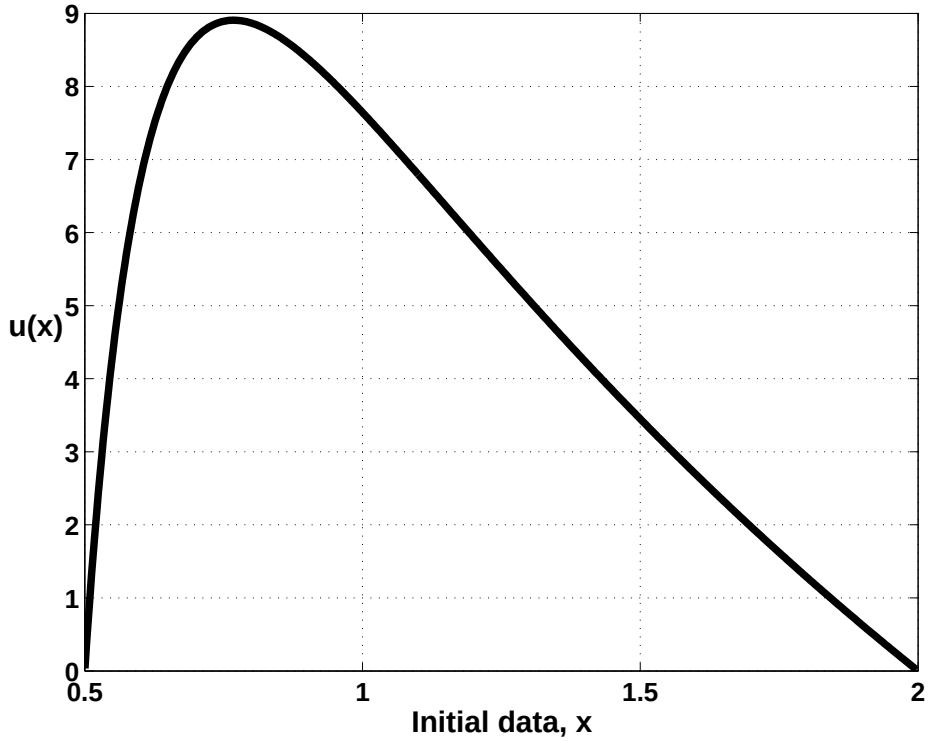


Figure 11.2: The mean exit time function $u(x)$ in (11.4).

Both features can be explained by arguing that the error from checking paths only at discrete time points $\{t_i\}$ is dominating the statistical sampling error. The confidence interval relates to the sample mean as an approximation to the mean exit time for the discrete process $\{t_i, X(t_i)\}$, but this quantity is a relatively poor overestimate of the mean exit time for the continuous-time process $\{t, X(t)\}$. This suggests that to improve the overall accuracy we should decrease the stepsize, Δt, rather than increase the number of sample paths, M. Reducing Δt to 10^{-3} produced a sample mean of 7.7137 with an approximate 95% confidence interval of $[7.6641, 7.7634]$. Further reducing Δt to 10^{-4} produced a sample mean of 7.6688 with an approximate 95% confidence interval of $[7.6200, 7.7177]$. These findings are consistent with the explanation above.

Next, we attempt to quantify the convergence rate of this method. Continuing with $\mu = 0.5$, $\sigma = 0.2$, $a = 0.5$, $b = 2$ and $X_0 = 1.5$, we fix the number of sample paths at $M = 5 \times 10^5$ and apply the algorithm with four different stepsizes; $\Delta t = 10^{-1}$, 10^{-2}, 10^{-3} and 10^{-4}. The picture on the left of Figure 11.4 shows the sample means as crosses with confidence intervals as vertical bars. We have added the exact value, $T_{\text{exit}}^{\text{mean}} = 0.5993$, as a dashed line.

