Elementary Calculus of Financial Mathematics

Mathematical Modeling and Computation

About the Series

The SIAM series on Mathematical Modeling and Computation draws attention to the wide range of important problems in the physical and life sciences and engineering that are addressed by mathematical modeling and computation; promotes the interdisciplinary culture required to meet these large-scale challenges; and encourages the education of the next generation of applied and computational mathematicians, physical and life scientists, and engineers.

The books cover analytical and computational techniques, describe significant mathematical developments, and introduce modern scientific and engineering applications. The series will publish lecture notes and texts for advanced undergraduate- or graduate-level courses in physical applied mathematics, biomathematics, and mathematical modeling, and volumes of interest to a wide segment of the community of applied mathematicians, computational scientists, and engineers.

Appropriate subject areas for future books in the series include fluids, dynamical systems and chaos, mathematical biology, neuroscience, mathematical physiology, epidemiology, morphogenesis, biomedical engineering, reaction-diffusion in chemistry, nonlinear science, interfacial problems, solidification, combustion, transport theory, solid mechanics, nonlinear vibrations, electromagnetic theory, nonlinear optics, wave propagation, coherent structures, scattering theory, earth science, solid-state physics, and plasma physics.

A. J. Roberts, *Elementary Calculus of Financial Mathematics*

James D. Meiss, *Differential Dynamical Systems*

E. van Groesen and Jaap Molenaar, *Continuum Modeling in the Physical Sciences*

Gerda de Vries, Thomas Hillen, Mark Lewis, Johannes Müller, and Birgitt Schönfisch, *A Course in Mathematical Biology: Quantitative Modeling with Mathematical and Computational Methods*

Ivan Markovsky, Jan C. Willems, Sabine Van Huffel, and Bart De Moor, *Exact and Approximate Modeling of Linear Systems: A Behavioral Approach*

R. M. M. Mattheij, S. W. Rienstra, and J. H. M. ten Thije Boonkkamp, *Partial Differential Equations: Modeling, Analysis, Computation*

Johnny T. Ottesen, Mette S. Olufsen, and Jesper K. Larsen, *Applied Mathematical Models in Human Physiology*

Ingemar Kaj, *Stochastic Modeling in Broadband Communications Systems*

Peter Salamon, Paolo Sibani, and Richard Frost, *Facts, Conjectures, and Improvements for Simulated Annealing*

Lyn C. Thomas, David B. Edelman, and Jonathan N. Crook, *Credit Scoring and Its Applications*

Frank Natterer and Frank Wübbeling, *Mathematical Methods in Image Reconstruction*

Per Christian Hansen, *Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion*

Michael Griebel, Thomas Dornseifer, and Tilman Neunhoeffer, *Numerical Simulation in Fluid Dynamics: A Practical Introduction*

Khosrow Chadan, David Colton, Lassi Päivärinta, and William Rundell, *An Introduction to Inverse Scattering and Inverse Spectral Problems*

Charles K. Chui, *Wavelets: A Mathematical Tool for Signal Analysis*

Elementary Calculus of Financial Mathematics

A. J. Roberts
University of Adelaide
Adelaide, South Australia, Australia

To Barbara, Sam, Ben, and Nicky
for their support over the years

Contents

Preface

Welcome! This book leads you on an introduction into the fascinating realm of financial mathematics and its calculus. Modern financial mathematics relies on a deep and sophisticated theory of random processes in time. Such randomness reflects the erratic fluctuations in financial markets. I take on the challenge of introducing you to the crucial concepts needed to understand and value financial options among such fluctuations. This book supports your learning with the bare minimum of necessary prerequisite mathematics.

To deliver understanding with a minimum of analysis, the book starts with a graphical/numerical introduction to how to adapt random walks to describe the typical erratic fluctuations of financial markets. Then simple numerical simulations both demonstrate the approach and suggest the symbology of stochastic calculus. The finite steps of the numerical approach underlie the introduction of the binomial lattice model for evaluating financial options.

Fluctuations in a financial environment may bankrupt businesses that otherwise would grow. Discrete analysis of this problem leads to the surprisingly simple extension of classic calculus needed to perform stochastic calculus. The key is to replace squared noise by a mean drift: in effect, $dW^2 = dt$. This simple but powerful rule enables us to differentiate, integrate, solve stochastic differential equations, and to triumphantly derive and use the Black–Scholes equation to accurately value financial options.

The first two chapters deal with individual realizations and simulations. However, some applications require exploring the distribution of possibilities. The Fokker–Planck and Kolmogorov equations link evolving probability distributions to stochastic differential equations (SDEs). Such transformations empower us not only to value financial options but also to model the natural fluctuations in biology models and to approximately solve differential equations using stochastic simulation.

Lastly, the formal rules used previously are justified more rigorously by an introduction to a sound definition of stochastic integration. Integration in turn leads to a sound interpretation of Ito's formula that we find so useful in financial applications.

Prerequisites

Basic algebra, calculus, data analysis, probability and Markov chains are prerequisites for this course. There will be many times throughout this book when you will need the concepts and techniques of such courses. Be sure you are familiar with those, and have appropriate references on hand.

Computer simulations

Incorporated into this book are MATLAB/SCILAB scripts to enhance your ability to probe the problems and concepts presented and thus to improve learning. You can purchase MATLAB from the Mathworks company, http://www.mathworks.com. SCILAB is available for free via http://www.scilab.org.

A. J. Roberts

List of Algorithms

Chapter 1

Financial Indices Appear to Be Stochastic Processes

Contents

Earnings momentum and visibility should continue to propel the market to new highs. *A forecast issued by E. F. Hutton, the Wall Street brokerage firm, moments before the stock market plunged on October* 19, 1987.

There is overwhelming evidence that share prices, financial indices, and currency values fluctuate wildly and randomly. An example is the range of daily cotton prices shown in Figure 1.1; another is the range of wheat prices shown in Figure 1.2. These figures demonstrate that randomness appears to be an integral part of the dynamics of the financial market. In a dissertation that was little appreciated at the time, the all-pervading randomness in finance was realized by Bachelier circa 1900. Prices evolving in time form a *stochastic process* because of these strong random fluctuations. This book develops the elementary theory of such stochastic processes in order to underpin the famous Black–Scholes equation for valuing financial options, here described by Tony Dooley:

> . . . the most-used formula in all of history is the Black–Scholes formula. Each day, it is used millions of times in millions of computer programs, making it more used than Pythagoras' theorem!
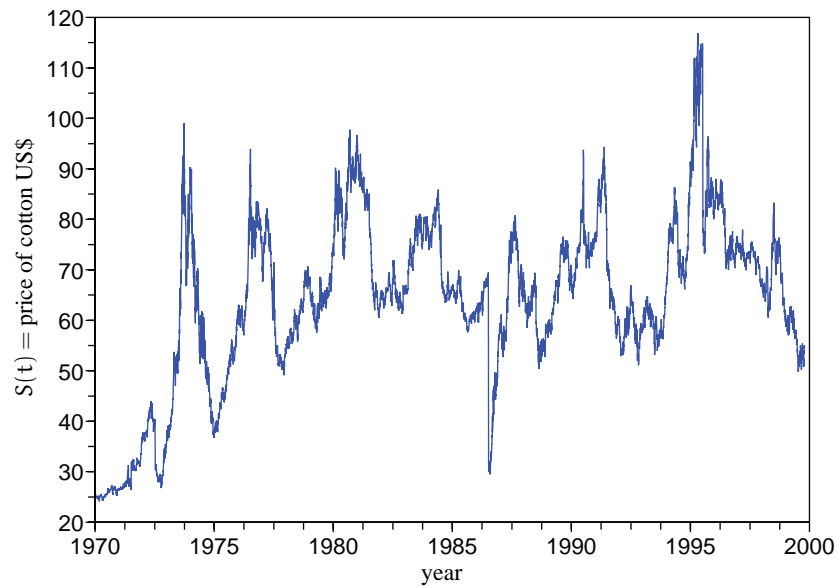
Figure 1.1. *Opening daily prices* $S(t)$ *of cotton over nearly* 30 *years showing the strong fluctuations typical of financial markets (note that there are about* 21 *trading days per month, that is,* 250 *per year).*
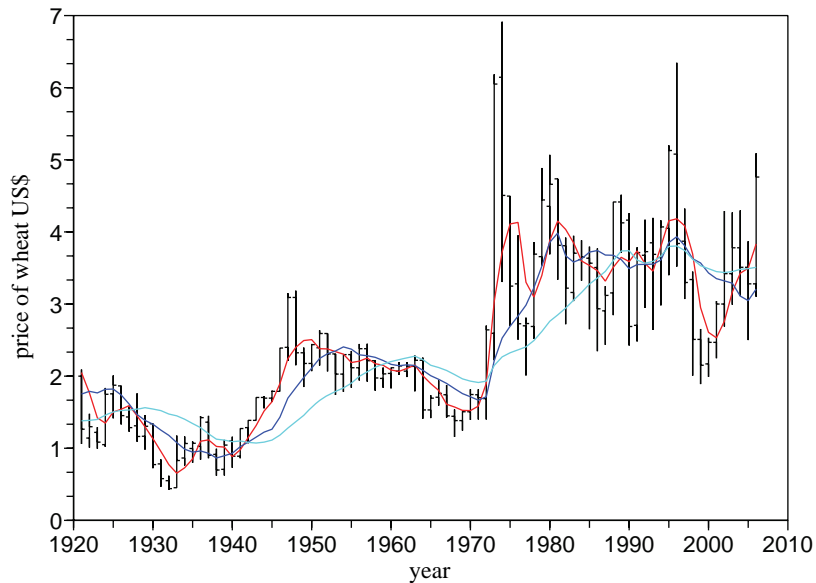
Figure 1.2. *The range of wheat prices over nearly* 90 *years. The vertical bars give the range in each year; the tick on the left of each bar is the opening price; and the tick on the right is the closing price. The red curve is the 4-year average, blue is the 9-year average, and cyan is the* 18*-year average.*

The financial world is not the only example of significant random fluctuations. In biology we generate differential equations representing the interactions of predators and prey, for example, foxes and rabbits. Such models are often written as ordinary differential equations (ODEs) of the form $dx/dt = \cdots$ and $dy/dt = \cdots$, where the right-hand side dots denote life and death interactions. However, biological populations live in an environment with random events such as drought, flood, or meteorite impact. There are fluctuations in the populations due to such unforeseen events. Random events are especially dangerous for populations of endangered species in which there are relatively few individuals.

Engineers may also need to analyze problems with random inputs. A truck driving along a road shakes from a variety of causes, one of which is travelling over the essentially random bumps in the road. In many aspects the truck's design must account for such stochastic vibrations.

These examples show that the study of stochastic differential equations (SDEs) is worthwhile. Note that the use of SDEs is an admission of ignorance of the nature of fluctuations. There are a multitude of unknown processes which influence the phenomena of interest. Under the central limit theorem, we assume that these unknown influences accumulate to become normally distributed—alternatively called Gaussian distribution.

Suggested activity: Do at least Exercise 1.1.

1.1   Brownian motion is also called a Wiener process

> Nothing in Nature is random ... A thing appears random only through the incompleteness of our knowledge.   *Spinoza*

A starting point to describe stochastic processes such as a stock price is Brownian motion or, more technically, a Wiener process. Figure 1.3 shows an example of this process. See the roughly qualitative similarities to the cotton prices shown in Figure 1.1 (though the cotton prices appear to fluctuate more) and to the wheat prices in Figure 1.2.

Algorithm 1.1 MATLAB/SCILAB code to plot m realizations of a Brownian motion/ Wiener process as shown in Figure 1.3. In SCILAB use `rand(.,.,"n")` instead of `randn(.,.)` for $N(0,1)$ distributed random numbers.

```
m=5;
n=300;
t=linspace(0,1,n+1)';
h=diff(t(1:2));
dw=sqrt(h)*randn(n,m);
w=cumsum([zeros(1,m);dw]);
plot(t,w)
```

Where do fluctuations in the financial indices come from? Many economic theorists assert that fluctuations reflect the random arrival of new knowledge. As new knowledge is made known to the traders of stocks and shares, they almost instantly assimilate the knowledge and buy and sell accordingly. However, according to these graphs it would seem
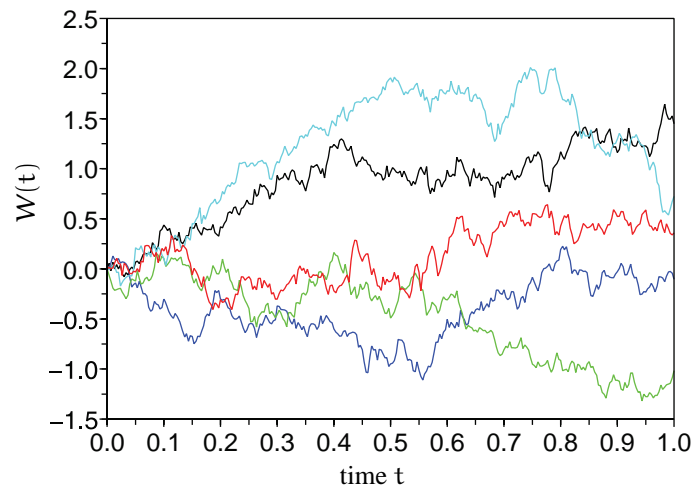
Figure 1.3. *Five realizations of Brownian motion (Wiener process) $W(t)$ gener-*
ated by Algorithm 1.1. Most figures in this book plot five realizations of a random process
to hint that stochastic processes, such as this Wiener process, are a composite of an un-
countable infinity of individual realizations.

that at least 95% of the "new knowledge" in a day is self-contradictory! The independent
fluctuations from one day to the next say to me that whatever a person discovers one day
is meaningless the next. This means to me that any such knowledge is worthless and
refutes the notion of genuine "new knowledge."[1] But there is an alternative model for the
fluctuations. Recent computer simulations[2] show that when many interacting agents try to
outsmart each other to obtain assets the result is mayhem. Competing agents individually
generate widely fluctuating valuations of the asset, such as seen in real share prices. No
new knowledge need be hypothesized, just greed. Fortunately, it appears that the *average*
valuation of the asset is realistic, though no proof of this is known. Thus a recent view of
financial prices is that the average valuation is reasonable, with fluctuations due to agents
competing with each other.

   Many agents look at financial indices and see trends and patterns on which they base
their recommendations. One recommendation which I heard in a public presentation was
based on Fibonnaci numbers. I view this sort of trend analysis as gibberish. Random
fluctuations often seem to have short-term patterns, as you see in Figure 1.3. The problem
is that humans are very susceptible to seeing patterns which do not exist. Remember that a
random sequence must, accidentally, have some chance patterns. Despite all the analysis by
"financial experts," sound statistical analysis shows that there are very few, if any, patterns
over time in the fluctuations of the stock market. The fluctuations are, to a very good
approximation, independent from day to day. We assume independence in our development
of suitable mathematics.

---

[1]See *Why economic theory is out of whack* by Mark Buchanan in the *New Scientist*, 19 July 2008, to read
a little on financial markets' lack of reaction to news.

[2]Read about the "El Farol" problem by W.B. Arthurs, *Amer. Econ. Assoc. Pap. Proc.*, 84, p. 406 (1994).

The Wiener process

Now we return to our main task, which is to understand the nature of Brownian motion (Weiner process). The stochastic process shown in Figure 1.3 is named after the British botanist Robert Brown, who first reported the motion in 1826 when observing in his microscope the movement of tiny pollen grains due to small but incessant and random impacts of air molecules. Wiener formalized its properties in the 20th century. Algorithm 1.1 generates the realizations shown in Figure 1.3 by dividing time into small steps of length $\Delta t = h$ so that the jth time step reaches time $t_j = jh$ (assuming $t = 0$ is the start of the period of interest). Then the algorithm determines W_j, the value of the process at time t_j, by adding up many independent and normally distributed increments:[3]

$$W_{j+1} = W_j + \sqrt{h}Z_j \quad \text{where} \quad Z_j \sim N(0,1) \quad \text{and with} \quad W_0 = 0.$$

We generate a Brownian motion, or Wiener process, $W(t)$ in the limit as the step size $h \to 0$ so that the number of steps becomes large, e.g., $1/h$. The random increments to W, namely $\Delta W_j = \sqrt{h}Z_j$, are chosen to scale with $\sqrt{h}$ because it eventuates that this scaling is precisely what is needed to generate a reasonable limit as $h \to 0$, as we now see. Consider the process W at some fixed time $t = nh$, that is, we take n steps of size h to reach t, and hence $W(t)$ is approximated by n random increments of variance h:

$$\begin{aligned}
W_n &= W_{n-1} + \sqrt{h}Z_{n-1} \\
&= W_{n-2} + \sqrt{h}Z_{n-2} + \sqrt{h}Z_{n-1} \\
&= \cdots \\
&= W_0 + \sum_{j=0}^{n-1} \sqrt{h}Z_j.
\end{aligned}$$

Now firstly $W_0 = 0$ and secondly we know that a sum of normally distributed random variables is a normally distributed random variable with mean given by the sum of means, and variance given by the sum of the variances. Since all the increments $Z_j \sim N(0,1)$, then $\sqrt{h}Z_j \sim N(0,h)$, and the sum of n such increments is

$$W(t) = \sum_{j=0}^{n-1} \sqrt{h}Z_j \sim N(0,nh) = N(0,t).$$

Thus we deduce that $W(t) \sim N(0,t)$.

By taking random increments $\propto \sqrt{h}$ we find that the distribution of W at any time is fixed—that is, independent of the number of discrete steps we took to get to that time. Thus considering the step size $h \to 0$ is a reasonable limit. Similar arguments show that the change in W over any time interval of length t, $W(t+s) - W(s)$, is also normally distributed $N(0,t)$ and, *further*, is independent of any details of the process that occurred for times before time s.

---

[3]Recall that saying a random variable $Z \sim N(a, b)$ means that Z is normally distributed with mean a and variance b (standard deviation $\sqrt{b}$).

This last property of independence is vitally important in various crucial places in the development of SDEs. To see this normal distribution, suppose time has been discretized with steps $\Delta t = h$, the time s corresponds to step $j = \ell$, and the time $s + t$ corresponds to step $j = \ell + n$. Then

$$
\begin{aligned}
W_{\ell+n} &= W_{\ell+n-1} + \sqrt{h}Z_{\ell+n-1} \\
&= W_{\ell+n-2} + \sqrt{h}Z_{\ell+n-2} + \sqrt{h}Z_{\ell+n-1} \\
&= \cdots \\
&= W_\ell + \sum_{j=\ell}^{\ell+n-1} \sqrt{h}Z_j .
\end{aligned}
$$

This sum of n increments is distributed $N(0,t)$ as before, and so

$$
W(t+s) - W(s) = W_{\ell+n} - W_\ell \sim N(0,t).
$$

Now, further, see that the sum of the increments above *only* involve random variables Z_j for $\ell \le j < \ell + n$ which are completely *independent* of the random increments $\sqrt{h}Z_j$ for $j < \ell$, and hence are completely independent of the details of $W(t)$ for times $t \le s$, that is, after time $t = s$ the changes in the process W from $W(s)$ are independent of $W(t)$ for earlier times $t \le s$.

Brownian motion has a prime role in stochastic calculus because of the central limit theorem. In this context, if the increments in W follow some other distribution on the smallest "micro" times, then provided the variance of the microincrements is finite, the process always appears Brownian on the macroscale. This is assured because the sum of many random variables with finite variance tends to a normal distribution. Indeed, in Section 1.4 and others following we approximate the increments as the binary choice of either a step up or a step down. In effect we choose $Z_j = \pm 1$, each with probability $\frac{1}{2}$; as the mean and variance of such a Z_j are zero and one, respectively, the cumulative sum $\sum_{j=0}^{n-1} \sqrt{h}Z_j$ has the appearance on the macroscale of an $N(0,nh)$ random variable, as required.

Definition 1.1. Brownian motion *or* Wiener process *or* random walk*, usually denoted* $W(t)$*, satisfies the following properties:*

- $W(t)$ *is continuous;*

- $W(0) = 0$*;*

- *the change* $W(t+s) - W(s) \sim N(0,t)$ *for* $t, s \ge 0$*; and*

- $W(t+s) - W(s)$ *is independent of any details of the process for times earlier than* s.

Continuous but not differentiable

Wiener proved that such a process exists and is unique in a stochastic sense. The only property that we have not seen is that $W(t)$ is continuous, so we investigate it now. We know that $W(t+s) - W(s) \sim N(0,t)$, so we now imagine t as small and write this as

$$
W(t+s) - W(s) = \sqrt{t}Z_t,
$$

where the random variables $Z_t \sim N(0,1)$. Now although Z_t will vary with t, it comes from a normal distribution with mean 0 and variance 1, so as $t \to 0$, then almost surely the right-hand side $\sqrt{t}Z_t \to 0$. Thus, almost surely $W(t+s) \to W(s)$ as $t \to 0$, and hence W is continuous (almost surely).

Although it is continuous we now demonstrate that a Wiener process, such as that shown in Figure 1.4 (left column), is too jagged to be differentiable.[4] As Figure 1.4 (right column) shows, recall that for a smooth function such as $f(t) = e^t$ we generally see a linear variation near any point; for example,

$$f(t+s) - f(s) = e^{t+s} - e^s = (e^t - 1)e^s \approx te^s,$$

or more generally, $f(t+s) - f(s) \approx tf'(s)$. Thus we are familiar with $f(t+s) - f(s)$ decreasing linearly with t, and upon this is based all the familiar rules of differential and integral calculus. In contrast, in the Wiener process, and generally for solutions of SDEs, Figure 1.4 (left) shows $W(t+s) - W(s)$ decreasing more slowly, like $\sqrt{t}$. Thus Wiener processes are much steeper and vastly more jagged than smooth differentiable functions. Notionally the Wiener process has "infinite slope" and is thus nowhere differentiable. This feature generates lots of marvelous new effects that make stochastic calculus enormously intriguing.

Example 1.2. Pick a normally distributed random variable $Z \sim N(0,1)$, then define $W(t) = Z\sqrt{t}$. Is $W(t)$ a Wiener process?

Solution: No; although $W(t) \sim N(0,t)$, it does not satisfy all the properties of a Wiener process. Consider each property in turn:

- true, $W(t)$ is clearly continuous;

- true, $W(0) = 0$ is satisfied;

- but $W(t+s) - W(s) = Z(\sqrt{t+s} - \sqrt{s})$; which has variance $(\sqrt{t+s} - \sqrt{s})^2 = t - 2\sqrt{s(t+s)} \neq t$, and so $W(t)$ does not satisfy the third property of a Wiener process (nor the fourth, incidentally). ∎

Example 1.3. Let $W(t)$ and $\widetilde{W}(t)$ be independent Wiener processes and ρ a fixed number $0 < \rho < 1$. Show that the linear combination $X(t) = \rho W(t) + \sqrt{1 - \rho^2}\widetilde{W}(t)$ is a Wiener process.

Solution: Look at the properties in turn.

- $X(t)$ is clearly continuous as W and $\widetilde{W}$ are continuous and the linear combination maintains continuity.

- $X(0) = \rho W(0) + \sqrt{1 - \rho^2}\widetilde{W}(0) = \rho \cdot 0 + \sqrt{1 - \rho^2} \cdot 0 = 0$.

---

[4]See the amazing deep zoom of the Wiener process shown by Algorithm A.2 in Appendix A with its unexplored features. Contrast this with a corresponding zoom, Algorithm A.1, of the exponential function which is boringly smooth.
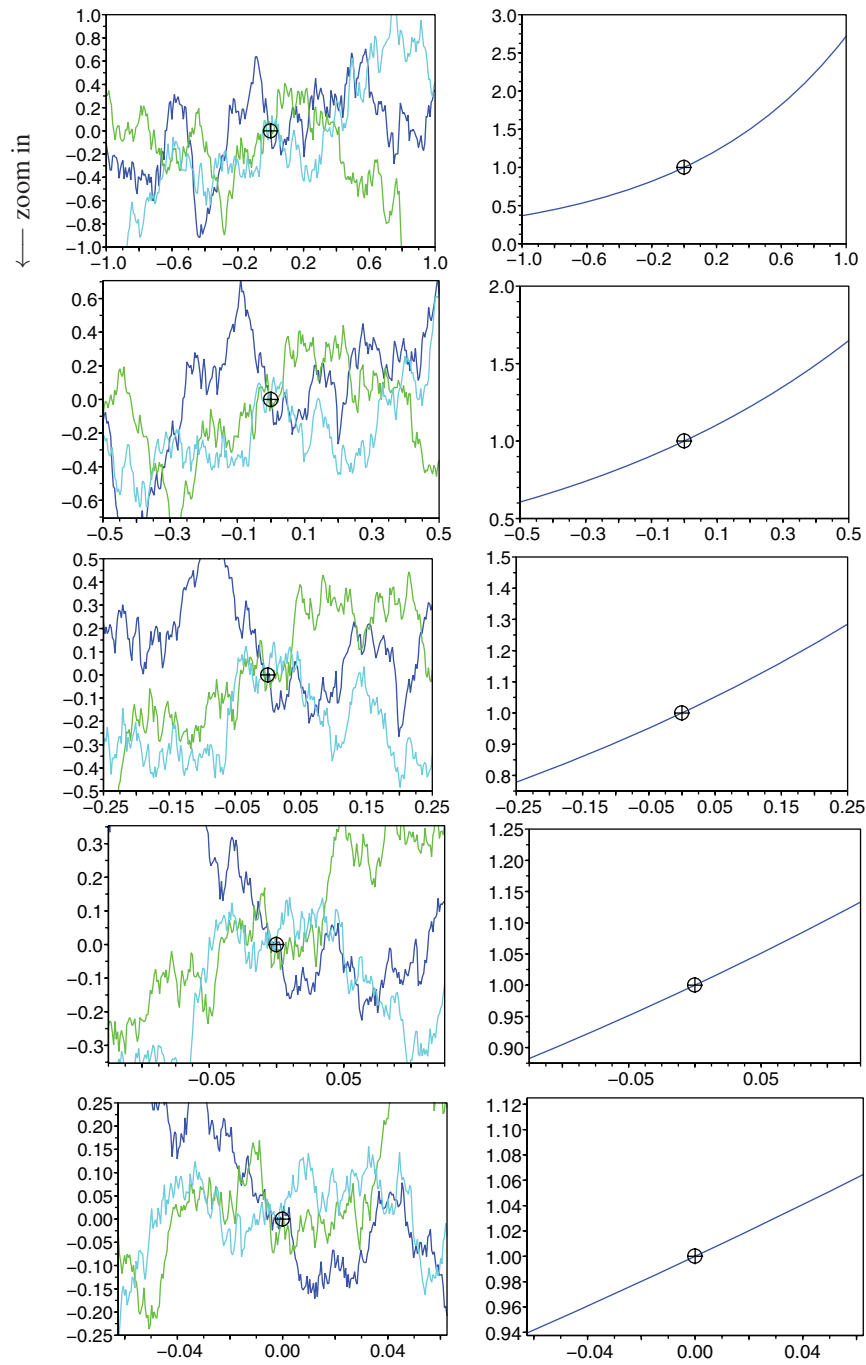
Figure 1.4. *Left: From top to bottom, zoom into three realizations of a Wiener process showing the increasing level of detail and jaggedness. Right: From top to bottom, zoom into the smooth exponential e^t. Time is plotted horizontally.*

- From its definition and from the properties of scaling and adding normally distributed independent random variables,

$$X(t+s) - X(s) = \rho W(t+s) + \sqrt{1-\rho^2}\widetilde{W}(t+s) - \rho W(s) - \sqrt{1-\rho^2}\widetilde{W}(s)$$

$$= \rho \underbrace{\underbrace{[W(t+s) - W(s)]}_{\sim N(0,t)}}_{\sim N(0,\rho^2 t)} + \underbrace{\sqrt{1-\rho^2}\underbrace{[\widetilde{W}(t+s) - \widetilde{W}(s)]}_{\sim N(0,t)}}_{\sim N(0,(1-\rho^2)t)}$$

$$\underbrace{}_{\sim N(0,\rho^2 t + (1-\rho^2)t) = N(0,t)} \quad .$$

- Also from its definition, $X(t+s) - X(s) = \rho[W(t+s) - W(s)] + \sqrt{1-\rho^2}[\widetilde{W}(t+s) - \widetilde{W}(s)]$, but neither $W(t+s) - W(s)$ nor $\widetilde{W}(t+s) - \widetilde{W}(s)$ depends on the earlier details of W or $\widetilde{W}$, so neither does $X(t+s) - X(s)$, and hence this increment cannot depend upon the earlier details of X. ∎

Summary

The Wiener process $W(t)$ is *the* basic stochastic process from which we build an understanding of system with fluctuations. The Wiener process is continuous but not differentiable; its independent random fluctuations, ΔW, scale with the square root of the time step, $\sqrt{\Delta t}$.

1.2 Stochastic drift and volatility are unique

Applying the basic Wiener process, Figure 1.3, to the prices of assets needs refinement for three reasons:

1. Different assets have different amounts of fluctuation;

2. risky assets generally have a positive *expected* return, whereas the expected value of a Wiener process $E[W(t)] = 0$ as it is distributed $N(0,t)$;

3. the Wiener process assumes that any step, ΔW, is independent of the magnitude of W, whereas we expect that any change in the value of an asset, ΔS, would be proportional to S—investors expect, for example, to get twice the return from a doubling in investment.

Deal with each of these in turn:

1. Increase the size of the fluctuations by scaling the Wiener process to the asset value $S(t) = \sigma W(t)$, where the scaling factor σ is called the *volatility* of the asset. That is, any increment in asset value $\Delta S = \sigma \Delta W$ is distributed $N(0, \sigma^2 \Delta t)$, where Δt is the time step (like h). Symbolically we write this as

$$dS = \sigma \, dW.$$

Interpret this differential equation as meaning $S_{j+1} = S_j + \sigma\sqrt{\Delta t}Z_j$ for $Z_j \sim N(0,1)$ as before.
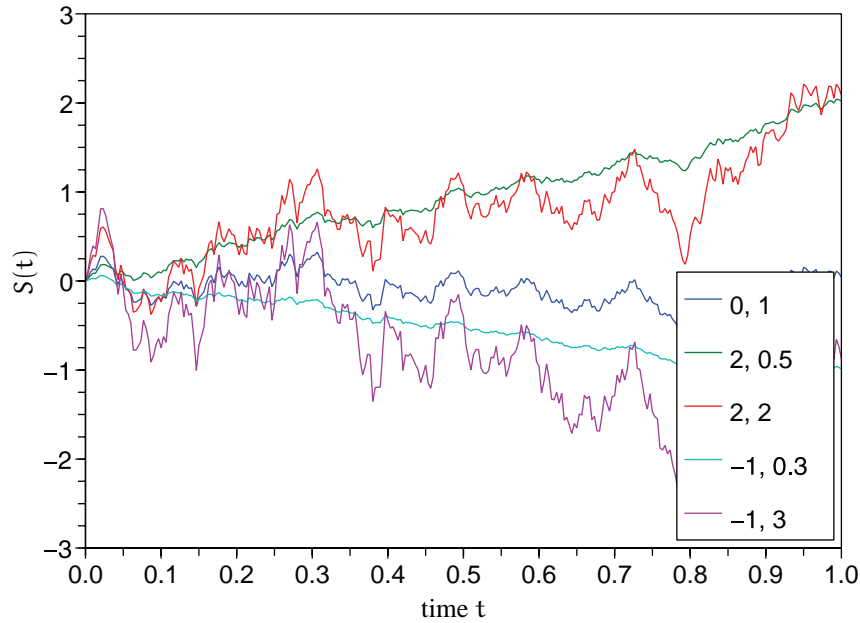
Figure 1.5. *Algorithm* 1.2 *draws this realization of a Wiener process,* $\mu = 0$ *and* $\sigma = 1$*, with four other transformed versions, labeled "*μ, σ*," with various drift* μ *and volatility* σ *but for the same realization of the Wiener process.*

2. We expect (hope) the asset will increase in value in the long term. One way to model such growth is to ensure that the increments have a nonzero mean. Simply add a little to each increment:

$$S_{j+1} = S_j + \mu \Delta t + \sigma \sqrt{\Delta t} Z_j,$$

   where the parameter μ is called the *drift*. The factor Δt is used so that μ is interpreted as the expected rate of growth of $S(t)$; for example, in the absence of fluctuations (the volatility $\sigma = 0$) the price $S(t) = S_0 + \mu t$ exactly, whence μ is the slope of the increase in price. Figure 1.5 shows a realization of a Wiener process and various $S(t)$ derived from it with different drifts and volatilities, as computed by Algorithm 1.2. Equivalently we write $\Delta S = \mu \Delta t + \sigma \Delta W$, or in terms of infinitesimal differentials,

$$dS = \mu \, dt + \sigma \, dW. \tag{1.1}$$

   Equation (1.1) symbolically records the general *stochastic differential equation*[5] when we allow, as we do next, the drift μ and the volatility σ to vary with price S and/or t instead of being constant.

---

[5]In the physics and engineering communities, SDEs such as (1.1) are written with the appearance of ordinary differential equations, called a Langevin equation, namely $dS/dt = \mu(t, S) + \sigma(t, S)\xi(t)$, recognizing that $\xi(t)$ represents what is called white noise. This is an SDE in a different disguise. However, physicists and engineers usually interpret it in the Stratonovich sense, which is subtly different from the Ito interpretation of SDEs developed in this chapter.

Algorithm 1.2 MATLAB/SCILAB code to draw five stochastic processes, all scaled from the one Wiener process, with different drifts and volatilities.

```
n=300;
t=linspace(0,1,n+1)';
h=diff(t(1:2));
dw=sqrt(h)*randn(n,1);
w=cumsum([0;dw]);
plot(t,t*[0 2 2 -1 -1]+w*[1 0.5 2 0.3 3])
legend('0, 1','2, 0.5','2, 2','-1, 0.3','-1, 3',4);
```

3. In (1.1) μ is the *absolute* rate of return per unit time. However, financial investors require a percentage rate of return. That is, we expect increments *relative* to the current value to be the same. For example, for an asset with no volatility (treasury bonds, perhaps) we obtain exponential growth (compound interest) such as $S(t) = S_0 e^{rt}$, where r is the (continuously compounded) *interest rate*. Differentiating this gives $dS/dt = rS$, whence $dS/S = r\,dt$, and so the relative increment in the price of the asset is $\Delta S/S = r\,\Delta t$. For a stochastic quantity the immediately generalize to assume that the relative increment has both a deterministic component, as above, and an additional stochastic component,

$$\frac{\Delta S}{S} = \alpha\,\Delta t + \beta\,\Delta W,$$

for some constants α and β called the *stock drift* and the *stock volatility*, respectively. Figure 1.6 plots $\Delta S/S$ for cotton prices from which one could roughly estimate the magnitude of α and β; Algorithm 1.3 provides code to numerically estimate α and β. Rearranging and writing in terms of differentials, we suppose throughout our applications to finance that assets satisfy the SDE

$$dS = \alpha S\,dt + \beta S\,dW \tag{1.2}$$

Algorithm 1.3 In MATLAB/SCILAB, given h is the time step, this code estimates the stock drift and stock volatility from a times series of values s. In SCILAB use $ instead of end.

```
dx=diff(s)./s(1:end-1)
alpha=mean(dx)/h
beta=std(dx)/sqrt(h)
```
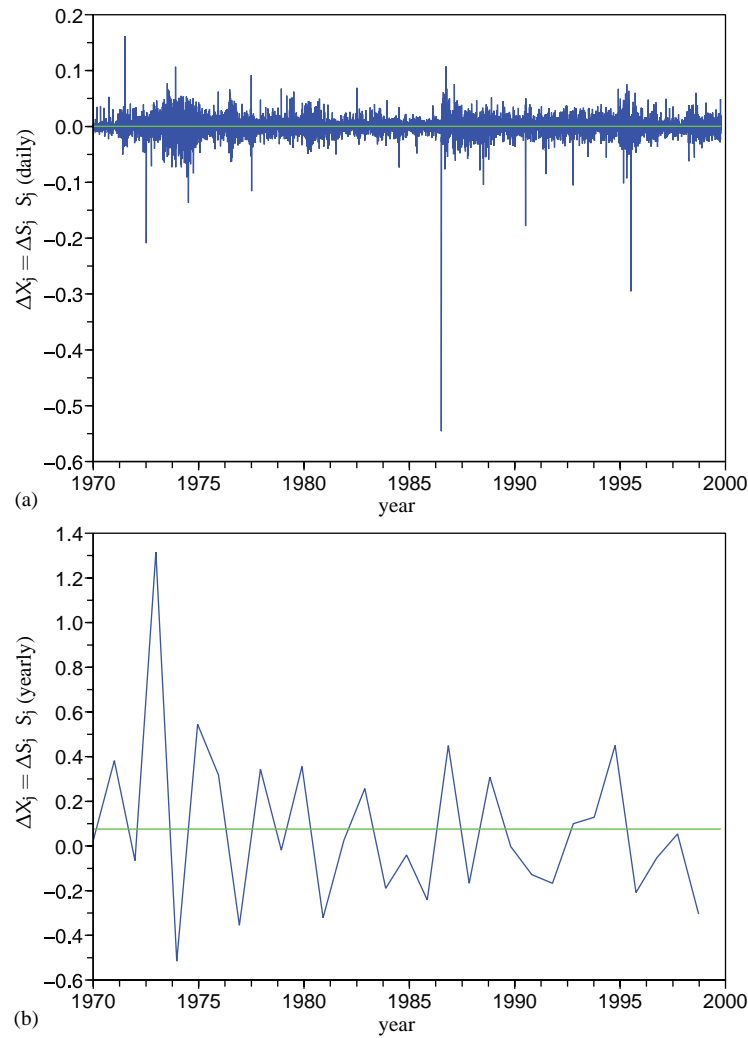
Figure 1.6. *Relative increments,* $\Delta S/S$, *of the cotton prices of Figure* 1.1 *as a function of time:* (a) *shows a daily drift and volatility coefficients of* $\alpha = 2.77 \times 10^{-4}/day$ *and* $\beta = 0.0181/\sqrt{day}$, *respectively;* (b) *shows that this translates into* $\alpha = 6.9\%/year$ *and* $\beta = 28.6\%/\sqrt{year}$.

with drift $\mu = \alpha S$ and volatility $\sigma = \beta S$. The only straightforwardly observed departure from the model (1.2) is that financial indices generally have larger negative "jumps" than predicted by the Weiner dW: that is, rare falls in the financial market are too big for a Wiener process; such rare falls reflect financial "crashes." However, for most purposes the stochastic model (1.2) is sound. We use it throughout our exploration of finance.

Table 1.1. *Annual closing price of wheat and its relative increments. The mean and standard deviation of the last column estimates the stock drift and stock volatility.*

Year	Close S_j	ΔS_j	$(\Delta S_j)/S_j$
2001	$3.00	$0.42	0.14
2002	$3.42	$0.36	0.11
2003	$3.78	$-0.27	-0.07
2004	$3.51	$-0.23	-0.07
2005	$3.28	$1.48	0.45
2006	$4.76		

Example 1.4 (stock drift and volatility). The price of wheat in US$ per bushel closed at the prices shown in the second column of Table 1.1. Given this data, estimate the stock drift and stock volatility of the price of wheat.

Solution: Compute the changes in the price as shown in the third column of Table 1.1: $\Delta S_j = S_{j+1} - S_j$. Compute the relative changes in the price as shown in the fourth column of Table 1.1: $(\Delta S_j)/S_j = (S_{j+1} - S_j)/S_j$. The average of these five numbers in the fourth column estimates the stock drift: $\alpha \approx 0.11 = 11\%$ per year. The standard deviation of these five numbers in the fourth column estimates the stock volatility: $\beta \approx 0.21 = 21\%$ per $\sqrt{\text{year}}$.

Of course, these estimates are extremely crude because of the small amount of data supplied by Table 1.1. ∎

Definition 1.5. *An* Ito process *satisfies an* SDE *such as* (1.1).

This definition is ill-defined, as we have not yet pinned down precisely what is meant by $dS = \mu\,dt + \sigma\,dW$. Precision comes in Chapter 4. For the moment we continue to develop and work with an intuitive understanding of the symbols. Finally, the *Doob–Meyer decomposition* theorem (which I will not prove) asserts that any given Ito process $X(t)$ has a unique drift μ and volatility σ. Thus we are assured that there is a one-to-one correspondence between SDEs, $dX = \mu\,dt + \sigma\,dW$, and the stochastic process which is its solution $X(t)$.[6] Figure 1.7 shows that by zooming into an Ito process, we find that at the smallest scales the process looks like one with linear drift and constant volatility. Thus the decomposition of the process into the SDE form $dX = \mu\,dt + \sigma\,dW$ is justified because this form applies on small scales—we just imagine that a host of such little pictures can be "pasted" together to form the large scale process $X(t)$.

Summary

A stochastic process $X(t)$ satisfies an SDE, symbolically written $dX = \mu\,dt + \sigma\,dW$, with some drift μ and some volatility σ. Financial assets satisfy particular SDEs of the form $dS = \alpha S\,dt + \beta S\,dW$.