We may investigate further the data in Figure 11.4 using a log-log plot and a least squares fit; as described for Figure 7.2. Let us assume that the statistical

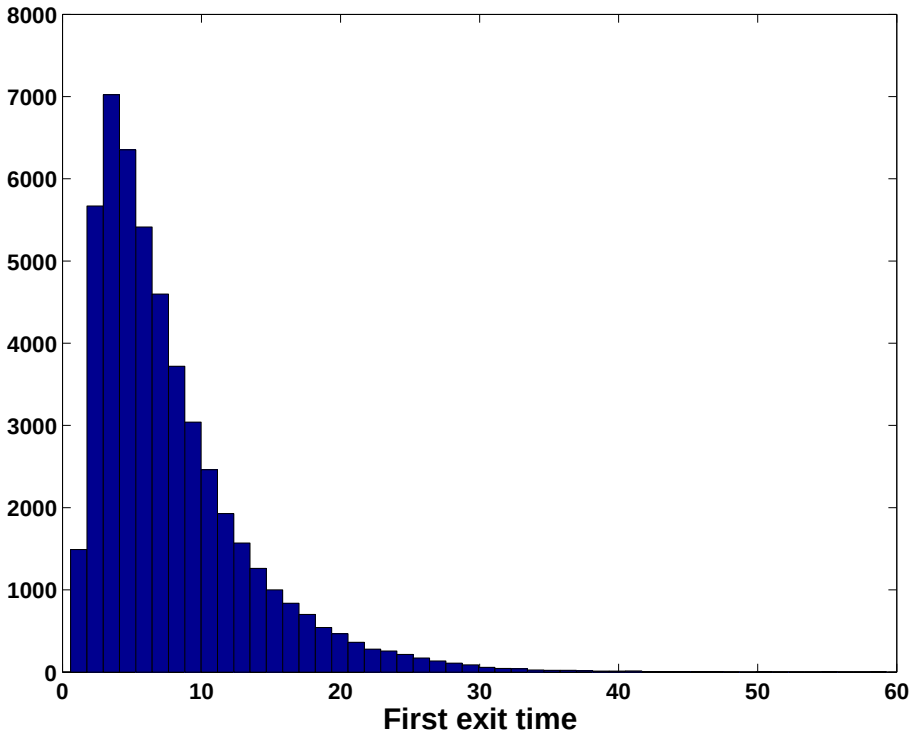


Figure 11.3: Histogram of mean exit times for $M = 5 \times 10^5$ sample paths.

sampling error is negligible and that the overall error, $\text{err}_{\Delta t} := |a_M - T_{\text{exit}}^{\text{mean}}|$, satisfies

$$\text{err}_{\Delta t} \approx C\Delta t^q, \tag{11.5}$$

for some constants C and q. On the right in Figure 11.4 we plot $\text{err}_{\Delta t}$ as a function of Δt, using a log-log scale. A dashed reference line of slope $\frac{1}{2}$ is included. The error decay seems to be consistent with this slope. A least squares fit to the power law $\log \text{err}_{\Delta t} = \log(C) + q \log(\Delta t)$ produced $q = 0.45$ with a residual of 0.1171.

This result is in agreement with the widely reported observation that the mean exit time based on a discrete grid gives an $O(\Delta t^{\frac{1}{2}})$ estimate of the mean exit time for the SDE; a result that continues to hold when the Euler–Maruyama method is used to approximate the SDE paths. See section 11.5 for references.

Under this $O(\Delta t^{\frac{1}{2}})$ error behavior, Exercise 11.7 asks you to argue that to optimize computational effort the number of sample paths, M, should be of the order Δt^{-1}. In Figure 11.5 we applied the Monte Carlo method with $\Delta t = 10^{-4}$ and $M = 10^4$, for 40 equally spaced X_0 values in the range $[a, b]$. Here we fixed $\mu = 0.1$ and $\sigma = 0.2$, with $a = 0.5$ and $b = 2$. The Monte Carlo sample means are plotted as crosses, with the corresponding 95% confidence intervals indicated by vertical lines. The exact mean exit times from (11.4) are plotted using a

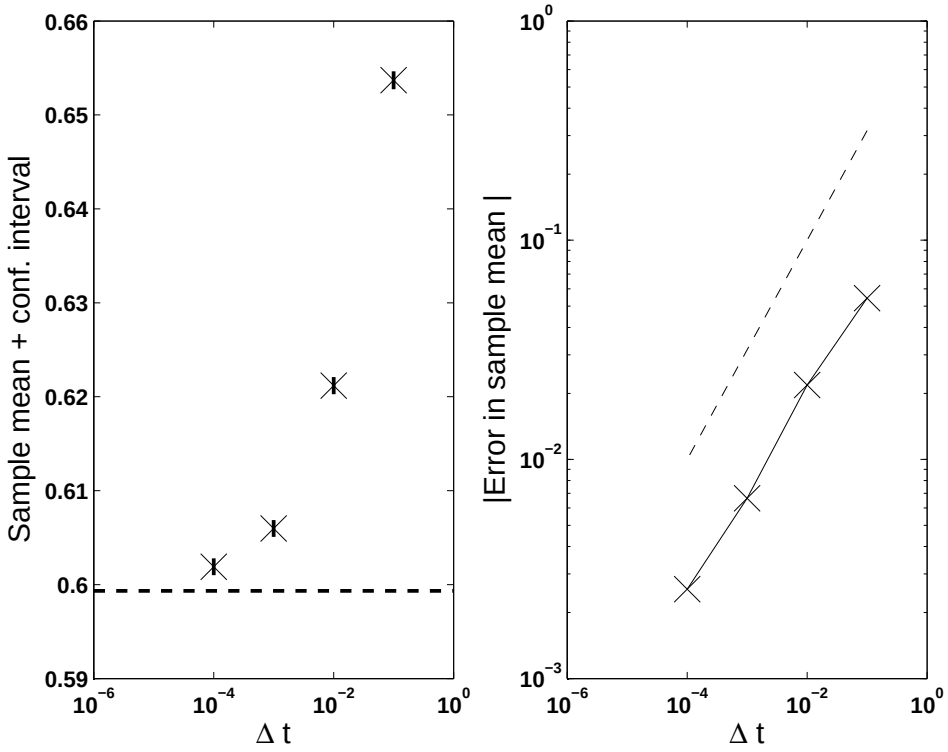


Figure 11.4: Total error in Monte Carlo approximation to mean exit time for a range of stepsizes, on a linear SDE.

dashed linetype. With these parameter values the Monte Carlo method does a reasonable job of providing one or two digits of accuracy, typically being more accurate for X_0 close to the extremes of a and b, where the variance in the exit time is smaller.

For our next example, we repeat the type of experiment in Figure 11.5 in the case of the mean-reverting square root process (5.8). We recall that this SDE is used to model the movement of interest rates, and in this context the mean exit time measures, on average, the time at which the interest rate drifts outside a given range. We mentioned in section 5.3 that, with $X_0 > 0$, the SDE solution is guaranteed to remain positive. This property does not hold for a numerical method, in general. Hence, to avoid computing the square root of a negative number, we implement the Euler–Maruyma method in the pseudocode of section 11.3 as

$$\text{Replace } X_n \text{ by } X_n + \Delta t\lambda(\mu - X_n) + \sqrt{\Delta t}\,\xi_n\sigma\sqrt{|X_n|}. \tag{11.6}$$