---

[6]Note that the term "process" applies to the entire ensemble of realizations of a stochastic function. A stochastic process is *not* just any one realization because each realization is markedly different. Indeed, an SDE gives rise to an infinitude of realizations. It is the ensemble of possible solutions with their probability of being realized that is included within the term "stochastic process."
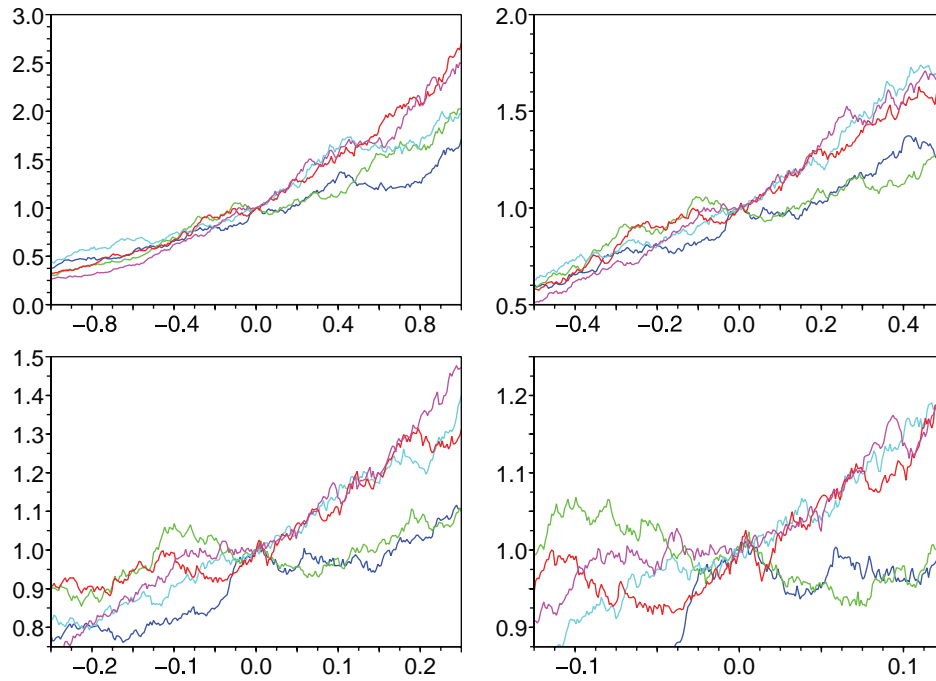
Figure 1.7. *Five realizations of the Ito process* $X = \exp(t + 0.2\,W(t))$ *shown at different levels of magnification (the* t *axis is horizontal and* $X(t)$ *is plotted vertically). The top left displays the largest scale showing the exponential growth in* $X(t)$*; the bottom right displays the smallest scale view of* $X(t)$ *showing that on small scales it looks just like a stochastic process with linear drift and constant volatility, as promised by the Doob–Meyer decomposition.*

1.3   Basic numerics simulate an SDE

Section 2.2 in Chapter 2 develops algebraic solutions of SDEs. Here we resort to a simple numerical technique to approximate solutions of SDEs. The numerical solution of SDEs such as that for assets (1.2) or the general SDE (1.1) is based on replacing the infinitesimal differentials with corresponding finite differences. In a sense this undoes the limit $h \to 0$ discussed in the previous section.

1.3.1   The Euler method is the simplest

Rewrite the general SDE (1.1) in terms of finite differences (increments)

$$\Delta S = \mu\,\Delta t + \sigma\,\Delta W,$$

and then evaluate it at the jth time step, recognizing that the drift and volatility are generally functions of S and t so that

$$\Delta S_j = S_{j+1} - S_j = \mu(S_j, t_j)\Delta t_j + \sigma(S_j, t_j)\Delta W_j,$$

Algorithm 1.4 Code for simulating five realizations of the financial SDE $dS = \alpha S\, dt + \beta S\, dW$. The for-loop steps forward in time. As a second subscript to an array, the colon forms a row vector over all the realizations.

```
alpha=1, beta=2
n=1000; m=5;
t=linspace(0,1,n+1)';
h=diff(t(1:2));
s=ones(n+1,m);
for j=1:n
  dw=randn(1,m)*sqrt(h);
  s(j+1,:)=s(j,:)+alpha*s(j,:)*h+beta*s(j,:).*dw;
end
plot(t,s)
```

where ΔW_j are independent random samples from $N(0, \Delta t_j)$. Rearranging for S_{j+1} gives the recursion

$$S_{j+1} = S_j + \mu(S_j, t_j)\Delta t_j + \sigma(S_j, t_j)\Delta W_j. \tag{1.3}$$

This form allows the time steps to be different. The Euler method (1.3) is a most general and simple method to numerically simulate the realizations of SDEs.

Usually, for simplicity, we choose a fixed step in time of $\Delta t_j = h$, whence $t_j = jh$ and $\Delta W_j = \sqrt{h}Z_j$ for random variables $Z_j \sim N(0,1)$ so that the Euler method (1.3) is typically invoked as

$$S_{j+1} = S_j + \mu(S_j, t_j)h + \sigma(S_j, t_j)\sqrt{h}Z_j. \tag{1.4}$$

Example 1.6 (exponential Brownian motion). The solutions of the SDE $dS = \alpha S\, dt + \beta S\, dW$ are collectively called exponential Brownian motion. For any given α and β, the Euler method (1.4) reduces to

$$S_{j+1} = S_j + \alpha S_j h + \beta S_j \sqrt{h}Z_j. \tag{1.5}$$

With the initial condition that $S_0 = 1$, Algorithm 1.4 shows how this method may be implemented in MATLAB/SCILAB to generate many different realizations.

For three different combinations of stock drift α and volatility β, Figures 1.8–1.10 plot five realizations of the solution $S(t)$. See that increasing stock drift α indeed increases the rate of growth, and that increasing the stock volatility β indeed increases the level of fluctuations.

Notice that in this last case with relatively large volatility, Figure 1.10, the realizations, apart from a couple of large excursions, generally seem to stay smallish in magnitude. This is very strange, since the deterministic part of the SDE indicates that the solutions should grow. However, it is not so. Noise can act to stabilize growth; in financial applications, high volatility can act to keep stocks low even if they would otherwise increase. This surprising result is supported by later algebraic results.

This simple Euler method (1.4) for an SDE looks just like a Markov chain, as discussed in stochastic process modelling (see, e.g., Kao 1997). The difference is that instead of discrete states, here we have a continuum of states parametrized by the asset price S.

Figure 1.8. *Simulation of the exponential Brownian motion of asset values with stock drift* α = 1 *and stock volatility* β = 0.5 *as generated by Algorithm* 1.4.

Figure 1.9. *Simulation of the exponential Brownian motion of asset values with larger stock drift* α = 2 *and stock volatility* β = 0.5 *as generated by Algorithm* 1.4.

Figure 1.10. *Simulation of the exponential Brownian motion of asset values with stock drift $\alpha = 1$ and larger stock volatility $\beta = 2$ as generated by Algorithm* 1.4.

But in the numerical code, time is still discrete. At times t_j the system is in the state of the asset having price S_j. At the next time t_{j+1}, the system makes a transition to a state S_{j+1}, a distance $\mu_j h + \sigma_j \sqrt{h} Z_j$ away, where the stochastic aspect enters through the normally distributed random variable Z_j. In stochastic process modeling we investigate probability distributions, whereas here for the moment we focus on realizations. Chapter 3 addresses probability distributions of solutions to SDEs.

Why use MATLAB/SCILAB?

True, the financial industry standard is to use spreadsheets for almost all numerical computation. However, spreadsheets are many orders of magnitude slower than script and programming languages (furthermore, common spreadsheet programs have notorious errors). I do not endorse crippling your power by limiting yourself to inefficient tools. Instead, I seek to empower you to manage large-scale numerical problems through MATLAB/SCILAB. I quote from David Chadwick, *Stop The Subversive Spreadsheet!*: [7]

> The presence of a spreadsheet application in an accounting system can subvert all the controls in all other parts of that system." So says Ray Butler of the Computer Audit Unit, HM Customs and Excise and Ray should know. For ten years he has been investigating errors in spreadsheets used by companies for calculating their VAT payments. Over this period, he has collected useful data on types and frequencies of errors as well as on the effectiveness of

---

[7] http://staffweb.cms.gre.ac.uk/~cd02/EUSPRIG

different audit methods not only those used by VAT officers but also those in use by other auditors. Ray has written extensively about the phenomenon of errors in spreadsheets (an excellent example is to be found in his article entitled 'The Subversive Spreadsheet'. It seems surprising that Ray would find such a problem in such a straightforward well-defined business application but as he is quick to point out "Even in a domain such as indirect taxation, which is characterised by relatively simple calculations, relatively high domain knowledge by developers, and generally well-documented calculation rules, the use of spreadsheet applications is fraught with danger and errors."

Ray is not alone in his interest of spreadsheet risks. Chris Conlong of the Business Modelling Group at KPMG Consulting is also only too aware of the problems and, when asked, frequently refers to the findings of a KPMG survey of financial models based on spreadsheets. The survey found that 95% of models were found to contain major errors (errors that could affect decisions based on the results of the model), 59% of models were judged to have 'poor' model design, 92% of those that dealt with tax issues had significant tax errors and 75% had significant accounting errors.

These figures are truly astounding and if extrapolated to all major organisations throughout the world hint at potential disaster scenarios just waiting to happen. A colleague recently remarked "Spreadsheet errors are a business time-bomb waiting to go off. They're a bit like the millennium bug — nobody knew the time-bomb was there until it was pointed out and then everybody knew and knew when it would happen. However, with the spreadsheet problem few people know that there is a time-bomb at all and none knows when their particular bomb may go off."

1.3.2   Convergence is relatively slow

As with any numerical method, we expect the numerical solution to approach the true solution as the time step $h \to 0$. This is true for the Euler method (1.3), but the rate of convergence is slow: for a deterministic differential equation the error of the Euler method is generally $\mathcal{O}(h)$, that is, the error decreases in proportion to h; for an SDE the error of the Euler method is the larger $\mathcal{O}(\sqrt{h})$. For example, for a deterministic differential equation with time steps of $h = 0.001$ we would expect an error of about 0.1%, whereas for an SDE the error would be about $\sqrt{0.001} = 3\%$. The relatively slow convergence of numerical solutions of SDEs is difficult to overcome.

For now we illustrate just the convergence. One crucial issue for SDEs is that different realizations of the Wiener process—the noise—generate quite different looking realizations of the solution $S(t)$. See the different realizations in each of the above figures, for example. Thus, to examine convergence *we must retain the one realization of the Wiener process as we vary the step size* h. Consequently, in the following example we generate the Wiener process first with the finest time step, then sample it with increasingly coarse time steps.

Example 1.7 (convergence to exponential Brownian motion). Superimpose plots of numerical solutions of the SDE $dS = S\,dt + 1.5S\,dW$ with initial condition $S_0 = 1$ for different time steps h to show qualitatively the relatively slow convergence as $h \to 0$.
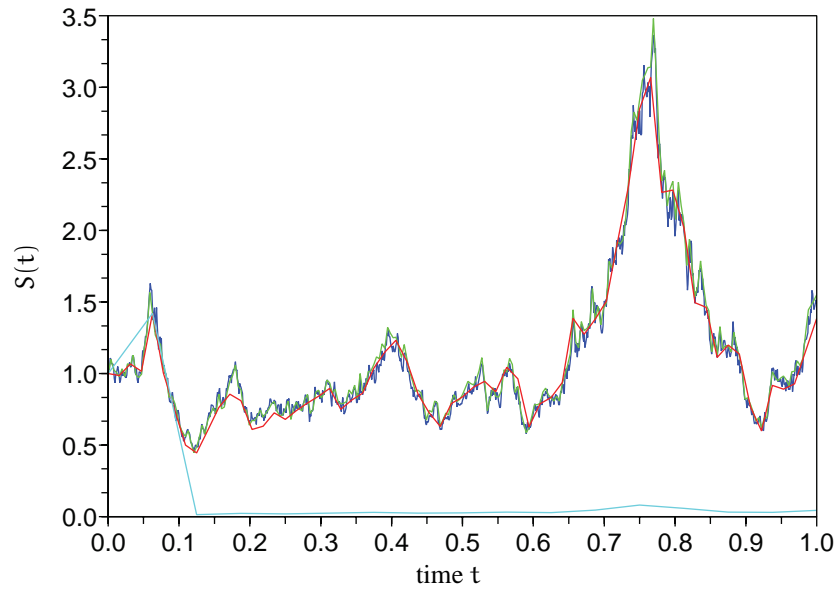
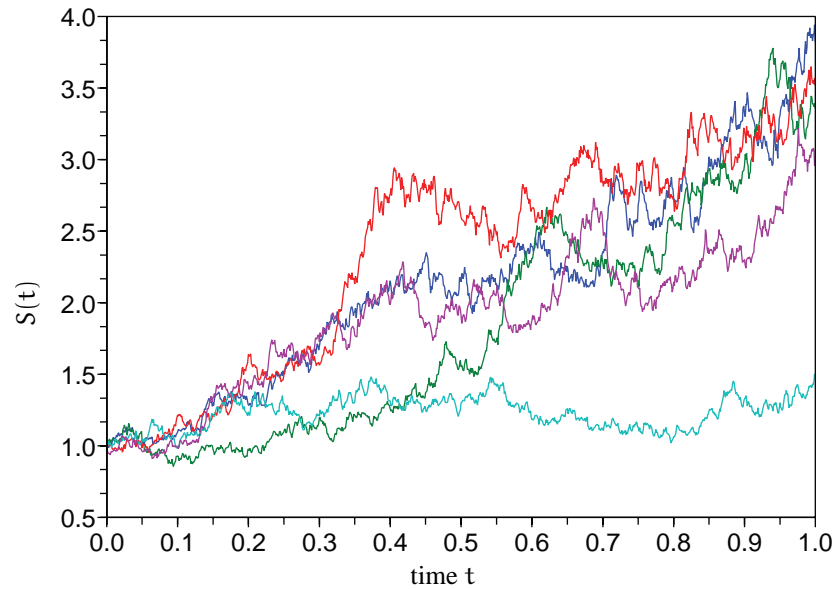**Figure 1.11.** *One realization of the exponential Brownian motion of asset values with stock drift α = 1 and stock volatility β = 1.5 as generated by Algorithm 1.5 with different time steps h = 1/16 (cyan), 1/64 (red), 1/256 (green), and 1/1024 (blue) to illustrate convergence as h → 0.*
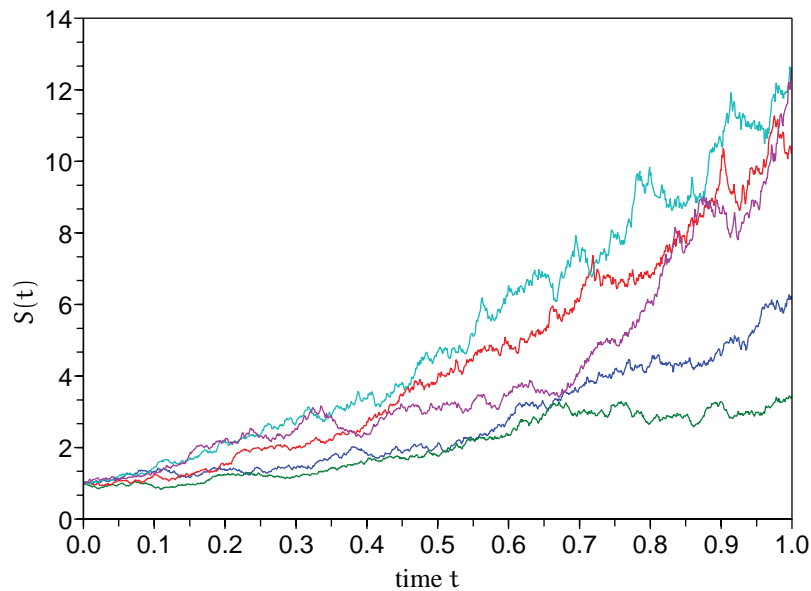

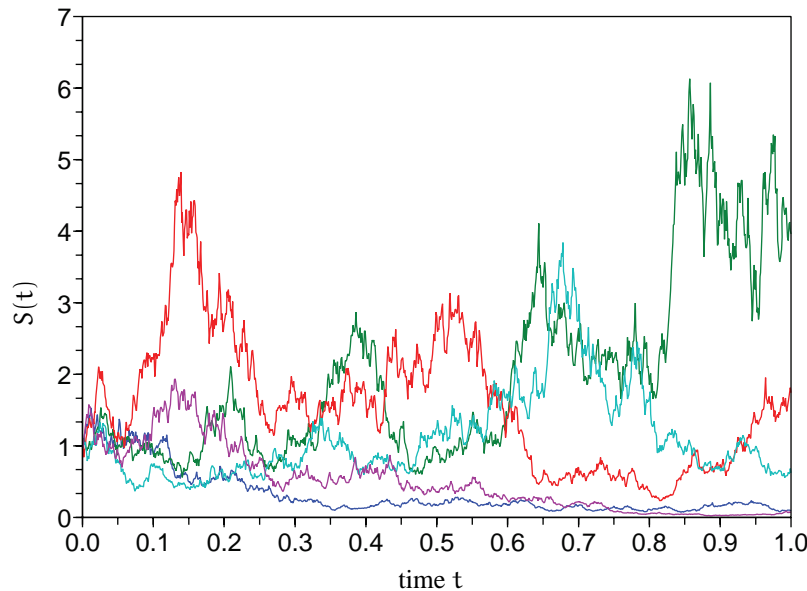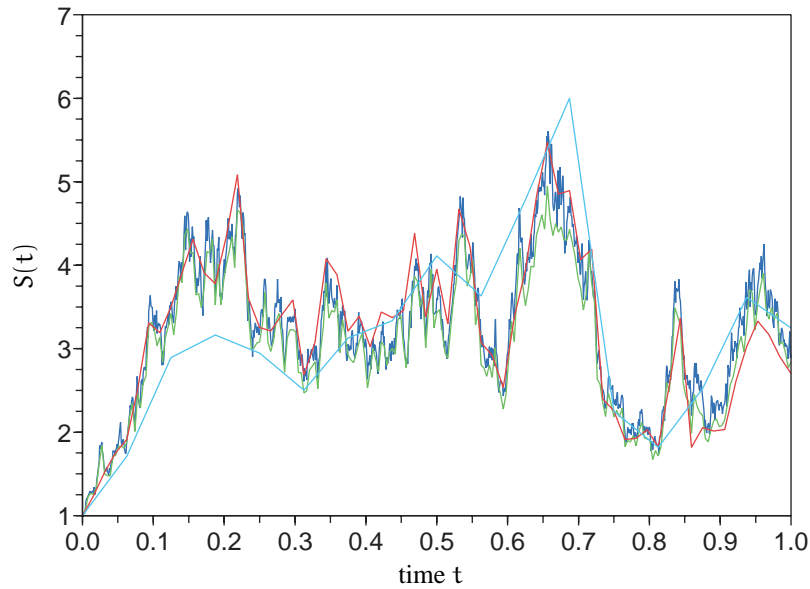**Figure 1.12.** *One realization of the exponential Brownian motion of asset values with stock drift α = 1 and stock volatility β = 1.5 as generated by Algorithm 1.5 with different time steps h = 1/16 (cyan), 1/64 (red), 1/256 (green), and 1/1024 (blue) to illustrate convergence as h → 0.*

Algorithm 1.5 Code for starting *one* realization with very small time step, then repeat with time steps four times as long as the previous.

```
alpha=1, beta=1.5
n=4^5;
t=linspace(0,1,n+1)';
w=[0;cumsum(sqrt(diff(t)).*randn(n,1))];
for col=['b' 'g' 'r' 'c']
    dt=diff(t); dw=diff(w);
    s=ones(n+1,1);
    for j=1:n
        s(j+1)=s(j)+alpha*s(j)*dt(j)+beta*s(j)*dw(j);
    end
    plot(t,s,col), hold on
    t=t(1:4:end); w=w(1:4:end); n=n/4;
end
```

Adapt the code used before. Algorithm 1.5 finds numerical solutions on $0 \le t \le 1$ with n steps of length $h = 1/n$; first with $n = 4^5 = 1024$, then increasing the step length h by a factor of 4 and correspondingly decreasing n by a factor of 4 each time we draw a new approximation. We generate the Wiener process, $W(t)$, with the smallest step length, and then for each approximation we sample the *one realization* of $W(t)$ with a step four times larger than the previous approximation by subsampling statements such as `w=w(1:4:end)`.

Executing Algorithm 1.5 twice generated Figures 1.11–1.12, one for each of the two different realizations of the Wiener process $W(t)$. In each figure *the numerical solutions do seem to converge, albeit a little slowly.* ∎

Suggested activity: Do at least Exercise 1.7(1).

Summary

The Euler method (1.4) numerically computes realizations of SDE (1.1) provided the time step is small—generally use a time step $h \approx 0.001$ or less. In a free online article, Higham (2001) presents more information on such basic methods for SDEs and their convergence and points to more sophisticated and accurate schemes.

1.4   A binomial lattice prices call option

We use the principle that there cannot be risk-free profit, called arbitrage, to price options in stochastic finance. The arguments employed here to introduce the rational pricing of options looks like a numerical approximation of SDEs. But we simplify even more than previously by not only discretizing time but also discretizing asset prices. The analysis presented here is based on research by Black, Scholes, and Merton (circa 1973) which later was simplified, as we now see.[8]

---

[8]Read *A calculus of risk* by Gary Stix in the May 1998 issue of *Scientific American*. Chapters 2 and 3 of the book by Stampfli and Goodman (2001) provide reading to supplement this section.

Interestingly, the application in Example 1.12 shows that such valuation of options fully justifies taking action on global warming almost *irrespective of the actual probability* that global warming is occurring!

Definition 1.8. *A European* call option *gives the buyer the right but not the obligation to purchase an asset from the seller at a previously agreed price on a particular date.*[9] *The agreed price is called the* exercise price *or the* strike price.

For definiteness we refer to the seller of a call option as Alice and to the buyer as Bob.[10] The call option: is sold by Alice so she obtains money to invest elsewhere; is exercisable by Bob at the fixed price at the specified date, but can be abandoned by Bob without penalty.

Insurance is a form of call option

Bob buys an insurance policy for his car from Alice, the insurance company. If, during the ensuing year, Bob damages his car so that it is worth less, then he makes an insurance claim—Bob exercises the option. Alternatively, if during the year Bob's car remains undamaged so it is worth the same (ignoring normal wear and tear), then Bob makes no claim—Bob does not exercise the option and Alice keeps the price of the insurance policy as profit. That Bob makes a claim on an insurance policy (or not) is closely analogous to Bob exercising an option (or not).

Expiry value of a call option

At the expiry date, Bob is not going to buy the asset from Alice for a price above what he could pay on the open market. Thus if the listed price of the asset is ultimately below the strike price, then the option is worthless and Bob tosses it away. However, if the listed price is higher than the strike price of the call option, then Bob will exercise his option and buy the asset from Alice so that, if nothing else, he could then dispose of the asset at a profit. The value of the call option to Bob *at the expiry date* is thus

$$C = \max\{0, S - X\},$$

where S is the price of the asset and X is the exercise (strike) price. A call option has some nonnegative value because, depending on the vagaries of market fluctuations, it may be worth either nothing or something, but is never a liability.

Example 1.9 (trade a call option). Alice holds ten units of Telecom shares, an asset, at the start of the year; we say altogether they are worth \$35. She sells to Bob at \$4 per call option, allowing him to buy the ten shares at the end of the year at the strike price of $X = \$38.50$.

---

[9]An American call option is the same but additionally allows the buyer the right to purchase the asset before the particular date. Throughout this section we discuss only the easier European call option and the corresponding put option.

[10]Alice and Bob first made a name for themselves in helping to solve problems in quantum physics. They made their contribution in the area of quantum cryptography, where they developed methods of foiling that dastardly eavesdropping spy, known only by her code name, Eve. Since then Alice and Bob have worked for us in financial theory.

If the price of the ten shares on the open market goes up 25% to $S = \$43.75$, then Bob will exercise his right to buy the shares from Alice for $38.50 and sell them immediately for $43.75, which gives him $1.25 profit because he originally paid Alice $4 for the call option. Alternatively, if the shares drop in price by 20% to $S = \$28$, then Bob will not exercise his option because he would lose money if he did (over and above the $4 he originally paid for the option). ∎

Questions: In this example, is the $4 that Bob paid to Alice a fair price? Should Alice have asked for more, or should Bob have insisted on less? How can we decide?

We answer these questions by a model that is a Markov chain and hence looks like a numerical solution of an SDE. Section 2.3 examines the continuous time, continuous state version using more theory of SDEs. The principle we use is that of *arbitrage*: *there cannot exist the opportunity for risk-free profit*. In this context, "profit" means an amount over and above that which is obtained by investing in secure bonds. We discover that if the call option is valued too highly, then Alice could form a portfolio of the asset and sell options so that she is guaranteed to profit. Alternatively, if the call option is valued too low, then Bob could buy the options with money obtained through selling the asset so that he is guaranteed to profit. It is only when the call option is precisely and correctly valued that nobody is *guaranteed* to do better than invest in bonds. With a correct valuation of options, people must accept risk in order to better their return from bonds.

1.4.1 Arbitrage value of forward contracts

Before proceeding to the interesting case of options, let us introduce arbitrage in its application to valuing forward contracts.

Example 1.10 (Telecom shares). Instead of giving the *option* to Bob of buying the Telecom shares of Example 1.9, suppose Alice and Bob write a *forward contract*: The forward contract states that Bob will definitely purchase the asset, the ten Telecom shares, from Alice for $38.50 at the end of one year. However, the contract involves risk: If the market value goes above $38.50, then Bob gains because he obtains the asset at a cheaper price than the open market price, and Alice loses; if the market value stays below $38.50, then Bob loses because he must buy at a higher price than the open market price, and Alice correspondingly gains. Who should pay what to whom at the start of the year for this forward contract? Arbitrage determines the value of the forward contract.

Solution: Suppose Bob pays Alice F_0 for the forward contract (negative F_0 means Alice pays Bob). Alice avoids risk by purchasing the asset at the start of the year for $35 with her own funds; she then has the asset of the shares ready for sale at the agreed price at the end of the year. Since Bob pays her F_0 for the forward contract, Alice invests only $35 − F_0 into this portfolio of the asset and the forward contract. Alice makes this investment to receive $38.50 at the end of the year—risk free.

Arbitrage asserts that any risk-free return must be the same as investing in bonds. Alternatively, Alice could invest this amount, $35 − F_0, in risk-free bonds paying, say, 12% interest over the year. Thus Alice could alternatively obtain a risk-free return of $1.12(\$35 − F_0)$ by investing in bonds. Thus arbitrage asserts $\$38.50 = 1.12(\$35 − F_0)$. Rearranging $F_0 = \$35 − \$38.50/1.12 = \$0.625$. That is, Bob should pay Alice 62.5 cents for this forward contract. ∎

$$\boxed{Sd = 28} \longleftarrow \boxed{S = 35} \longrightarrow \boxed{Su = 43.75}$$
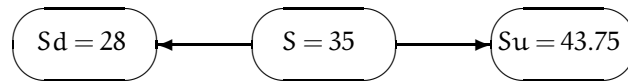
Figure 1.13. *State transition diagram for a simple binomial model of asset prices over one period: The prices either go up by 25% or down by 20%.*

$$Su = 43.75$$
$$C_u = 5.25$$

$$S = 35$$
$$C = ?$$

$$Sd = 28$$
$$C_d = 0$$
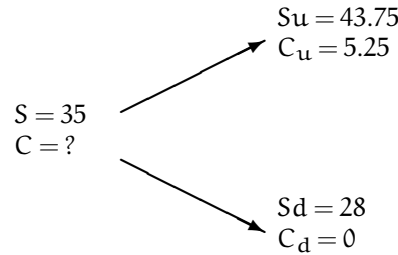
Figure 1.14. *Rotate the states of Figure* 1.13 90° *and plot time horizontally to give a simple binomial tree model of asset prices over one period.*

Now use arbitrage to value a general forward contract. Suppose Alice and Bob contract for Bob to purchase an asset from Alice at time T for some agreed price X. Let Bob and Alice value this forward contract at F_0 at the start of the period, that is, Bob pays Alice F_0 to sign the contract. To avoid risk, Alice could purchase the asset immediately for its known current price S_0. She would invest $S_0 - F_0$ of her funds to do this and would then receive X at expiry time T from Bob—risk free. Being risk free, this return must be the same as what Alice would get by investing the same amount, $S_0 - F_0$, in bonds. At the expiry time T the bonds would be worth $R(S_0 - F_0)$ for some R reflecting the interest rate of the bonds. Arbitrage asserts $X = R(S_0 - F_0)$, which rearranged determines

$$F_0 = S_0 - X/R \tag{1.6}$$

to be the value of a forward contract.

Analogous reasoning values options: we need to value Alice's investment and to discount the expiry value by the bond rate.

1.4.2 A one step binomial model

Our initial simple analysis rests on a Markov chain with just three states representing the possible prices of an asset now and in a year's time. For definiteness we continue with the numbers used in Example 1.9; Figure 1.13 shows the state transition diagram. From the current state of the asset price being $S = \$35$, here we restrict our attention to the two possibilities that the price after a year has either risen by 25% or fallen by 20%; that is, the asset price is multiplied by a factor of $u = 1.25$ or $d = 0.80$, respectively. I do not record any probabilities for the transitions from the current price S because the probabilities of the price increasing or decreasing are irrelevant!

It is usual in financial applications to arrange the states vertically and to plot time horizontally as shown in Figure 1.14. Additionally we include in this diagram the value of the call option on the asset at the end of the year: if the asset price goes up, the call

option is worth $C_u = \$43.75 - \$38.50 = \$5.25$, as the strike price agreed to between Alice and Bob was $X = \$38.50$, whereas if the asset price goes down the call option is worthless, $C_d = \$0$, because Bob, if he were to buy the asset, would prefer to purchase it at the open market rate of $S = \$28$ rather than from Alice for $X = \$38.50$. But how do we determine the value, C, of the call option at the start of the year?

A risk-free profit from overpriced options

Suppose Alice wants to buy the asset of the Telecom shares. One possibility way to get the requisite \$35 is to invest \$23 of her own money and to sell three call options to Bob at the valuation of \$4 each. After one year, if the asset price goes down, then the options are worthless and Alice's asset, and hence her hypothetical portfolio, is worth \$28; whereas if the asset price goes up, she has to buy two more lots of Telecom shares, costing $\$43.75 \times 2$, in order to satisfy Bob's demand for the three lots of shares, for which Alice receives $\$38.50 \times 3$, and hence at the end of the year Alice will have $\$38.50 \times 3 - \$43.75 \times 2 = \$28$. The hypothetical portfolio's value of \$28, irrespective of whether the asset price rises or falls, is better than what Alice could have gained if she had invested her \$23 in bonds at, say, 12% interest, because at the end of the year that would have only been worth \$25.76. Alice makes a risk-free profit!

This risk-free profit shows that, at \$4, the call options are overpriced. The opportunity for risk-free profit would result in traders clammering to sell options, which would flood the market and hence reduce the price of the option. We soon find that the market forces the call option to have a definite price.

But before we do that, note that Alice's choice of selling three call options is the result of a careful balancing act.

- Suppose Alice had only sold two call options and invested \$27 of her own money to buy the asset; then if the market goes up, her portfolio is worth $\$38.50 \times 2 - \$43.75 = \$33.25$; whereas if the market goes down, it is worth only \$28. This latter figure is a markedly lower return than she could get from investing her \$27 in bonds, which at 12% interest would be worth \$30.24 at the end of the year. If the asset price increases, then Alice has done well, but the point is that this good outcome is *only achieved by taking the risk* that the portfolio will fall in value relative to bonds.

- Conversely, suppose Alice had sold four call options and invested \$19 of her own money to buy the asset; then if the asset price goes down, her portfolio is worth \$28; whereas if the asset price goes up, her portfolio is worth only \$22.75. Although these are both higher than the bond return, when the options are valued correctly, the imbalance between the two cases results in a potential loss for Alice, and *there is risk*.

Only by selling an appropriate number of options does Alice create a risk-free portfolio.

Hedge against fluctuations

We endeavour to discover under what conditions it is impossible to make a profit (above the bond rate) without risk. Look at the issue from Alice's point of view. Alice could construct a hypothetical portfolio of a unit of the asset and H call options sold to Bob, initially costing

Alice a net investment of $S - HC$. If the price of the asset goes down, then the call options are worthless, and so the value of the portfolio at the end of the year is just that of the asset, namely $Sd = \$28$. By the definition of a risk-free portfolio, this value must be the same as the value of the portfolio even if the asset rises in price—a rise in price increases the value of the asset, but also increases the value of the call option Alice sold to Bob. If the price of the asset goes up, then Bob will elect to exercise the call option, and at the end of the year Alice will have to sell H units of the asset to Bob at a price of X, receiving HX, but at the expense of having to buy $H - 1$ units of the asset at a price Su. Thus at the end of the year the portfolio is worth

$$HX - (H-1)Su = Su - H(Su - X) = Su - HC_u,$$

where $C_u = \$5.25$ is the net value of each call option if the asset price has gone up. In order for the portfolio to have the same value at the end of the year irrespective of whether the asset price has gone up or down, we must then have $Su - HC_u = Sd$, that is, $\$43.75 - H \times \$5.25 = \$28$. Solve this linear equation immediately to give us the magic result that Alice must sell $H = 3$ call options to insulate her portfolio from price fluctuations and ensure the value of $28 at the end of the year.

A fair price for the option

Now Alice would like this investment to be larger than the return from investing the initial capital, $S - HC = \$35 - 3C$, in bonds at 12% interest. Thus Alice wants

$$\$28 \geq 1.12 \times (\$35 - 3C) \quad \Rightarrow \quad C \geq \frac{1}{3}(\$35 - \$28/1.12) = \$\frac{10}{3} = \$3.33.$$

But if the price of the call option is higher than \$3.33, then the market will be flooded with people trying to sell such options. Thus the arbitrage price of the call option must be $C = \$3.33$. This is sensibly lower than the overpriced \$4 we used earlier.

Value an option first by finding how one could hedge against risk, and second by equating such a risk-free portfolio to investment in bonds.

Suggested activity: Do at least Exercise 1.11.

Example 1.11 (Bob wants to buy cheap options). In the converse of Alice's argument, Bob may want to buy a call option to protect himself from fluctuations when he *sells* an asset.

Solution: The argument is easier if we phrase it as Bob buying x units of the asset, where x will turn out to be negative to represent that he actually sells the asset. So at the start of the year Bob buys the three call options at price C from Alice and x units of the asset at $S = \$35$ each; this costs him $\$35x + 3C$ which, if Bob had not bought the call options and invested in bonds instead, would be worth $1.12 \times (\$35x + 3C)$ at 12% interest; if the asset price goes down, Bob's portfolio will be worth just $\$28x$, as the call options are worthless; whereas if the asset price goes up, the portfolio is worth $\$43.75x + 3 \times \$5.25$. To be protected against the fluctuations, that is, to be risk free, these last two outcomes have to be equal:

$$\$28x = \$43.75x + 3 \times \$5.25 \quad \Rightarrow \quad x = -1.$$

Thus Bob sells one unit of the asset as well as buying the three call options. To be worthwhile, this guaranteed return has to be at least as large as the return he would obtain by investing in bonds; thus

$$\$28x \geq 1.12(\$35x + 3C) \quad \Rightarrow \quad C \leq \$3.33.$$

Thus Bob would only be interested in buying call options if they were priced less than $3.33; whereas Alice was only interested in selling them for more than $3.33. The rational price for the call option must then be $C = \$3.33$.

Value a general option

We now derive the general formula for the one-step binomial price of a call option from the point of view of the seller, Alice. The above example shows that the buyer and seller of call options have opposing interests that balance at one price only.[11] There are two parts to the analysis: determining the number of options to sell, and determining the fair price for the options.

The number of options to sell Refer to the binary tree given earlier but only pay attention to the algebra. An asset bought by Alice at a cost of S at the start time may increase in price by a factor u to a price Su, or it may fall in price by a factor d to a price Sd. By selling H call options at the start of the year for some strike price X, Alice can protect herself against fluctuations—ensuring a risk free outcome. At the end of the period each call option is worth C_u or C_d depending on whether the asset price has risen or fallen, respectively (so far we have seen only $C_u = Su - X$ and $C_d = 0$, but more generally these have other values). To be risk free the portfolio must have the same value at the end no matter which outcome eventuated: thus $Su - HC_u = Sd - HC_d$, which, rearranged, gives the hedge ratio

$$H = \frac{Su - Sd}{C_u - C_d}. \tag{1.7}$$

Observe the reasonably intuitive result that, in order to make each outcome have the same value, the number of call options sold must be the ratio of the range of the asset price to the range of the option. Thus H is called the *hedge ratio*.

The fair price Recall that for a riskless portfolio there can only be rational buyers and sellers to establish the portfolio if the investment returns the same as the risk-free bond interest rate. Let $R = e^r$ be the factor by which bonds increase in price over the time period so that r is the equivalent continuously compounded interest rate over that time period (we used $R = 1.12$ in the earlier example, that is, $r = 0.1133 = 11.33\%$). Then the original investment by Alice of $S - Hc$, if put into bonds, would be worth $R(S - Hc)$. Setting the equivalent return in bonds equal to the return of the risk-free portfolio gives $R(S - Hc) = Su - HC_u (= Sd - HC_d)$ which, rearranged, gives the fair price of a call option as

$$C = \frac{S(R - u) + HC_u}{HR}.$$

---

[11]This balance between the two different views of a call option is rather like the beautiful relation between a linear programming problem and its dual that you see in operations research.

However, we obtain a more appealing form of this expression after substituting (1.7) and rearranging to

$$C = \frac{1}{R}[pC_u + (1-p)C_d], \tag{1.8}$$

$$\text{where} \quad p = \frac{R-d}{u-d} \quad \text{and} \quad 1-p = \frac{u-R}{u-d}.$$

In the earlier example, $p = 0.71$ and $1-p = 0.29$. See that the price of the call option is just a convex combination of the possible values at the end of the period discounted by the bond factor R to give a value for the start of the period.[12]

This argument is just used to set the prices of call options. People would not actually create a risk-free portfolio as given above because if they did, their portfolio would not do any better than the return from bonds. In general, people buy and sell options just as they buy and sell assets, namely, based on their own assessment of the future movement of market prices.

Example 1.12 (support research into global warming). Let us present the case for financing research projects as an option to ensure a future for humanity. In particular, we explore justifying a research project into the contentious issue of global warming.[13] Suppose a researcher in climatology, called JR for short, submits to a country's government a project proposal costing 1M$ in personnel and equipment, and promising increased knowledge to ameliorate any significant global warming. Should the government fund the project? How is the project to be valued?

We value the project as a European (put) option: the government is the buyer of the option (that is, pays for JR's project), and JR is the seller of the option. Interestingly, the project has this value *independent* of the probability that global warming actually occurs! This standard financial instrument of options values the project irrespective of the validity of the doubts and criticisms of the nay-sayers.

Now proceed with the option valuation, with almost entirely invented figures. Suppose the government funds JR's project by forming a portfolio of the funded project together with an investment into the country's economy. Say the economy is currently valued at 16 T$.[14] Suppose that the government's (potential) portfolio is this "bought" option with a (tiny) fraction ϕ (phi) of the economy invested in the economy.[15] Look a decade, say, into the future to value this portfolio. In the spirit of our binomial modeling, either one of two things will happen:

---

[12]The weights appearing in this convex combination appear as probabilities. But p and $1-p$ have nothing to do with the actual probability of an up or down step in the asset price. Nonetheless, the theory of transformations between different probability distributions, the Radon–Nikodym derivative, is based on the view that such weights do act like probabilities for some purposes.

[13]Although global warming is now generally accepted, and its dangers estimated, for the past 30 years or more global warming was a very contentious issue. Our valuation of research projects applies throughout this history, but such valuation is only recently recognized.

[14]1 T$ is one tera-dollar, more commonly known as a trillion dollars. 1 G$ denotes one giga-dollar, otherwise known as a billion dollars. 1 M$ denotes one mega-dollar, that is, a million dollars. I prefer this scientific notation.

[15]The variable ϕ is the reciprocal of the hedge ratio H used earlier. The reason is that now the portfolio is one option and some fraction of the asset (the economy), whereas previously the portfolio was one asset and any number of options.

- If global warming is a chimera, global temperatures will revert to normal, the economy will grow, and the knowledge of JR's project is worthless. Say the economy grows to 24 T\$ over the decade. The value of the portfolio is then simply $P_u = \phi 24\,\text{T\$}$.