In other words, we take the absolute value before applying the square root func-

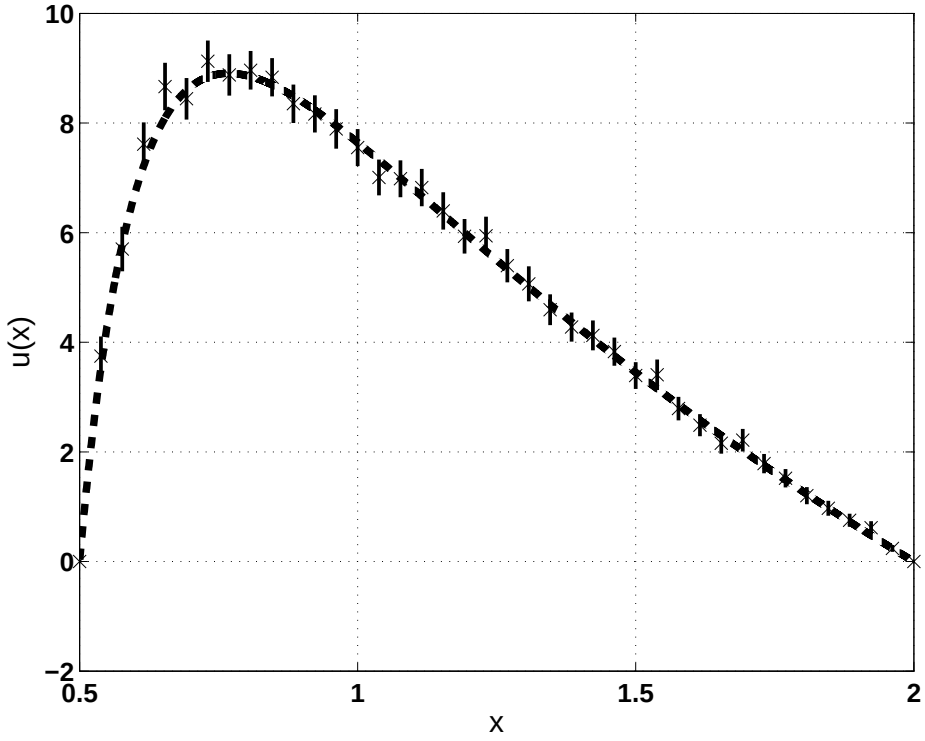


Figure 11.5: Dashed line: mean exit time function $u(x)$ in (11.4). Crosses: Monte Carlo approximations, with confidence intervals indicated.

tion[2].

Setting $\lambda = 1$, $\mu = 0.5$ and $\sigma = 0.3$, with $a = 1$ and $b = 2$, we applied the Monte Carlo method with $M = 10^3$ sample paths and a stepsize of $\Delta t = 10^{-3}$. Results are shown in Figure 11.6. As in Figure 11.5 the sample means are shown as crosses with 95% confidence intervals indicated by vertical lines, and the dashed line shows the "exact" mean exit time curve. In this case, we used MATLAB's built-in function `bvp4c` to solve the relevant boundary value ODE problem (11.2). An analytical solution is available, but is rather messy to compute with; see Exercise 11.9. We see that the accuracy from the Euler–Maruyama based Monte Carlo method is no worse than that in Figure 11.5, where the exact SDE solution was available, and a smaller Δt and larger M were used.

As a final example, we consider the SDE (5.18) for a particle in a double-well potential with additive noise. We take $a = -3$ and $b = 1$. We recall from section 5.3 that without noise, solutions must evolve in a way that lowers the potential. It is then interesting to see how the additive noise term allows solution paths to break this rule. Note that the potential function takes a much lower

[2]An alternative is to apply the standard EM method and then re-define X_n as $|X_n|$.

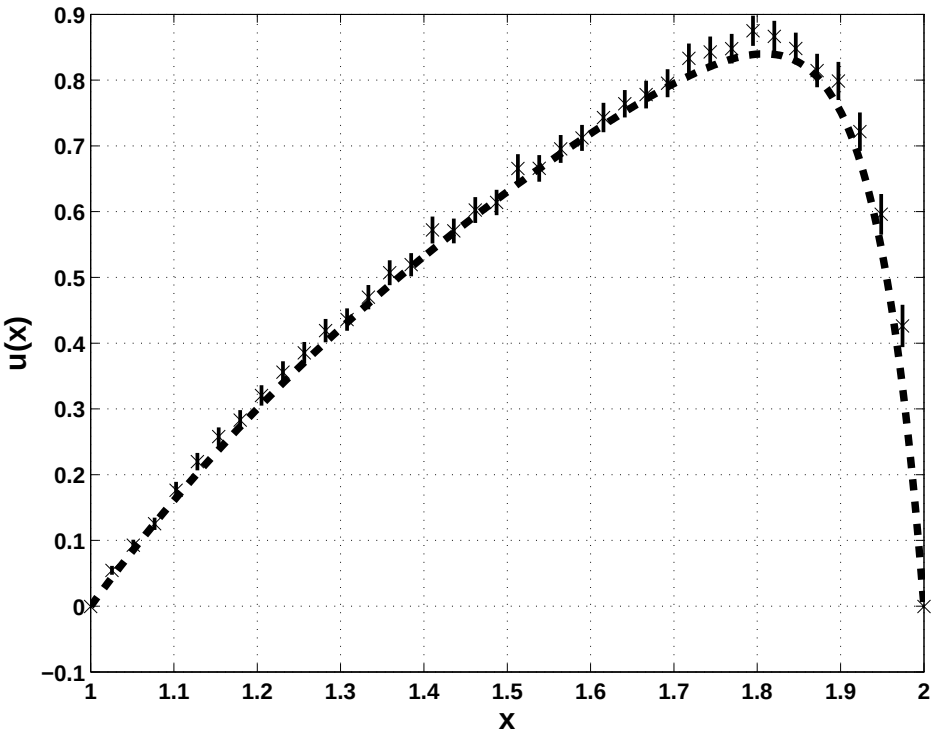


Figure 11.6: Dashed line: mean exit time function $u(x)$ for square root process (5.8). Crosses: Monte Carlo approximations, with confidence intervals indicated.

value at $x = b$ than at $x = a$, so, for small noise, an exit via b is much more likely than an exit via a. In other words, we are essentially measuring the mean time taken for the particle to climb over the hump and visit the right-hand well in Figure 5.3. In Figure 11.7 we show the mean exit time function $u(x)$ for three different noise levels, $\sigma = 2$ (solid) and $\sigma = 4$ (dashed) and $\sigma = 6$ (dash-dotted), computed from MATLAB's `bvp4c` routine. For $\sigma = 2$ we see that the mean exit time is very flat for initial values $-3 < x \leq -1$ other than those very close to 3. Here, only the upper limit $b = 1$ is relevant—despite starting near -3, solution paths rapidly drop down the potential well and are very likely to climb to the value $b = 1$ before $a = -3$. The function $u(x)$ thus has a *boundary layer* at the left-hand end. For $\sigma = 4$ and $\sigma = 6$ we are adding more noise, allowing more paths to reach $a = -1$ from a nearby starting point, which smooths out the boundary layer. Increasing σ also lowers all the mean exit times. Fixing $X_0 = 0$ and $\sigma = 4$, in Figure 11.8 we plot the overall error, $\mathrm{err}_{\Delta t} := |a_M - T_{\mathrm{exit}}^{\mathrm{mean}}|$, in the Euler–Maruyama based Monte Carlo method for $M = 5 \times 10^5$ paths and stepsizes of $\Delta t = 10^{-2}$, 10^{-3}, 10^{-4} and 10^{-5}. As in Figure 11.4, we show the confidence intervals in the left-hand picture and give a log-log error plot on the right-hand

picture. In this case a least-squares fit gave a power of $q = 0.4561$ and a residual of 0.0174.