- If global warming is real, global temperatures continue to climb, sea levels rise, flooding productive land and valuable housing, and the economy suffers badly and falls to a value of 12 T\$. But the knowledge gained by JR's project empowers the government to take remedial action to limit the damage caused by global warming; suppose the knowledge gained by JR's project saves 3 G\$. Consequently, by calling on the knowledge gained by the research project, the government then values the portfolio at $P_d = 3\,\text{G\$} + \phi 12\,\text{T\$}$.

Risk-free portfolio

One generates a risk-free portfolio by choosing the (reciprocal) hedge ratio ϕ so that these values at the end of the decade are identical: $P_u = P_d$, that is, $\phi 24\,\text{T\$} = 3\,\text{G\$} + \phi 12\,\text{T\$}$, which rearranged gives the hedge ratio $\phi = \frac{1}{4} \times 10^{-3}$. At the end of the decade the risk-free portfolio would thus have value 6 G\$, independent of the likelihood of global warming occurring.

Arbitrage gives initial value

Now suppose that over the decade, bonds will increase by $33\frac{1}{3}\%$ (roughly 3% per year). That is, bonds will increase in value by a factor of $R = 4/3$. Since the portfolio is risk free, it must have the same value as investing in bonds. Thus the initial value of the portfolio is its final value divided by R, namely $P = 6\,\text{G\$}/(4/3) = 4.5\,\text{G\$}$. Let the present value of JR's project be V; then the portfolio would cost $V + \phi 16\,\text{T\$} = V + 4\,\text{T\$}$. Consequently, as this cost must equal the initial value of the option $P = 4.5\,\text{G\$}$, the value of JR's research project is $V = 0.5\,\text{G\$} = 500\,\text{M\$}$.

Since JR offers to do the project, generating the knowledge, for 1 M\$, the government should value this project, as an option, as 500 times its cost. Research into global warming is an extremely good value for the money as precautionary insurance, independent of the probability of the likelihood of global warming. ∎

Analogous valuation applies to all such precautionary research projects.

Suggested activity: Do at least Exercise 1.12.

1.4.3   Use a multiperiod binomial lattice for accuracy

Make your analysis more accurate by dividing the relevant time period into more steps and by correspondingly increasing the number of possible asset prices. The previous two examples assumed that the time period over which the call option operates is divided into just one time step. It also assumed that the price of an asset went up or down only in one of two large steps. Section 2.3 applies SDEs by making the time step infinitesimally small and asset prices continuous. As an interim stage, we decide now, for example, to model the asset price movement with quarterly time steps ($\Delta t = h = 1/4$ year). Over a year the
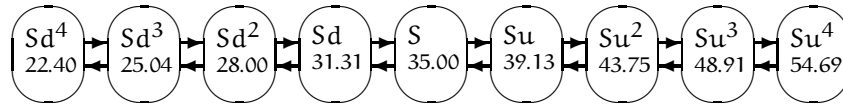
Figure 1.15. *State transition diagram for a simple Markov chain model of asset prices over multiple time steps. The prices either go up or down by, for example, a factor of* $u = 1.1180$ *or* $d = 1/u = 0.8944$.

asset price then has four increments, so we model the domain of asset prices by a nine state Markov chain: we decide on having four prices on either side of the current price S, as shown in Figure 1.15, to allow for the possibility that the four increments of the price are either all up or all down. Although by no means necessary, it is convenient to have the selection of prices be at a constant ratio $u = 1/d$ to their neighboring prices. Thus, for example, an increment in price followed by a decrement in price will result in the price being exactly back at the starting value, rather than somewhere in between.

But why choose $u = 1/d = 1.1180$? Recall that for exponential Brownian motion (which is characteristic of asset prices), in each time step of the numerical approximation (1.5) the prices are multiplied by a factor

$$1 + \beta Z \sqrt{h} \approx e^{\beta Z \sqrt{h}}$$

(upon ignoring the drift term αh) where $Z \sim N(0,1)$. The rough approximation that prices only either rise or fall is equivalent to assuming that $Z = \pm 1$, whence the factors

$$u = e^{\beta \sqrt{h}} \quad \text{and} \quad d = 1/u = e^{-\beta \sqrt{h}}. \tag{1.9}$$

Now here we investigate a quarterly model so that $h = 1/4$, that is $\sqrt{h} = 1/2$. Hence here the ratio

$$u = e^{\beta/2} = \sqrt{e^\beta} = \sqrt{u_{yearly}} = \sqrt{1.25} = 1.1180.$$

The general rule is that the multiplicative factor of asset price increments should be proportional to some constant raised to the power of $\sqrt{h}$. This keeps the volatility comparable in the approximations over differently sized time steps.

A further reason for making all the increments in price use the same factors u and d is that it is then easier to apply the formula (1.8), as the parameters u and d being constant lead to a constant parameter p. However, p also depends upon the bond rate R which also varies with the time step h. Here we want the interest compounded over four quarters to be the same as the yearly interest. Thus $R_{quarterly}^4 = R_{yearly}$, and hence in this example,

$$R_{quarterly} = 1.12^{1/4} = 1.02873.$$

Using this value for R and the earlier values for the price increments, the parameter

$$p = 0.6007 \quad \text{and correspondingly} \quad 1 - p = 0.3993.$$

In general, since bonds growing by compound interest increase in the value like e^{rt} for some rate r, it follows that when sampled at time steps of size h, the appropriate factor is $(e^r)^h$. That is, the multiplicative factor for the rate of increase in the value of bonds should be proportional to a constant to the power of h.
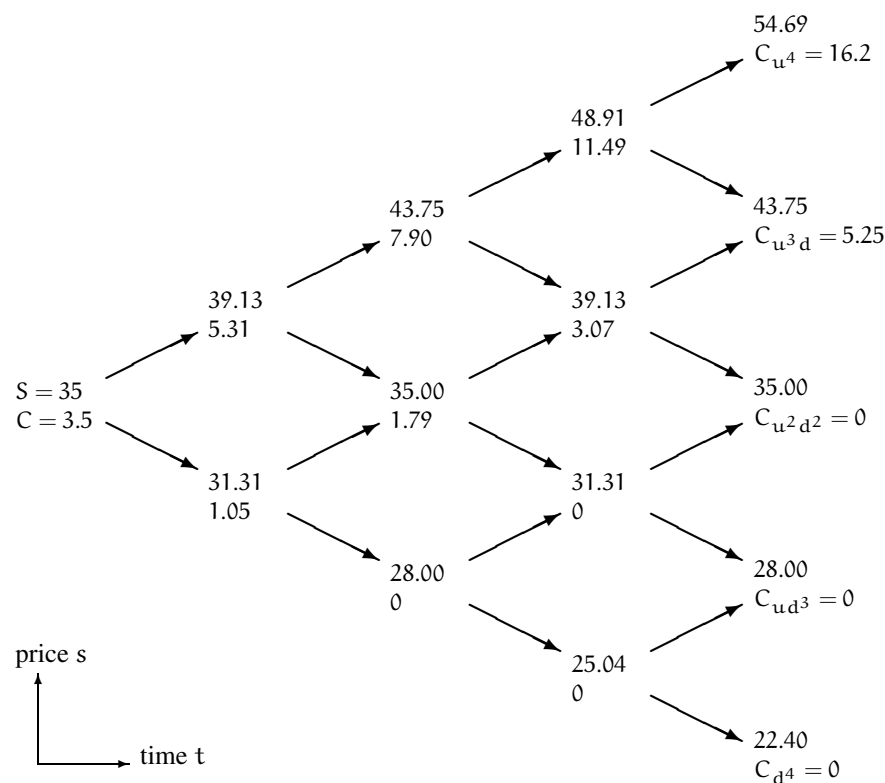
Figure 1.16. *Example of a four-step binary lattice to compute the value of a call option. The asset price is the top number of each pair, and the computed option value is the bottom number.*

Now investigate the four quarterly time steps. Lay out the states vertically with time horizontally, as we did for the single step and as shown in Figure 1.16. See that because at each time step the price is assumed to only either rise or fall, at each time step only half the prices are accessible, and thus only these are drawn at any time. Although included in the figure, as yet we do not know the value of the call option at each of the nodes. However, at the end of the year (the last column in the figure), we do know that the value of a call option is $C = \max\{0, S - X\}$ for the various asset prices S. These have been recorded and labeled so that $C_{u^k d^l}$ denotes the value of the call option after, in any order, k increments and l decrements in the price. Then determine the value of the call option at all other times by proceeding from right to left. First, consider the problem of determining the value of the call option at the start of the fourth quarter following three successive increments in the price of the asset. The state of the system is that the asset is priced at $Su^3 = 48.91$— the upper state in the second column from the right in Figure 1.16. In this quarter of the year the problem looks just like the simple one-step binomial model we analyzed earlier: consider just the top-right part of the lattice as in Figure 1.17. At the start of this final quarter the issues are exactly the same as outlined for the one-step binomial model: Alice has an asset with which she can adjust her risk-free portfolio by choosing the right number of call options valued at a strike price of $X = \$38.50$; because it is risk free, arbitrage asserts

$$54.69$$
$$C_{u^4} = 16.2$$

$$48.91$$
$$C_{u^3} = ?$$

$$43.75$$
$$C_{u^3 d} = 5.25$$
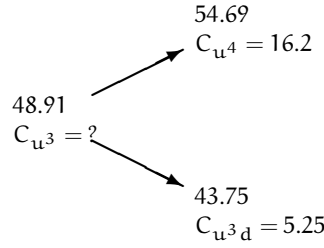
Figure 1.17. *Each step in using a multi-step binary lattice to value a call option looks like a simple one step binomial approximation, such as this extract from the top-right of Figure* 1.16. *Thus the previous arguments and formulae apply.*

it must return the same value as investments in bonds. Thus the formula (1.8) applies with an appropriate change in subscripts, namely,

$$
\begin{aligned}
C_{u^3} &= \frac{1}{R} [pC_{u^4} + (1-p)C_{u^3 d}] \\
&= [0.6007 \times \$16.19 + 0.3993 \times \$5.25]\,/1.0287 \\
&= \$11.49\,.
\end{aligned}
$$

Similar reasoning applies for all other values of the call option at the start of the last quarter. For example,

$$
C_{u^2 d} = \frac{1}{R} [pC_{u^3 d} + (1-p)C_{u^2 d^2}] = \$3.06\,.
$$

Once the values of the call option are determined at the start of the last quarter, the same reasoning and formula give the values at the start of the third quarter—and so on from right to left across the tableau to give the values already appearing in Figure 1.16. Thus a four-step binomial lattice model estimates that the call option with strike price $X = \$38.50$ on the asset is worth $C = \$3.50$ at the start of the year.

Suggested activity: Do at least Exercise 1.15.

Summary

Use expiration values of the call option to determine the values at the third step; use third-step values for those at the second; use second-step values for those at the first; and finally, compute the initial call option value.[16] Algorithm 1.6 shows that such computations are easily done in MATLAB/SCILAB. Each step is based upon the risk-free return of a hypothetical portfolio of options and assets.

---

[16]The repeated application of (1.8) makes it look like a numerical approximation to a partial differential equation (PDE) for the value of the call option as a function of asset price s and time t, namely $C(t, s)$. If over a time step of 1 (year or whatever) $R = e^r$ and $u = e^\beta$, then for many small time steps of size h, $R = e^{rh} \approx 1 + rh$; $u = e^{\beta\sqrt{h}} \approx 1 + \beta\sqrt{h}$; $d = e^{-\beta\sqrt{h}} \approx 1 - \beta\sqrt{h}$; and $p \approx (1 + r\beta\sqrt{h})/2$. Then expanding $C(t, s)$ for nearby values in s and t into Taylor series transforms (1.8) into the PDE $\frac{\partial c}{\partial t} - rc + rs\frac{\partial c}{\partial s} + \frac{1}{2}\beta^2 s^2 \frac{\partial^2 c}{\partial s^2} = 0$. We later derive this directly from an SDE.

Algorithm 1.6 MATLAB/SCILAB code for a four-step binomial lattice estimate of the value of a call option.

```
u=1.25^(1/2), d=1/u
r=1.12^(1/4)
x=38.50
s=35*u.^(-4:2:4)
p=(r-d)/(u-d)
c=max(0,s-x)
for i=1:4
    c=(p*c(2:end)+(1-p)*c(1:end-1))/r
end
```

1.5   Summary

- Many fluctuating, noisy signals, such as financial indices, are modeled using a Weiner process, $W(t)$, which has the following properties: $W(t)$ is continuous; $W(0) = 0$; the change $W(t+s) - W(s) \sim N(0,t)$ for $t, s \geq 0$; and the change $W(t+s) - W(s)$ is independent of any details of the process for times earlier than s.

- An Ito process, $S(t)$, satisfies an SDE in the form

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

  where μ is called the drift and σ the volatility. Financial prices are assumed to have drift $\mu = \alpha S$ and volatility $\sigma = \beta S$.

- Obtain numerical solutions by discretizing the SDE

$$\Delta S_j = \mu(S_j, t_j)\Delta t_j + \sigma(S_j, t_j)\Delta W_j$$

  for (very) small time steps Δt_j.

- Value financial options by first determining a hedge that would ensure a risk-free result, and second requiring arbitrage that the risk-free result is the same as that obtained from secure bonds. A one-step binomial model applies these principles to value an option as

$$C = \frac{1}{R}[pC_u + (1-p)C_d],$$

$$\text{where} \quad p = \frac{R-d}{u-d} \quad \text{and} \quad 1-p = \frac{u-R}{u-d},$$

  where R is the bond factor, u and d the factors of increase and decrease in the price of the underlying asset, and C_u and C_d the corresponding values of the option at the end of the period.

- A more accurate multistep binomial lattice model repeatedly applies the above formula backward in time from expiry date to the current time to determine the current value of an option.

Exercises

1.1. Get your friends and family to play this simple little game that illustrates a key aspect of stochastic mathematics in application to finance.

1. Draw on a big sheet of paper a sequence of 30 squares and label them consecutively 0 (bankrupt), 1, 1, 2, 2, 3, 3, 4, ..., 14, 14, 15 (millionairedom). Each of these squares represents one state in a 30-state Markov chain. Imagine each state represents the value of some asset such as the value of a small business that each player is managing.

2. Give each player a token and a six-sided die.

3. At the start place each token on the second "2", the fifth state. Imagine this corresponds to the small business having an initial value of \$200,000.

4. Each turn in the game corresponds to, say, one year in time. In each year the business may be poor or may grow. Thus in each turn each player rolls his/her die and moves according to the following rules:

   - If a player rolls a 1 or 2, then he/she moves down some states;
   - if a player rolls a 3, he/she stays in the same state;
   - if a player rolls a 4, 5, or 6, he/she moves up some states.

   But the number of states (squares) a player moves is given by the number written in each square. Thus in the first move, because the fifth square/state is a "2," a player moving up moves from the fifth square to the seventh square, and a player moving down moves to the third square.

   That the number written in each square is (roughly) proportional to the position of the square in the sequence corresponds to the financial reality that small businesses usually grow/shrink by small amounts, whereas large companies grow/shrink by large amounts. Investors expect returns in proportion to their investments.

5. Each player continues to role his/her die and move until reaching 0 or 15. That is, players continue to operate their businesses until they either go bankrupt or reach millionairedom.

Questions:

1. Why do you *expect* each business to grow? That is, why do you *expect* each player to reach the "millionairedom" state?

2. When you play the game, roughly what proportion of players reach millionairedom? What proportion go bankrupt?

3. How do you explain the actual results?

1.2. In MATLAB/SCILAB write a script to use `randn` and `cumsum` to generate, say, five realizations of a Wiener process—approximate it by taking $n = 1000$ time steps over $0 < t \leq 1$. Plot the realizations. Introduce a drift μ and volatility σ and plot realizations of the resultant process; for example, investigate $\mu \in \{0, \pm 1, \pm 3\}$ and $\sigma \in \{1/3, 1, 3\}$.

1.3. Estimate (albeit very crudely) the stock drift and stock volatility of the price of silver (US\$/ounce) from the data in Table 1.2.

Table 1.2. *Annual closing prices of silver (US$/ounce).*

Year	Close
2000	$4.595
2001	$4.650
2002	$4.790
2003	$5.960
2004	$6.845
2005	$8.910

Table 1.3. *Annual closing value of the Japanese Nikkei Index (yen).*

Year	Close
1992	16925.0000
1993	17417.1992
1994	19723.0996
1995	19868.1992
1996	19361.3008
1997	15258.7002
1998	13842.1699
1999	18934.3398
2000	13785.6904
2001	10542.6201
2002	8578.9502
2003	10676.6396
2004	11488.7598
2005	16111.4297
2006	17225.8300

1.4. Estimate (albeit crudely) the stock drift and stock volatility of the value of the Japanese Nikkei Index (yen) from the data in Table 1.3. Plot the Index as a function of time and copy and paste the data into MATLAB/SCILAB or an equivalent program, and compute estimates of the stock drift and stock volatility.

1.5. Implement the Euler code of Example 1.6. Plot numerical solutions for $\alpha = 1$ and $\beta = 2$ over $0 \leq t \leq 1$. Now generate simultaneously $m = 300$ realizations of the solution of the SDE (for $\alpha = 1$ and $\beta = 2$) and draw a histogram of their values at time $t = 1$, say, using bins of width 0.25 from 0 to 5; see that almost all solutions have a smallish numerical value (less than one). Nonetheless, compute the mean of the values at time $t = 1$ and see that it is roughly $e^1 = 2.718$!

1.6. Modify the Euler code of Example 1.6 to numerically solve and plot five different realizations of the (Ornstein–Uhlenbeck process) solutions of the SDE: $dS = -S\,dt + dW$ with initial conditions $S_0 = 0, 1, 2, 3,$ and 4. Perhaps integrate over $0 \leq t \leq 4$.

1.7. Modify the Euler code of Example 1.6 to numerically solve and plot five different realizations of the solutions of the following SDEs:

1. $dX = \left[\frac{2X}{1+t} + (1+t)^2\right]dt + (1+t)^2 dW$ with $X_0 = 0$ and $X_0 = 1$;

2. $dX = dt + 2\sqrt{X}\,dW$ with $X_0 = 1$;

3. $dX = -\frac{1}{2}e^{-2X}dt + e^{-X}dW$ with $X_0 = 0$;

4. $dX = \left[\frac{1}{2}X + \sqrt{1+X^2}\right]dt + \sqrt{1+X^2}\,dW$ with $X_0 = 0$;

5. $dZ = Z^3 dt - Z^2 dW$ with $Z(0) = 1$.

1.8. Quantitatively investigate convergence as time step $h \to 0$ at time $t = 1$ for a range of realizations of exponential Brownian motion (use relative differences). Summarize your findings.

1.9. What is the value of a forward contract on an asset when

1. its current value is $60, the agreed price after a year is $57, and the bond rate is 4% per year?

2. its current value is $35, the agreed price after two years is also $35, and the bond rate is 5% per year?

1.10. Reconsider Example 1.10 of the valuation of a forward contract on Telecom shares. What bond rate would cause the current value of the forward contract to be precisely $1?

1.11. Use a one-period binomial model to estimate that the hedge ratio $H = 2$ and the fair value of the call option is $5.57 when the initial asset price is $35, the exercise price of the call option is also $35, the rate of interest for bonds is 10%, and the period is one year. Assume that the asset price moves up or down by 25% per year. See that a portfolio of the asset and either one call option or three call options sold at this price are subject to fluctuations in the price of the asset that could result in the investor losing.

1.12. Use a one-period binomial model to estimate the hedge ratio and the fair value of the call option when the initial asset price is $50, the exercise price of the call option is also $50, the rate of interest for bonds is 10%, and the period is one year. Assume that the asset price moves up by 25% or down by 20% per year.

1.13. What interest rate for bonds would be needed for the call option of Example 1.9 to be correctly valued at $4 ?

1.14. An asset is initially priced at $50. The exercise price of a call option on this asset is also $50. Assume that the asset price moves up by 25% or down by 20% per year. The call option is offered for sale at $8. Use a one-period binomial model to estimate the interest rate for bonds that would make this the correct value for the call option.

1.15. Suppose that volatility corresponds to $+25\%$ and -20% per year, and that the bond interest rate is 12% per year. What would be the corresponding factors u, d, and R for each monthly step of a twelve-step monthly binomial lattice model?

1.16. Modify the MATLAB/SCILAB code of Algorithm 1.6 to use an n-step binomial lattice for estimating the value of the call option (where n is a supplied parameter). Show that it is not until $n \geq 512$ (or thereabouts) that the estimated initial value of the call option converges to $3.40 to the nearest cent.
Use this modified code to refine the value of the call option in the situations described in Exercises 1.11–1.12.

1.17. Alice buys a put option from Bob (see Definition 1.13), strike price $X = \$57$, for some fair price P which you are to determine by arguments similar to those employed for call options. Develop one step of a binomial model to estimate the value of the put option.

> **Definition 1.13.** *A European* put option *gives the buyer (of the option) the right but not the obligation to sell an asset at a previously agreed strike price on a particular date.*

>   1. First, show that at the expiry of the put option, when the asset has price S, the put option has value $P = \max\{0, X - S\}$ by considering the cases $X < S$ and $X > S$.

>   2. Second, at the start of a year Alice constructs a portfolio of one bought put option, value P, and hedges by buying ϕ units of the asset (selling if ϕ is negative), each of price $S = \$60$. Explain the value of the portfolio at the end of the year if the asset goes up in price by 25%, and if the asset goes down in price by 20%. Deduce the risk-free ???hedge ratio $\phi = \frac{1}{3}$.

>   3. Lastly, use the principle of arbitrage to determine the fair price $P = \$4$ for the option; assume bonds increase in value by $4\frac{1}{6}\%$ over the year.

1.18. Alice buys a put option from Bob (see Definition 1.13), strike price $X = \$39$, for some fair price P which you are to determine by arguments similar to those employed for call options. Develop one step of a binomial model estimating the value of the put option.

>   1. First, show that at the expiry of the put option, when the asset has price S, the put option has value $P = \max\{0, X - S\}$ by considering the cases $X < S$ and $X > S$.

>   2. Second, at the start of a year Alice constructs a portfolio of one bought put option, value P, and hedges by buying ϕ units of the asset (selling if ϕ is negative), each of price $S = \$60$. Explain the value of the portfolio at the end of the year if the asset may either double or halve in value. Deduce the risk-free hedge ratio ϕ.

>   3. Lastly, use the principle of arbitrage to show the fair price P for the option; assume inflation is rampant and that bonds increase in value by 20% over the year.

1.19. Alice buys a put option from Bob (see Definition 1.13), strike price X, for some fair price P which you are to determine by arguments similar to those employed for call options. Investigate one step of a binomial model for determining the value of the put option, then apply your result to a multiperiod binomial lattice.
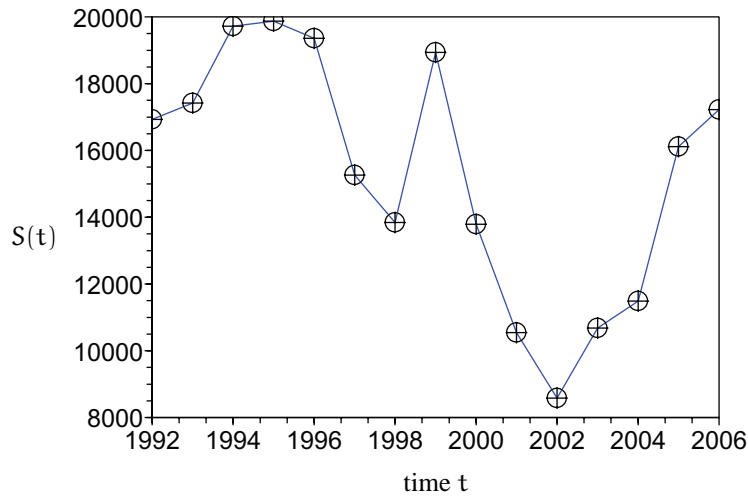
Figure 1.18. *Japanese Nikkei Index (yen) as a function of year.*

1. First, argue that at the expiry of the put option when the asset has price S, the option has value $P = \max\{0, X - S\}$ by considering the cases $X < S$ and $X > S$. See that a put option, unlike a call option, becomes more valuable the lower the asset price.

2. Second, at the start of a time interval Alice constructs a portfolio of one bought put option, value P, and hedges with ϕ units of the asset, each of price S. At the end of the time interval the value of the portfolio will be $P_u + \phi Su$ if the asset goes up in price by a factor of u and $P_d + \phi Sd$ if the asset goes down in price by the factor d. Argue that for a risk-free result the hedge ratio $\phi = (P_d - P_u)/(Su - Sd)$.

3. Lastly, argue that if such a risk-free portfolio is to give the same return as bonds, which increase in value by a factor of R over the interval, then the value of the put option at the start of the interval is the same expression as (1.8), namely

$$P = \frac{1}{R}\left[\frac{R-d}{u-d}P_u + \frac{u-R}{u-d}P_d\right].$$

4. Hence estimate a fair value of a put option when the initial asset price is \$50, the exercise price of the put option is also \$50, the rate of interest for bonds is 10%, and the period is one year. Assume that the asset price moves up by 25% or down by 20% per year.

5. Revise the MATLAB/SCILAB code of Exercise 1.16 to estimate the value of the above put option using an $n = 4$, 32, 128, and 512 step binomial lattice.

1.20. Clearly explain the crucial features of a Wiener process that empower us to model noisy, fluctuating dynamics. Explain how and why a Wiener process is transformed to model general noisy, fluctuating signals.

Answers to selected exercises

1.1.   1. On *average*, a player moves up 16% of the time. That is, we could make a business case that shows 16% growth per year.

     2. Only about 1 in 3 reach millionairedom; the rest go bankrupt.

     3. Stochastic fluctuations ruin the expected growth.

1.3.   Stock drift $\alpha \approx 15\%$ per year; stock volatility $\beta \approx 13\%$ per $\sqrt{\text{year}}$.

1.4.   See Figure 1.18. Stock drift $\alpha \approx 2.2\%$ per year; stock volatility $\beta \approx 21\%$ per $\sqrt{\text{year}}$.

1.9.   1. \$5.19.

     2. \$3.25 .

1.10. $R = 1.1324$, that is, 13.24%.

1.12.   \$7.58.

1.13.   Approximately 22%.

1.14. $R = 1.1236$, that is, an interest rate of 12.36% over the period.

1.16.   \$4.83 and \$6.89, respectively.

1.18.   Hedge ratio $\phi = 1/10$; price $P = \$4$.

1.19.   4. \$3.03.

     5. \$2.07, \$2.31, \$2.34, and \$2.34 .

Chapter 2

Ito's Stochastic Calculus Introduced

Contents

This chapter investigates how to manipulate symbolically SDEs. We discover in the algebraic solutions of the SDEs features which are also discerned in the numerical solutions. The key is the discovery of a stochastic form of the chain rule for differentiation called Ito's formula. The formula is proved in Chapter 4 via a careful definition of stochastic integration. Here we use Ito's formula to derive the Black–Scholes PDE that values financial options.

2.1 Multiplicative noise reduces exponential growth

We first explain the solution of two simple SDEs. The second one is the solution for exponential Brownian motion and introduces the key ingredient we need for the description and development of Ito's formula.

2.1.1   Linear growth with noise

Consider the constant coefficient SDE $dX = \mu\,dt + \sigma\,dW$; in this short section the drift μ and volatility σ are constants. Numerically, we interpret this SDE to mean

$$\Delta X_j = \mu\,\Delta t_j + \sigma\,\Delta W_j,$$

when discretized with time steps $\Delta t_j = t_{j+1} - t_j$, which we usually take to be the constant time step h. Now sum this discretization over n steps starting from time $t_0 = 0$ when $W_0 = 0$:

$$\sum_{j=0}^{n-1}(X_{j+1} - X_j) = \sum_{j=0}^{n-1}\mu(t_{j+1} - t_j) + \sum_{j=0}^{n-1}\sigma(W_{j+1} - W_j).$$

Because the drift μ and volatility σ are constant, there is massive cancellation in the terms of this equation, leading to

$$\begin{aligned}
X_n - X_0 &= \mu(t_n - t_0) + \sigma(W_n - W_0) \\
\Rightarrow X_n &= X_0 + \mu t_n + \sigma W_n \quad \text{as } W_0 = t_0 = 0 \\
\Rightarrow X(t) &= X_0 + \mu t + \sigma W(t)
\end{aligned}$$

in the limit as $\max_j \Delta t_j \to 0$. We have algebraically solved our first SDE!

2.1.2   Exponential Brownian motion

The linear constant coefficient SDE is rather trivial. Proceed now to consider the exponential Brownian motion SDE that, for example, is claimed to govern stock prices:

$$dX = \underbrace{\alpha X}_{\mu}\,dt + \underbrace{\beta X}_{\sigma}\,dW.$$

As we interpreted before, this SDE means that evaluating the right-hand side at the current time forms a numerical approximation:

$$\Delta X_j = \alpha X_j\,\Delta t_j + \beta X_j\,\Delta W_j.$$

Divide by X_j to lead to a form with a right-hand side seen before:

$$\frac{\Delta X_j}{X_j} = \alpha\,\Delta t_j + \beta\,\Delta W_j. \tag{2.1}$$

The right-hand side looks like the constant drift and volatility that we summed so successfully before, but the left-hand side is problematic. However, recall that the derivative of $\log x$ is $1/x$, matching the $1/x_j$ in the left-hand side above, so perhaps differences of $\log X$ will lead to something useful. Indeed, they do, as we see below:

$$\begin{aligned}
\Delta \log X_j &= \log X_{j+1} - \log X_j \quad \text{by definition of the difference } \Delta \\
&= \log(X_j + X_{j+1} - X_j) - \log X_j \\
&= \log(X_j + \Delta X_j) - \log X_j
\end{aligned}$$

$$= \log\left(1 + \frac{\Delta X_j}{X_j}\right)$$

$$= \left(\frac{\Delta X_j}{X_j}\right) - \frac{1}{2}\left(\frac{\Delta X_j}{X_j}\right)^2 + \cdots \quad \text{by Taylor series of } \log(1+x)$$

$$= \left(\alpha \Delta t_j + \beta \Delta W_j\right) - \frac{1}{2}\left(\alpha \Delta t_j + \beta \Delta W_j\right)^2 + \cdots \quad \text{by the SDE (2.1)}$$

$$= \alpha \Delta t_j + \beta \Delta W_j - \frac{1}{2}\alpha^2 \Delta t_j^2 - \alpha\beta \Delta t_j \Delta W_j - \frac{1}{2}\beta^2 \Delta W_j^2 + \cdots.$$

Here and later the ellipses $\cdots$ denotes the higher order terms in the Taylor series. Now sum the right- and left-hand sides (using $t_0 = W_0 = 0$), where for simplicity we assume the time step is constant h; then

$$\log X_n - \log X_0 = \alpha t_n + \beta W_n - \frac{1}{2}\alpha^2 h t_n - \alpha\beta h W_n - \frac{1}{2}\beta^2 t_n + \cdots,$$

where we have magically simplified the form of the quadratic terms—a crucial step which we justify a little later. Now, taking the limit as the time step $h \to 0$ and assuming the higher order terms $\to 0$ in this limit, the above expression becomes

$$\log X(t) - \log X_0 = \alpha t + \beta W(t) - \frac{1}{2}\beta^2 t,$$

which rearranges into the remarkable analytic solution

$$X(t) = X_0 \exp\left[(\alpha - \tfrac{1}{2}\beta^2)t + \beta W(t)\right]. \tag{2.2}$$

See the astonishing feature, which we saw in the numerical solutions of Figure 1.10, that noise, parametrized by β, may act to stabilize what would otherwise exponentially grow by apparently reducing the growth rate from α to $\alpha - \frac{1}{2}\beta^2$.

Example 2.1. The ODE $dX = X\,dt$ has growing solutions $X = X_0 e^t$, but the same ODE with large enough multiplicative noise, say $dX = X\,dt + 2X\,dW$, has solutions (2.2) of $X = X_0 \exp[-t + 2W(t)]$. Since $W(t)$ only grows like $\sqrt{t}$, the dominant term in the argument of the exponential function is the $-t$, which shows that almost surely all solutions of the SDE decay to zero! This almost sure decay is supported by a histogram, Figure 2.1, of the values of $X(1)$ over many realizations. That there are a few realizations of $X(1)$ which are large is significant and will be discussed later. ∎

Now look at the detailed justification of the transformations of the three quadratic terms used above (note the time steps are the constant h).

- First,

$$\sum_{j=0}^{n-1} \Delta t_j^2 = \sum_{j=0}^{n-1} h^2 = nh^2 = ht_n \to 0 \text{ as } h \to 0.$$

- Second,

$$\sum_{j=0}^{n-1} \Delta t_j \Delta W_j = h \sum_{j=0}^{n-1} \Delta W_j = hW_n \sim N(0, h^2 t),$$

and hence almost surely $\to 0$ as $h \to 0$.
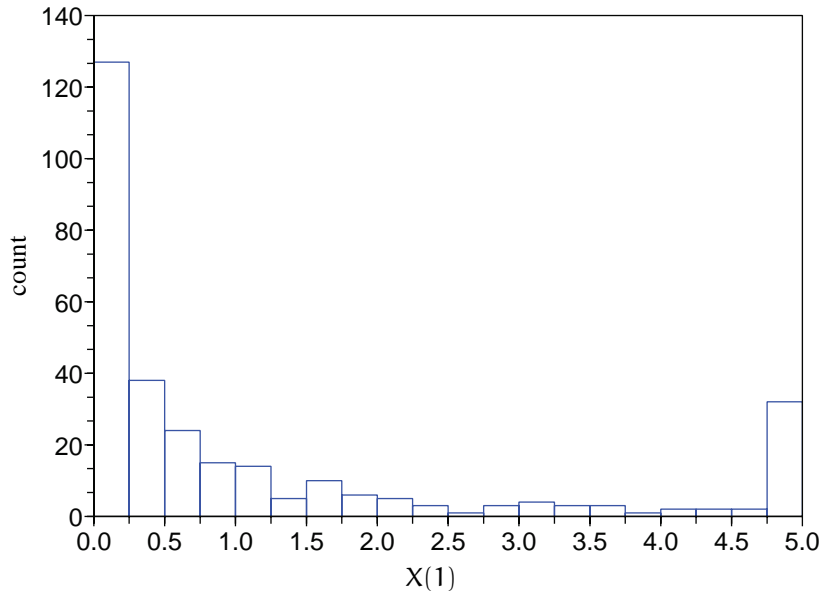
Figure 2.1. *Histogram of* 300 *realizations of* $X(1)$ *from the* SDE $dX = X\,dt + 2X\,dW$ *with* $X(0) = 1$ *(Figure* 1.10 *plots five realizations in time). This histogram shows most realizations decaying to zero, although some large excursions (shown gathered at* $X(1) = 5$*) significantly affect the statistics.*

- Lastly, and much more interestingly,

$$\sum_{j=0}^{n-1} \Delta W_j{}^2 = \sum_{j=0}^{n-1} \left(\sqrt{h}Z_j\right)^2$$

$$= h\sum_{j=0}^{n-1} Z_j^2$$

$$= h\sum_{j=0}^{n-1} 1 + h\sum_{j=0}^{n-1} (Z_j^2 - 1)$$

$$= t_n + Y,$$

where $Y = h\sum_{j=0}^{n-1}(Z_j^2 - 1)$ for some random variables $Z_j \sim N(0,1)$. Now, as derived in Exercise 2.1, $E[Z^p] = (p-1)(p-3)\dots 1$ for even p, and thus $E[Y] = h\sum_{j=0}^{n-1} E[Z_j^2 - 1] = E[Z^2] - 1 = 0$. Hence the value of $\sum_{j=0}^{n-1} \Delta W_j{}^2$ averages to t_n, as asserted earlier.

But the sum may fluctuate wildly due to Y—we now argue that its fluctuations are

insignificant by showing that the variance of Y vanishes as the time step $h \to 0$:

$$\begin{aligned}
\mathrm{Var}[Y] &= h^2 \sum_{j=0}^{n-1} \mathrm{Var}[Z_j^2 - 1] \quad \text{as } Z_j \text{ are independent} \\
&= h^2 n \, \mathrm{Var}[Z^2 - 1] \quad \text{as they are distributed identically} \\
&= h^2 n \, \mathrm{E}[(Z^2 - 1)^2] \quad \text{as } \mathrm{E}[Z^2 - 1] = 0 \\
&= h^2 n \, \mathrm{E}[Z^4 - 2Z^2 + 1] \\
&= h^2 n [3 - 2 + 1] \quad \text{using expectation of even powers of } Z \\
&= 2h^2 n \\
&= 2h t_n \\
&\to 0 \quad \text{as } h \to 0 \text{ for all } t_n.
\end{aligned}$$

Because the variance of Y tends to 0 and $\mathrm{E}[Y] = 0$, then $Y \to 0$ almost surely as $h \to 0$. Consequently $\sum_{j=0}^{n-1} \Delta W_j^2 \to t_n$ almost surely.

Since $\sum_{j=0}^{n-1} \Delta W_j^2$ is almost surely $\sum_{j=0}^{n-1} \Delta t_j$, this is as if "$\Delta W_j^2 = \Delta t_j$"[17] which suggests the novel symbolic rule introduced below.

Summary

For stochastic calculus, the following symbolic rules for infinitesimals effectively apply: $dt^2 = 0$ *and* $dt\, dW = 0$ *(as for ordinary calculus), but that surprisingly* $dW^2 = dt$. *These symbolic rules will be used extensively. They derive, for example, that* $X(t) = X_0 \exp\left[(\alpha - \frac{1}{2}\beta^2)t + \beta W(t)\right]$ *solves the financial SDE* $dX = \alpha X\, dt + \beta X\, dW$.

2.2   Ito's formula solves some SDEs

Recall that when you were first introduced to integration in your earlier courses it was presented as antidifferentiation. That is, you used differentiation rules to infer integration formulae. The same is true for the solution of some SDEs that we develop in this section. The basic rule of stochastic differentiation is Ito's formula with the identity $dW^2 = dt$ at its heart. We first present a simple version of Ito's formula and then present the full version.[18]

2.2.1   Simple Ito's formula

Let $f(t, w)$ be a smooth function of two arguments—smooth means that it is differentiable at least several times and that Taylor's theorem applies. Figures 2.2 and 2.3 show two example surfaces of two such smooth functions. Then consider the Ito process

---

[17]This last point is perhaps not too surprising because we know $\mathrm{E}[\Delta W_j^2] = \Delta t_j \, \mathrm{E}\left[\left(\Delta W_j / \sqrt{\Delta t_j}\right)^2\right] = \Delta t_j$. But importantly we have is that we have shown that the fluctuations about this expectation are negligible in the limit of continuous time.

[18]Chapters 5 and 6 of the book by Stampfli and Goodman (2001) provide appropriate reading to supplement this section.
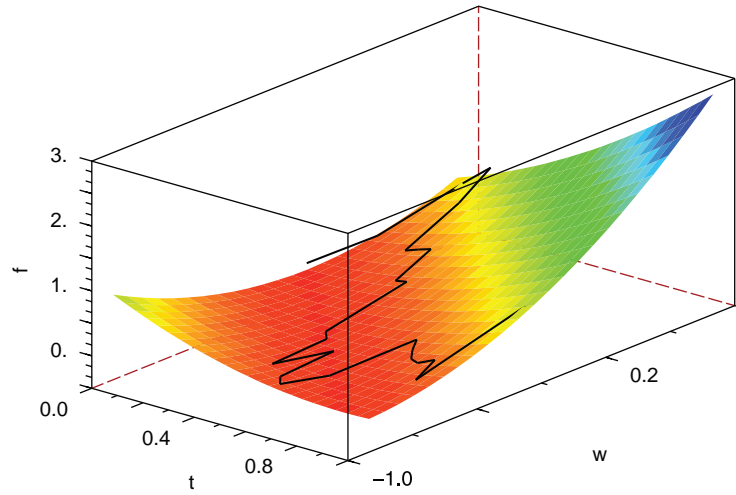
Figure 2.2. *The smooth surface* $f = (t+w)^2$ *with one realization of a Wiener process* $w = W(t)$ *evolving on the surface: the Ito process* $X = f(t, W(t))$ *is the height of the curve as a function of time and changes due to direct evolution in time and through evolution of* w.
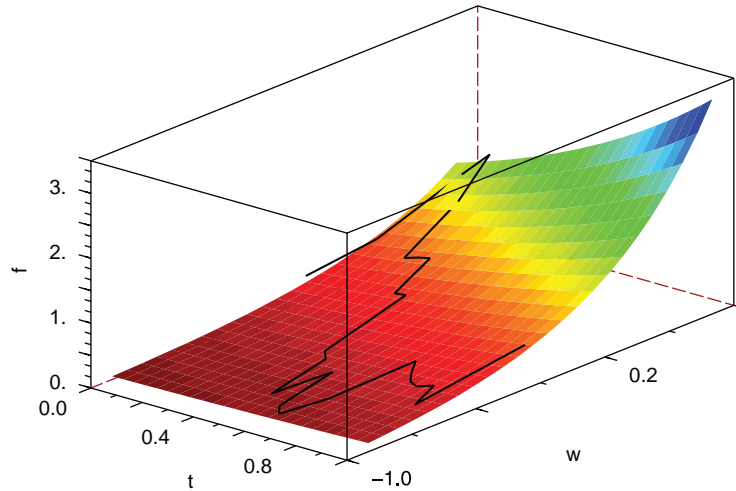
Figure 2.3. *The smooth surface* $f = 0.3 \exp(t + 2w)$ *with one realization of a Wiener process* $w = W(t)$ *evolving on the surface: The Ito process* $X = f(t, W(t))$ *is the height of the curve as a function of time and changes due to direct evolution in time and through evolution of* w.