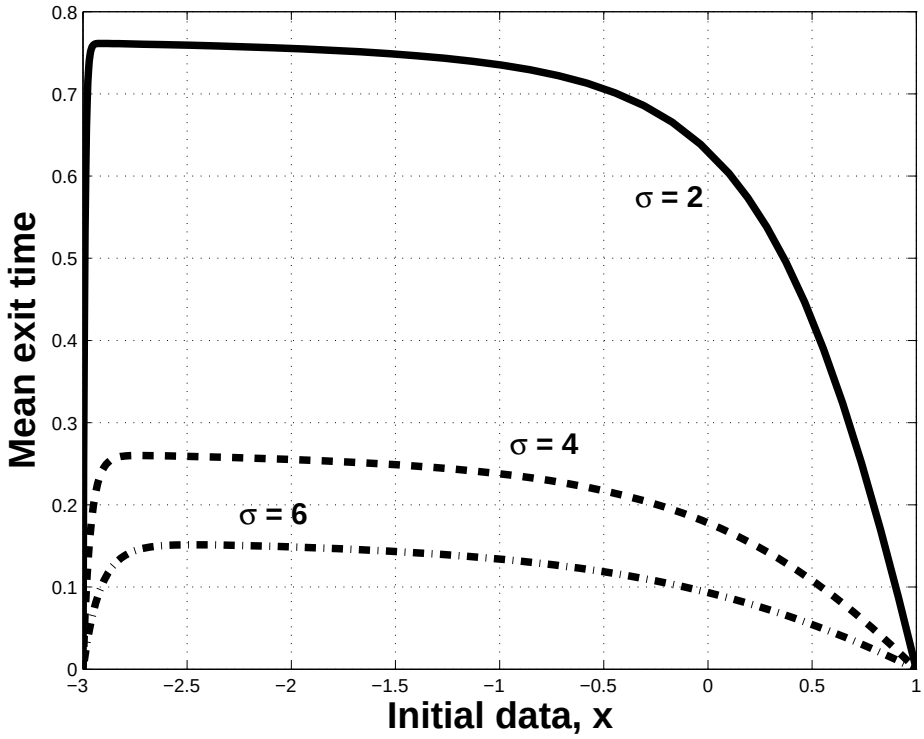


Figure 11.7: The mean exit time function $u(x)$ from (11.2) for the double-well potential SDE (5.18), with three different additive noise levels.

11.5 Notes and References

The texts [19] and [37] offer accessible introductions to the theory of mean exit times, and in particular show how the boundary value ODE (11.2) arises. There are many interesting variations of this problem. For example a boundary may *reflect* rather than *absorb* or the quantity to be computed may be the probability that the solution reaches a before reaching b.

A more comprehensive numerical treatment of the double-well mean exit time problem can be found in [32].

Results concerning the *density* of the exit time for the mean-reverting square root process can be found in [41].

Several authors, including [4, 9, 32, 43], have noted the $O(\Delta t^{\frac{1}{2}})$ error behaviour in the type of simulations of this chapter. This convergence rate is disappointing: we saw in Chapter 8 that the Euler–Maruyama method can