$X(t) = f(t, W(t))$, where $W(t)$ is a Wiener process—see the black curves wiggling across the surfaces in Figures 2.2 and 2.3. We explore these examples later: $f(t, w) = (t + w)^2$, whence $X = (t + W(t))^2$, and $f(t, w) = \exp(at + bw)$, whence $X = \exp(at + bW(t))$. Note that f itself is smooth; the stochastic part of $X(t)$ comes only via the use of the

Wiener process in the evaluation of f. Now consider the change in X that occurs over some small change in time, Δt, through the direct dependence upon t and indirectly through the dependence upon the Wiener process W. Denoting $X(t+\Delta t) = X + \Delta X$ and $W(t+\Delta t) = W + \Delta W$, observe (using a multivariable Taylor series of f) that

$$
\begin{aligned}
X + \Delta X &= f(t + \Delta t, W + \Delta W) \\
&= f(t, W) + \frac{\partial f}{\partial t}\Delta t + \frac{\partial f}{\partial w}\Delta W \\
&\quad + \frac{1}{2}\frac{\partial^2 f}{\partial t^2}\Delta t^2 + \frac{\partial^2 f}{\partial t \partial w}\Delta t\,\Delta W + \frac{1}{2}\frac{\partial^2 f}{\partial w^2}\Delta W^2 + \cdots \\
&\approx X + \frac{\partial f}{\partial t}\Delta t + \frac{\partial f}{\partial w}\Delta W + \frac{1}{2}\frac{\partial^2 f}{\partial w^2}\Delta t,
\end{aligned}
$$

as by the earlier rules for differentials $\Delta t^2 = \Delta t \Delta W = \cdots = 0$ and $\Delta W^2 = \Delta t$. In the limit as $\Delta t \to 0$ the differences become differentials, and thus

$$
dX = \left(\frac{\partial f}{\partial t} + \frac{1}{2}\frac{\partial^2 f}{\partial w^2}\right)dt + \frac{\partial f}{\partial w}\,dW, \tag{2.3}
$$

where the partial derivatives are evaluated at (t, W). This is the simple version of *Ito's formula* and gives the differential of an Ito process X which depends directly and smoothly upon a Wiener process.

Example 2.2. Determine the differential of $X(t) = (t + W(t))^2$, and hence deduce an SDE which $X(t)$ satisfies.

Solution: Here $f(t, w) = (t + w)^2$ (see Figure 2.2), so $f_t = f_w = 2(t + w)$ and $f_{ww} = 2$; thus

$$
dX = [2(t + W) + 1]dt + 2(t + W)dW.
$$

Recognizing $t + W = \sqrt{X}$, rewrite this as the SDE[19]

$$
dX = [2\sqrt{X} + 1]dt + 2\sqrt{X}\,dW. \qquad \blacksquare
$$

Example 2.3. Determine the differential of $X(t) = c\exp[at + bW(t)]$, and hence solve the SDE $dX = \alpha X\,dt + \beta X\,dW$.

Solution: Here $f(t, w) = ce^{at+bw}$ (see Figure 2.3), whence $f_t = ace^{at+bw}$, $f_w = bce^{at+bw}$ and $f_{ww} = b^2 ce^{at+bw}$. Thus Ito's formula asserts

$$
\begin{aligned}
dX &= \left(ace^{at+bW} + \tfrac{1}{2}b^2 ce^{at+bW}\right)dt + bce^{at+bW}dW \\
&= ce^{at+bW}\left[\left(a + \tfrac{1}{2}b^2\right)dt + b\,dW\right].
\end{aligned}
$$

Rewritten as $dX = (a + \tfrac{1}{2}b^2)X\,dt + bX\,dW$ this is the same as the given SDE provided $\alpha = a + \tfrac{1}{2}b^2$ and $\beta = b$. Thus it is the solution of the SDE provided $b = \beta$ and

---

[19]Both of these expressions of the stochastic dynamics are correct. Prefer the second for many purposes as the right-hand side has no occurrences of the Wiener process $W(t)$ except for the differential dW. In applications where SDEs arise, we normally formulate an SDE model in such a form where the *only* direct appearance of "noise" is in the differential dW. Hence prefer this second SDE.

$a = \alpha - \frac{1}{2}\beta^2$, and hence the solution is

$$X(t) = c\exp[(\alpha - \tfrac{1}{2}\beta^2)t + \beta W(t)],$$

as discussed in Section 2.1.2.   ■

Example 2.4. Derive the drift and volatility of the stochastic process $Y = W(t)^3$; express in terms of Y only.

    Solution: Here $Y = f(t, W)$, where $f(t, w) = w^3$ so that $f_t = 0$, $f_w = 3w^2$ and $f_{ww} = 6w$. Thus Ito's formula asserts

$$\begin{aligned}
dY &= 0\,dt + 3W^2\,dW + \tfrac{1}{2}6W\,dW^2 \\
&= 3W\,dt + 3W^2\,dW \\
&= 3Y^{1/3}dt + 3Y^{2/3}dW.
\end{aligned}$$

Hence the stochastic process Y has drift $\mu = 3Y^{1/3}$ and volatility $\sigma = 3Y^{2/3}$.   ■

2.2.2   Ito's formula

The simple form (2.3) of Ito's formula rests upon the Taylor series of $f(t, w)$ and the rule that the only quadratic differential to retain is $dW^2 = dt$. The general form of Ito's formula, sometimes referred to modestly as Ito's lemma, is the same and addresses the differential of a function of a stochastic process in general rather than just a function of a Wiener process.

Theorem 2.5 (Ito's formula). *Let $f(t, x)$ be a smooth function of its arguments and X be an Ito process with drift $\mu(t, X)$ and volatility $\sigma(t, X)$, that is, $dX = \mu\,dt + \sigma\,dW$; then $Y(t) = f(t, X(t))$ is also a Ito process with differential*

$$dY = \frac{\partial f}{\partial t}dt + \frac{\partial f}{\partial x}dX + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}dX^2 \tag{2.4}$$

where in dX^2 retain only $dW^2 \mapsto dt$.

and the partial derivatives are evaluated at (t, X). Equivalently, expanding expression (2.4) *gives*

$$dY = \left(\frac{\partial f}{\partial t} + \mu\frac{\partial f}{\partial x} + \tfrac{1}{2}\sigma^2\frac{\partial^2 f}{\partial x^2}\right)dt + \sigma\frac{\partial f}{\partial x}dW. \tag{2.5}$$

    Although the formula (2.5) is more explicit, we believe (2.4) is a more memorable form of Ito's formula. The rigorous proof of this formula requires a well-defined stochastic integral and so is deferred to Chapter 4.

Example 2.6. Example 2.2 shows that if $X(t) = (t + W(t))^2$, then $dX = [2\sqrt{X} + 1]dt + 2\sqrt{X}\,dW$; hence deduce dY for $Y = e^X$.

    Solution: Use Ito's formula with $f(t, x) = e^x$ for which $f_t = 0$ and $f_x = f_{xx} = e^x$:

$$dY = f_x dX + \tfrac{1}{2} f_{xx} dX^2$$
$$= e^X \left\{ \left[2\sqrt{X} + 1 \right] dt + 2\sqrt{X} \, dW \right\}$$
$$+ \tfrac{1}{2} e^X \left\{ \left[2\sqrt{X} + 1 \right] dt + 2\sqrt{X} \, dW \right\}^2$$
$$= e^X \left\{ \left[2\sqrt{X} + 1 \right] dt + 2\sqrt{X} \, dW + \tfrac{1}{2} 4 (\sqrt{X})^2 dW^2 \right\}$$
$$= e^X \left\{ \left[1 + 2\sqrt{X} + 2X \right] dt + 2\sqrt{X} \, dW \right\}. \qquad\blacksquare$$

Example 2.7 (product rule). Recall that in ordinary calculus the product rule for differentiation is $d(fg) = f\,dg + g\,df$. In stochastic calculus there is an extra term. Let $X(t)$ and $Y(t)$ be stochastic processes with differentials

$$dX = \mu\,dt + \sigma\,dW \quad \text{and} \quad dY = \nu\,dt + \rho\,dW.$$

Then one may argue, using our symbolic rules, that

$$d(XY) = (X + dX)(Y + dY) - XY$$
$$= X\,dY + Y\,dX + dX\,dY$$
$$= X\,dY + Y\,dX + \sigma\rho\,dt$$

by retaining only the $dW^2 = dt$ term in the product $dX\,dY$. This is indeed correct, but now deduce it using Ito's formula.

    Solution: The difficulty is that the product XY is a function of two Ito processes, whereas Ito's formula applies only to a function of one Ito process. However, write $XY = \tfrac{1}{2}[Z^2 - X^2 - Y^2]$, where $Z(t) = X + Y$ is a new stochastic process that, because it is simply the sum of X and Y, has differential $dZ = (\mu + \nu)dt + (\sigma + \rho)dW$. Then use Ito's formula separately on each of the components in the right-hand side of $d(XY) = \tfrac{1}{2}[d(Z^2) - d(X^2) - d(Y^2)]$:

$$d(X^2) = 0\,dt + 2X\,dX + \tfrac{1}{2} 2\,dX^2$$
$$= (2X\mu + \sigma^2)dt + 2X\sigma\,dW;$$
$$\text{similarly} \quad d(Y^2) = (2Y\nu + \rho^2)dt + 2Y\rho\,dW;$$
$$\text{and} \quad d(Z^2) = \left[2Z(\mu + \nu) + (\sigma + \rho)^2 \right] dt + 2Z(\sigma + \rho)dW.$$

Substitute these, with $Z = X + Y$, into

$$d(XY) = \tfrac{1}{2} \left[d(Z^2) - d(X^2) - d(Y^2) \right]$$
$$= \tfrac{1}{2} \big[(2X\mu + 2Y\mu + 2X\nu + 2Y\nu + \sigma^2 + 2\sigma\rho + \rho^2)dt$$
$$+ (2X\sigma + 2Y\sigma + 2X\rho + 2Y\rho)dW$$
$$- (2X\mu + \sigma^2)dt - 2X\sigma\,dW - (2Y\nu + \rho^2)dt$$
$$- 2Y\rho\,dW]$$
$$= X(\nu\,dt + \rho\,dW) + Y(\mu\,dt + \sigma\,dW) + \sigma\rho\,dt$$
$$= X\,dY + Y\,dX + \sigma\rho\,dt$$

as required. $\qquad\blacksquare$
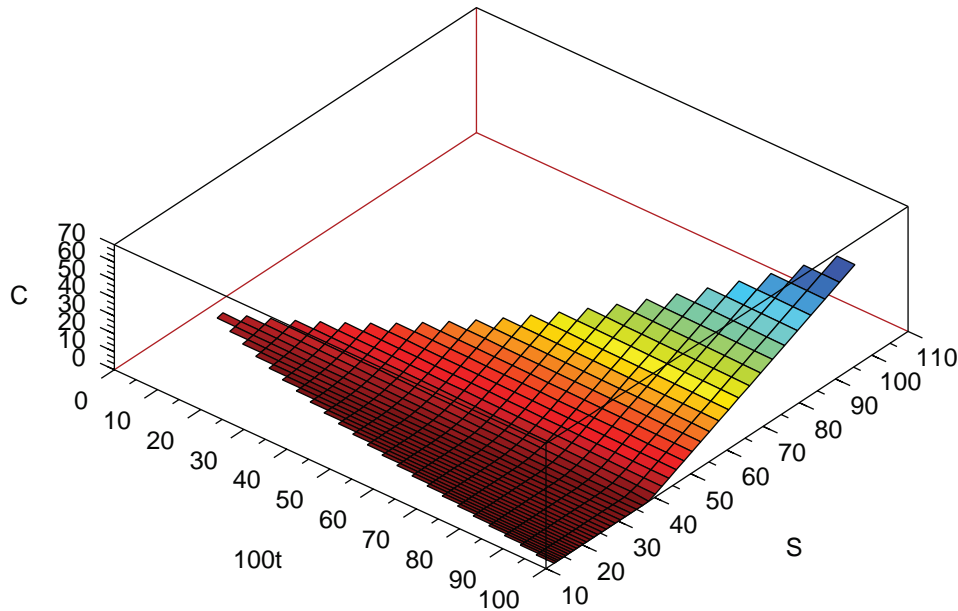
Figure 2.4. *Value* $C(t, S)$ *of the call option for Example* 1.9 *demonstrating that the value* C *is a smooth function of time* t *and asset price* S.

Suggested activity: Do at least Exercise 2.10.

Summary

Ito's formula is a form of chain rule for stochastic processes to empower all stochastic calculus. It is the same as the deterministic chain rule except for the crucial addition of a quadratic differential and the recognition that "$dW^2 = dt$": if $Y = f(t, X)$, then the differential $dY = \frac{\partial f}{\partial t} dt + \frac{\partial f}{\partial x} dX + \frac{1}{2} \frac{\partial^2 f}{\partial x^2} dX^2$.

2.3   The Black–Scholes equation prices options accurately

One application of Ito's formula is to pricing options based on an asset. This section develops the analysis that Section 1.4 began using binomial lattices to price call options. We adopt the same line of argument. Recall that the argument is first to form a portfolio of the asset, together with call options, which is risk free, then to sensibly require this risk-free portfolio to give the same return as investing in risk-free bonds.

Recall that the value of a call option varies smoothly in both time t and the asset price S, as seen in Figure 2.4. Hence, the derivatives $\frac{\partial C}{\partial t}$, $\frac{\partial C}{\partial S}$, and $\frac{\partial^2 C}{\partial S^2}$ are all well defined and smoothly varying. Consequently, *in any realization*, the value of a call option is an Ito process, say $C(t) = C(t, S(t))$, through the dependence upon the Ito process $S(t)$, the realization of the asset value. Recall that the stochastic model of an asset's value is that its differential $dS = \alpha S\, dt + \beta S\, dW$. Hence Ito's formula tells us how the option value C fluctuates in time through the fluctuating changes in asset value.

Now find a hypothetical, risk-free portfolio that has the same return as bonds. Construct a portfolio of one call option sold and some number ϕ (phi) of units of the asset. In Section 1.4 we held one asset and sought the appropriate number of call options—here it is preferable to do the complement, with ϕ as the reciprocal of the earlier H. This portfolio has a value $\Pi = -C(t, S) + \phi S$, which is itself also an Ito process as it is a function of the stochastic asset value S. Thus over a small time interval dt the value of the portfolio changes by an amount deduced via Ito's formula (2.4):[20]

$$d\Pi = -dC + \phi\, dS \qquad \text{as } \Pi = -C + \phi S$$

$$= -\frac{\partial C}{\partial t} dt - \frac{\partial C}{\partial S} dS - \frac{1}{2}\frac{\partial^2 C}{\partial S^2} dS^2 + \phi\, dS \qquad \text{as } C = C(t, S) \text{ by Ito}$$

$$= -\left[\frac{\partial C}{\partial t} + \alpha S\frac{\partial C}{\partial S} + \frac{1}{2}\beta^2 S^2\frac{\partial^2 C}{\partial S^2}\right] dt$$

$$\quad - \beta S\frac{\partial C}{\partial S} dW + \phi\alpha S\, dt + \phi\beta S\, dW \qquad \text{as } dS = \alpha S\, dt + \beta S\, dW$$

$$= \left[-\frac{\partial C}{\partial t} - \alpha S\frac{\partial C}{\partial S} - \frac{1}{2}\beta^2 S^2\frac{\partial^2 C}{\partial S^2} + \alpha S\phi\right] dt$$

$$\quad + \beta S\left[-\frac{\partial C}{\partial S} + \phi\right] dW.$$

A portfolio is risk free when it has no stochastic fluctuations, that is, when it has zero volatility. Make the volatility, $\beta S[-\frac{\partial C}{\partial S} + \phi]$, of this portfolio zero by choosing a portfolio with $\phi = \frac{\partial C}{\partial S}$ units of the asset per sold call option.[21]

Setting $\phi = \frac{\partial C}{\partial S}$, the portfolio changes in price according to the residual drift term in $d\Pi$, namely,

$$d\Pi = \left[-\frac{\partial C}{\partial t} - \frac{1}{2}\beta^2 S^2\frac{\partial^2 C}{\partial S^2}\right] dt.$$

Because this portfolio is risk-free, its return must equal the return of investing the same in bonds. Given the value of the portfolio is $-C + \frac{\partial C}{\partial S}S$ and the interest rate r (so that $R = e^{rt}$), the corresponding investment in bonds returns

$$r\left[-C + \frac{\partial C}{\partial S}S\right] dt = d\Pi = \left[-\frac{\partial C}{\partial t} - \frac{1}{2}\beta^2 S^2\frac{\partial^2 C}{\partial S^2}\right] dt.$$

Equating the coefficients of dt and rearranging leads to the *Black–Scholes equation* for the value $C(t, S)$ of the call option,

$$\frac{\partial C}{\partial t} + rS\frac{\partial C}{\partial S} + \frac{1}{2}\beta^2 S^2\frac{\partial^2 C}{\partial S^2} = rC. \tag{2.6}$$

Involving derivatives in both time t and asset value S, this is a PDE for the option value C. The next section discusses numerical methods of solving the Black–Scholes PDE (2.6); such numerical methods relate closely to the binomial lattice model for valuing options.

---

[20]The first line of this derivation is justified properly by Section 2.3.2.

[21]$\phi = \frac{\partial C}{\partial S}$ is analogous to the earlier hedge formula as $1/H = (C_u - C_d)/(Su - Sd) \approx \Delta C/\Delta S$.

Example 2.8 (forward contract). The Black–Scholes PDE (2.6) straightforwardly values forward contracts to be exactly as in (1.6). To apply the valuation to forward contracts, solve the Black–Scholes PDE (2.6) with expiry value $C(T,S)$ chosen appropriately for a forward contract.[22]

Solution: Recall from Section 1.4.1 that a forward contract is a binding agreement by Alice to sell to Bob an asset at an agreed price X at some expiry time T. At expiry Bob values the forward contract at $C(T,S) = S - X$ because if the asset value $S(T) > X$, then Bob buys the asset more cheaply than the open market price, whereas if the asset value $S(T) < X$, then the contract commits Bob to buying the asset at a more expensive price than the open market price. This expiry value of $C(T,S) = S - X$ is linear in the asset value S. Being linear, it leads to a simple solution of the Black–Scholes PDE (2.6).

Seek solutions of the Black–Scholes PDE (2.6) that are linear in asset value S, but which have a general variation in time. That is, seek solutions in the form $C(t,S) = a(t)S + b(t)$. Later we will use the information that we know at expiry that $a(T) = 1$ and $b(T) = -X$ in order for the expiry value to be $C(T,S) = S - X$. Substitute $C(t,S) = a(t)S + b(t)$ into the Black–Scholes PDE (2.6) to deduce

$$\frac{da}{dt}S + \frac{db}{dt} + rSa + 0 = raS + rb.$$

The marvelous simplification here is that the second derivative term vanishes, which enables this case to be straightforward. Now this equation has to hold for all asset values S; thus equate the coefficients of the terms in S and the terms constant in S to determine

$$\frac{da}{dt} = 0 \quad \text{and} \quad \frac{db}{dt} = rb.$$

- Since $da/dt = 0$, $a(t)$ must be constant, and since at expiry $a(T) = 1$, then $a(t) = 1$ for all time.

- Since $db/dt = rb$, then $b(t) = \text{constant} \times e^{rt}$. The expiry condition that $b(T) = -X$ determines the constant to be $-Xe^{-rT}$ so that $b(t) = -Xe^{r(t-T)}$. Write this as $b(t) = -X/e^{r(T-t)}$, as $T - t$ is the time remaining to expiry.

That is, the value of the forward contract for all time up to the expiry time T and for all asset values S is $C(t,S) = S - X/e^{r(T-t)}$. Consequently, at the start of the time period, $t = 0$, Alice values the forward contract as $C(0,S_0) = S_0 - X/e^{rT}$. Recalling $R = e^{rT}$, this valuation is identical to the earlier (1.6). ∎

More general algebraic solutions may be found for the Black–Scholes equation (2.6) when applied to some options. However, we do not explore this here, as it is more valuable to *qualitatively* check numerical solutions of complicated practical options.

Interpret terms to graphically solve

Before attempting to quantitatively solve such a PDE, we must learn to interpret the effects of the various terms that appear—the interpretations are analogues of those discussed in

---

[22]In this example, we use C as the symbol for the value of the forward contract in order to be consistent with the symbology of the Black–Scholes equation (2.6). In Section 1.4.1 we used the symbol F for the value of the forward contract. The change in variable name is insignificant.

classical one-dimensional continuum mechanics; see, e.g., Roberts (1994). The following interpretations of its mathematical symbols empower us to *qualitatively solve* the Black–Scholes equation (2.6).

- First, the term $rS\frac{\partial C}{\partial S}$ is an "advection" term, as it carries information in S-space. For example, in the absence of the other terms, the equation $\frac{\partial C}{\partial t} + rS\frac{\partial C}{\partial S} = 0$ asserts that on characteristics $dS/dt = rS$, that is, on curves $S = \text{constant} \times e^{rt}$,

$$\frac{dC}{dt} = = \frac{\partial C}{\partial t} + \frac{\partial C}{\partial S}\frac{dS}{dt} \qquad \text{by the chain rule}$$
$$= \frac{\partial C}{\partial t} + rS\frac{\partial C}{\partial S} \qquad \text{as } dS/dt = rS \text{ on each curve}$$
$$= 0 \qquad \text{by the equation.}$$

  Thus on the characteristics $S \propto e^{rt}$ we would find that C is constant—in effect the value C is carried along these characteristic curves. These characteristics are those corresponding to the value of an investment in bonds.

- Second, the term rC on the right-hand side of the Black–Scholes equation (2.6) acts as a *source* of value for the option in time. Including the source term rC, three of the four terms in the Black–Scholes equation (2.6) form $\frac{\partial C}{\partial t} + rS\frac{\partial C}{\partial S} = rC$. Along each characteristic curve $S \propto e^{rt}$ this PDE asserts that the value of the option grows like $\frac{dC}{dt} = rC$, that is, C also grows exponentially along with any bonds, as $C \propto e^{rt}$.

- Lastly, the term $\frac{1}{2}\beta^2 S^2\frac{\partial^2 C}{\partial S^2}$ implies that in addition to these effects, the price of an option "diffuses" in the S-space. This diffusion $\frac{\partial^2 C}{\partial S^2}$ depends directly upon the volatility, βS, of the underlying asset so that large fluctuations in a risky asset cause large diffusion—large "blurring"—of the value of an option. However, and *most importantly*, remember that this term represents *negative* diffusion, on its own $\frac{\partial C}{\partial t} = -\frac{1}{2}\beta^2 S^2\frac{\partial^2 C}{\partial S^2}$, and so the Black–Scholes equation (2.6) *must be solved backward in time*.

Summary

We solve the Black–Scholes equation (2.6) with "initial" conditions specified at the future expiry date of the option, of $C(T,S) = \max\{0, S - X\}$ for a call option; we expect the value of the option to be steadily discounted at the bond rate as we work backward in time (in proportion to e^{rt}); and the value to smooth out by the fluctuation induced "diffusion."

Example 2.9 (knock out). Exotic options often just change the boundary conditions for the Black–Scholes equation (2.6). For example, a "knock out" is a call or put option that additionally expires if the underlying security value S reaches a predetermined "barrier price." There are two varieties of knock outs: "down and outs" expire if the underlying security falls to the barrier price, whereas "up and outs" expire if the underlying security rises to the barrier price.

In a "down and out," the holder of the knock out option receives a rebate c if ever the asset value S falls below some barrier, S_{min} say. For asset values $S > S_{min}$, the arguments
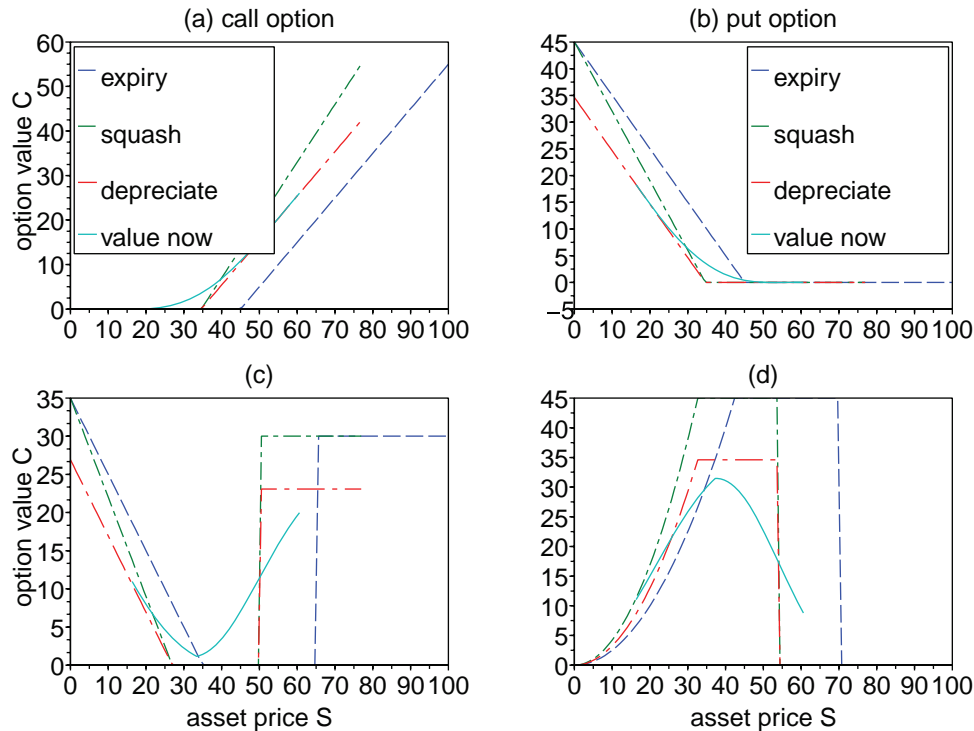
Figure 2.5. *Qualitative solution of the Black–Scholes equation* (2.6) *to predict the value of an option. (a) a call option; (b) a put option; (c), (d) two other special concocted options. The dashed line is the agreed value of the option at expiry depending upon the asset price* S, *and the solid line is a rough estimate of the value of the option at some earlier time.*

behind the Black–Scholes equation (2.6) still apply. Thus one values a knock out option by solving the Black–Scholes equation for asset values $S > S_{min}$. At the asset value $S = S_{min}$ we additionally know that the knock out option has value c, its rebate. Thus when solving the Black–Scholes equation, we keep supplying this value c at the barrier $S = S_{min}$ as a boundary condition to the values in the domain $S > S_{min}$. ∎

Given the enormous variety of possible options, such as the knock out, it is remarkably useful to solve the Black–Scholes equation (2.6) qualitatively. In Figure 2.5 we follow the three qualitative steps on page 51 for any given final valuation of an option plotted as a function of the asset price S (dashed line):

- First, "squash" the option value to the left by a factor R (dot-dashed line) corresponding to the "advection" in the Black–Scholes equation;

- second, deflate/depreciate the valuation C by a factor R, that is, squash vertically (dotted line);

- third, smooth the corners and curves in the line to account for the "blurring" of the option's value by the stochastic nature of the asset price (solid line) to give the option value at some earlier time—the longer the time period or the larger the volatility, the more you smooth the curve.

Qualitative solutions such as these valuably check numerical or algebraic solutions of the Black–Scholes equation (2.6).

You may want to combine and confirm the first two of the above steps by the analysis of Exercise 2.11.

2.3.1 Discretizations form a trinomial model

The Black–Scholes equation (2.6) is intimately connected with the binomial lattice approximation. To connect the two views, first observe that all derivatives with respect to S in the Black–Scholes equation (2.6) are multiplied by the corresponding power of S; thus these are in the form of an Euler–Cauchy differential equation.[23] Simplify such terms by transforming S into $x = \log S$, as then the derivatives simplify: $S\frac{\partial}{\partial S} = \frac{\partial}{\partial x}$ and $S^2\frac{\partial^2}{\partial S^2} = \frac{\partial^2}{\partial x^2} - \frac{\partial}{\partial x}$. The chain rule derives first $\frac{\partial C}{\partial S} = \frac{\partial x}{\partial S}\frac{\partial C}{\partial x} = \frac{1}{S}\frac{\partial C}{\partial x}$, and then $\frac{\partial^2 C}{\partial x^2} = \left(S\frac{\partial}{\partial S}\right)^2 C = S\frac{\partial}{\partial S}\left(S\frac{\partial C}{\partial S}\right) = S\frac{\partial C}{\partial S} + S^2\frac{\partial^2 C}{\partial S^2}$. Such a transformation from S to $x = \log S$ also ensures that the geometric sequence of asset prices used in the binomial lattice, a factor of $u = 1/d$ from each other, becomes a straightforward arithmetic sequence in x, of constant spacing $\Delta x = \log u$. The Black–Scholes equation (2.6) then becomes

$$\frac{\partial C}{\partial t} + r\frac{\partial C}{\partial x} + \frac{1}{2}\beta^2\left[\frac{\partial^2 C}{\partial x^2} - \frac{\partial C}{\partial x}\right] = rC. \tag{2.7}$$

We numerically solve the tranformed Black–Scholes PDE (2.7) on a grid in the tS-plane. Create a grid, such as that seen in Figure 2.4 and implicit in the multiperiod lattice of Figure 1.16, of spacing $\Delta t = h$ in time and spacing $\Delta x = \delta$ in asset value.[24] Let $C_{i,j}$ denote the value of C at the jth time $t_j = jh$ and ith location $x_i = i\delta$ (say). Then a difference approximation to the time derivative $\frac{\partial C}{\partial t} \approx \frac{C_{i,j} - C_{i,j-1}}{h}$, whereas centered approximations to the x derivatives are $\frac{\partial C}{\partial x} \approx \frac{C_{i+1,j} - C_{i-1,j}}{2\delta}$ and $\frac{\partial^2 C}{\partial x^2} \approx \frac{C_{i+1,j} - 2C_{i,j} + C_{i-1,j}}{\delta^2}$. Thus a finite difference approximation[25] to the transformed Black–Scholes PDE (2.7) which is backward in time and centered in log-price is thus

$$\frac{C_{i,j} - C_{i,j-1}}{h} + r\frac{C_{i+1,j} - C_{i-1,j}}{2\delta}$$
$$+ \frac{1}{2}\beta^2\frac{C_{i+1,j} - 2C_{i,j} + C_{i-1,j}}{\delta^2} - \frac{1}{2}\beta^2\frac{C_{i+1,j} - C_{i-1,j}}{2\delta} = rC_{i,j}.$$

---

[23]Read about Euler–Cauchy ODEs in many standard texts (see, e.g., Kreyszig 1999, §2.6).

[24]Of course one may also discretize the Black–Scholes equation (2.6) with equal ΔS rather than equal $\Delta \log S$, but then the former seems less natural and also confuses the comparison with the earlier binomial model.

[25]Issues associated with the numerical solution of such a PDE are dealt with in detail in many texts (see, e.g., Kreyszig 1999).

All option values referred to in this equation are at the jth time t_j, except for the time derivative where $C_{i,j-1}$ appears. Rearranging the equation for the value $C_{i,j-1}$ of the call option at the earlier time t_{j-1} gives

$$C_{i,j-1} = \frac{\beta^2 h}{\delta^2} \left(\frac{1}{2} + \frac{\delta(2r - \beta^2)}{4\beta^2} \right) C_{i+1,j}$$

$$+ \left(1 - rh - \frac{\beta^2 h}{\delta^2} \right) C_{i,j}$$

$$+ \frac{\beta^2 h}{\delta^2} \left(\frac{1}{2} - \frac{\delta(2r - \beta^2)}{4\beta^2} \right) C_{i-1,j}. \tag{2.8}$$

This looks like a "trinomial" lattice approximation to the pricing of an option. Indeed, if we choose the increments in the log-price so that the term in $C_{i,j}$ vanishes, namely $\frac{\beta^2 h}{\delta^2} = 1 - rh$, then (2.8) looks even more like the binomial lattice approximation,

$$C_{i,j-1} = (1 - rh) \left[pC_{i+1,j} + (1 - p)C_{i-1,j} \right],$$

where $p = \frac{1}{2} + \frac{\delta(2r - \beta^2)}{4\beta^2}$ and with the multiplication by $1 - rh$ playing the role of the division by the bond factor R.

Example 2.10. Revisit Example 1.9 which valued a call option on an asset with initial price $S = 35$ over one year with volatility $\beta = \log 1.25 = 0.2231$, bond interest rate $r = \log 1.12 = 0.1133$, and strike price $X = 38.50$. Perform an n-step approximation, time step $h = 1/n$ years, with $\delta \approx \beta \sqrt{2h}$ so that the coefficient of $C_{i,j}$ is approximately 0.5 to reasonably ensure stability.[26] With $n = 4$ time steps, Algorithm 2.1 estimates the value of the call option to be $C = 3.49$, whereas with $n = 8$ we find $C = 3.40$, and we find $C = 3.41$ for $n = 16$ and above. Using this discretisation of the Black–Scholes equation (2.6) determines the value of an option with just $n = 16$ time steps, whereas with the binomial model we needed $n = 256$. ∎

Suggested activity: Do at least Exercise 2.15.

2.3.2   Self-financing portfolios

One question you may have had about the derivation of the Black–Scholes equation (2.6) is the following: Why did I not treat the hedge ratio ϕ as a stochastic Ito process? After all, the hedge ratio varies with the stochastic asset price S $\left(\text{recall } \phi = \frac{\partial C}{\partial S} \right)$ and so should be a stochastic process. The answer comes from investigating more completely the details of purchasing or selling the asset and, in particular, how such trades are financed by cash held in bonds.

Imagine we have sold a call option on an asset and seek a time varying portfolio of ϕ units of the asset and ψ bonds. We transfer money from and to bonds as needed to buy

---

[26]A useful rule of thumb to ensure stability of the trinomial model (2.8) is to choose time step h and x-step δ so that none of the three coefficients of $C_{i,j}$ and $C_{i,j\pm1}$ on the right-hand side of (2.8) are negative.

Algorithm 2.1 Code for determining the value of the call option in Example 2.10. Observe the use of nan to introduce unspecified boundary conditions that turn out to be irrelevant on the selected grid just as they are for the binomial model. In SCILAB use %nan instead of nan.

```
n=4
strike=38.50
beta=log(1.25)
r=log(1.12)
h=1/n
delta=beta*sqrt(2*h) % perhaps double
rt=beta^2*h/delta^2
p=0.5+delta*(2*r-beta^2)/(4*beta^2)
x=log(35)+(-n:n)*delta;
s=exp(x);
c=max(0,s-strike);
i=2:2*n;
for j=n:-1:1
    c=[nan c(i)*(1-r*h-rt)+c(i+1)*rt*p+c(i-1)*rt*(1-p) nan]
end
estimated_value=c(n+1)
```

and sell units of the asset. The whole portfolio then has value $\Pi = -C + \phi S + \psi B$ at any time t, where $B(t)$ is the value of a cash bond; we assume B grows exponentially, like e^{rt}, though more general models may also be analyzed. Now discretize time into small intervals of length Δt_j (perhaps each interval is one day):

- over the jth interval (the jth day) we hold ϕ_j units and ψ_j bonds with value from the start of the day of $\Pi_j = -C_j + \phi_j S_j + \psi_j B_j$;

- at the very start of the next $(j+1)$th interval (the opening of the next day) we find the portfolio has value $\Pi_{j+1} = -C_{j+1} + \phi_j S_{j+1} + \psi_j B_{j+1}$;

- at the start of the $(j+1)$th interval (the start of the day's trading) we immediately adjust the portfolio to make it risk free over the forthcoming $j+1$th interval by trading to then hold ϕ_{j+1} units and ψ_{j+1} bonds with value $-C_{j+1} + \phi_{j+1} S_{j+1} + \psi_{j+1} B_{j+1}$; as the portfolio is to be self-financing, this must have the same value as Π_{j+1}, since we just trade cash bonds for units of the asset; consequently, for a self-financing portfolio,

$$-C_{j+1} + \phi_{j+1} S_{j+1} + \psi_{j+1} B_{j+1} = -C_{j+1} + \phi_j S_{j+1} + \psi_j B_{j+1}$$
$$\Rightarrow S_{j+1}(\phi_{j+1} - \phi_j) + B_{j+1}(\psi_{j+1} - \psi_j) = 0$$
$$\Rightarrow S_{j+1}\Delta\phi_j + B_{j+1}\Delta\psi_j = 0.$$

The change in value of the portfolio from one time step to the next (from one day to the next) is thus

$$\begin{aligned}
\Delta\Pi_j &= -C_{j+1} + \phi_{j+1}S_{j+1} + \psi_{j+1}B_{j+1} - (-C_j + \phi_j S_j + \psi_j B_j)\\
&= -\Delta C_j + \phi_j(S_{j+1} - S_j) + \psi_j(B_{j+1} - B_j)\\
&\quad + (\phi_{j+1} - \phi_j)S_{j+1} + (\psi_{j+1} - \psi_j)B_{j+1}\\
&= -\Delta C_j + \phi_j \Delta S_j + \psi_j \Delta B_j + \underbrace{S_{j+1}\Delta\phi_j + B_{j+1}\Delta\psi_j}_{=\,0 \text{ as self-financing}}.
\end{aligned}$$

Write this in terms of infinitesimals to see that the changes in the value of such a managed, self-financing portfolio is

$$d\Pi = -dC + \phi\,dS + \psi\,dB. \tag{2.9}$$

That is, obtain the change in value of the portfolio as if the number of units of each component is held constant—just as we did in deriving the Black–Scholes equation (2.6). Exercise 2.17 asks you to recover the Black–Scholes equation (2.6) from an ensuing more sophisticated analysis based upon (2.9).

Summary

Ito's formula identifies risk-free portfolios that underpin the Black–Scholes equation (2.6) for valuing options. Discretizations of the Black–Scholes equation empower accurate numerical valuation of options, but need to be checked qualitatively.

2.4   Summary

- In stochastic calculus, differentials dW, dt, and $dW^2 = dt$ are retained, whereas $dt^2 = dt\,dW = 0$, as are higher order products.

- Ito's formula is that if $Y = f(t, X(t))$, then

$$dY = \frac{\partial f}{\partial t}dt + \frac{\partial f}{\partial x}dX + \tfrac{1}{2}\frac{\partial^2 f}{\partial x^2}dX^2,$$

  where dX^2 is to be simplified according to the above rules.

- The Black–Scholes PDE, for the value of an option at time t when the asset has price S,

$$\frac{\partial C}{\partial t} + rS\frac{\partial C}{\partial S} + \tfrac{1}{2}\beta^2 S^2 \frac{\partial^2 C}{\partial S^2} = rC,$$

  is an advection-diffusion equation to be solved backward in time from the expiry values of the option $C(T, S)$.

- Numerical discretizations of the Black–Scholes equation are effective in valuing options.

- Self-financing portfolios of ϕ and ψ units, respectively, of assets with prices S and B satisfy $S\,d\phi + B\,d\psi = 0$.

Exercises

2.1. Use the probability distribution function $p(z) = \exp(-z^2/2)/\sqrt{2\pi}$ for $N(0,1)$ random variables Z and integration by parts to deduce that $E[Z^p] = (p-1)(p-3)\ldots 1$ for even integers p. Hence deduce for Wiener processes $W(t)$ that $E[W(t)^p] = (p-1)(p-3)\ldots 1 \cdot t^{p/2}$ (for even integers p) by writing, at any given time, $W = Z\sqrt{t}$ where $Z \sim N(0,1)$.

2.2. Show that *almost surely* $\sum_{j=0}^{n-1} (\Delta W_j)^4 \to 0$ as the time step $\Delta t_j = h \to 0$ for a fixed final time $T = nh$. HINT: Use that $E[|Z|^4] = 3$ for random variates $Z \sim N(0,1)$.

2.3. Argue that *almost surely* $\sum_{j=0}^{n-1} |\Delta W_j|^p \to 0$ as $h \to 0$ for $p \geq 3$. HINT: Use that $E[|Z|^p]$ is finite for $Z \sim N(0,1)$.

2.4. Use Ito's formula to deduce the differential dX for the stochastic process $X(t) = 2 + t + \exp(W(t))$.

2.5. Use Ito's formula to find an SDE satisfied by the Ito process $X = \cos(t^2 + W)$. Write the SDE in terms of only t, X, and the differentials dt and dW.

2.6. Use Ito's formula to show that the following are the solutions to the SDEs you solved numerically in Exercise 1.7:

1. $X = (1+t)^2(X_0 + t + W(t))$ satisfies

$$dX = \left[\frac{2X}{1+t} + (1+t)^2 \right] dt + (1+t)^2 dW;$$

2. $X = (1 + W(t))^2$ satisfies $dX = dt + 2\sqrt{X}\, dW$ with $X_0 = 1$;

3. $X = \log(1 + W(t))$ satisfies $dX = -\frac{1}{2}e^{-2X}dt + e^{-X}dW$ with $X_0 = 0$;

4. $X = \sinh(t + W(t))$ satisfies $dX = \left[\frac{1}{2}X + \sqrt{1+X^2} \right] dt + \sqrt{1+X^2}\, dW$ with $X_0 = 0$; and

5. $Z = 1/(1 + W(t))$ satisfies $dZ = Z^3 dt - Z^2 dW$ with $Z_0 = 1$.

What features that you saw in the numerical solutions are explained by these analytic solutions?

2.7. Let $I_0(t) = 1$, $I_1(t) = W(t)$, $I_2(t) = W(t)^2 - t$, $I_3(t) = W(t)^3 - 3tW(t)$, and $I_4(t) = W(t)^4 - 6tW(t)^2 + 3t^2$. Use Ito's formula to show $dI_n = nI_{n-1}\, dW$. Describe the analogue with classic calculus. Use guesswork, checked with Ito's formula, to determine corresponding I_5 and I_6.[27]

2.8. Let an Ito process $Y(t) = 1/(t + X(t))$ in terms of the Ito process $X(t) = W(t)^2$, where $W(t)$ is a Wiener process. Use Ito's formula to deduce an expression for dY in terms of only t, dt, Y, and dW.

2.9. Use the simple version (2.3) of Ito's formula to generate, by substituting a variety of functions $f(t, W)$, a variety of SDEs whose analytic solutions you know.

2.10. Use Ito's formula (2.4) for the following:

---

[27]These $I_n(t)$ are closely related to the Hermite polynomials; see (Kreyszig 1999, pp. 246–247) or (Abramowitz and Stegun 1965, Chap. 22).

- Since $X = t + W(t)$ satisfies $dX = dt + dW$, confirm that $Y = \sinh X$ satisfies $dY = \left[\frac{1}{2}Y + \sqrt{1 + Y^2}\right] dt + \sqrt{1 + Y^2}\, dW$.

- For the above Ito process Y, deduce dZ for Ito process $Z = Y^3$.

Recall that $\frac{d}{dx}\sinh x = \cosh x$, $\frac{d}{dx}\cosh x = \sinh x$, and $\cosh^2 x = 1 + \sinh^2 x$.

2.11. Consider the Black–Scholes equation (2.6) with zero volatility:

$$\frac{\partial C}{\partial t} + rS\frac{\partial C}{\partial S} = rC.$$

Show by algebraic differentiation that $C = f(Se^{-rt})e^{rt}$ satisfies this PDE for any differentiable function f.

Now suppose you are considering some option with specified value at the expiry time T, say $C_T(S)$. Deduce that the above function $f(S) = C_T(RS)/R$, where $R = e^{rT}$. Hence argue that all points on the specified curve (S_T, C_T) (where S_T denotes the value of the asset at expiry) are mapped by the above PDE to points (S, C) at time $t = 0$ by the simple contraction mapping $S = S_T/R$ and $C = C_T/R$. Thus comment on how the Black–Scholes equation (2.6) with nonzero volatility predicts that the initial value of an option is just a smoothed version of the final value, but scaled by the factor $1/R$.

2.12. In November 2007 the company Wesfarmers purchased the Coles–Myer group. As part of the settlement, Coles–Myer shares were transformed into Wesfarmers *Partially Protected Shares* (WPPS). With a Floor Price specified to be $36, a Cap Price of $45, and the Lapse Date meaning the date four years from the Issue Date, the specifications of the WPPS included the following: *On a Lapse Date determined by Wesfarmers,*

1. *each Partially Protected Share will be reclassified into one Ordinary Share; and*

2. *subject to clause 8, each Partially Protected Shareholder will be issued an additional number of Ordinary Shares for each Partially Protected Share held on that date, in accordance with the following:*

   (a) *if the* MVWAP *is equal to or more than the Cap Price, no additional Ordinary Shares;*

   (b) *if the* MVWAP *is equal to or less than the Floor Price, 0.25 Ordinary Shares; and*

   (c) *if the* MVWAP *is between the Floor Price and the Cap Price, the number of Ordinary Shares calculated using the following formula:*

   $$\frac{\textit{Cap Price}}{\text{MVWAP}} - 1$$

   where MVWAP *means the* VWAP *for the period of two months immediately preceding, but not including, the date of the Lapse Notice.*

Instead of part 2(c), for simplicity treat the MVWAP as the sale price of Ordinary Shares sold on the Australian Stock Exchange on the Lapse Date. A challenge is to find the initial value of these WPPS using the Black–Scholes equation; for now just
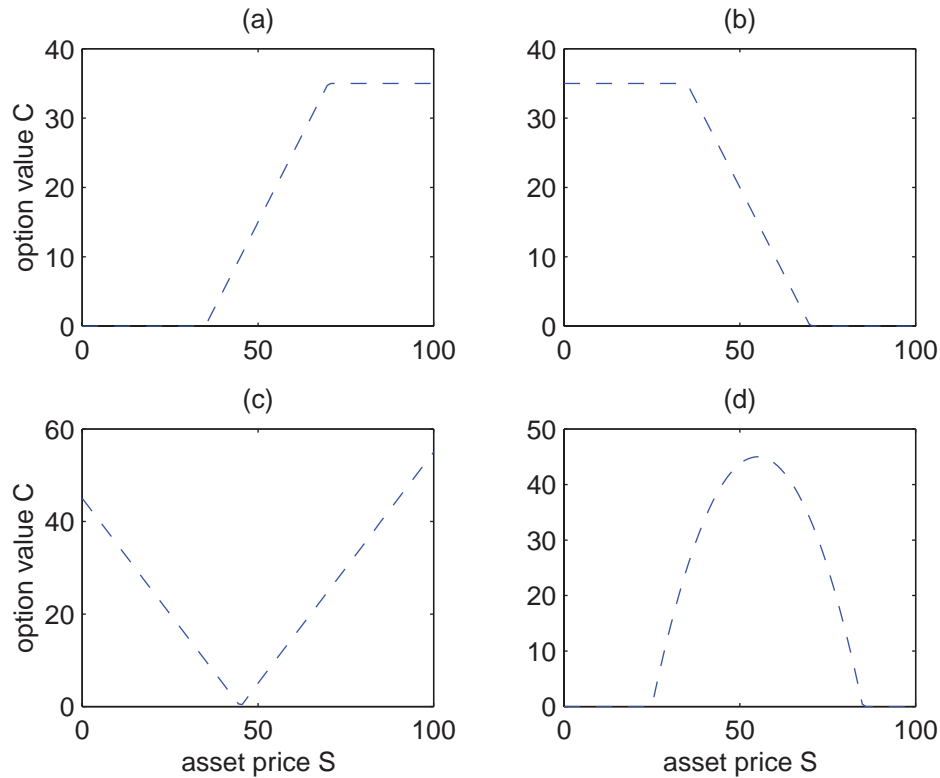
Figure 2.6. *The expiry value of four different options on an asset.*

do the following. *Translate the above* WPPS *conditions into an expiry condition for the Black–Scholes equation (a condition that depends upon the sale price).*

2.13. When one purchases, as an asset, shares in a company, one also expects some regular dividends as part of the financial benefit of holding the shares. Generalize the arguments of Section 2.3 to derive a modified Black–Scholes equation when the asset in the portfolio also returns dividends at a known rate δ. That is, for each share held in the portfolio, over a small time interval Δt the dividend contributes $\delta \Delta t$ to increasing the value of the portfolio.

2.14. *This exercise is challenging.* Extend the previous modification to the case where the dividend has both a deterministic and stochastic component; over a small time interval the dividend's contribution to increasing the value of the portfolio would be $\delta \Delta t + \rho \Delta W$.

2.15. Use the trinomial model to value the following option. Comment on the match with the qualitative graphical solution.

You, a bookmaker, have a client, Bob, who wants to bet $100 that Telstra shares will rise to be above $5 in one year's time. The share's current value is $4. Suppose the interest rate for bonds is 10% per year, and the volatility of the shares is 20% per $\sqrt{\text{year}}$. Deduce that you charge Bob at least about $20 to make this bet.

2.16. Modify the trinomial MATLAB/SCILAB code for Example 2.10 to value a knock out option (Example 2.9) which is as for Example 2.10 but additionally pays a rebate of $1 if the asset price ever falls to $30.

2.17. Using a self-financing portfolio and assuming the interest rate on bonds is the constant r, rederive the Black–Scholes equation (2.6) from (2.9).

2.18. Consider in turn each of the four options whose expiration values are plotted in Figure 2.6. Use the qualitative arguments of Section 2.3 to sketch by hand the option value, as a function of S, as predicted by the Black–Scholes equation at a significantly earlier time.

Answers to selected exercises

2.4. $dX = (1 + \frac{1}{2}e^W)dt + e^W dW = \frac{1}{2}(X - t)dt + (X - t - 2)dW.$

2.5. $dX = -\left[\frac{1}{2}X + 2t\sqrt{1 - X^2}\right]dt - \sqrt{1 - X^2}\,dW.$

2.8. $dX = dt + 2\sqrt{X}\,dW$, and hence $dY = 2Y^2(1 - 2tY)dt - 2Y^2\sqrt{\frac{1}{Y} - t}\,dW.$

2.16. The knock out value is $3.59.

Chapter 3

The Fokker–Planck Equation Describes the Probability Distribution

Contents

Previous chapters concentrated on the properties of individual realizations of a stochastic process: that they have drift and volatility, and that simple numerics converge, albeit slowly. The alternative is to discover the statistics of stochastic processes: their mean and variance, or more generally, the probability distribution of the realizations. Compare this with Markov chains where, instead of running many simulations of actual realizations of the process, we typically discuss the probability distribution over the states of the chain (see, e.g., Kao 1997, Chap. 4). In Markov chains, the transition matrix guides the evolution of the probabilities; for SDEs, the Fokker–Planck equation describes the evolution of the probability distribution.

Example 3.1 (Wiener processes diffuse). Consider an ensemble of realizations of a Wiener process $W(t)$; for example, see five of the realizations shown in Figure 1.3, and the ten realizations in Figure 3.1. As time increases, these realizations spread out over w-space, that is, they diffuse. We know that at any time t, a Wiener process is distributed as an $N(0, t)$ random variable. Thus its probability distribution function (PDF) is the Gaussian, normal distribution $p(t, w) = \exp(-w^2/2t)/\sqrt{2\pi t}$: as shown in Figure 3.2, this Gaussian spreads in time corresponding to the spreading realizations. Recall that this Gaussian distribution satisfies the PDE for diffusion (see, e.g., Kreyszig 1999, §11.5–6):

$$\frac{\partial p}{\partial t} = \frac{1}{2}\frac{\partial^2 p}{\partial w^2}.$$
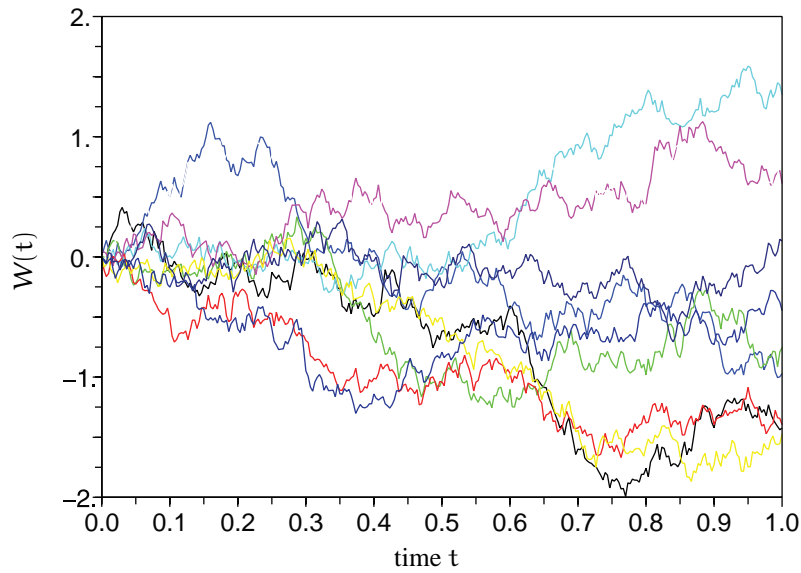
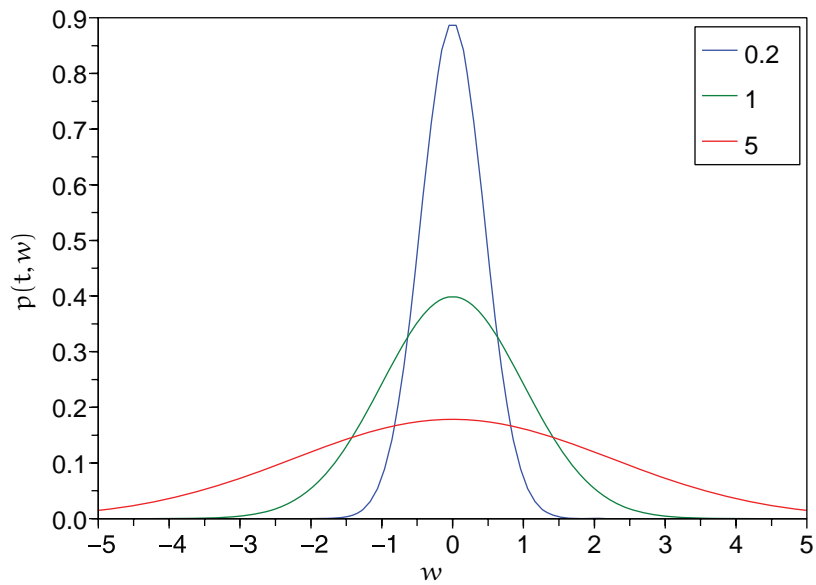**Figure 3.1.** *Ten realizations of a Wiener process $W(t)$.*

Figure 3.2. *The spreading Gaussian probability distribution function $p(t, w)$ of a Wiener process at three times:* $t = 0.2$, $t = 1$, *and* $t = 5$.

This PDE governing the evolution of the PDF $p(t, w)$ is the simplest example of a Fokker–Planck equation describing the evolution of the distribution of a stochastic process.

However, this view of the Wiener process is limited in that, for example, it does not discern the nature of the increments nor the continuity of the stochastic process. Consider a random function $X(t) = Z\sqrt{t}$, where Z is a random variable distributed $N(0, 1)$ so $X(t)$ is a square-root curve with a randomly chosen coefficient; then $X(t)$ has *exactly the same* Gaussian PDF as the Wiener process but is quite different in nature. The same comment holds if Z is instead generated randomly and independently at each time t. ∎

Although the PDF does not discriminate between the possibilities discussed at the end of this example, useful statistics are obtained from the PDF and sometimes show overall features missed by individual realizations of the SDE! Further, when we cannot solve the SDE sometimes, we can solve the Fokker–Planck equation, and hence make at least partial analytic progress.

Example 3.2 (expect important but rare large fluctuations).   Section 2.1.2 describes how the solution of the SDE $dX = \alpha X \, dt + \beta X \, dW$ is the stochastic process $X(t) = \exp[(\alpha - \frac{1}{2}\beta^2)t + \beta W(t)]$. This process almost always decays to zero provided $\alpha - \frac{1}{2}\beta^2 < 0$ (even when the drift rate $\alpha > 0$). However, this remarkable statement misleads as rare fluctuations are significant.

Determining the expected value of $X(t)$ from its PDF illustrates the significance of the rare fluctuations. First, we determine that

$$E\left[e^{\beta W(t)}\right] = e^{\beta^2 t/2},\tag{3.1}$$

using the PDF of the Wiener process. From the definition of expectation in terms of the PDF

$$
\begin{aligned}
E[e^{\beta W(t)}] &= \int_{-\infty}^{\infty} e^{\beta w} p(t, w) \, dw \\
&= \int_{-\infty}^{\infty} e^{\beta w} \frac{1}{\sqrt{2\pi t}} e^{-w^2/2t} \, dw \\
&= \int_{\infty}^{\infty} \frac{1}{\sqrt{2\pi t}} \exp\left[-\frac{w^2}{2t} + \beta w\right] dw \qquad \text{then upon completing the square} \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi t}} \exp\left[-\frac{1}{2t}(w - t\beta)^2 + \frac{\beta^2 t}{2}\right] dw \\
&= e^{\beta^2 t/2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi t}} e^{-(w - t\beta)^2/(2t)} \, dw \\
&= e^{\beta^2 t/2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi t}} e^{-u^2/(2t)} \, du \qquad \text{substituting } u = w - t\beta \\
&= e^{\beta^2 t/2},
\end{aligned}
$$

as the integral of the Gaussian is 1. This derivation confirms (3.1).

Second, we conclude that the solutions to exponential Brownian motion, although almost always tending to zero for large enough β, nonetheless has a growing expectation

whenever $\alpha > 0$:

$$
\begin{aligned}
E[X(t)] &= E\left[X_0 e^{(\alpha - \beta^2/2)t + \beta W(t)}\right] \\
&= E\left[X_0 e^{(\alpha - \beta^2/2)t} e^{\beta W(t)}\right] \\
&= X_0 e^{(\alpha - \beta^2/2)t} E\left[e^{\beta W(t)}\right] \\
&= X_0 e^{(\alpha - \beta^2/2)t} e^{\beta^2 t/2} \\
&= X_0 e^{\alpha t}.
\end{aligned}
$$

For example, the almost surely decaying realizations $X(t) = e^{-t + 2W(t)}$ of the SDE $dX = X\,dt + 2X\,dW$ (such that $X(0) = 1$) nonetheless have an exponentially growing expectation $E[X(t)] = e^t$. Thus very rare, very large fluctuations in $X(t)$ must occur so that the expectation can grow even though almost all realizations decay to zero. The histogram of Figure 2.1 hinted at this significance of the few large fluctuations.

Further, as a specific instance of Exercise 3.1, you will also find here that $\text{Var}[X(t)] = e^{6t} - e^{2t}$, which grows alarmingly quickly even though, again, almost all realizations decay to zero! ∎

Example 3.3 (random walkers solve Laplace's equation). One way that we can approximately solve Laplace's equation $\nabla^2 u = 0$ in some domain (see, e.g., Kreyszig 1999, §11.9–11) is to choose a point of interest (x_0, y_0) and release from that point a large number of random walkers (drunks!) who execute Brownian motion in both x and y: that is, $x = W_1(t)$ and $y = W_2(t)$ for two independent Wiener processes. These (drunken) walkers stick to the boundary when they first hit the boundary as shown in Figure 3.3 (see the simulation of Algorithm A.3). Then when enough walkers have become stuck to the boundary, one simply takes the average of the boundary values $f(P)$ to estimate $u(x_0, y_0)$. Amazingly, this stochastic procedure finds the solution of the deterministic Dirichlet problem $\nabla^2 u = 0$ such that $u = f(P)$ on the boundary.[28]

One way to see the connection between the random walkers and Laplace's equation is to recognize that the PDF of the random walkers satisfy the two-dimensional diffusion equation $\frac{\partial p}{\partial t} = \frac{1}{2}\nabla^2 p$; then when they remain stuck to the boundary, the time derivative becomes zero, and hence the walkers have somehow solved $0 = \nabla^2 p$, which is Laplace's equation. Section 3.2 explores this useful connection between SDEs and deterministic differential equations. ∎

The Fokker–Planck equation establishes a useful transformation between the solution of SDEs and certain PDEs. One great advantage of using this connection for solving PDEs is that you do not have to compute the solution everywhere if all you need is the solution at one or a few points. Another great advantage is that it easily handles complex shaped domains, as it needs no underlying grid. This technique is incredibly useful for solving problems in very high dimensional domains. For example, in investigating quantum dynamics (of Bose–Einstein condensates) with 1000 atoms, say, one wants to solve PDEs in a vector space with over 10^{1000} dimensions! The stochastic solution involving

---

[28] An estimate of the error is the sample standard deviation divided by the square root of the number of walkers who reached the boundary.
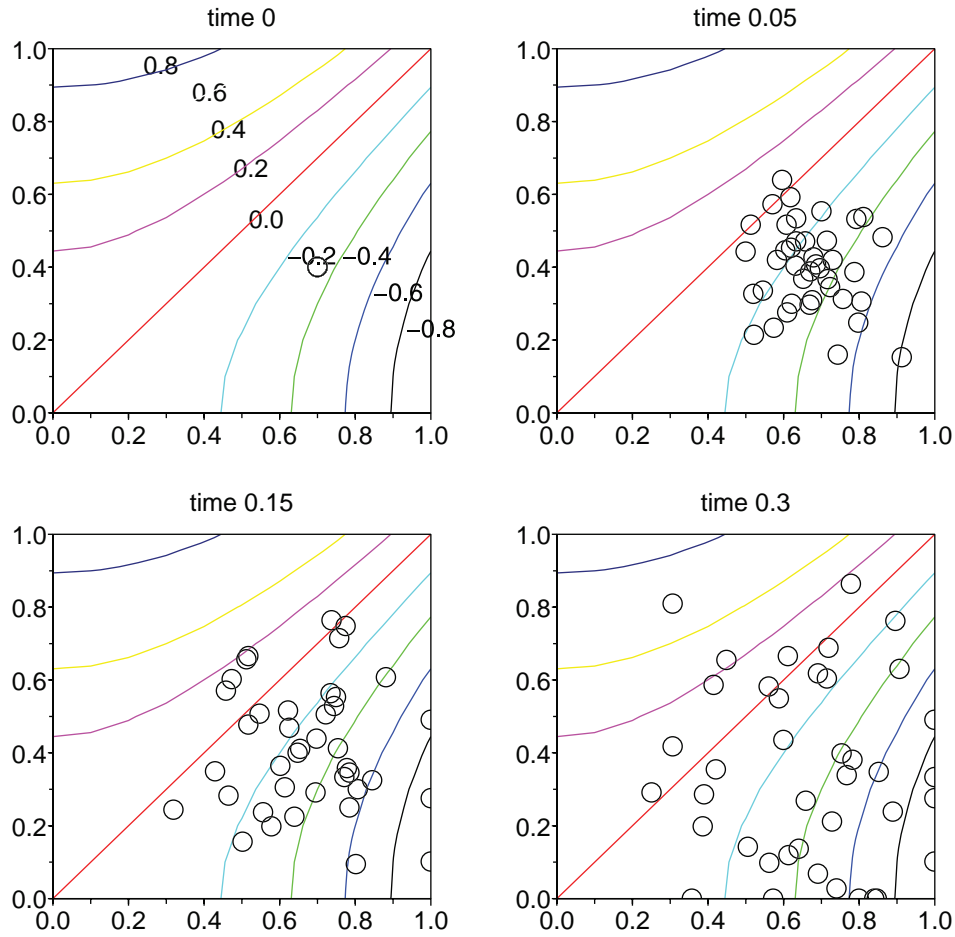
Figure 3.3. *Forty random walkers (circles) released from* $(0.7, 0.4)$, *spread out stochastically in time until they hit the boundaries (the unit square) where they stick to collectively estimate* $u(0.7, 0.4)$. *The contour plot shows the true solution of Laplace's equation: its value on the boundary gives the contribution of each random walker to the estimate of* $u(0.7, 0.4)$.

many realizations of just 1000 atoms is accessible via supercomputers. That there is this connection between PDEs and SDEs is established by the Fokker–Planck equation for the PDFs of the stochastic process.

3.1 The probability distribution evolves forward in time

We now discover how to describe the evolution of the probability distribution function $p(t, x)$ for a general Ito process with some drift and volatility. For simplicity we just consider one stochastic process dependent upon one Wiener process—the generalization to

Figure 3.4. *Left: either an up or down step reaches the state* x *at time* $t + h$. *Right: small intervals of length* $d\xi$ *reach a small interval of length* dx *over one time step.*

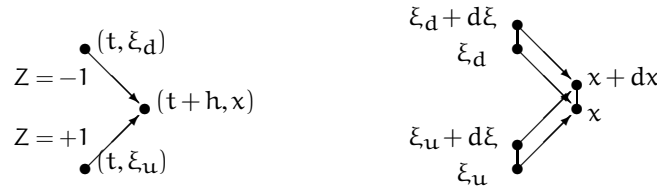multiple coupled stochastic processes involving multiple independent Wiener processes is analogous and of much practical use, but we will not explore it in this book. This section ends by showing how to model biological populations with stochastic effects representative of those endemic in the environment.

Analogue with Markov chains

The probability distribution function $p(t, x)$, namely, the probability of realizations being near x at time t, is analogous to the vector of probabilities $\mathbf{p}(t)$, where the jth component, $p_j(t)$, is the probability being in a state j at time t discussed in Markov chains such as birth and death processes (see, e.g., Kao 1997, §4.1).

Theorem 3.4 (Fokker–Planck equation). *Consider the Ito process* $X(t)$ *with drift* $\mu(X)$ *and volatility* $\sigma(X)$, *and hence satisfying the* SDE $dX = \mu\,dt + \sigma\,dW$, *where for simplicity we restrict our attention to the autonomous case of no direct time dependence in the drift and volatility. Then the* PDF *of the ensemble of realizations* $p(t, x)$ *satisfies the* Fokker–Planck equation, *alternatively called the* Kolmogorov forward equation,

$$\frac{\partial p}{\partial t} = -\frac{\partial}{\partial x}[\mu p] + \frac{\partial^2}{\partial x^2}\left[\tfrac{1}{2}\sigma^2 p\right]. \tag{3.2}$$

Analogue

The Fokker–Planck equation is analogous to the evolution of probability distributions in Markov chains: $\mathbf{p}(t+1) = \mathbf{p}(t)P$ for some transition matrix P. In the Fokker–Planck equation (3.2), the x derivatives involving the drift and volatility operate exactly like the transition matrix P—they operate to dictate how probability distributions evolve in time.

Proof. We derive this Fokker–Planck equation for the case of constant volatility σ (variable volatility makes the details much more complicated).[29] Throughout the derivation, dashes denote $\partial/\partial x$, and the drift μ is evaluated at x unless otherwise specified.

We investigate the probability that the realization at time $t + h$ is near the value x, that is, within the interval $[x, x + dx]$ for some small (infinitesimal) interval length dx. In terms of the PDF, the probability of being in this interval is

$$\Pr\{X(t+h) \in [x, x + dx]\} = p(t+h, x)\,dx.$$

---

[29]Exercises 3.7–3.8 ask you to deal with specific cases of variable volatility. Appendix B.1 gives a proof for general volatility.

The realization reaches x from a variety of possible values of $X(t)$ depending upon the Wiener increment $\Delta W = \sqrt{h}Z$. That Z is a normal random variable is largely immaterial—so long as Z has zero mean and unit variance, its cumulative sum, over many small time steps, will quickly become normal—thus for simplicity and to the same effect assume $Z = \pm 1$ each with probability $\frac{1}{2}$. This assumption is rather like the binary model of asset price movement. As Figure 3.4 (left) shows, the process reaches x if it starts from ξ_u with $Z = +1$, $\Delta W = +\sqrt{h}$, or if it starts from ξ_d if $Z = -1$, $\Delta W = -\sqrt{h}$, where the values for ξ are determined from

$$x = \xi + \mu(\xi)h \pm \sigma\sqrt{h}.$$

This is a pair of implicit equations for the two ξ values. Solve the pair approximately by iteration after rearranging to $\xi = x - \mu(\xi)h \mp \sigma\sqrt{h}$ and starting from[30]

$$\xi \approx x;$$
$$\xi \approx x - \mu(x)h \mp \sigma\sqrt{h};$$
$$\xi = x - \mu h \mp \sigma\sqrt{h} + \cdots.$$

Differentiating this shows how intervals of X at time t become slightly stretched or compressed in the evolution to time $t + h$, as shown in Figure 3.4 (right):

$$d\xi = \left[1 - \mu'h + \cdots\right]dx.$$

Then because $Z = \pm 1$, each with probability $\frac{1}{2}$,

$$\begin{aligned}
p(t+h,x)dx &= \Pr\{X(t+h) \in [x, x+dx]\} \\
&= \tfrac{1}{2}\Pr\{X(t) \in [\xi_u, \xi_u + d\xi_u]\} + \tfrac{1}{2}\Pr\{X(t) \in [\xi_d, \xi_d + d\xi_d]\} \\
&= \tfrac{1}{2}p(t, \xi_u)d\xi_u + \tfrac{1}{2}p(t, \xi_d)d\xi_d \\
&= \tfrac{1}{2}p(t, x - \mu h - \sigma\sqrt{h})[1 - \mu'h]dx \\
&\quad + \tfrac{1}{2}p(t, x - \mu h + \sigma\sqrt{h})[1 - \mu'h]dx + \cdots.
\end{aligned}$$

Divide by the infinitesimal dx and expand p in Taylor series about $p(t, x)$:

$$\begin{aligned}
p(t+h,x) &= \tfrac{1}{2}\left[p - (\mu h + \sigma\sqrt{h})p' + \tfrac{1}{2}(\mu h + \sigma\sqrt{h})^2 p''\right][1 - \mu'h] \\
&\quad + \tfrac{1}{2}\left[p + (-\mu h + \sigma\sqrt{h})p' + \tfrac{1}{2}(-\mu h + \sigma\sqrt{h})^2 p''\right][1 - \mu'h] \\
&\quad + \cdots \\
&= \left[p - \mu h p' + \tfrac{1}{2}\sigma^2 h p''\right][1 - \mu'h] + \cdots \\
&= p - \mu h p' - \mu'h p + \tfrac{1}{2}\sigma^2 h p'' + \cdots.
\end{aligned}$$

Putting p onto the left and dividing by h leads to

$$\frac{p(t+h,x) - p(t,x)}{h} = -(\mu p)' + \tfrac{1}{2}\sigma^2 p'' + \cdots.$$

---

[30] Recall that the ellipsis "$\cdots$" denotes small terms of higher order in time step h that we neglect. Here such terms are typically in $h^{3/2}$, h^2, and so on.

Take the limit as $h \to 0$ to deduce

$$\frac{\partial p}{\partial t} = -\frac{\partial}{\partial x}(\mu p) + \frac{1}{2}\sigma^2 \frac{\partial^2 p}{\partial x^2}.$$

This is the Fokker–Planck equation (3.2) for the case of constant volatility. Appendix B.1 gives an alternate and more general proof.   □

Interpretation

The Fokker–Planck equation (3.2) has the following physical interpretation following from the modeling of motion in a one-dimensional continuum (see, e.g., Roberts 1994). Rewrite equation (3.2) in "conservative" form

$$\frac{\partial p}{\partial t} + \frac{\partial}{\partial x}\left[\mu p - \frac{\partial}{\partial x}(Dp)\right] = 0,$$

where $D(x) = \frac{1}{2}\sigma(x)^2$. This equation describes the conservation of probability distribution $p(t,x)$ as it "'moves" along the x-axis with a flux $q = \mu p - \frac{\partial}{\partial x}(Dp)$. The second component of the flux, $-\frac{\partial}{\partial x}(Dp)$, is a diffusive term induced by the volatility of the SDE. However, physical diffusion normally appears in the flux as $-D\frac{\partial p}{\partial x}$, with $D(x)$ being the effective diffusion coefficient. Here the additional part of the second term $-\frac{\partial}{\partial x}(Dp)$ is $-\frac{\partial D}{\partial x}p = \sigma\sigma' p$ and is best grouped with the first term $-\mu p$. Thus write the flux as

$$q = (\mu - \sigma\sigma')p - D\frac{\partial p}{\partial x},$$

and interpret this flux as probability distribution being carried by a mean velocity $\mu - \sigma\sigma'$ due to the drift and to the asymmetry in the noise, and as probability distribution diffusing with coefficient $D = \frac{1}{2}\sigma^2$.[31]

3.1.1   Steady state probability distributions

Example 3.5 (the Ornstein–Uhlenbeck process). Imagine a car: when you press down on it for a short time, its suspension reacts and subsequently lifts the car back to its equilibrium height above the road. The natural decay in the dynamics of the suspension brings the car back to its equilibrium. Now drive the car along a bumpy road: the car moves up and down in a complex response to the bumps and to the dynamics of its suspension. Consider the bumps as a stochastic forcing of the suspension; then we would model the combined dynamics by an SDE. The height of the car above the road is around about its normal height, but the bumps cause stochastic fluctuations in its height. Let us see such behavior in mathematics.

---

[31] If the random noise is absent, the Fokker–Planck equation reduces to the *Liouville equation* $\frac{\partial p}{\partial t} = -\frac{\partial}{\partial x}(\mu p)$, which describes the dynamics of probability distribution for a deterministic differential equation. This view of the dynamics is sometimes sought when either the initial conditions are stochastic or because chaos in the dynamics, such as fluid turbulence, causes randomness to be effectively generated in the dynamics of the system.
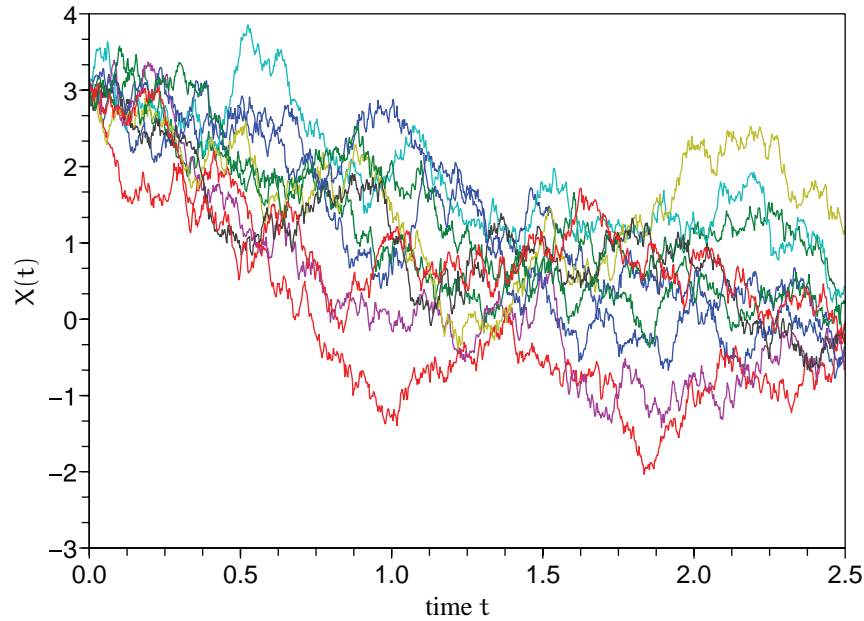
Figure 3.5. *Ten realizations of an Ornstein–Uhlenbeck process* $X(t)$ *with parameters* $\alpha = 1$ *and* $\sigma = \sqrt{2}$, *and all realizations starting with* $X(0) = 3$.

The so-called *Ornstein–Uhlenbeck process*, met in Exercise 1.6, combines deterministic exponential decay (modeling the car's suspension) with an additive noise (modeling the forcing of the bumps). The form of the SDE is generally

$$dX = -\alpha X \, dt + \sigma \, dW \tag{3.3}$$

for some constants α and σ measuring the rate of deterministic decay and the level of stochastic forcing, respectively. Figure 3.5 shows ten realizations, and Figure 3.6 shows how the corresponding PDF evolves. Find the steady state PDF for this process, that is, the density of realizations in Figure 3.5 *after* the initial transients, namely the shape that is appearing at large times in Figure 3.6.

Solution: The Fokker–Planck equation (3.2) for the PDF is

$$\frac{\partial p}{\partial t} = -\frac{\partial}{\partial x}(-\alpha x p) + \frac{\partial^2}{\partial x^2}\left(\tfrac{1}{2}\sigma^2 p\right).$$

Obtain the differential equation for the steady state PDF by setting $\partial p / \partial t = 0$; then this Fokker–Planck equation becomes

$$0 = \frac{\partial}{\partial x}\left[\alpha x p + \tfrac{1}{2}\sigma^2 \frac{\partial p}{\partial x}\right]$$

$$\Rightarrow \text{constant} = \alpha x p + \tfrac{1}{2}\sigma^2 \frac{\partial p}{\partial x} \qquad \text{upon integrating.}$$
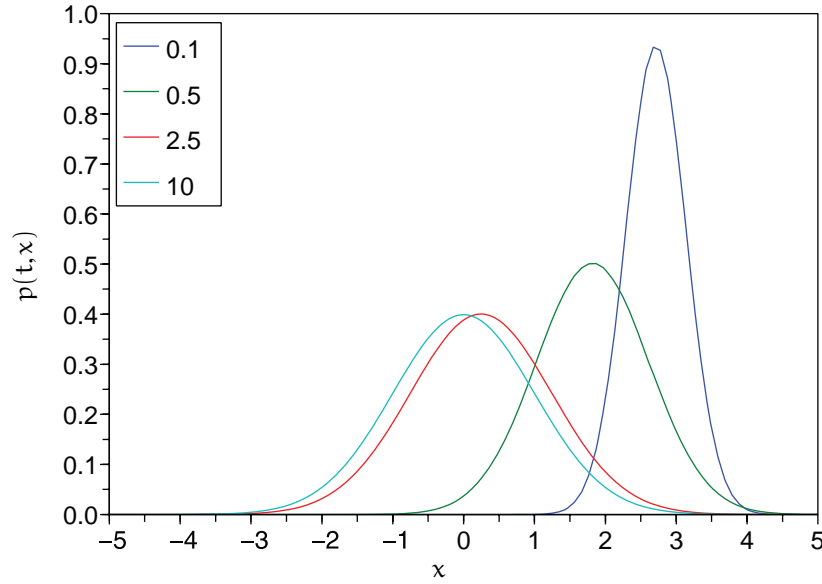
Figure 3.6. *Evolving* PDF $p(t,x)$ *for the Ornstein–Uhlenbeck process* $dX = -X\,dt + \sqrt{2}\,dW$ *with initial condition* $X(0) = 3$ *and plotted at times* $t = 0.1$, *0.5, 2.5, and* 10. *Compare with the realizations in Figure* 3.5.

The constant on the left-hand side is zero as the PDF p and its derivative have to vanish, for large enough x, in order for the integral of p to be one. Rearranging it as a separable ODE (see, e.g., Kreyszig 1999, §1.3–4) leads to

$$\tfrac{1}{2}\sigma^2 \int \frac{dp}{p} = \int -\alpha x\,dx$$

$$\Rightarrow \log p = -\alpha x^2/\sigma^2 + \text{constant}$$

$$\Rightarrow p = Ae^{-\alpha x^2/\sigma^2}$$

for some integration constant A. This is a Gaussian distribution centered on $x = 0$ and with width proportional to $\sigma/\sqrt{\alpha}$. The constant of proportionality is then well known to be $A = \sqrt{\alpha/\pi}/\sigma$ in order to ensure that the total probability, the area under $p(x)$, is one. The additive noise of the Ornstein–Uhlenbeck process has just "smeared out" the stable fixed point at $X = 0$. The width of the smearing is proportional to the strength of the noise, σ, and proportional to the inverse square root of the rate of attraction of the fixed point, $1/\sqrt{\alpha}$. ∎

Analogue

The steady state probability distribution $p(x)$ is analogous to the steady state distributions π found for discrete state stochastic models. The difference here is the continuum of possible states. Just as you normalize steady state distributions, so we normalize here: For a finite number of states $\sum_j \pi_j = 1$, here $\int p(x)\,dx = 1$.
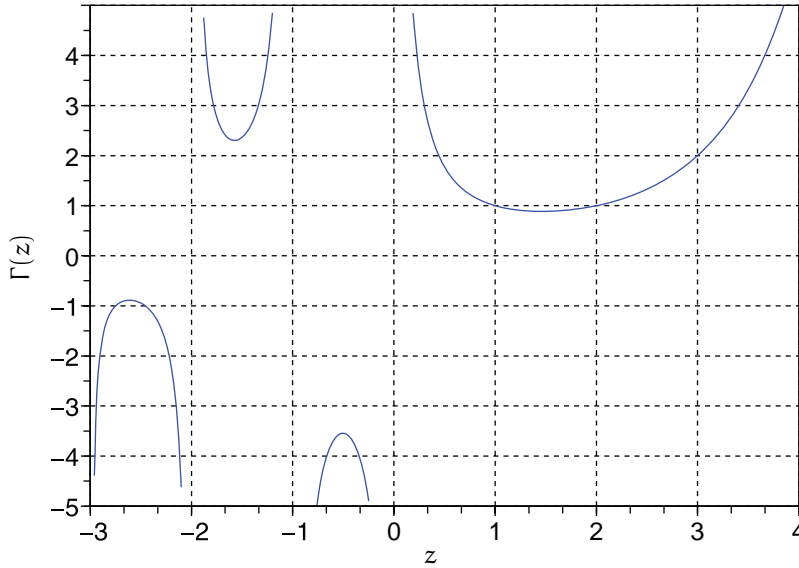
Figure 3.7. *The Gamma function* (3.4) *for real argument z.*

Gamma function

In the next example we need to use the *Gamma function*, $\Gamma(z)$, which you have possibly already met in other studies (see, e.g., Kreyszig 1999, pp. A54–55),

$$\Gamma(z) = \int_0^\infty x^{z-1} e^{-x} \, dx, \qquad (3.4)$$

as plotted in Figure 3.7 for real z. Integration by parts shows that $\Gamma(z+1) = z\Gamma(z)$. Hence, for example, since $\Gamma(1) = 1$ then for integer n, $\Gamma(n+1) = n!$. Other special values are $\Gamma(\frac{1}{2}) = \sqrt{\pi}$, and hence $\Gamma(\frac{3}{2}) = \frac{1}{2}\sqrt{\pi}$.