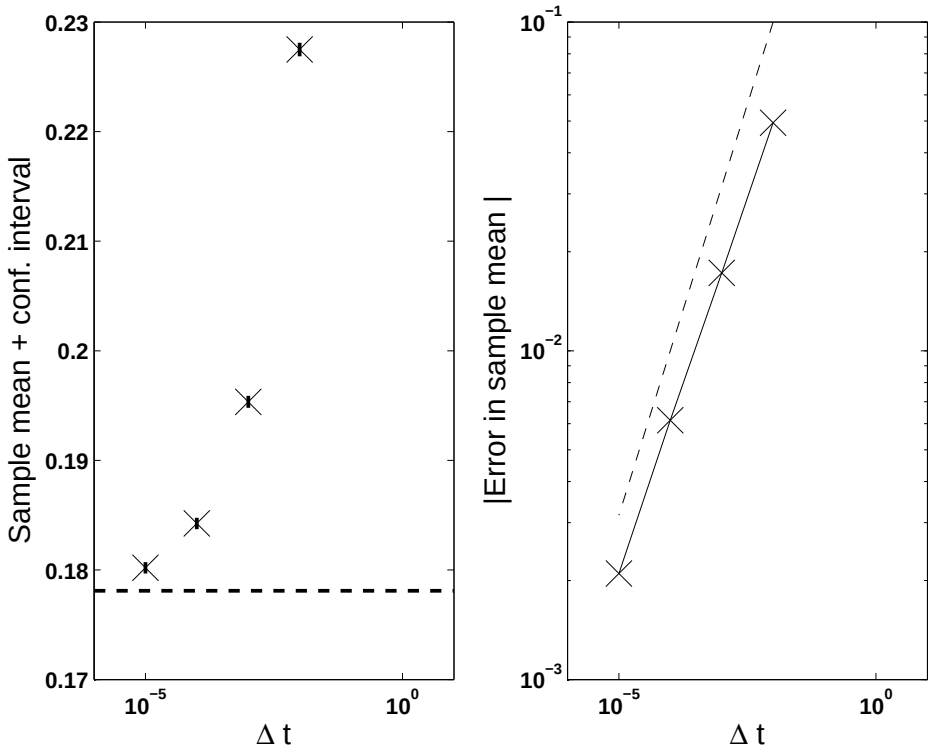


Figure 11.8: Total error in Monte Carlo approximation to mean exit time for a range of stepsizes, on a double-well potential.

approximate expected values of functions of the solution—quantities such as $\mathbb{E}[\Phi(X(T))]$—to an accuracy of $O(\Delta t)$. There are several ways to improve the accuracy of this basic algorithm.

- Adaptively reduce the stepsize when a path approaches a boundary,
- After each step, calculate the probability that an exit was missed. Then draw from a uniform $(0, 1)$ random number generator in order to decide whether to record an exit. Details for some simple SDEs can be found in [43].
- Use random stepsizes from a suitable exponential distribution, as described in [32, 33].

We are not aware of a rigorous proof that, ignoring the sampling issue, the error is $O(\Delta t^{\frac{1}{2}})$ for computations such as those in Figures 11.4 and 11.8, but analysis for the related *killed diffusion* problem has been given in [22].

Perhaps the most efficient weak simulation algorithms for mean exit times are the random walk based methods described in [47, Chapter 6].

Exercises

11.1. ⋆⋆ Confirm that (11.4) solves (11.3).

11.2. ⋆⋆ Show that there is a single point $x^\star$ over $[a, b]$ where $u(x)$ in (11.4) is maximum, and find $u(x^\star)$.

11.3. ⋆ For $x = \frac{1}{2}(a + b)$, show that $u(x)$ in (11.4) satisfies $u(x) \to 0$ as $a \to b$ from below. Interpret this result.

11.4. ⋆⋆ Show that $u(x)$ in (11.4) satisfies $u(x) = C(x - a) + O(x - a)^2$ as $x - a$ tends to zero from above, and find the constant C.

11.5. ⋆⋆ Show that $u(x)$ in (11.4) satisfies $u(x) = D(b - x) + O(b - x)^2$ as $b - x$ tends to zero from above, and find the constant D.

11.6. ⋆⋆⋆ The expression for $u(x)$ in (11.4) appears to break down when $\mu = \frac{1}{2}\sigma^2$. By letting $\mu = \frac{1}{2}\sigma^2 + \epsilon$ and considering the limit $\epsilon \to 0$, show that a sensible mean exit time formula can be established in this special case.

11.7. ⋆ This exercise should be compared with Exercise 7.3. Refering to the Monte Carlo algorithm outlined in section 11.3, suppose that the mean exit time for the discrete numerical solution differs from the mean exit time for the SDE by an amount that behaves like $O(\Delta t^{\frac{1}{2}})$ as $\Delta t \to 0$. Argue that the number of sample paths, M, used in the Monte Carlo method should be chosen so that M is proportional to Δt^{-1}.