Example 3.6 (a two humped camel). Investigate the steady state PDF of the SDE $dX = (3X - X^3) dt + X \, dW$ and relate it to the deterministic dynamics. Compare it with the steady state of the SDE $dX = (3X - X^3) dt + 2X \, dW$ that has twice the volatility.

Solution: First, investigate the deterministic dynamics of $dX = (3X - X^3) dt$ as done in courses on ODEs (see, e.g., Kreyszig 1999, §3.3–3.5). This ODE has fixed points (equilibria) where $3X - X^3 = 0$, namely $X = 0$ and $X = \pm\sqrt{3}$; the fixed point at $X = 0$ is unstable as the linearized dynamics are $dX = 3X \, dt$ with exponentially growing solutions, whereas the fixed points at $X = \pm\sqrt{3}$ are stable as the local dynamics, say $X = \pm\sqrt{3} + Y(t)$, are $dY = -6Y \, dt$ with exponentially decaying solutions. Thus all deterministic trajectories evolve to one or another of the fixed points $X = \pm\sqrt{3}$.

Second, consider the steady state PDF $p(x)$ of the SDE $dX = (3X - X^3) dt + X \, dW$. It satisfies the time-independent Fokker–Planck equation

$$0 = -\frac{\partial}{\partial x}\left[(3x - x^3)p\right] + \frac{\partial^2}{\partial x^2}\left[\tfrac{1}{2}x^2 p\right].$$

One integral with respect to x leads to

$$-(3x - x^3)p + \frac{\partial}{\partial x}\left[\tfrac{1}{2}x^2 p\right] = \text{constant},$$

but this constant has to be zero, as p and its derivatives must vanish for large enough x. Thus by expanding, rearranging, and recognizing that the ODE is separable, we obtain

$$\tfrac{1}{2}x^2\frac{\partial p}{\partial x} = (2x - x^3)p$$

$$\Rightarrow \int \frac{dp}{p} = \int \frac{4x - 2x^3}{x^2}\,dx$$

$$\Rightarrow \log p = \int \frac{4}{x} - 2x\,dx$$

$$\Rightarrow \log p = 4\log|x| - x^2 + \text{constant}$$

$$\Rightarrow p = Ax^4 e^{-x^2}.$$

Thus the steady state PDF, as shown in Figure 3.8, is zero near $x = 0$, increases away from $x = 0$ by the x^4 factor, but soon is brought back to zero by the rapid decay of the e^{-x^2} factor. The two humps of the probability distribution correspond to the two stable fixed points of the deterministic ODE.

Determine the integration constant A by requiring that the area under the PDF be one: using symmetry, integration over half the domain requires

$$\tfrac{1}{2} = \int_0^\infty Ax^4 e^{-x^2}\,dx$$

$$= \frac{A}{2}\int_0^\infty u^{3/2}e^{-u}\,du \quad \text{upon substituting } u = x^2$$

$$= \frac{A}{2}\Gamma(5/2) = \frac{3\sqrt{\pi}}{8}A,$$

and thus $A = 4/(3\sqrt{\pi})$.

Lastly, a perusal of the deterministic part of $dX = (3X - X^3)dt + 2X\,dW$ again suggests that there should be two humps in the PDF near the two stable deterministic fixed points $X = \pm\sqrt{3}$. However, the steady solutions of the corresponding Fokker–Planck equation,

$$0 = -\frac{\partial}{\partial x}\left[(3x - x^3)p\right] + \frac{\partial^2}{\partial x^2}\left[2x^2 p\right],$$

are derived via

$$-(3x - x^3)p + \frac{\partial}{\partial x}\left[2x^2 p\right] = 0$$

$$\Rightarrow 2x^2\frac{\partial p}{\partial x} = (-x - x^3)p$$

$$\Rightarrow \int \frac{dp}{p} = \int \frac{-x - x^3}{2x^2}\,dx$$
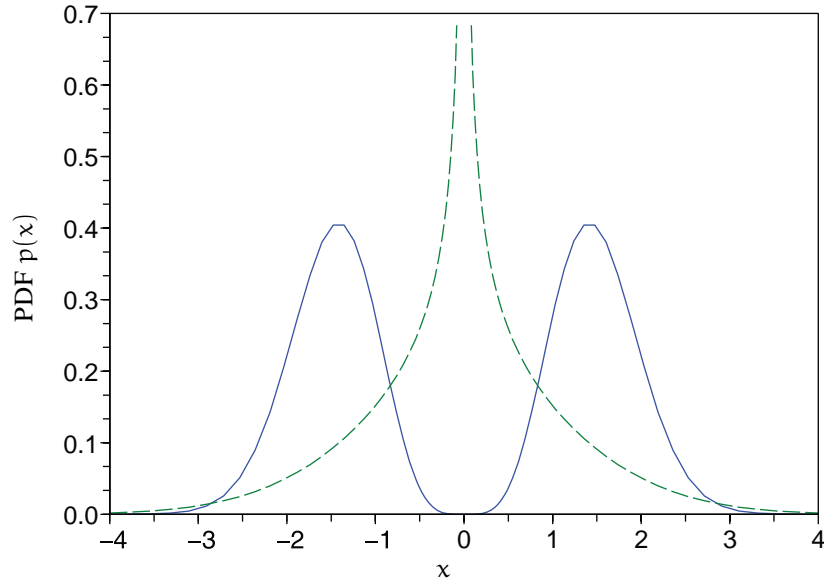
Figure 3.8. *Steady state probability distributions for Example* 3.6. *Solid blue line:* $dX = (3X - X^3)dt + X dW$ *with lower volatility has two humps; dashed green line:* $dX = (3X - X^3)dt + 2X dW$ *with larger volatility peaks at the origin.*

$$\Rightarrow \log p = \int -\frac{1}{2x} - \tfrac{1}{2}x \, dx$$

$$\Rightarrow \log p = -\tfrac{1}{2}\log|x| - \tfrac{1}{4}x^2 + \text{constant}$$

$$\Rightarrow p = \frac{A}{\sqrt{|x|}}e^{-x^2/4}.$$

Figure 3.8 shows that doubling the level of the multiplicative noise, now $2X \, dW$, effectively stabilizes the fixed point at the origin. The large spike of probability at the origin has finite area as $1/\sqrt{x}$ is integrable. Requiring the total area under the PDF to be one implies

$$\tfrac{1}{2} = \int_0^\infty Ax^{-1/2}e^{-x^2/4}dx$$

$$= \frac{A}{\sqrt{2}}\int_0^\infty u^{-3/4}e^{-u}du \quad \text{upon substituting } u = x^2/4$$

$$= \frac{A}{\sqrt{2}}\Gamma(1/4),$$

and thus $A = 1/(\sqrt{2}\Gamma(1/4)) = 0.1950$. ∎

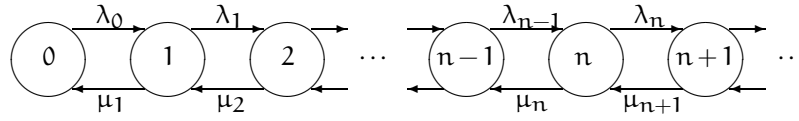**Figure 3.9.** *Transitions between the numbers of individuals in a population.*

3.1.2   Modeling large birth and death processes

Consider the modeling of populations by a birth and death process where we track each and every birth and death, as shown in the states and transitions of Figure 3.9. But in large populations we only need to know aggregated births and deaths: for example, how many *thousands* of people are infected in a flu epidemic? In large populations it is grossly inefficient to track each and every event. The Fokker–Planck equation empowers us to transform a single event model, such as a birth and death model, into an efficient aggregate SDE model via the evolution of the probability distribution. Here we introduce the key idea via one example.

Example 3.7 (births/deaths in a large population). Malthus proposed one of the first mathematical models of biological populations: The number of animals $N(t)$ grows in time according to $dN/dt = (\alpha - \beta)N$, where α is the birth rate per individual and β is the death rate per individual. When the birth rate is greater than the death rate, $\alpha > \beta$, then inevitably the population grows exponentially. Question: What happens in our uncertain fluctuating environment? Answer: The population fluctuates as plotted in Figure 3.10.

   This example introduces a scheme to describe the number of individuals in a population, $N(t)$, as a Markov birth and death process. This is then approximated as the Fokker–Planck equation (3.2) for a stochastic version of the Malthusian model $dN/dt = (\alpha - \beta)N$. Let n range over the total number of individuals in the population (a nonnegative integer),[32] and let $p_n(t)$ denote the probability that there are n individuals in the population at time t. Then changes to the number of individuals in the population are due to births at some rate λ_n and deaths at some rate μ_n, as shown in Figure 3.9. For a biological population with no constraints on the number of individuals, we expect that the birth and death rates will be proportional to the number of individuals: $\lambda_n = \alpha n$ and $\mu_n = \beta n$ for some constants α and β. These are the key parameters and variables of the stochastic Malthusian model.

   How does the number of individuals evolve? The time rate of change of the probability of there being n individuals is decreased by a birth to $n+1$ individuals or by a death to $n-1$ individuals, or is increased by a death among $n+1$ individuals or a birth among $n-1$ individuals:

$$\frac{dp_n}{dt} = \lambda_{n-1}p_{n-1} - (\lambda_n + \mu_n)p_n + \mu_{n+1}p_{n+1}$$

$$= \alpha(n-1)p_{n-1} - (\alpha+\beta)np_n + \beta(n+1)p_{n+1}$$

then upon assuming p_n varies smoothly in n, Taylor expanding all $p_{n\pm1}$ terms about n, and using dashes to denote $\partial/\partial n$

---

[32]For sexual animals, the usual practice is to count only the female of the species, as they are most closely involved in reproduction.
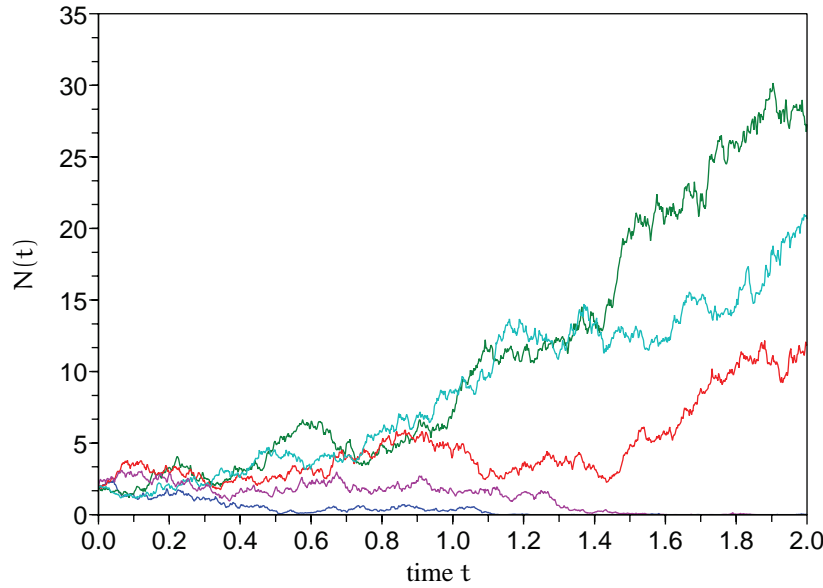
Figure 3.10. *Five realizations of the stochastic population model* (3.6) *with growth rate* $\alpha = 2$ *and death rate* $\beta = 1$. *Stochastic fluctuations potentially cause large variations in the population growth.*

$$
= \alpha(n-1)\left[p - p' + \tfrac{1}{2}p'' - \cdots\right] - (\alpha + \beta)np
$$
$$
+ \beta(n+1)\left[p + p' + \tfrac{1}{2}p'' + \cdots\right]
$$
$$
= (\beta - \alpha)p + [\beta n + \beta - \alpha n + \alpha]p'
$$
$$
+ \tfrac{1}{2}[\beta n + \beta + \alpha n - \alpha]p'' + \cdots
$$
$$
= \frac{\partial}{\partial n}\left[(\beta - \alpha)np\right] + \frac{\partial^2}{\partial n^2}\left\{\tfrac{1}{2}[(\alpha + \beta)n + (\beta - \alpha)]p\right\} + \cdots.
$$

That is, neglecting the higher order terms (in the ellipses),

$$
\frac{\partial p}{\partial t} \approx \frac{\partial}{\partial n}\left[(\beta - \alpha)np\right] + \frac{\partial^2}{\partial n^2}\left\{\tfrac{1}{2}[(\alpha + \beta)n + (\beta - \alpha)]p\right\}. \tag{3.5}
$$

To confirm this form of the right-hand side, start with it and work backward, expanding the derivatives using the product rule, until you match all the earlier terms.

The PDE (3.5) has the form of a Fokker–Planck equation (3.2) with drift $\mu = (\alpha - \beta)n$ and diffusion

$$
D = \tfrac{1}{2}\sigma^2 = \tfrac{1}{2}[(\alpha + \beta)n + (\beta - \alpha)] \approx \tfrac{1}{2}(\alpha + \beta)n
$$

for a large population. Such a Fokker–Planck-like equation governs the PDF of some SDE. Thus we could equivalently model the population by realizations of the stochas-

tic differential equation $dN = \mu\,dt + \sigma\,dW$, namely

$$dN = (\alpha - \beta)N\,dt + \sqrt{(\alpha + \beta)N}\,dW. \tag{3.6}$$

The first term in this SDE, $(\alpha - \beta)N\,dt$, models Malthusian growth; whereas the second, $\sqrt{(\alpha + \beta)N}\,dW$, models the fluctuations typical of a large number of random events. Such stochastic models realistically describe natural fluctuations in populations that we see in the realizations of Figure 3.10. ∎

Suggested activity: Do at least Exercises 3.2 and 3.4(1).

Summary

The Fokker–Planck equation (3.2) empowers us to not only predict steady state distributions, but to also transform individual-based models into aggregate SDEs that incorporate realistic fluctuations. Modern cell biochemistry also finds such transformations useful (Higham 2008).

3.2   Stochastically solve deterministic differential equations

We now head back toward our earlier claim that the random walkers in a domain somehow solve Laplace's equation. The trick is to show that at any time, the average field u experienced by the random walkers at that time is always the value at the release point. For simplicity, we address only one spatial dimension. But for wide applicability we permit the "random walkers" to undergo a quite general stochastic process.

From this, a strong link is cemented between PDEs and SDEs. This link helps solve the Black–Scholes equation (2.6) for the value of options in finance.

Theorem 3.8 (Feynman–Kac formula). *Let* $u(t,x)$ *satisfy the* PDE

$$\frac{\partial u}{\partial t} + \mu(x)\frac{\partial u}{\partial x} + D(x)\frac{\partial^2 u}{\partial x^2} = 0, \tag{3.7}$$

where as usual $D(x) = \frac{1}{2}\sigma(x)^2$. *Let* $X(t)$ *be the ensemble of solutions to the* SDE $dX = \mu(X)\,dt + \sigma(X)\,dW$ *with initial condition* $X(s) = y$, *then the specific initial value* $u(s,y) = E[u(t,X(t))]$ *for all* t. *This identity is sometimes called the* Feynman–Kac formula.

Example 3.9. The function $u(t,x) = x^2 - t$ is one of the solutions to the diffusion PDE $\frac{\partial u}{\partial t} = \frac{1}{2}\frac{\partial^2 u}{\partial x^2}$. Show first that the Feynman–Kac formula holds from the initial point $(t,x) = (1,2)$, and then that it holds from a general initial point $(t,x) = (s,y)$.

   Solution: The diffusion PDE $\frac{\partial u}{\partial t} = \frac{1}{2}\frac{\partial^2 u}{\partial x^2}$ has zero drift, $\mu = 0$, and unit volatility, $\sigma = 1$, so the corresponding SDE is just $dX = dW$ with solution of a Wiener process starting from the point $(t,x) = (1,2)$, namely $X(t) = W(t-1) + 2$.

Now at any later time $t > 1$, the expectation on the right-hand side of the Feynman–Kac formula is

$$
\begin{aligned}
E\left[u(t,X(t))\right] &= E\left[\{W(t-1)+2\}^2 - t\right] \\
&= E\left[W(t-1)^2 + 4W(t-1) + 4 - t\right] \\
&= E\left[W(t-1)^2\right] + 4E\left[W(t-1)\right] + (4-t) \\
&= \operatorname{Var}\left[W(t-1)\right] + 4E\left[W(t-1)\right] + (4-t) \\
&= (t-1) + 4 \times 0 + (4-t) = 3,
\end{aligned}
$$

as, by definition, $W(t-1)$ is distributed $N(0, t-1)$. The left-hand side of the Feynman–Kac formula is $u(1,2) = 2^2 - 1 = 3$ in agreement.

Similar algebra applies for any initial point (s,y). The solution of the corresponding SDE starting from the point $(t,x) = (s,y)$ is $X(t) = W(t-s) + y$. At any later time $t > 1$, the expectation on the right-hand side of the Feynman–Kac formula is then

$$
\begin{aligned}
E\left[u(t,X(t))\right] &= E\left[\{W(t-s)+y\}^2 - t\right] \\
&= E\left[W(t-s)^2 + 2yW(t-s) + y^2 - t\right] \\
&= E\left[W(t-s)^2\right] + 2y\,E\left[W(t-s)\right] + (y^2 - t) \\
&= \operatorname{Var}\left[W(t-s)\right] + 2y\,E\left[W(t-s)\right] + (y^2 - t) \\
&= (t-s) + 4 \times 0 + (y^2 - t) = y^2 - s,
\end{aligned}
$$

as, by definition, $W(t-s)$ is distributed $N(0, t-s)$. The left-hand side of the Feynman–Kac formula is $u(s,y) = y^2 - s$ in agreement. ∎

Proof.   Let $p(t,x|s,y)$ be the PDF for the general SDE of Theorem 3.8. It must satisfy the corresponding Fokker–Planck equation:

$$
\frac{\partial p}{\partial t} = -\frac{\partial}{\partial x}[\mu p] + \frac{\partial^2}{\partial x^2}[Dp].
$$

Now consider the expected value of u over all realizations released from $X = y$ at time s:

$$
E\left[u(t,X(t))\right] = \int u(t,x)p(t,x|s,y)\,dx,
$$

where the implicit limits of integration are over all x from $-\infty$ to $+\infty$, both here and below. We show that this expectation does not change in time:

$$
\begin{aligned}
\frac{\partial}{\partial t} E\left[u(t,X(t))\right] &= \frac{\partial}{\partial t} \int u(t,x)p(t,x|s,y)\,dx \\
&= \int \frac{\partial u}{\partial t}p + u\frac{\partial p}{\partial t}\,dx \\
&= \int \frac{\partial u}{\partial t}p - u\frac{\partial}{\partial x}[\mu p] + u\frac{\partial^2}{\partial x^2}[Dp]\,dx \quad \text{by Fokker–Planck,}
\end{aligned}
$$

then integrate the last two terms by parts to

$$= \underbrace{\left[-u\mu p + u\frac{\partial}{\partial x}[Dp]\right]}_{= 0 \text{ since } p \to 0 \text{ as } x \to \pm\infty} + \int \frac{\partial u}{\partial t}p + \mu p\frac{\partial u}{\partial x} - \frac{\partial u}{\partial x}\frac{\partial}{\partial x}[Dp]\,dx$$

and integrate the last term by parts again

$$= \underbrace{\left[-\frac{\partial u}{\partial x}Dp\right]}_{= 0 \text{ since } p \to 0 \text{ as } x \to \pm\infty} + \int \frac{\partial u}{\partial t}p + \mu p\frac{\partial u}{\partial x} + \frac{\partial^2 u}{\partial x^2}Dp\,dx$$

$$= \int \underbrace{\left(\frac{\partial u}{\partial t} + \mu\frac{\partial u}{\partial x} + D\frac{\partial^2 u}{\partial x^2}\right)}_{= 0 \text{ by the given PDE (3.7)}} p\,dx$$

$$= 0.$$

That is, the expectation $E[u(t,X(t))]$ is constant in time t; it must be always the same as its initial value of $E[u(s,X(s))] = E[u(s,y)] = u(s,y)$. □

Example 3.10 (solve the Black–Scholes equation stochastically). Recall that the Black–Scholes equation (2.6) for the value of a call option is

$$\frac{\partial C}{\partial t} + rS\frac{\partial C}{\partial S} + \frac{1}{2}\beta^2 S^2\frac{\partial^2 C}{\partial S^2} = rC,$$

where the bond rate is r and the asset stock volatility is β. This Black–Scholes equation is not in the form of the PDE (3.7) because of the source terms on the right-hand side. We change variables to x and $u(t,x)$ such that $C(t,S) = u(t,x)e^{rt}$ and $x = S$;[33] then $u(t,x) = C(t,S)e^{-rt}$ has the meaning of the value of the call option discounted by the bond rate to the current time $t = 0$. With this change, since $\frac{\partial C}{\partial t} = \frac{\partial u}{\partial t}e^{rt} + rC$, it is straightforward to see that the discounted value $u(t,x)$ satisfies the PDE

$$\frac{\partial u}{\partial t} + rx\frac{\partial u}{\partial x} + \frac{1}{2}\beta^2 x^2\frac{\partial^2 u}{\partial x^2} = 0,$$

which is in the requisite form (3.7).

As in the example shown in Figure 3.11, suppose at time zero we release stochastic particles, with path $x = Y(t)$, to evolve according to the corresponding SDE[34]

$$dY = rY\,dt + \beta Y\,dW \quad \text{with} \quad Y(0) = S_0, \tag{3.8}$$

where S_0 is the current asset price. Then according to Theorem 3.8 the current discounted value of the option is $u(0,S_0) = E[u(t,Y(t))]$ for all time t. In particular, $u(0,S_0) =$

---

[33]We unnecessarily change names from asset value S to abstract x only to match closely with the PDE (3.7).

[34]We use Y here only because X denotes the strike price for an option.

$E[u(t,Y(t))]$ will hold up to the expiry time, say $t = T$, of the option. Upon expiry we know the value of the option, namely $C(T,S) = \max\{S - X, 0\}$, where S is the asset price, denoted by Y for the solutions of the SDE. Hence the discounted value of each of the stochastic particles is

$$u(T,Y(T)) = e^{-rT}\max\{Y(T) - X, 0\}.$$

Thus using $u(0, S_0) = E[u(t, Y(t))]$, the current value of the option is

$$C(0, S_0) = u(0, S_0) = e^{-rT}E[\max\{Y(T) - X, 0\}].$$

Algorithm 3.1 lists this stochastic solution of the Black–Scholes equation. Figure 3.11 plots 10 realizations of $Y(t)$ —there is nothing particularly unusual to observe. Averaging over $m = 1000$ realizations, the estimated value of the call option varies over a range of approximately 3.3–3.5. The exact value for the option of \$3.40 is comfortably within this range.

Algorithm 3.1 Example MATLAB/SCILAB code using (3.8) to stochastically estimate the value of a call option, from Example 1.9, on an asset with initial price $S_0 = 35$, strike price $X = 38.50$ after one year in which the asset fluctuates by a factor of 1.25 and with a bond rate of 12%.

```
r=log(1.12)
b=log(1.25)
x=38.50
s0=35
m=1000;
n=1000;
t=linspace(0,1,n+1);
h=diff(t(1:2));
y=s0*ones(1,m);
for j=1:n
    y=y+r*y*h+b*y.*randn(1,m)*sqrt(h);
end
estimated_value=exp(-r)*mean(max(y-x,0))
```

To estimate this value to the nearest few cents we need to increase the accuracy of the stochastic estimate by a factor of 10. Such accuracy could only be obtained by using 100 times as many realizations, that is, $m \approx 100,000$ realizations! This is rather too many realizations for a practical method when the alternative of solving the Black–Scholes equation (2.6) is so quick.[35] ■

The previous example reinforces the connection between PDEs and SDEs. Recall that at the start of this chapter we claimed that random walkers could solve Laplace's equation by averaging over the values observed by the walkers when they first contacted the boundary. We now use the above theory to show that a similar technique will work quite

---

[35]This stochastic solution also opens up the possibility of using noise processes different from the Wiener process to estimate the value of an option. However, our developed theory does not yet justify such use!
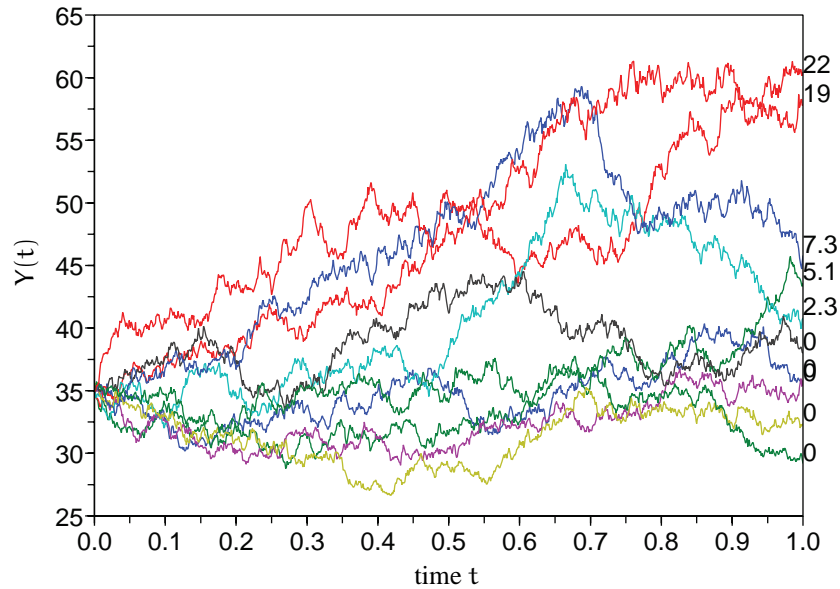
Figure 3.11. *Ten example realizations* $Y(t)$ *of the* SDE *to stochastically solve the Black–Scholes equation: the value of the option at the expiry time* $t = 1$ *is shown at the right. Average such values to estimate the value of the option at the initial time* $t = 0$.

generally. For example, Figure 3.12 shows ten realizations $X(t)$ (the analogue of the random walkers) initiated by being released from $x = 2$ and then stochastically walking until they hit the boundaries at $x = 0$ or $x = 3$; the value of some field u at $x = 2$ is then the appropriately weighted average of these two boundary values. Here we restrict our attention to solutions of boundary value problems of ODEs.

Theorem 3.11. *Let* $u(x)$ *in some domain* $a < x < b$ *satisfy the* ODE

$$\mu(x)\frac{\partial u}{\partial x} + D(x)\frac{\partial^2 u}{\partial x^2} = 0, \tag{3.9}$$

subject to the boundary conditions $u(a) = f(a)$ *and* $u(b) = f(b)$, *where as usual* $D(x) = \frac{1}{2}\sigma(x)^2$. *Let* $X(t)$ *be the ensemble of solutions to the* SDE $dX = \mu(X)\,dt + \sigma(X)\,dW$ *with initial condition* $X(0) = y$, *and let* τ *be the* first *exit time from* $a < X < b$ *of each realization. Then* u *at the release point is the average value of the realizations that have "stuck" to one or other of the boundaries:*[36]

$$u(y) = E[f(X(\tau))]. \tag{3.10}$$

---

[36]One may also obtain the solution of the forced ODE $\mu u_x + D u_x x = g(x)$: the formula $u(y) = E[f(X(\tau))] - E\left[\int_0^\tau g(X(t))\,dt\right]$.
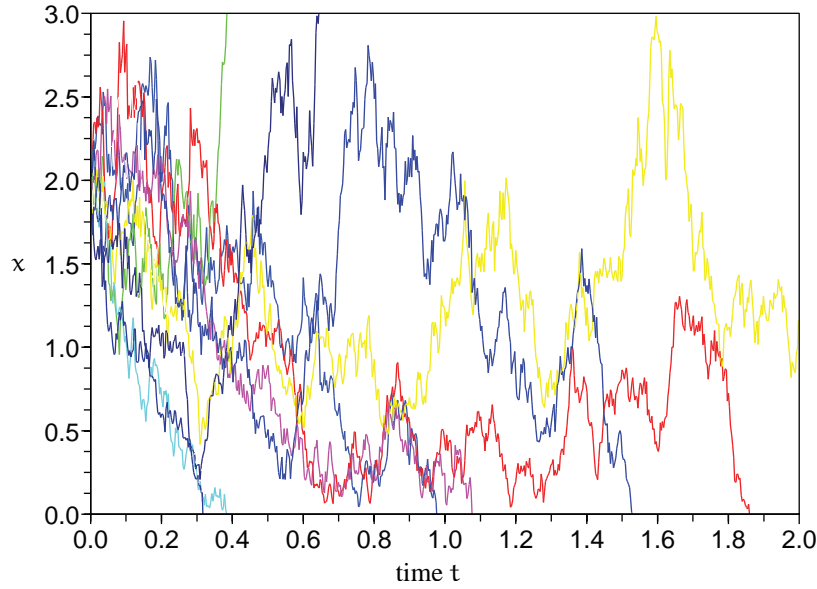
Figure 3.12. *Ten realizations* $X(t)$ *initiated from* $x = 2$, *then stochastically evolving until they hit the boundaries at* $x = 0$ *or* $x = 3$; *the value of some field* u *at position* $x = 2$ *is then a weighted average of its two boundary values.*

Proof.   A little loosely, require that the drift and volatility of the SDE are as required internally in the domain $a < x < b$, but reduce to zero outside and in particular at the end points $x = a$ and b. Then the correspondence between the ODE and the SDE is maintained throughout the domain. The only difference is that the solutions of the SDE "stick" to the boundary as soon as they reach it, as there the drift and volatility are both zero. Now let T denote a time so large that almost all solutions of the SDE have reached a boundary. Theorem 3.8, restricted to time-independent differential equations, then shows that the solution at the release point is

$$
\begin{aligned}
u(y) &= E\left[u(X(T))\right] \\
&= E\left[u(X(\tau))\right] \quad \text{as } X(T) = X(\tau) \text{ for each realization} \\
&= E\left[f(X(\tau))\right],
\end{aligned}
$$

as $u = f$ on each of the boundaries. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ □

Example 3.12.   Consider the ODE $-x\frac{du}{dx} + \frac{1}{2}(1 + x^2)\frac{d^2 u}{dx^2} = 0$ with boundary conditions $u(0) = 1$ and $u(3) = 5$. Solve it analytically, and also numerically estimate $u(1)$ and $u(2)$ via its corresponding SDE.

Solution: Analytically, when written in terms of $v = du/dx$, the ODE becomes separable:

$$-xv + \tfrac{1}{2}(1+x^2)\frac{dv}{dx} = 0$$

$$\Rightarrow \tfrac{1}{2}(1+x^2)\frac{dv}{dx} = xv$$

$$\Rightarrow \int \frac{dv}{v} = \int \frac{2x}{1+x^2}\,dx$$

$$\Rightarrow \log v = \log(1+x^2) + \text{constant}$$

$$\Rightarrow v = A(1+x^2)$$

$$\Rightarrow u = \int v\,dx = A\left(x + \tfrac{1}{3}x^3\right) + B\,.$$

With boundary conditions $u(0) = 1$ and $u(3) = 5$, determine $A = \tfrac{1}{3}$ and $B = 1$ so the analytic solution is

$$u = \tfrac{1}{3}\left(x + \tfrac{1}{3}x^3\right) + 1\,.$$

Thus we now know that $u(1) = 13/9 = 1.4444$ and $u(2) = 23/9 = 2.5556$.

Algorithm 3.2 MATLAB/SCILAB code to solve a boundary value problem ODE by its corresponding SDE. Use `find` to evolve only those realizations within the domain $0 < X < 3$. Continue until all realizations reach one boundary or the other. Estimate the expectation of the boundary values using the conditional vectors `x<=0` and `x>=3` to account for the number of realizations reaching each boundary.

```
m=100; % realizations
h=0.001; % time step
x=ones(1,m); % initial release
i=1:m;
while m>0
   x(i)=x(i)-x(i)*h+sqrt(1+x(i).^2).*randn(1,m)*sqrt(h);
   i=find((x>0)&(x<3)); % track those as yet unstuck
   m=length(i);
end
estimated_u=mean(1*(x<=0)+5*(x>=3))
```

Stochastically, recognize that the ODE comes from an Ito process with drift $\mu = -x$ and volatility $\sigma = \sqrt{1+x^2}$. To estimate $u(1)$, adapt Algorithm 3.1 to that of Algorithm 3.2: this version numerically solves the SDE $dX = -X\,dt + \sqrt{1+X^2}\,dW$ with initial condition $X(0) = 1$. Figure 3.12 plots ten realizations. With just $m = 100$ realizations we get significant fluctuations in the estimates of $u(1)$ about the true value, for example, 1.72, 1.28, 1.36, 1.32, and 1.40. Similar fluctuations are obtained for estimates of $u(2)$ such as 2.44, 2.48, 2.36, 2.56, and 2.20. Using many more realisations gives more accuracy. But at least we roughly approximate the answer quite quickly with relatively few realizations.

The binomial distribution estimates errors

In this example, as is typical, each of the m realizations reached either one boundary or the other, $x = 0$ or $x = 3$. Let $Y_i = 1$ if the ith realization reaches the $x = 3$ boundary, and conversely $Y_i = 0$ if the ith realization reaches the $x = 0$ boundary. Here, as in any problem, there is a probability p that the ith realization reaches the right-hand boundary for which $Y_i = 1$; here the estimate $\hat{p} \approx 0.4$. Now consider the value of the solution at the release point $x = 2$: $u(2) = E[f(X(\tau))]$. The boundary values $f(X(\tau))$ are either 1 or 5 depending upon whether the realization reached the left or the right boundary, respectively. That is, $f(X(\tau)) = 4Y_i + 1$. Consequently, $u(2) = E[f(X(\tau))] = E[4Y_i + 1]$, and hence

$$\hat{u}(2) = \frac{1}{m} \sum_{i=1}^{m} (4Y_i + 1) = \frac{4}{m} \sum_{i=1}^{m} Y_i + 1 = \frac{4}{m} \mathcal{Y} + 1,$$

where $\mathcal{Y} = \sum_{i=1}^{m} Y_i$ is distributed $\mathrm{bin}(m, p)$; here $\mathcal{Y} \sim \mathrm{bin}(m, 0.4)$.

- Now we know $E[\mathcal{Y}] = mp$ from the properties of the binomial distribution. Hence, $E[u(2)] = E[\frac{4}{m} \mathcal{Y} + 1] = \frac{4}{m} mp + 1 = 4p + 1$ which, with $\hat{p} = 0.4$, estimates $\hat{u}(2) \approx 2.6$ as before. The expectation of the binomial agrees with the earlier mean.

- Additionally we also know the variance of a binomial distribution: $\mathrm{Var}[\mathcal{Y}] = mp(1 - p)$. Thus we know $\mathrm{Var}[u(2)] = \mathrm{Var}[\frac{4}{m} \mathcal{Y} + 1] = \frac{16}{m^2} \mathrm{Var}[\mathcal{Y}] = \frac{16}{m^2} mp(1 - p) = 16p(1 - p)/m$ which, with $\hat{p} = 0.4$, estimates $\mathrm{Var}[\hat{u}(2)] \approx 0.04$. This provides the error in our estimate of $u(2)$.

Thus with $m = 100$ realizations we estimate $u(2) = 2.6 \pm 0.2$, a 10% error. Because the variance decreases with m, the error decreases with $\sqrt{m}$, and so we need $m \approx 10{,}000$ realizations to estimate $u(2)$ to a 1% error. ∎

    As these last two examples show, we need enormously many realizations to solve a PDE at all accurately via its SDE. It is not a practical method for PDEs in small numbers of dimensions. However, suppose we wish to solve a PDE in, say, 100 dimensions, as easily arises in finance when each day introduces another dimension. Then a finite difference solution would require a fantastic 10^{100} grid points to just resolve the domain with 10 grid points in each of the 100 dimensions, let alone try to solve the equation. In contrast, the stochastic solution is practical because we just solve 100 coupled SDEs for, say, a few thousand realizations to obtain a workable approximation. The SDE solution of PDEs is practical in problems of high dimension.

Suggested activity: Do at least Exercise 3.12.

Summary

For a given differential equation for unknown u, the *appropriate* SDE has realizations whose expectation of u does not change. Hence we are empowered to stochastically estimate the unknown u at some initial point. This provides a method to stochastically value a financial option.

3.3   The Kolmogorov backward equation completes the picture

To date we either sought the steady state probability distribution, the PDF $p(x)$, or we assumed that the initial state of the Ito process was specified. By investigating the evolution from a general initial state, we discover a fully informative function that is analogous to powers of the transition matrix of Markov chains.

Example 3.13 (Wiener process).   The Wiener process in Example 3.6 starts from $W(0) = 0$, that is, all the probability is lumped at $w = 0$ at time $t = 0$. More generally, we might know that at some time s the process has some value y: $W(s) = y$, or equivalently, all the probability is lumped at $x = y$ at time $t = s$. Because of the homogeneous nature of the Wiener process, we write down the more general PDF just by shifting the origin in time t and value w. Thus the PDF for the Wiener process, given that we know it was at y at time s, is

$$p(t, x | s, y) = \frac{1}{\sqrt{2\pi(t-s)}} \exp\left[-(x-y)^2 / 2(t-s) \right].$$

The use of the vertical bar in $p(t, x | s, y)$ is a convention to remind us that this function p is the PDF conditional on being at y at time s.   ∎

Example 3.14 (Ornstein–Uhlenbeck process).   Remarkably, the conditional PDF for an Ornstien–Uhlenbeck process is also always a Gaussian: the Gaussian's parameters just have exponentially decaying transients. Recall that Figure 3.5 plots 10 realizations showing how the realizations evolve; Figure 3.6 shows how the corresponding PDF evolves toward a Gaussian steady state. For simplicity, I record the conditional PDF for the specific SDE $dX = -X\,dt + \sqrt{2}\,dW$: it is

$$p(t, x | s, y) = \frac{1}{\sqrt{2\pi\left(1 - e^{-2(t-s)}\right)}} \exp\left[-\frac{\left(x - y e^{-(t-s)}\right)^2}{2\left(1 - e^{-2(t-s)}\right)} \right].$$

This probability distribution is that for a normal random variable with mean $x = y e^{-(t-s)}$ decaying from y at time $t = s$ to 0 exponentially quickly, and with variance $1 - e^{-2(t-s)}$ growing from 0 at time $t = s$ to saturate exponentially quickly at 1 for large times. The appearance of the steady state PDF matches the steady state found in Example 3.5.

Recall that Exercise 1.6, using many realizations of numerical simulations, also shows the same exponential approach to the steady state distribution from a variety of initial conditions.   ∎

This general conditional PDF has much more information about the underlying stochastic process—it is the analogue of powers of the transition matrix of a Markov chain. This is in contrast to $p(t, x)$ which, for example, cannot distinguish between a Wiener process and $Z\sqrt{t}$! Here $p(t, x | s, y)$ "knows" about the process when started with any value y at any time s. Thus, for example, at a small time step h into the future from time s, $p(s+h, x | s, y)$ will be some sharply peaked probability distribution: the peak will be at some $\bar{x}$ near y, and so $\mu = (\bar{x} - y)/h$ is an estimate of the drift at the locale (s, y); whereas the spread
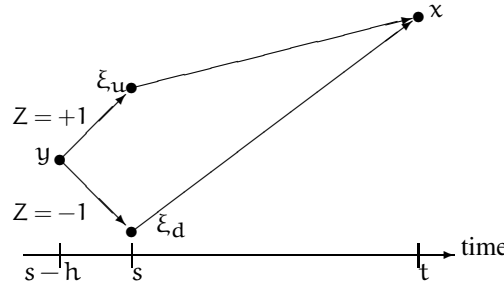
Figure 3.13. *To derive the Kolmogorov backward equation* (3.11)*, take a small time step from* $(s-h, y)$ *to time* s*, and see how it affects the subsequent evolution to* (t, x)*.*

of the peak, its standard deviation σ_x, will determine the local volatility $\sigma = \sigma_x/\sqrt{h}$. By uniqueness of drift and volatility of an Ito process, the Doob–Meyer decomposition, we can in principle determine the SDE for any given $p(t, x|s, y)$. *The conditional* PDF *is fully informative of the underlying stochastic process.*

Analogue

The conditional probability distribution $p(t, x|s, y)$ is analogous to powers P^n of the transition matrix P in discrete stochastic systems. Recall that P^n takes the probability distribution, $\mathbf{p}(s)$, at some time s and maps it to the later probability distribution $\mathbf{p}(t) = \mathbf{p}(s)P^n$ at time $t = s + n$. With $t - s \sim n$, $x \sim i$, $y \sim j$, $p(t, x|s, y)$ is exactly analogous to the i, jth element of P^n.

Theorem 3.15 (Kolmogorov backward equation). *The general* PDF $p(t, x|s, y)$ *for the Ito process* $dX = \mu(X)dt + \sigma(X)dW$ *satisfies the* Kolmogorov backward equation,

$$\frac{\partial p}{\partial s} + \mu(y)\frac{\partial p}{\partial y} + \tfrac{1}{2}\sigma(y)^2\frac{\partial^2 p}{\partial y^2} = 0, \tag{3.11}$$

as well as the Fokker–Planck equation (3.2).[37]

It is straightforward, albeit tedious, to verify that the example PDFs given earlier satisfy (3.11). See that the Kolmogorov backwards equation is like an advection-diffusion equation but with negative diffusivity: this implies it must be solved backward from t to the starting time s. This reminds us of the Black–Scholes equation (2.6) for pricing options, and it helped establish in Example 3.10 the striking connection with the Black–Scholes equation.

Proof. Throughout this derivation, dashes denote $\partial/\partial y$ and the drift μ and the volatility σ are evaluated at y unless otherwise specified. Figure 3.13 indicates how we investigate the

---

[37]The Kolmogorov backward equation and the Fokker–Planck equation are closely related. Each is the *adjoint* of the other. Those who have met the definition, properties, and use of adjoints will appreciate that such a pair of adjoint equations and their solutions provides two complementary and beautiful alternative views of problems. The Kolmogorov backward equation is sometimes also known as the Chapman–Kolmogorov equation.

probability that the system arrives at (t,x) given that it starts from $(s-h,y)$ by taking the first small time step to (s,ξ) and combining this with $p(t,x|s,\xi)$ to give the probability of ultimately reaching (t,x). As for the proof of the Fokker–Planck equation (3.2), for simplicity assume that the Wiener increment over the time step h is $\Delta W = \sigma(y)\sqrt{h}Z$, where $Z = \pm 1$ each with probability $\frac{1}{2}$; thus the x value after the first small time step from $(s-h,y)$ is (Figure 3.13)

$$\xi = y + \mu(y)h \pm \sigma(y)\sqrt{h}.$$

Therefore the routes to (t,x) are either via ξ_u or ξ_d:

$$
\begin{aligned}
p(t,x|s-h,y) &= \tfrac{1}{2}p(t,x|s,\xi_u) + \tfrac{1}{2}p(t,x|s,\xi_d) \\
&= \tfrac{1}{2}p(t,x|s,y+\mu h+\sigma\sqrt{h}) + \tfrac{1}{2}p(t,x|s,y+\mu h-\sigma\sqrt{h}) \\
&= \tfrac{1}{2}\left[p + (\mu h+\sigma\sqrt{h})p' + \tfrac{1}{2}(\mu h+\sigma\sqrt{h})^2 p''\right] \\
&\quad + \tfrac{1}{2}\left[p + (\mu h-\sigma\sqrt{h})p' + \tfrac{1}{2}(\mu h-\sigma\sqrt{h})^2 p''\right] + \cdots \\
&= p + \mu h p' + \tfrac{1}{2}\sigma^2 h p'' + \cdots.
\end{aligned}
$$

Subtracting $p(t,x|s-h,y)$ from both sides and dividing by h leads to

$$0 = \frac{p(t,x|s,y) - p(t,x|s-h,y)}{h} + \mu p' + \tfrac{1}{2}\sigma^2 p'' + \cdots.$$

Take the limit as $h \to 0$ to deduce

$$0 = \frac{\partial p}{\partial s} + \mu\frac{\partial p}{\partial y} + \tfrac{1}{2}\sigma^2\frac{\partial^2 p}{\partial y^2}.$$

This is the Kolmogorov backward equation (3.11).   □

Appendix B.2 gives an alternate derivation.

Summary

The conditional PDF $p(t,x|s,y)$ can completely characterize a stochastic process: as a function of the initial condition, $X(s) = y$, the conditional PDF $p(t,x|s,y)$ satisfies the Kolmogorov backward equation (3.11).

3.4   Summary

- The PDF $p(t,x|s,y)$ gives the probability that a stochastic process has value x at time t given that it had value y at an earlier time s.

- The PDF of the SDE $dX = \mu(X)dt + \sigma(X)dW$ satisfies both the Fokker–Planck equation (3.2),

$$\frac{\partial p}{\partial t} = -\frac{\partial}{\partial x}[\mu p] + \frac{\partial^2}{\partial x^2}\left[\tfrac{1}{2}\sigma^2 p\right],$$

and the Kolmogorov backward equation (3.11),

$$\frac{\partial p}{\partial s} + \mu(y)\frac{\partial p}{\partial y} + \tfrac{1}{2}\sigma(y)^2\frac{\partial^2 p}{\partial y^2} = 0.$$

- The PDFs of SDEs may be determined from the corresponding Fokker–Planck equation.

- The Kolmogorov backward equation enables some PDEs to be solved stochastically.

Exercises

3.1. Show that $\mathrm{Var}[X(t)] = X_0^2\big[e^{(2\alpha+\beta^2)t} - e^{2\alpha t}\big]$ for $X(t) = X_0\exp[(\alpha - \tfrac{1}{2}\beta^2)t + \beta W(t)]$. Give an example showing that even when the expectation decays to zero, the variance may nonetheless grow in time.

3.2. Consider in turn each of the stochastic processes illustrated in Figures 3.14–3.17. Each figure shows ten realizations of some Ito process, a different Ito process for each figure. Assuming each Ito process is settling upon some steady state PDF $p(x)$; sketch the PDF.

3.3. The realizations plotted in Exercise 3.2 all come from Ito processes you have met or will soon meet. Identify as best you can which Ito process shown in Exercise 3.4 corresponds to each of the four figures.

3.4. Use the Fokker–Planck equation to investigate the steady state PDFs for the following stochastic processes:

    1. $dX = (2X - X^2)dt + X\,dW$ for $X(t) \geq 0$;

    2. $dX = (3X - 2X^2)dt + 2X\,dW$ for $X(t) \geq 0$;

    3. $dX = \beta\sqrt{1+X^2}\,dW$;

    4. $dX = -X\,dt + \sqrt{1+X^2}\,dW$;

    5. $dX = X\,dt + (1+X^2)\,dW$;

    6. $dX = -X\,dt + \sqrt{2}(1+X^2)^{1/4}dW$.

3.5. Use the Fokker–Planck equation to find the structure of the steady state PDF of solutions $X(t)$ to the SDE $dX = -2X\,dt + \sqrt{1+X^2}\,dW$.

3.6. Consider a general SDE $dX = \mu(X)dt + \sigma(X)dW$. Assuming a steady state PDF exists for the ensemble of realizations, show that the steady state PDF may be written as

$$p(x) = \frac{A}{D(x)}\exp\left[\int\frac{\mu(x)}{D(x)}dx\right],$$

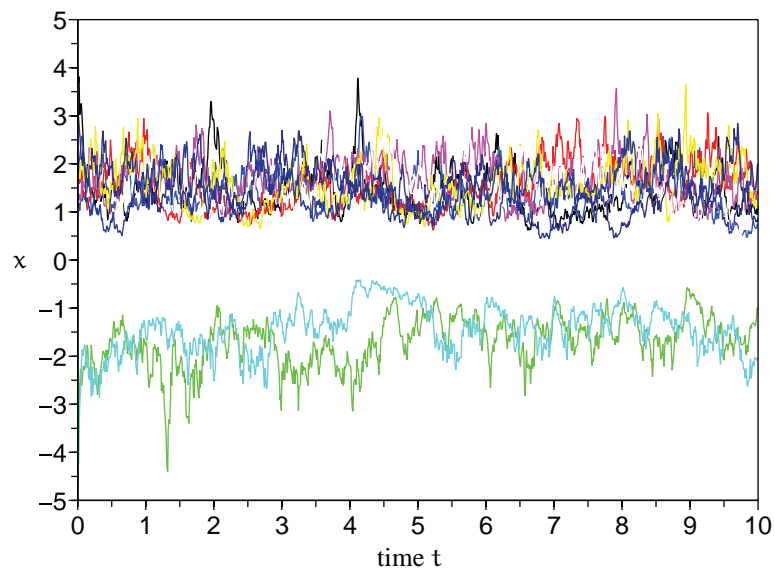where $D(x) = \tfrac{1}{2}\sigma(x)^2$ and A is a normalization constant.

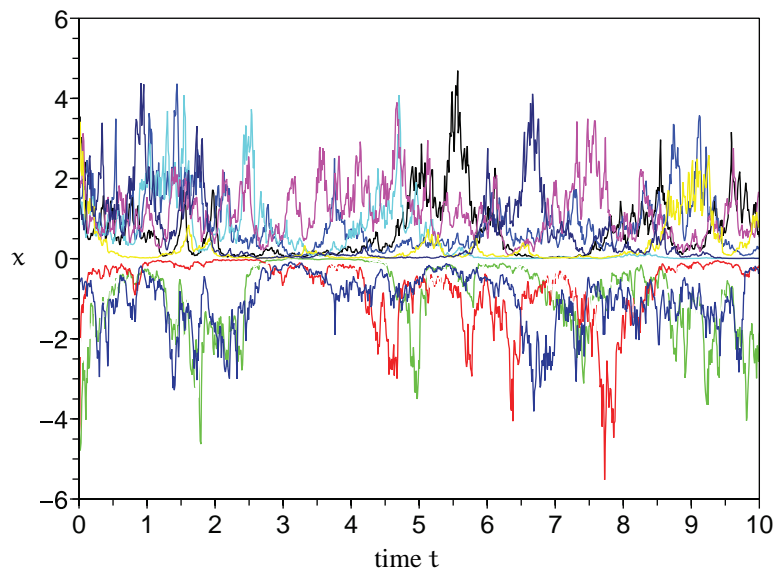**Figure 3.14.** *Ten realizations of an Ito process.*

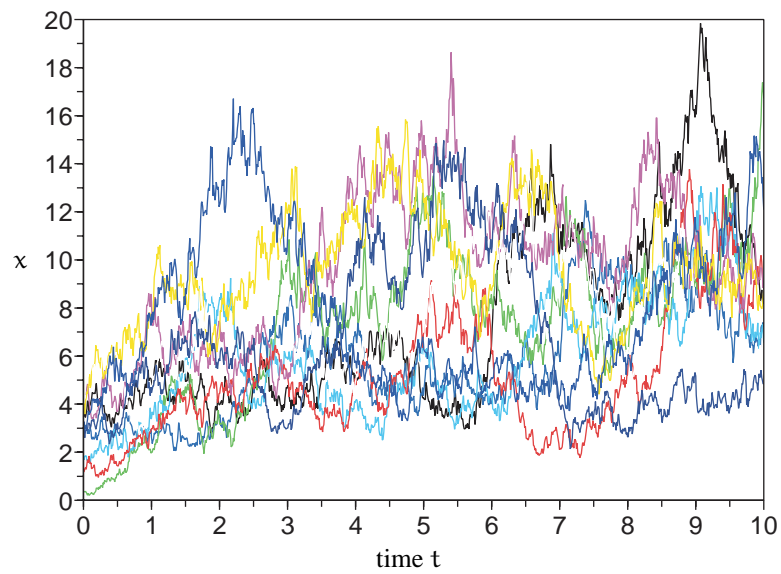**Figure 3.15.** *Ten realizations of an Ito process.*

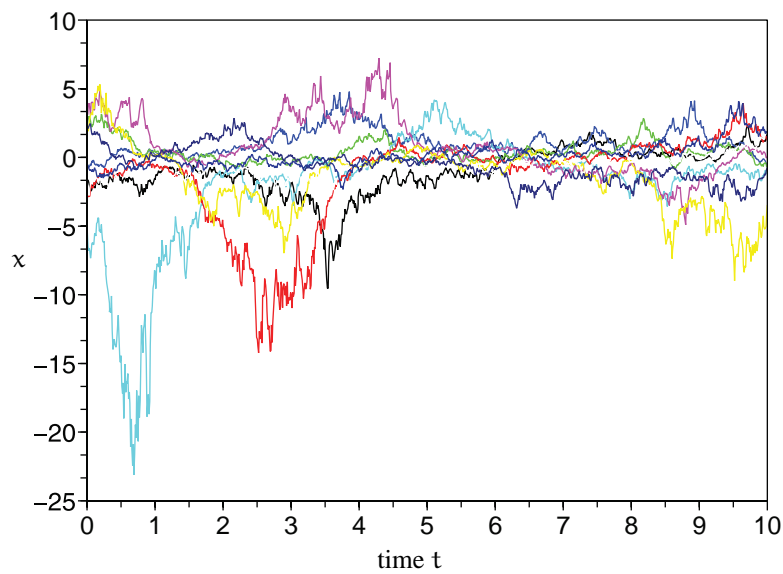**Figure 3.16.** *Ten realizations of an Ito process.*

Figure 3.17. *Ten realizations of an Ito process.*

3.7. Derive from first principles the Fokker–Planck equation

$$\frac{\partial p}{\partial t} = \frac{\partial^2}{\partial x^2}(\tfrac{1}{2}x^2 p)$$

for the PDF $p(t,x)$ of solutions $X(t)$ of the SDE

$$dX = X\,dW$$

(this SDE is the case of the financial SDE $dX = \alpha X\,dt + \beta X\,dW$ for no stock drift, $\alpha = 0$, and unit stock volatility, $\beta = 1$).

- First, assume small time steps of size $\Delta t = h$ and approximate the Wiener process by the up/down binomial steps $\Delta W = \pm\sqrt{h}$. Show that for the stochastic system to reach $(t+h,x)$ it must have come from (t,ξ) for $\xi = x/(1\pm\sqrt{h})$.
- Second, use Taylor series about $p(t,x)$ to deduce

$$p(t+h,x) = p + hp + 2xh\frac{\partial p}{\partial x} + \tfrac{1}{2}x^2 h\frac{\partial^2 p}{\partial x^2} + \cdots,$$

  where the right-hand side is evaluated at (t,x). You may use that $(1\pm\sqrt{h})^{-1} = 1\mp\sqrt{h}+h+\cdots$.
- Lastly, rearrange and take the limit as the time step $h \to 0$ to derive the corresponding Fokker–Planck equation.

3.8. Generalize the proof of the Fokker–Planck equation (3.2) to include variable volatility $\sigma(x)$. This generalization is quite difficult because of the need to keep track of all the details, and also you will need to show

$$\xi = x \mp \sigma\sqrt{h} + (\sigma\sigma' - \mu)h + \cdots.$$

3.9. Reanalyze Example 3.7, but suppose instead that the death rate $\mu_n = \beta n^2$ due to increased competition for limited resources by larger populations of individuals. Use its Fokker–Planck equation to then argue that the corresponding SDE is *approximately*

$$dN = (\alpha N - \beta N^2)dt + \sqrt{\alpha N + \beta N^2}\,dW.$$

3.10. Consider the PDE $\frac{\partial u}{\partial t} + \frac{\partial u}{\partial x} + \frac{1}{2}\frac{\partial^2 u}{\partial x^2} = 0$. One solution is $u(t,x) = x^2 - 2xt + t^2 - t$. Write down the corresponding SDE, and its solution $X(t)$, and then verify the Feynman–Kac formula for solutions with initial condition $X(s) = y$.

3.11. Corresponding to the Black–Scholes equation is the SDE (3.8). Use many realizations of its solution to numerically estimate the options of Exercises 1.11, 1.12, and 1.19. Roughly what errors do you expect in your estimates?

3.12. Solve analytically $2\frac{du}{dx} + x\frac{d^2 u}{dx^2} = 0$ on $1 < x < 5$ such that $u(1) = 4$ and $u(5) = 0$. Use the corresponding SDE to estimate $u(2)$, $u(3)$ and $u(4)$ and their errors. Compare. Say we use $m = 100$ realizations with a time step of $h = 0.001$: discuss how the errors improve with increasing m and/or decreasing h.

3.13. Verify, perhaps using computer algebra, that the PDFs for the Wiener and Ornstein–Uhlenbeck processes satisfy the Kolmogorov backward equation (3.11).

Answers to selected exercises

3.4.   1. $p \propto x^2 e^{-2x}$ (you determine the constant of proportionality in this and other answers); see the hump in probability near $x = 2$ corresponding to the deterministic fixed point;

2. $p \propto e^{-x}/\sqrt{x}$, but here the deterministic fixed point at $x = 3/2$ has been washed out by the noise which instead has stabilized the origin and hence generated an integrable peak in probability at $x = 0$;

3. $p \propto 1/(1 + x^2)$; see that for small x the system spreads by a random walk, $dX \propto dW$, but this spreading is arrested by the stabilizing effect of multiplicative noise for large x, where $dX \propto X \, dW$, generating this "Cauchy distribution" which has long tails (curiously β has no influence on the steady state distribution);

4. $p \propto 1/(1 + x^2)^2$ shows that by stabilizing the origin the previous distribution has much smaller tails;

5. $p \propto \frac{1}{(1+x^2)^2} e^{-1/(1+x^2)}$, whereas here the exponential factor in the solution does not affect the distribution very much so the PDF is much like the previous one but generated by a deterministically unstable origin which is stabilized by noise that grows quadratically with x;

6. $p \propto \frac{1}{\sqrt{1+x^2}} e^{-\sqrt{1+x^2}}$; for small x this process is like an Ornstein–Uhlenbeck process, $dX \approx -X \, dt + \sqrt{2} \, dW$, but the PDF decays a little faster for large x because the noise increases like $\sqrt{X}$.

3.5. $p \propto 1/(1 + x^2)^3$.

3.12. Analytic solution is $u = 5/x - 1$.

Chapter 4

Stochastic Integration Proves
Ito's Formula

Contents

So far we have played with stochastic differentials, such as

$$dX = \mu(t,X)dt + \sigma(t,x)dW, \tag{4.1}$$

and invoked rules for their manipulation that seem to make sense in the context of stochastic Ito processes. But the derivation of the symbolic rules has been largely intuitive rather than rigorous. In this chapter we put stochastic calculus on a firm footing. *The best treatment of stochastic differential forms such as* (4.1) *asserts that they are just shorthand for integrals such as*

$$[X(t)]_a^b = X(b) - X(a) = \int_a^b \mu(t,X(t))dt + \int_a^b \sigma(t,X(t))dW.$$

In turn, interpret these integrals as being approximated by the sums

$$\int_a^b \mu(t,X(t))dt \approx \sum_{j=0}^{n-1} \mu_j \Delta t_j \quad \text{and} \quad \int_a^b \sigma(t,X(t))dW \approx \sum_{j=0}^{n-1} \sigma_j \Delta W_j.$$

But what do these integrals really mean? Can we really approximate them like this? What properties do they have? How are they related to ordinary integrals?

Ordinary integrals

Interpret integrals such as $\int_a^b \mu(t,X(t))dt$ as ordinary integrals in the sense of ordinary calculus (they are Riemann–Lebesgue integrals). All the properties that we are familiar
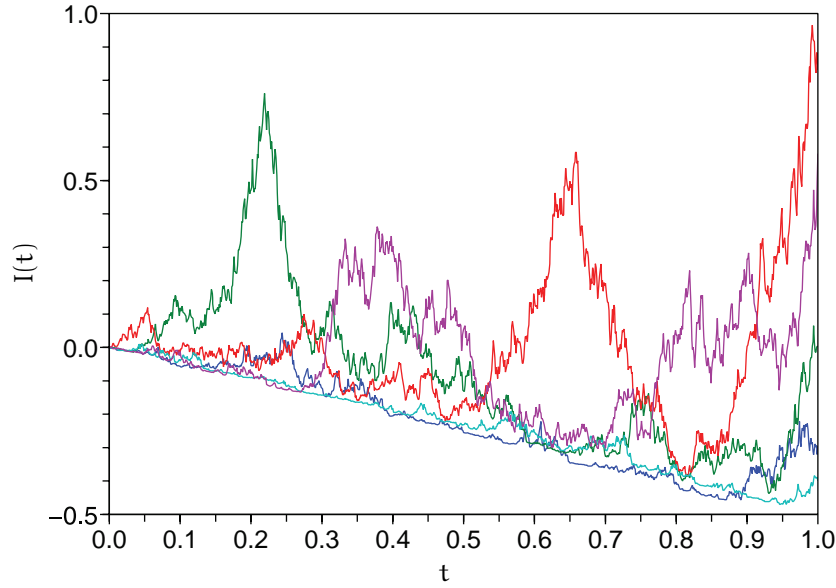
Figure 4.1. *Five realizations of the Ito integral* $I(t) = \int_0^t W(s)\,dW(s)$ *for* $0 < t < 1$ *as computed by Algorithm* 4.1.

with follow provided the integrand is a smooth function of any Ito process. Thus for these integrals there is nothing that we really need to develop other than perhaps efficient methods for their numerical evaluation. In this chapter we focus exclusively upon Ito integrals of the form $\int_a^b \sigma(t, X(t))\,dW$.

Example 4.1. Use Ito's formula to deduce $I = \int_a^b W\,dW$. See in the numerical evaluations of Figure 4.1 that the integral is *not* $[\frac{1}{2}W^2]_a^b$ because the realizations of $I(t) = \int_0^t W(s)\,dW(s)$ are often negative, whereas $\frac{1}{2}W^2$, being a square, is always positive.

Solution: The indefinite stochastic integral $I = \int W\,dW$ is synonymous with the SDE $dI = W\,dW$. We guess that despite the above comments, the indefinite integral will somehow contain $\frac{1}{2}W^2$, so use the simple form (2.3) of Ito's formula to see

$$d\left(\tfrac{1}{2}W^2\right) = 0\,dt + W\,dW + \tfrac{1}{2}\cdot 1\,dW^2 = W\,dW + \tfrac{1}{2}dt.$$

The $W\,dW$ term on the right-hand side is the desired differential; the dt term is some adjustment. Move the $\frac{1}{2}dt$ to the left-hand side to obtain

$$dI = W\,dW = d\left(\tfrac{1}{2}W^2\right) - \tfrac{1}{2}dt = d\left(\tfrac{1}{2}W^2 - \tfrac{1}{2}t\right).$$

Thus upon "summing," the indefinite integral $\int W\,dW = \frac{1}{2}W^2 - \frac{1}{2}t$. As seen in Figure 4.1, the integral is a fluctuating positive component superimposed upon $-t/2$.

Thus, further, the definite integral $\int_a^b W\,dW = \frac{1}{2}\left[W^2 - t\right]_a^b$. ∎

Algorithm 4.1 MATLAB/SCILAB code (essentially an Euler method) to numerically evaluate $\int_0^t W(s)\,dW(s)$ for $0 < t < 1$ to draw Figure 4.1.

```
m=5;
n=1000;
t=linspace(0,1,n+1)';
dw=randn(n,m)*sqrt(diff(t(1:2)));
w=[zeros(1,m);cumsum(dw,1)];
x=[zeros(1,m);cumsum(w(1:end-1,:).*dw,1)];
plot(t,x)
```

4.1 The Ito integral $\int_a^b f\,dW$

Here we *outline* the definition of an Ito integral $\int_a^b f\,dW$ and its properties. Øksendal (1998) [§3.1] and Kloeden and Platen (1992) [§3.1] give full details of the rigorous development of the integral. The development follows four steps:

1. Define integrals for piecewise constant stochastic integrands—called "step functions" $\phi(t, \omega)$—such as those illustrated in Figure 4.2;

2. use such step functions to approximate arbitrarily well any reasonable stochastic integrand $f(t, \omega)$;

3. for a sequence of step functions $\phi_n \to f$ as $n \to \infty$, then define the Ito integral of f to be the $\lim_{n\to\infty} \int_a^b \phi_n\,dW$;

4. Ito integral properties then follow from those for integrals of step functions.

We modify notation slightly: hereafter we use ω to denote the family of realizations of a Wiener process $W(t, \omega)$. This more explicit notation allows a more general dependence upon the Wiener process to appear in the integrands, such as a dependence upon the past history. Including the past history of the process is essential if the theory of integrals is to apply to SDEs such as $dX = X\,dW$ which we interpret as $X = \int X\,dW$ where the integrand $X(t, \omega)$ depends upon all the previous values of the Wiener process W. In contrast, in $\int W\,dW$ the increment $W\,dW$ to the integral depends only upon the current value of W.

Definition 4.2. *The class of stochastic functions that we may integrate (over the interval $[a, b]$) is denoted by $\mathcal{V}$. It is defined to be composed of functions $f(t, \omega)$ such that*

- *the expected square integral* $E[\int_a^b f(t, \omega)^2\,dt] < \infty$;

- $f(t, \omega)$ *depends only upon the past history of the Wiener process,* $W(s, \omega)$, *for $s \leq t$; for example,* $f = W(t, \omega)^2$, $f = e^{W(t,\omega)}$, *or* $f = \int_0^t W(s, \omega)\,ds$, *but not* $f = W(t+1, \omega)$.[38]

---

[38]Such functions that only depend upon the previous history of the Wiener process are called variously $\mathcal{F}_t$-measurable, $\mathcal{F}_t$-adapted, or $\mathcal{F}_t$-previsible. Slight technical differences exist between these terms. However, we endeavour to work entirely in a manner so that the differences in the terms are immaterial; thus herein we just refer to a dependence only upon the past history.
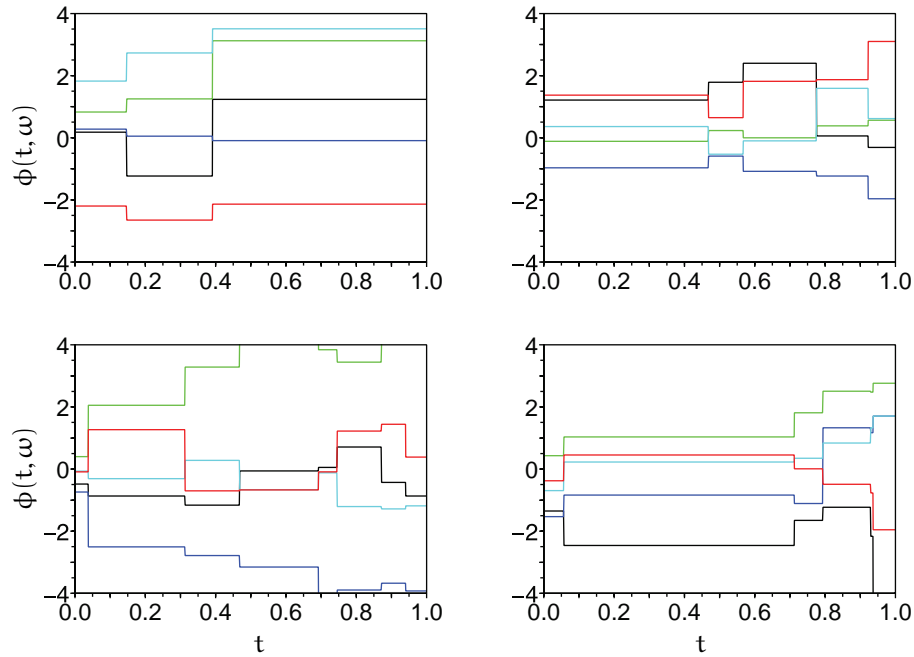
Figure 4.2. *Five realizations (different colors) of four different stochastic step functions $\phi(t, \omega) \in \mathcal{S}$: different step functions $\phi(t, \omega)$ have different partitions; different realizations of the one-step function $\phi(t, \omega)$ have the same partition but different values distributed according to some probabilities.*

Definition 4.3. *Let the class of* step functions $\mathcal{S} \subset \mathcal{V}$ *be the class of piecewise constant functions. That is, for each step function $\phi(t, \omega) \in \mathcal{S}$, there exists a finite partition $a = t_0 < t_1 < t_2 < \cdots < t_n = b$ such that $\phi(t, \omega) = \phi(t_j, \omega)$ for $t_j \leq t < t_{j+1}$. Figure 4.2 plots four different members of $\mathcal{S}$.*

Given any such partition, we often use ϕ_j or $\phi_j(\omega)$ to denote $\phi(t_j, \omega)$, just as we used W_j to denote $W(t_j, \omega)$.

First task: The Ito integral for step functions

Our first task, as in the outline given at the start of this chapter, is to define the Ito integral for step functions, such as those drawn in Figure 4.2, and investigate its key properties.

Definition 4.4. *For any step function $\phi(t, \omega)$ define the* Ito integral

$$I(\omega) = \int_a^b \phi(t, \omega) \, dW(t, \omega) = \int_a^b \phi \, dW = \sum_{j=0}^{n-1} \phi_j \Delta W_j. \qquad (4.2)$$

From this definition two familiar and desirable properties immediately follow, linearity and union (proofs are left as exercises), as does a third "no anticipation" property:

- *Linearity*:

$$\int_a^b (\alpha\phi + \beta\psi)\,dW = \alpha\int_a^b \phi\,dW + \beta\int_a^b \psi\,dW;\tag{4.3}$$

- *union*:

$$\int_a^b \phi\,dW + \int_b^c \phi\,dW = \int_a^c \phi\,dW;\tag{4.4}$$

- the integral $\int_a^b \phi\,dW$ depends only upon the history of the Wiener process up to time b (the integral is $\mathcal{F}_b$-adapted/measurable/previsible).

Two important properties of the Ito integral arise from its stochastic nature. These properties are so important we codify them as named theorems.

Theorem 4.5 (martingale property). *For the Ito integral* (4.2) *of a step function ϕ, its mean, average, or expected value is* always *zero:*

$$\mu_I = E\left[\int_a^b \phi(t,\omega)\,dW(t,\omega)\right] = 0.\tag{4.5}$$

Example 4.6. Recall from Example 4.1 that

$$I(\omega) = \int_0^t W(s,\omega)\,dW(s,\omega) = \tfrac{1}{2}\left[W(t,\omega)^2 - t\right].$$

Verify $E[I(\omega)] = 0$.

    Solution: From the expression for the integral,

$$\begin{aligned}
E[I] &= E\left[\tfrac{1}{2}(W(t,\omega)^2 - t)\right]\\
&= \tfrac{1}{2}E[W(t,\omega)^2] - \tfrac{1}{2}E[t]\\
&= \tfrac{1}{2}t - \tfrac{1}{2}t \quad \text{as } E[W(t,\omega)^2] = \mathrm{Var}[W(t,\omega)] = t\\
&= 0.
\end{aligned}$$

Although we only prove this martingale property[39] for step functions here, they do hold generally, and hence we have given an example which is not within the class of step functions. ∎

---

[39]The term *martingale* refers to stochastic processes, such as the Wiener process, for which, in symbols, $E[X(t,\omega)\mid\mathcal{F}_a] = X(a,\omega)$ for $t > a$. The symbol $\mathcal{F}_a$, called a *filtration*, denotes all the history of the process up to time a. This statement says that if you know the history of the process up to time a (that is, conditional on $\mathcal{F}_a$), then the process $X(t,\omega)$ is a called a martingale if the expected value thereafter is just $X(a,\omega)$. For example, a Wiener process is a martingale because, given *only* the history up to time a, the independence from the earlier history of subsequent increments in the Wiener process implies that the expected value of the Wiener process does not change from $W(a,\omega)$. The Ito integral may be considered a stochastic function of the end point b, and we see that as b varies the expected value of the integral is always the value it has for $b = a$, namely zero. Thus an Ito integral is a martingale.

Proof.   Recall the linearity of the expectation and that ϕ_j is independent of ΔW_j, as $\phi_j = \phi(t_j, \omega)$ depends only upon the earlier history of the Wiener process, while the increment ΔW_j is independent of the earlier history. Thus for the relevant partition, such as one of those in Figure 4.2,

$$
\begin{aligned}
E\left[\int_a^b \phi\, dW\right] &= E\left[\sum_{j=0}^{n-1} \phi_j \Delta W_j\right] \quad \text{by definition} \\
&= \sum_{j=0}^{n-1} E\left[\phi_j \Delta W_j\right] \quad \text{by linearity} \\
&= \sum_{j=0}^{n-1} E\left[\phi_j\right] \underbrace{E\left[\Delta W_j\right]}_{=0} \quad \text{by independence} \\
&= 0.
\end{aligned}
$$
$\quad\quad\quad\square$

Remarkably, irrespective of what integrand we choose in an Ito integral, we always know the mean value of the Ito integral, namely zero! But even more remarkably (see the next theorem), we can also readily determine the variance, that is, the spread, of an Ito integral. The two most important aspects of the stochastic $I(\omega) = \int_a^b \phi\, dW$, its mean and variance, can be determined *without* ever actually computing the Ito integral!

Theorem 4.7 (Ito isometry). *For the Ito integral* (4.2) *of a step function* $\phi(t, \omega)$, *its variance is the integral of the expectation of the squared integrand:*[40]

$$
\sigma_I^2 = \text{Var}\left[\int_a^b \phi(t, \omega)\, dW(t, \omega)\right] = E\left[\left(\int_a^b \phi\, dW\right)^2\right] = \int_a^b E\left[\phi(t, \omega)^2\right] dt. \quad (4.6)
$$

Example 4.8. Recall from Example 4.1 that the Ito integral

$$
I(\omega) = \int_0^t W(s, \omega)\, dW(s, \omega) = \tfrac{1}{2}[W(t, \omega)^2 - t].
$$

Verify the Ito isometry for this integral.
    Solution: Directly from the analytic solution,

$$
\begin{aligned}
\text{Var}[I(\omega)] &= \text{Var}\left[\tfrac{1}{2}(W(t, \omega)^2 - t)\right] \\
&= \tfrac{1}{4} E\left[(W^2 - t)^2\right] \quad \text{as } E[I(\omega)] = 0 \\
&= \tfrac{1}{4} E\left[W^4 - 2W^2 t + t^2\right] \\
&= \tfrac{1}{4}\left(3t^2 - 2 \cdot t \cdot t + t^2\right) \\
&= \tfrac{1}{2} t^2,
\end{aligned}
$$

---

[40]As if the square can be taken inside the integration and $dW^2 = dt$.

whereas from the Ito isometry (4.6),

$$\text{Var}[I(\omega)] = E\left[\left(\int_0^t W(s,\omega)\,dW(s,\omega)\right)^2\right]$$

$$= \int_0^t E\left[W(s,\omega)^2\right]ds$$

$$= \int_0^t s\,ds$$

$$= \tfrac{1}{2}t^2.$$

Since these are equal, we verify the Ito isometry for this Ito integral. ∎

Proof. Again we use the properties of the expectation and that the zero mean increment ΔW_j is independent of any earlier stochastic quantity. Thus for the relevant partition, as in Figure 4.2,

$$E\left[\left(\int_a^b \phi\,dW\right)^2\right] = E\left[\left(\sum_{j=0}^{n-1} \phi_j \Delta W_j\right)^2\right]$$

$$= E\left[\left(\sum_{j=0}^{n-1} \phi_j \Delta W_j\right)\left(\sum_{k=0}^{n-1} \phi_k \Delta W_k\right)\right]$$

$$= E\left[\sum_{j,k=0}^{n-1} \phi_j \Delta W_j \phi_k \Delta W_k\right]$$

$$= \sum_{j,k=0}^{n-1} E\left[\phi_j \Delta W_j \phi_k \Delta W_k\right].$$

Now within this double sum three cases arise: If $j < k$, then ΔW_k is independent of ϕ_j, ΔW_j, and ϕ_k—as ΔW_k depends only upon times $t > t_k$ whereas ϕ_j, ΔW_j and ϕ_k depend only upon times $t \leq t_k$—so the expectation may be split

$$E\left[\phi_j \Delta W_j \phi_k \Delta W_k\right] = E\left[\phi_j \Delta W_j \phi_k\right]\underbrace{E\left[\Delta W_k\right]}_{=0},$$

and so all such terms vanish; whereas if $j > k$, then ΔW_j is independent of ϕ_k, ΔW_k, and ϕ_j so the expectation may be split,

$$E\left[\phi_j \Delta W_j \phi_k \Delta W_k\right] = E\left[\phi_j \Delta W_k \phi_k\right]\underbrace{E\left[\Delta W_j\right]}_{=0},$$

and so all such terms vanish, leaving just the $j = k$ terms. Thus

$$E\left[\left(\int_a^b \phi \, dW\right)^2\right] = \sum_{j=0}^{n-1} E\left[\phi_j^2 \Delta W_j^2\right]$$

$$= \sum_{j=0}^{n-1} E\left[\phi_j^2\right] E\left[\Delta W_j^2\right] \quad \text{by independence}$$

$$= \sum_{j=0}^{n-1} E\left[\phi_j^2\right] \Delta t_j \quad \text{by variance of } \Delta W_j$$

$$= \int_a^b E\left[\phi^2\right] dt$$

by the definition of ordinary integration for piecewise constant integrands such as $E\left[\phi(t,\omega)^2\right]$. ◻

Suggested activity: Do at least Exercises 4.1 and 4.4.

Second task: Step functions approximate

The second task is to prove that the class of step functions $\mathscr{S}$ can be used to approximate arbitrarily well any given Ito process $f(t,\omega) \in \mathcal{V}$. For example, Figure 4.3 shows a sequence of four step functions that approximate the given Ito process increasingly well; they approximate the Ito process in all realizations.

Although this second task involves considerable uninteresting material, which we omit, we introduce the interesting and important notion of a Cauchy sequence as a means to approximation. You are familiar with this issue of approximability every time you work with real numbers. Recall that we approximate real numbers by rationals; for example, we often use $22/7$ to approximate π. For another example, in a computer all floating point numbers are stored and manipulated as an integer divided by a power of 2. But these approximations are only valid to some prescribed finite error. Rigorous mathematics establishes approximability to *every* error, no matter how small.

Sequences provide a mechanism for such arbitrarily accurate approximation. For example, Newton iteration $x' = \frac{1}{2}(x + 2/x)$ provides a sequence of rational approximations to the irrational $\sqrt{2}$: one such sequence is 1, 3/2, 17/12, 577/408, $\frac{665857}{470832}$, and so on. Similarly, a well-known approximation to the irrational $\pi/4$ is the sequence of partial sums of the series $1 - 1/3 + 1/5 - 1/7 + \cdots$, namely 1, 2/3, 13/15, 76/105, ... Such approximation of reals by a sequence of rationals works because there are rational numbers that are arbitrarily close to any specified real number; we chose a suitable sequence of such rational numbers that converge to the required real number. *Similarly, we can choose a sequence of stochastic step functions, such as those shown in Figure* 4.3*, that converge to any given Ito process.*

The notion of convergence is a problem

In earlier courses you will have discussed convergence in terms of the distance between an element of the sequence and its limit; for example, $\sqrt{2} - 1$, $\sqrt{2} - 3/2$, $\sqrt{2} - 17/12$,
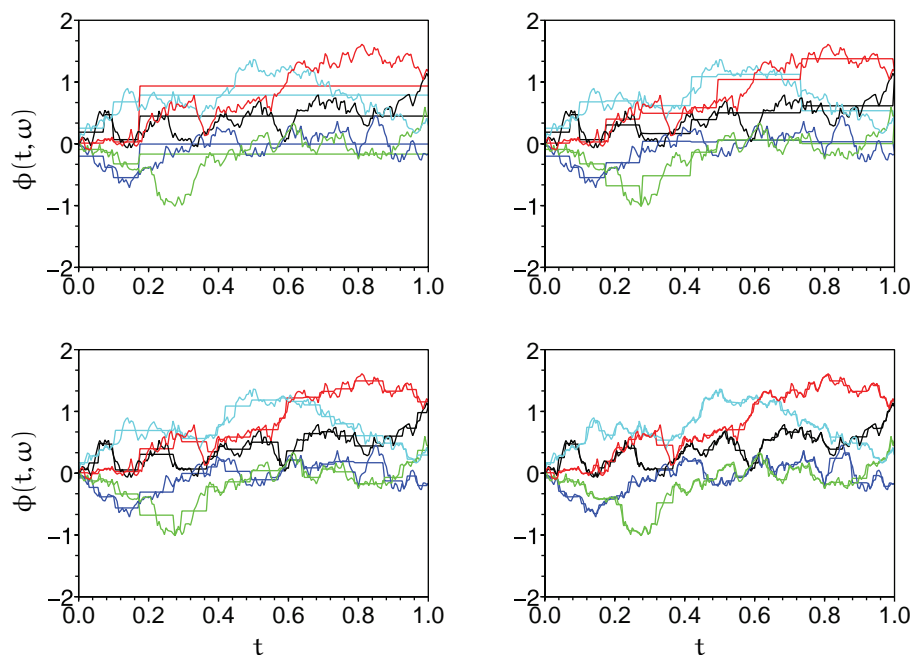
Figure 4.3. *Five realizations of an Ito process plotted once in each subfigure. Superimposed are four successively refined approximations by step functions. As the number of partitions increases from top-left to bottom-right, the corresponding realizations of the step functions increasingly closely approximate the realizations of the Ito process.*

$\sqrt{2} - 577/408$. Then you will have proved this distance tends to zero and concluded that the sequence converges to the limit. But here the limit is not in the class of things that we can as yet integrate! We can as yet only integrate step functions, not yet general Ito processes. Thus *we use a different definition of convergence:* instead of investigating the distance between elements of the sequence with the supposed limit, we investigate the distance between pairs of elements of the sequence. Table 4.1 gives an example. A sequence I_n is termed a *Cauchy sequence* if, loosely, the differences $\|I_n - I_m\| \to 0$ as $n, m \to \infty$, or more precisely, if for all $\epsilon > 0$ there exists an N such that $\|I_n - I_m\| < \epsilon$ for all $n, m > N$ (see, e.g., Kreyszig 1999, §14.1). If a sequence is Cauchy, then the sequence must converge to some limit, *even if we do not know what that limit is.*

The next subsubsection argues that Ito integrals of converging step functions form a Cauchy sequence, and hence the Ito integrals converge to something. The advantage of using the notion of a Cauchy sequence in a test for convergence is that it only involves operations of the elements of the sequence and does not involve comparisons with the actual limit. This is just what we need here because we have defined Ito integrals only for step functions, so we have no limit for comparison. But we need to clearly define distances between stochastic quantities.