11.8. ⋆⋆ The SDE $dX(t) = \sqrt{X(t)\,(1 - X(t))}dW(t)$ is a special case of *Wright-Fisher diffusion*, which arises in population dynamics; see, for example, [37, Section 12.2]. For $0 = a < X(0) < b = 1$, show that the ODE (11.2) for the mean exit time has solution $u(x) = -2\,((1 - x)\log(1 - x) + x \log x)$. This function is plotted in Figure 11.9. [Hint: recall that $\log x$ integrates to $x \log x - x$.]

11.9. ⋆⋆⋆ Consider the case where the SDE (11.1) is the square root process (5.8). With $0 < a < b$, find an analytical solution to the ODE (11.2). Express the solution in a form that is amenable to numerical evaluation.

PROG of CHAPTER: Fixed Δt and M version.
PROG Ex 1: Try weak EM. Same convergence rate?
PROG Ex 2: Sqrt process with EM.

Quotes:

At any given time, a large portion of the world's computers
are doing Monte Carlo simulations.
Don Estep, (during a lecture at the International Conference on Computational and Mathematical Methods in Science and Engineering, 2005).

Those who trust to chance must abide by the results of chance.

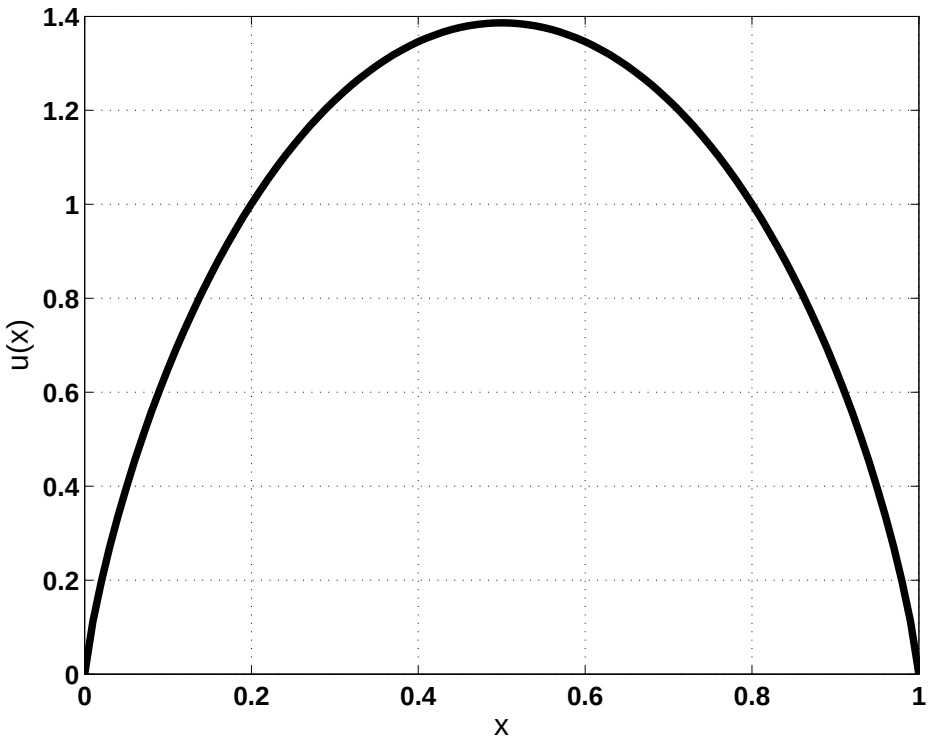


Figure 11.9: The mean exit time function $u(x)$ for the SDE in Exercise 11.8.

Calvin Coolidge (1872–1933), source: `http://www.quotedb.com/quotes/3169`.

The purpose of models is not to fit the data
but to sharpen the questions.
Samuel Karlin, 11th R. A. Fisher Memorial Lecture (Royal Society), 1983.

All models are wrong, but some are useful.
George E. P. Box, source
`http://www-personal.umich.edu/~dronis/box.htm`