Definition 4.9. *Measure the* distance *between two stochastic quantities by*

$$\text{dist}[I(\omega), J(\omega)] = E\left[(I(\omega) - J(\omega))^2\right]. \tag{4.7}$$

Table 4.1. *A sequence* I_n *approximating* $\sqrt{2}$ *and the distance between pairs of elements in the sequence,* $|I_m - I_n|$. *This distance tends to zero as* both $m, n \to \infty$ *and so is a Cauchy sequence.*

	n	1	2	3	4	5
m	$I_m \backslash I_n$	1	3/2	17/12	577/408	$\frac{665857}{470832}$
1	1	0	0.5	0.4167	0.4142	0.4142
2	3/2	0.5	0	0.0833	0.0858	0.0858
3	17/12	0.4167	0.0833	0	0.0025	0.0025
4	577/408	0.4142	0.0858	0.0025	0	2.E-06
5	$\frac{665857}{470832}$	0.4142	0.0858	0.0025	2.E-06	0

If this distance is zero, then I *and* J *are* almost surely *the same—realizations can only differ with probability zero. Analogously, measure the distance between two stochastic* processes *over a time interval* $[a, b]$ *by*

$$\text{dist}[I(t, \omega), J(t, \omega)] = \int_a^b E\left[(I(t, \omega) - J(t, \omega))^2\right] dt. \tag{4.8}$$

Establishing this ability of step functions to approximate Ito processes uses techniques that do not illuminate any significant properties of stochastic integration, and so we omit all details. Suffice it to say that Øksendal (1998) uses the norm $\|f\|^2 = \int_a^b E[f^2] dt$ on $\mathcal{V}$ and proceeds in three steps: step functions may approximate bounded continuous functions; bounded continuous functions may approximate bounded functions; and bounded functions may approximate the Ito processes in $\mathcal{V}$. Hence Øksendal concludes that step functions $\mathcal{S}$ may approximate arbitrarily well any Ito process in $\mathcal{V}$. We skip these uninteresting technicalities.

Third task: The limit exists

The third task defines the Ito integral of any $f(t, \omega) \in \mathcal{V}$ to be the limit of Ito integrals of step functions. The nontrivial task here is to deduce that the limit exists and is unique.

- Given some stochastic function $f(t, \omega)$ in $\mathcal{V}$.

- Find a sequence of step functions $\phi_n(t, \omega)$ that tends to $f(t, \omega)$ in the norm, that is,

$$\text{dist}(\phi_n, f) = \int_a^b E\left[(\phi_n - f)^2\right] dt \to 0 \quad \text{as } n \to \infty.$$

For example, Figure 4.3 plots five realizations of one such sequence of step functions approximating an $f(t, \omega)$.

- Compute the Ito integrals $I_n(\omega) = \int_a^b \phi_n(t, \omega) \, dW(t, \omega)$ of these step functions.

For example, Table 4.2 records the five realizations of the Ito integral of the step functions shown in Figure 4.3.

Table 4.2. *The five realizations of the Ito integral* $I_n(\omega)$ *of the step functions shown in Figure* 4.3. *The color matches that of the plot.*

	ω_1	ω_2	ω_3	ω_4	ω_5
I_1	0.3700	0.1287	−0.0399	−0.0274	1.0372
I_2	0.3446	0.0208	0.0237	0.0511	0.7501
I_3	0.5084	0.0105	0.0022	0.0751	0.6465
I_4	0.3454	−0.0756	−0.0930	−0.0589	0.5133
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$I(\omega)$	0.0520	−0.5076	−0.4369	−0.4164	0.2014

- The Ito integrals should converge to some number $I(\omega)$, which is stochastic because it depends upon the realization; call this the integral and denote it by $\int_a^b f\,dW$. For example, it is plausible that each column of Table 4.2 converges to five different realizations of the integral $I(\omega)$—recall that numerical approximations to stochastic integration converge slowly.

Lemma 4.10. *The Ito integrals* $I_n(\omega) = \int_a^b \phi_n(t,\omega)\,dW(t,\omega)$ *form a stochastic Cauchy sequence, in that* $E[(I_n - I_m)^2] \to 0$ *as* $n, m \to \infty$, *and hence must converge.*

Proof. Use the properties of Ito integrals for step functions and see their pairwise distance

$$
\begin{aligned}
\mathrm{dist}(I_n, I_m) &= E\left[(I_n - I_m)^2\right] \\
&= E\left[\left(\int_a^b \phi_n - \phi_m\,dW\right)^2\right] \quad \text{by linearity} \\
&= \int_a^b E\left[(\phi_n - \phi_m)^2\right]dt \quad \text{by Ito isometry} \\
&= \int_a^b E\left[((\phi_n - f) + (f - \phi_m))^2\right]dt \\
&\qquad \text{and since } (a+b)^2 \le 2a^2 + 2b^2 \text{ by Exercise 4.5} \\
&\le \int_a^b E\left[2(\phi_n - f)^2 + 2(f - \phi_m)^2\right]dt \\
&= 2\,\mathrm{dist}(\phi_n, f) + 2\,\mathrm{dist}(f, \phi_m) \\
&\to 0 \text{ as } n, m \to \infty \quad \text{by convergence to } f(t,\omega). \qquad \blacksquare
\end{aligned}
$$

Thus the Ito integrals, $I_n(\omega)$, of the sequence of step functions are a (stochastic) Cauchy sequence. This Cauchy sequence converges almost surely to some value—we call this value the *Ito integral* of $f(t,\omega)$, denoted $I(\omega) = \int_a^b f(t,\omega)\,dW(t,\omega)$.

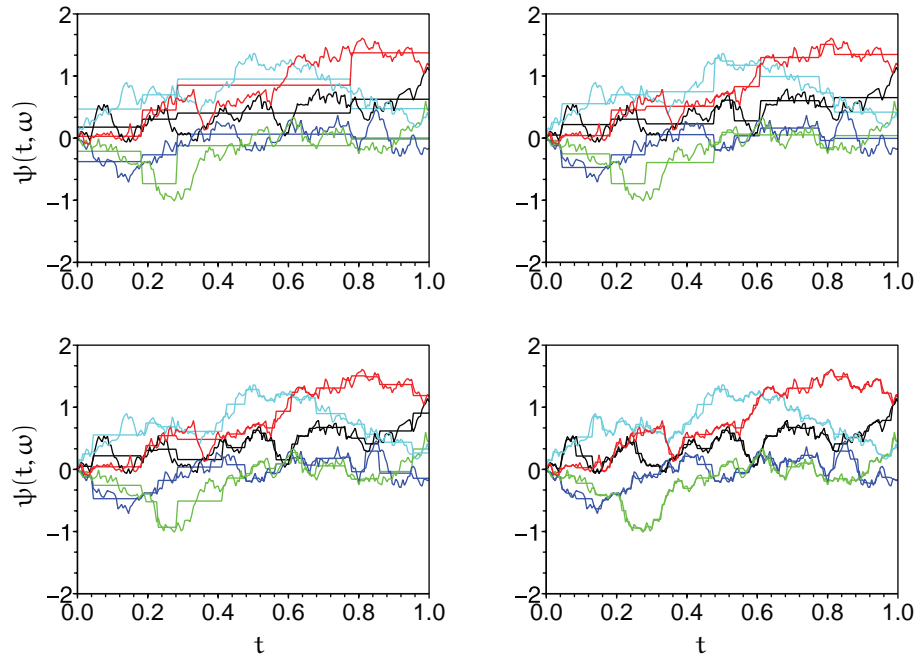**Figure 4.4.** *The same five realizations of an Ito process as in Figure* 4.3, *plotted with a different sequence of four successively refined approximations by step functions* $\psi(t,\omega)$.

Table 4.3. *The five realizations of the Ito integral* $J_n(\omega)$ *of the step functions shown in Figure* 4.4. *The color matches that of the plot.*

	ω_1	ω_2	ω_3	ω_4	ω_5
J_1	0.5346	0.1007	0.3329	0.1693	0.5758
J_2	0.5896	-0.0269	0.1393	-0.0905	0.6371
J_3	0.4704	0.0871	0.1088	-0.0060	0.6719
J_4	0.4306	-0.0624	-0.0418	-0.1413	0.5318
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$J(\omega)$	0.0520	-0.5076	-0.4369	-0.4164	0.2014

Lemma 4.11. *A stochastic integral* $I(\omega) = \int_a^b f(t,\omega)\, dW(t,\omega)$ *is unique; that is, it is almost always one definite value in each realization, though different for different realizations.*

Proof.   Consider any other sequence of step functions $\psi_n(t,\omega)$ that converge to the integrand $f(t,\omega)$; for example, see those plotted in Figure 4.4 which use a different sequence

of partitions than that shown in Figure 4.3. Suppose the Ito integrals $J_n(\omega)$ of this sequence converge to the value $J(\omega)$. For example, Table 4.3 gives integrals $J_n(\omega)$ for the step function sequence shown in Figure 4.4; see that these integrals are different from the integrals $I_n(\omega)$ and so potentially could converge to something different, namely $J(\omega)$ instead of $I(\omega)$. However, the two potential values of the integral, $I(\omega)$ and $J(\omega)$, are almost always the same as their distance

$$
\begin{aligned}
\operatorname{dist}(I, J) &= \mathrm{E}\left[(I-J)^2\right] \\
&= \mathrm{E}\left[((I-I_n)+(I_n-J_n)+(J_n-J))^2\right] \\
&\le \mathrm{E}\left[3(I-I_n)^2+3(I_n-J_n)^2+3(J_n-J)^2\right] \quad \text{by Exercise 4.5} \\
&= 3\mathrm{E}\left[(I-I_n)^2\right]+3\mathrm{E}\left[(I_n-J_n)^2\right]+3\mathrm{E}\left[(J_n-J)^2\right] \\
&\to 0 \text{ as } n \to \infty,
\end{aligned}
$$

as the first and last terms tend to zero by the definition of the values $I(\omega)$ and $J(\omega)$ of the integrals, and the middle term also tends to zero because

$$
\begin{aligned}
\mathrm{E}\left[(I_n-J_n)^2\right] &= \mathrm{E}\left[\left(\int_a^b \phi_n - \psi_n\, dW\right)^2\right] \quad \text{by definition and linearity} \\
&= \int_a^b \mathrm{E}\left[(\phi_n - \psi_n)^2\right] dt \quad \text{by Ito isometry} \\
&= \int_a^b \mathrm{E}\left[(\phi_n - f + f - \psi_n)^2\right] dt \\
&\le \int_a^b \mathrm{E}\left[2(\phi_n - f)^2 + 2(f - \psi_n)^2\right] dt \quad \text{by Exercise 4.5} \\
&= 2\int_a^b \mathrm{E}\left[(\phi_n - f)^2\right] dt + 2\int_a^b \mathrm{E}\left[(f - \psi_n)^2\right] dt \\
&\to 0 \text{ as } \phi_n, \psi_n \to f \text{ in this norm.}
\end{aligned}
$$

Thus the Ito integral $\int_a^b f(t,\omega)\, dW(t,\omega)$ exists and is almost surely unique. $\qquad\square$

Suggested activity: Do at least Exercise 4.5.

Fourth task: Properties then follow

The final task is to show that the integral properties for the Ito integral of functions $f(t,\omega) \in \mathcal{V}$ follow from those for step functions. There is little of interest in the technicalities, so a proof is omitted. From the five main properties of integrals of step functions, the following are the five main properties of general stochastic integrals:

- linearity,
$$
\int_a^b (\alpha f + \beta g)\, dW = \alpha \int_a^b f\, dW + \beta \int_a^b g\, dW; \tag{4.9}
$$

- union,

$$\int_a^b f\,dW + \int_b^c f\,dW = \int_a^c f\,dW;\tag{4.10}$$

- the Ito integral $\int_a^b f\,dW$ depends only upon the history of the Wiener process up to time b (the integral is $\mathcal{F}_b$-adapted/measurable/previsible);

- the martingale property; that is, the mean, average, or expected value is *always* zero:

$$E\left[\int_a^b f\,dW\right] = 0;\tag{4.11}$$

- Ito isometry, the variance is the ordinary integral of the expectation of the squared integrand:

$$\mathrm{Var}\left[\int_a^b f\,dW\right] = E\left[\left(\int_a^b f\,dW\right)^2\right] = \int_a^b E\left[f^2\right]dt.\tag{4.12}$$

We saw some of these properties in previous examples, for example, in the integral $\int W\,dW$.

Summary

This section put stochastic integrals on a strong footing. As for ordinary integrals, the basis is the approximation by sums over small increments in time. Consequently, stochastic integrals have many of the usual properties such as linearity and union. But they also have additional important properties, such as the martingale property, that their expectation is always zero, and the Ito isometry, that their variance may be computed as an ordinary integral. Having now put integration on a firm mathematical base, we now proceed to more rigorous analysis of stochastic processes.

4.2   The Ito formula

So far we have seen the Ito formula (2.4)–(2.5), enabling us to differentiate stochastic Ito processes and hence solve some SDEs. However, the integral form of the Ito formula, stated below, is now to be rigorously established because we now have a carefully defined Ito integral. The Ito formula also may be used to determine some integrals, as integration forms a relatively simple subset of the cases of solving differential equations.

Theorem 4.12. *Let* $f(t,x)$ *be a smooth function of its arguments and* $X(t,\omega)$ *be an Ito process with drift* $\mu(t,X)$ *and volatility* $\sigma(t,X)$, *that is,* $dX = \mu\,dt + \sigma\,dW$, *then* $Y(t,\omega) = f(t,X(t,\omega))$ *is also an Ito process such that*

$$[f(t,X(t,\omega))]_a^b = \int_a^b\left(\frac{\partial f}{\partial t} + \mu\frac{\partial f}{\partial x} + \tfrac{1}{2}\sigma^2\frac{\partial^2 f}{\partial x^2}\right)dt + \int_a^b \sigma\frac{\partial f}{\partial x}dW,\tag{4.13}$$

for all intervals $[a,b]$.

Before proceeding to prove the Ito formula, we here look at its role in determining Ito integrals. Take the view that transforming an Ito integral, $\int \cdots dW$, into an ordinary integral, $\int \cdots dt$, is effectively a solution of the integral. We seek to simplify $\int \cdots dW$ at the acceptable cost of introducing $\int \cdots dt$.

Example 4.13. We have seen that $\int W \, dW = \frac{1}{2}W(t,\omega)^2 - \frac{1}{2}t$, but what is $I = \int X \, dX$?[41]
 Solution: If $I(\omega)$ was an ordinary integral, we would immediately write down the integral as $\frac{1}{2}X^2$. So for this stochastic integral we guess $I(\omega)$ has $\frac{1}{2}X^2$ in it, apply Ito's formula (4.13) to the function $f(t,X) = \frac{1}{2}X^2$, and see what eventuates:

$$\left[\tfrac{1}{2}X^2\right]_a^b = \int_a^b \left(0 + \mu X + \tfrac{1}{2}\sigma^2 1\right) dt + \int_a^b \sigma X \, dW$$

$$= \int_a^b X \underbrace{(\mu \, dt + \sigma \, dW)}_{dX} + \tfrac{1}{2}\int_a^b \sigma^2 \, dt$$

$$\Rightarrow \int_a^b X \, dX = \left[\tfrac{1}{2}X^2\right]_a^b - \tfrac{1}{2}\int_a^b \sigma^2 \, dt.$$

This reduces to $\int W \, dW$ for which $\sigma = 1$ in the right-hand side. ■

In Example 4.13 the limits of integration, a and b, were carried along in the working and made no real difference to the algebraic manipulations other than increasing the level of detail. It is simpler to neglect to include the limits a and b and symbolically work entirely with indefinite integrals, as we often do hereafter.

Example 4.14. Determine $I(\omega) = \int t \, dW$ in terms of an ordinary integral.
 Solution: We might expect $tW(t,\omega)$ to appear in the answer, so we use the Ito formula (4.13) with $f(t,W) = tW(t,\omega)$, for which $f_t = W$, $f_W = t$, and $f_{WW} = 0$, and we also remember that $\mu = 0$ and $\sigma = 1$ for a Wiener process $W(t,\omega)$. Applying (4.13) leads to

$$tW(t,\omega) = \int \left(W + 0.t + \tfrac{1}{2}.0\right) dt + \int 1.t \, dW$$

$$\Rightarrow \int t \, dW = tW(t,\omega) - \int W \, dt.$$

See that this looks just like ordinary integration by parts. ■

---

[41] Assume $X(t,\omega)$ is an Ito process with drift $\mu(t,\omega)$ and volatility $\sigma(t,\omega)$; thus view the integral as $I(\omega) = \int X\mu \, dt + \int X\sigma \, dW$.

Example 4.15. Determine as far as possible $\int W^2\,dW$.

Solution: Guess that $\frac{1}{3}W^3$ might appear, so consider $f = \frac{1}{3}W^3$ in the integral form of Ito's formula (4.13):

$$\tfrac{1}{3}W^3 = \int \left(0 + 0 \cdot W^2 + \tfrac{1}{2} \cdot 2W\right) dt + \int 1 \cdot W^2\,dW$$

$$= \int W\,dt + \int W^2\,dW$$

$$\Rightarrow \int W^2\,dW = \tfrac{1}{3}W^3 - \int W\,dt. \qquad\blacksquare$$

Now we proceed to prove the integral form (4.13) of Ito's formula (2.5); Øksendal (1998) [§4.1] gives more details than we present. We do this now because it is only now that we have properly defined integration of stochastic functions, and because the differential form we used before is viewed as symbolically equivalent to the integral form. Throughout, observe the crucial role of the independence of Wiener increments ΔW_j from earlier events.

Proof. Consider the case where the drift μ and volatility σ are step functions (piecewise constant) with a common partition $a = t_0 < t_1 < \cdots < t_n = b$ with $h = \max_j \Delta t_j$ (so that $h \to 0$ means that the partition is everywhere made of smaller and smaller pieces). Then arbitrary reasonable μ and σ are approximated by a sequence of such step functions, and so the same formula holds.

For brevity let an overdot denote $\partial/\partial t$ and a dash denote $\partial/\partial x$ so that, for example, $\dot{f}_j = \frac{\partial f}{\partial t}\big|_{(t_j, X_j)}$. Then the left-hand side of the Ito formula (4.13) transforms as follows:

$$[f]_a^b = \sum_{j=0}^{n-1} \Delta f_j$$

$$= \sum_{j=0}^{n-1} \left[f(t_{j+1}, X_{j+1}) - f(t_j, X_j)\right]$$

$$= \sum_{j=0}^{n-1} \left[f(t_j + \Delta t_j, X_j + \Delta X_j) - f(t_j, X_j)\right]$$

$$= \sum_{j=0}^{n-1} \left[\dot{f}_j \Delta t_j + f_j' \Delta X_j + \tfrac{1}{2}\ddot{f}_j \Delta t_j{}^2 + \dot{f}_j' \Delta t_j \Delta X_j + \tfrac{1}{2}f_j'' \Delta X_j{}^2 + R_j\right]$$

by expanding $f(t,x)$ in a Taylor series, assuming $f(t,x)$ is differentiable (even though the process $X(t,\omega)$ need not be) and where the residual term $R_j = \mathcal{O}\left(|\Delta t_j|^3 + |\Delta X_j|^3\right)$. Then in the next few steps we use that the drift μ and volatility σ are step functions, are thus constant on each element on the partition, and so $\Delta X_j = \mu_j \Delta t_j + \sigma_j \Delta W_j$ exactly. Then consider in each of the following dot points each term in turn on the right-hand side of the above,

- $\sum_{j=0}^{n-1} \dot{f}_j \Delta t_j \to \int_a^b \dot{f}\,dt$ as $h \to 0$ (provided we refine the partition in such a way that μ and σ are step functions in every partition) to give the first term in Ito's integral formula;

- two more terms are obtained from

$$\sum_{j=0}^{n-1} f'_j \Delta X_j = \sum_{j=0}^{n-1} f'_j \mu_j \Delta t_j + \sum_{j=0}^{n-1} f'_j \sigma_j \Delta W_j \quad \text{as step functions}$$

$$\to \int_a^b f' \mu \, dt + \int_a^b f' \sigma \, dW \quad \text{as } h \to 0;$$

- with the last term in Ito's integral formula coming from the last of the quadratic terms in the sum, as we now show the other quadratic terms are zero starting with

$$\left| \sum_{j=0}^{n-1} \tfrac{1}{2} \ddot{f}_j \Delta t_j^2 \right| \le \sum_{j=0}^{n-1} \tfrac{1}{2} |\ddot{f}_j| \Delta t_j^2$$

$$\le \tfrac{1}{2} h \sum_{j=0}^{n-1} |\ddot{f}_j| \Delta t_j \quad \text{as } h = \max_j \Delta t_j$$

$$\to \tfrac{1}{2} h \int_a^b |\ddot{f}| \, dt$$

$$\to 0 \quad \text{as } h \to 0,$$

provided this integral exists (which we assume for $f(t, x)$ of interest) and hence this term must contribute nothing;

- the next quadratic term contributes nothing because

$$\sum_{j=0}^{n-1} \dot{f}'_j \Delta t_j \Delta X_j = \sum_{j=0}^{n-1} \dot{f}'_j \Delta t_j (\mu_j \Delta t_j + \sigma_j \Delta W_j)$$

$$= \sum_{j=0}^{n-1} \mu_j \dot{f}'_j \Delta t_j^2 + \underbrace{\sum_{j=0}^{n-1} \sigma_j \dot{f}'_j \Delta t_j \Delta W_j}_{=Y, \text{ say}},$$

and the first term here vanishes by the previous case (assuming $\int_a^b |\mu \dot{f}'| \, dt$ exists) whereas the second term, Y, is more delicate because it is stochastic, but we determine

$$E[Y^2] = E\left[\left(\sum_{j=0}^{n-1} \sigma_j \dot{f}'_j \Delta t_j \Delta W_j \right) \left(\sum_{k=0}^{n-1} \sigma_k \dot{f}'_k \Delta t_k \Delta W_k \right) \right]$$

$$= E\left[\sum_{j,k=0}^{n-1} \sigma_j \dot{f}'_j \Delta t_j \Delta W_j \sigma_k \dot{f}'_k \Delta t_k \Delta W_k \right]$$

$$= \sum_{j,k=0}^{n-1} E\left[\sigma_j \dot{f}'_j \Delta W_j \sigma_k \dot{f}'_k \Delta W_k \right] \Delta t_j \Delta t_k,$$

but all terms in this sum vanish except those for which $j = k$ because,[42] for example, if $j < k$ then the factor ΔW_k is independent of all other factors inside the expectation, and so the expectation factors to

$$E\left[\sigma_j \dot{f}_j' \Delta W_j\, \sigma_k \dot{f}_k' \Delta W_k\right] = E\left[\sigma_j \dot{f}_j' \Delta W_j\, \sigma_k \dot{f}_k'\right] \underbrace{E[\Delta W_k]}_{=0},$$

and similarly for $j > k$, thus leaving only the terms $k = j$ in

$$E[Y^2] = \sum_{j=0}^{n-1} E\left[\sigma_j^2 \dot{f}_j'^2 \Delta W_j^2\right] \Delta t_j^2$$

as ΔW_j is independent of earlier history

$$= \sum_{j=0}^{n-1} E\left[\sigma_j^2 \dot{f}_j'^2\right] \underbrace{E\left[\Delta W_j^2\right]}_{=\Delta t_j} \Delta t_j^2$$

$$\leq h^2 \sum_{j=0}^{n-1} E\left[\sigma_j^2 \dot{f}_j'^2\right] \Delta t_j$$

$$\to h^2 \int_a^b E\left[\sigma^2 \dot{f}'^2\right] dt$$

$$\to 0 \quad \text{as } h \to 0,$$

and thus almost surely the term $Y \to 0$ as $h \to 0$;

- the only significant quadratic term is

$$\sum_{j=0}^{n-1} \tfrac{1}{2} f_j'' \Delta X_j^2$$

$$= \sum_{j=0}^{n-1} \tfrac{1}{2} \left(\mu_j \Delta t_j + \sigma_j \Delta W_j\right)^2 f_j''$$

$$= \sum_{j=0}^{n-1} \tfrac{1}{2} \mu_j^2 f_j'' \Delta t_j^2 + \sum_{j=0}^{n-1} \mu_j \sigma_j f_j'' \Delta t_j \Delta W_j + \sum_{j=0}^{n-1} \tfrac{1}{2} \sigma_j^2 f_j'' \Delta W_j^2,$$

and of these three terms the first two vanish by almost identical arguments to the previous two cases which, setting $c_j = \tfrac{1}{2}\sigma_j^2 f_j''$ for brevity, leave the sum $\sum_{j=0}^{n-1} c_j \Delta W_j^2$ that we want to turn into the integral $\int_a^b c\, dt$ in the rather special way unique to stochastic calculus—we thus compare $\sum_{j=0}^{n-1} c_j \Delta W_j^2$ with $\sum_{j=0}^{n-1} c_j \Delta t_j$ by showing

---

[42] Note that this is exactly the same argument used to prove the Ito isometry, and which we use again later.

(similarly as before) the vanishing of

$$E\left[\left(\sum_{j=0}^{n-1} c_j \Delta W_j{}^2 - \sum_{j=0}^{n-1} c_j \Delta t_j\right)^2\right]$$

$$= E\left[\left(\sum_{j=0}^{n-1} c_j (\Delta W_j{}^2 - \Delta t_j)\right)^2\right]$$

$$= E\left[\left(\sum_{j=0}^{n-1} c_j (\Delta W_j{}^2 - \Delta t_j)\right)\left(\sum_{k=0}^{n-1} c_k (\Delta W_k{}^2 - \Delta t_k)\right)\right]$$

$$= E\left[\sum_{j,k=0}^{n-1} c_j c_k (\Delta W_j{}^2 - \Delta t_j)(\Delta W_k{}^2 - \Delta t_k)\right]$$

$$= \sum_{j,k=0}^{n-1} \underbrace{E\left[c_j c_k (\Delta W_j{}^2 - \Delta t_j)(\Delta W_k{}^2 - \Delta t_k)\right]}_{= 0 \text{ unless } k = j \text{ by independence of increments}}$$

$$= \sum_{j=0}^{n-1} E\left[c_j^2 (\Delta W_j{}^2 - \Delta t_j)^2\right]$$

as $\Delta W_j{}^2 - \Delta t_j$ independent of earlier history

$$= \sum_{j=0}^{n-1} E\left[c_j^2\right] E\left[(\Delta W_j{}^2 - \Delta t_j)^2\right]$$

$$= \sum_{j=0}^{n-1} E\left[c_j^2\right] E\left[\Delta W_j{}^4 - 2\Delta W_j{}^2 \Delta t_j + \Delta t_j{}^2\right]$$

$$= \sum_{j=0}^{n-1} E\left[c_j^2\right] \left(3\Delta t_j{}^2 - 2\Delta t_j \Delta t_j + \Delta t_j{}^2\right)$$

$$= 2 \sum_{j=0}^{n-1} E\left[c_j^2\right] \Delta t_j{}^2$$

$$\leq 2h \sum_{j=0}^{n-1} E\left[c_j^2\right] \Delta t_j$$

$$\to 2h \int_a^b E\left[c^2\right] dt$$

$$\to 0 \quad \text{as } h \to 0,$$

and thus almost surely,

$$\sum_{j=0}^{n-1} c_j \Delta W_j{}^2 \to \sum_{j=0}^{n-1} c_j \Delta t_j = \sum_{j=0}^{n-1} \tfrac{1}{2}\sigma_j^2 f_j'' \Delta t_j \to \int_a^b \tfrac{1}{2}\sigma^2 f'' \, dt$$

as $h \to 0$ to give the last term in Ito's integral formula;

- and for the very last task I simply claim that the residuals of cubic and higher order terms vanish, $\sum_{j=0}^{n-1} R_j \to 0$ as $h \to 0$ (part of the argument is Exercise 2.3), leaving just the integral form of the Ito formula. □

We terminate here this development of the basic theory that underpins SDEs and their applications. With this establishment of Ito's formula on a firm theoretical footing, and with the working concepts developed in earlier chapters, you are now able to not only solve practical problems such as the valuation of options, but you also have the understanding to study many of the more theoretical books and articles in financial mathematics and stochastic processes such as that by Øksendal (1998).

4.3   Summary

- An SDE such as $dX = \mu(t, X)dt + \sigma(t, x)dW$ is a convenient shorthand for the Ito integral equation

$$X(b, \omega) - X(a, \omega) = \int_a^b \mu(t, X(t, \omega)) dt + \int_a^b \sigma(t, X(t, \omega)) dW.$$

- Crucial properties of Ito integrals are

  - linearity,
  $$\int_a^b (\alpha f + \beta g) dW = \alpha \int_a^b f \, dW + \beta \int_a^b g \, dW;$$

  - union,
  $$\int_a^b f \, dW + \int_b^c f \, dW = \int_a^c f \, dW;$$

  - the Ito integral $\int_a^b f \, dW$ depends only upon the history of the Wiener process up to time b (it is $\mathcal{F}_b$-adapted/measurable/previsible);

  - it is martingale; that is, the mean, average, or expected value is *always* zero:
  $$E\left[\int_a^b f \, dW\right] = 0;$$

  - it has Ito isometry, that is, the variance is the ordinary integral of the expectation of the squared integrand:
  $$\mathrm{Var}\left[\int_a^b f \, dW\right] = E\left[\left(\int_a^b f \, dW\right)^2\right] = \int_a^b E\left[f^2\right] dt.$$

- The integral form of Ito's formula is

$$[f(t, X(t, \omega))]_a^b = \int_a^b \left(\frac{\partial f}{\partial t} + \mu \frac{\partial f}{\partial x} + \frac{1}{2} \sigma^2 \frac{\partial^2 f}{\partial x^2} \right) dt + \int_a^b \sigma \frac{\partial f}{\partial x} dW$$

for a stochastic process X satisfying $dX = \mu \, dt + \sigma \, dW$.

- In proving all of the above, a crucial feature is the independence of increments from earlier in the history. The proofs also rest upon the Ito isometry.

Exercises

4.1. Prove the linearity (4.3) of the Ito integral for step functions.

4.2. Prove the union (4.4) of the Ito integral for step functions.

4.3. By considering the differential $d(W^3 - 3tW)$, deduce the Ito integral $I(\omega) = \int_0^T W(t, \omega)^2 - t \, dW(t, \omega)$, where $W(t, \omega)$ is a Wiener process. Hence verify the martingale property and Ito isometry for this Ito integral.

4.4. Argue from basic principles that

$$\int_0^T W(t, \omega) \, dW(t, \omega) = \frac{1}{2} \left[W(T, \omega)^2 - T \right]$$

by forming a partition over the time interval $[0, T]$, approximating the integral as $I_n(\omega) = \sum_{j=0}^{n-1} W_j \Delta W_j$, and showing that these almost surely tend to $I(\omega) = \frac{1}{2} \left[W(T, \omega)^2 - T \right]$ by considering $I - I_n$.

1. Argue that $I - I_n = \frac{1}{2} \sum_{j=0}^{n-1} \left[(\Delta W_j)^2 - \Delta t_j \right]$ by writing

$$\frac{1}{2} \left[W(T, \omega)^2 - T \right] = \frac{1}{2} \sum_{j=0}^{n-1} \left[\Delta (W_j^2) - \Delta t_j \right].$$

2. Then use some of the ideas appearing in the proof of the Ito isometry and in the steps leading to (2.2) to show that the distance between the integral $I(\omega)$ and the approximate $I_n(\omega)$, measured by $E[(I - I_n)^2]$, must tend to zero.

4.5. Prove that $(a + b)^2 \leq 2a^2 + 2b^2$ for any real numbers a and b by considering $(a + b)^2 + (a - b)^2$. Similarly prove $(a + b + c)^2 \leq 3a^2 + 3b^2 + 3c^2$ for any real a, b, and c.

4.6. Reconsider $E[W(t, \omega)^k]$ for a Wiener process $W(t, \omega)$. Consider $d(W(t, \omega)^k)$ by Ito's formula and the martingale property of Ito integrals to deduce

$$E[W(T, \omega)^k] = \frac{1}{2} k(k - 1) \int_0^T E[W(t, \omega)^{k-2}] \, dt \quad \text{for } k \geq 2.$$

Hence determine $E[W(t, \omega)^2]$, $E[W(t, \omega)^4]$, and $E[W(t, \omega)^6]$. Compare your solution with that for Exercise 2.1.

4.7. For a nonrandom function $g(t)$, that is, g has no dependence upon the realizations ω of a Wiener process $W(t, \omega)$, use Ito's formula to show

$$\int g(t) dW(t, \omega) = g(t) W(t, \omega) - \int W(t, \omega) \frac{dg}{dt} dt.$$

4.8. Let $I_n(t, \omega) = t^{n/2} H_n(W(t, \omega)/\sqrt{t})$, where H_n is the nth Hermite polynomial, (see (Kreyszig 1999, pp. 246–247) or (Abramowitz and Stegun 1965, Chap. 22)), and the first few Hermite polynomials are $H_0(x) = 1$, $H_1(x) = x$, $H_2(x) = x^2 - 1$, and $H_3(x) = x^3 - 3x$. What are $I_0(t, \omega), \ldots, I_3(t, \omega)$? The aim of this exercise is to show that under Ito integration the functions $I_n(t)$ are the stochastic analogue of powers under ordinary integration. Use Ito's formula to show that

$$\int_0^T I_{n-1}(t, \omega) dW(t, \omega) = \frac{1}{n} I_n(T, \omega).$$

See how Ito integration maps I_{n-1} to $\frac{1}{n} I_n$ just as ordinary integration analogously maps the power t^{n-1} to $\frac{1}{n} t^n$. Continue to hence deduce

$$\int \cdots \int_{0 \leq t_1 \leq \cdots \leq t_n \leq t} 1 \, dW(t_1, \omega) \, dW(t_2, \omega) \ldots dW(t_n, \omega)$$
$$= \frac{1}{n!} t^{n/2} H_n(W(t, \omega)/\sqrt{t}).$$

Note: $H_n(x)$, among other properties, satisfy the recurrences $H_n' = n H_{n-1}$ and $H_n'' - x H_n' + n H_n = 0$.

Answers to selected exercises

4.3. $I(\omega) = \frac{1}{3} W(T, \omega)^3 - T W(T, \omega)$ and $\sigma_I^2 = \frac{2}{3} T^3$.

4.5. Consider $(a + b + c)^2 + (a - b)^2 + (a - c)^2 + (b - c)^2$.

4.6. $E[W(t, \omega)^2] = t$, $E[W(t, \omega)^4] = 3t^2$, and $E[W(t, \omega)^6] = 15t^3$.

4.8. $I_0 = 1$, $I_1 = W(t, \omega)$, $I_2 = W^2 - t$, and $I_3 = W^3 - 3tW$.

Appendix A

Extra MATLAB/SCILAB Code

The following algorithms generate plots that evolve in time as they execute, and they generate little movies. This aspect distinguishes them from the algorithms listed in the body of the text which generate static graphs. Since they do not correspond to specific figures in the text, they are gathered here in an appendix.

Algorithm A.1 This algorithm draws a 20 second zoom into the exponential function to demonstrate the smoothness of our usual functions.

```
t=linspace(-1,1,1024);
plot(t,exp(t)),grid
title('exponential function')
t0=clock;t=0;
while t<20
  axis([0 0 1 1]+exp(-t/5)*[-1 1 -1 1])
  drawnow
  t=etime(clock,t0);
end
```

Algorithm A.2 This algorithm draws a 60 second zoom into Brownian motion. Use the Brownian bridge to generate a start curve and new data as the zoom proceeds. Force the Brownian motion to pass through the origin. Note the self-affinity as the vertical is scaled with the square root of the horizontal. Note the infinite number of zero crossings that appear near the original one.

```
tfin=60;
n=2^10+1; % increase for slower better resolution
t=linspace(-1,1,n);
h=diff(t(1:2));
x=cumsum([0,sqrt(h)*randn(1,n-1)]);
x=x-x((n+1)/2); % force through origin
j=1:2:n; i=2:2:n; k=(n+3)/4:(3*n+1)/4;
hand=plot(t,x);
title('Weiner process')
axis([-1 1 -1 1])
hold on,plot([-1.1 0 1.1],[0 0 0],'ro-'),hold off
pause
t0=clock;tt=0;
while tt<tfin
  width=exp(-tt/5);
  axis([-width width -sqrt(width) sqrt(width)])
  drawnow
  if h>width/(n/4) % interpolate new data
    h=h/2;
    x(j)=x(k); t(j)=t(k);
    t(i)=0.5*(t(i-1)+t(i+1));
    x(i)=0.5*(x(i-1)+x(i+1))+sqrt(h/2)*randn(size(i));
    set(hand,'xdata',t,'ydata',x)
  end
  tt=etime(clock,t0);
end
```

Algorithm A.3 This algorithm random walks from the specific point $(0.7, 0.4)$ to show that the walkers first exit locations given a reasonable sample of the boundary conditions. This algorithm compares the numerical and exact solution $u = x^2 - y^2$.

```
[x,y]=meshgrid(0:0.1:1);
clabel(contour(x,y,x.^2-y.^2))
hold on
m=200
t=0;
z0=[0.7 0.4];
z=z0(ones(m,1),:);
plot(z(:,1),z(:,2),'b.','markersize',20)
h=plot(z(:,1),z(:,2),'r.','erasemode','xor','markersize',20);
axis('equal'),axis([0 1 0 1])
hold off
pause

j=1:m;
t0=clock;
while length(j)>m/8
  dt=etime(clock,t0)-t;
  z(j,:)=z(j,:)+0.1*sqrt(dt)*randn(length(j),2);
  z=max(0,min(z,1));
  t=t+dt;
  set(h,'xdata',z(:,1),'ydata',z(:,2))
  drawnow
  j=find(max(abs(z'-0.5))<0.5-1e-7);
end

% pause to show all points, then discard nonexiting walks
pause(3)
j=find(max(abs(z'-0.5))>=0.5-1e-7);
set(h,'xdata',z(j,1),'ydata',z(j,2))
drawnow
actual_u=z0(1)^2-z0(2)^2
fp=z(j,1).^2-z(j,2).^2;
estimate_u=mean(fp)
error_u=std(fp)/sqrt(length(fp)-1)
```

Appendix B

Two Alternate Proofs

B.1 Fokker–Planck equation

This appendix provides an alternate proof of Theorem 3.4 that the Fokker–Planck equation (3.2) governs the evolution of a PDF $p(t,x)$ of a stochastic Ito process. This proof use some theory from Chapter 4 and some techniques used in continuum mechanics (see, e.g., Roberts 1994). The proof here is more general than that of Section 3.1 in that here we allow both drift and volatility to vary in both x and time t.

Proof. Let $f(x)$ denote some *arbitrary* smooth function. For the Ito process $X(t,\omega)$, define a new Ito process $Y(t,\omega) = f(X(t,\omega))$. Ito's formula then tells us that

$$dY = f'(X)dX + \tfrac{1}{2}f''(X)dX^2$$
$$= \left[f'(X)\mu(t,X) + \tfrac{1}{2}f''(X)\sigma^2(t,X)\right]dt + f'(X)\sigma(t,X)dW$$

as $dX = \mu(t,X)dt + \sigma(t,X)dW$ in general. Now integrate this over *any* time interval $[a,b]$ to get the Ito integral formula (4.12) for this process Y:

$$\int_a^b dY = \int_a^b f'(X)\mu(t,X) + \tfrac{1}{2}f''(X)\sigma^2(t,X)\,dt + \int_a^b f'(X)\sigma(t,X)\,dW.$$

Now $\int_a^b dY = Y(b,\omega) - Y(a,\omega) = f(X(b,\omega)) - f(X(a,\omega))$ from the definition $Y = f(X)$. Thus Ito's integral formula asserts

$$f(X(b,\omega)) - f(X(a,\omega)) = \int_a^b f'(X)\mu(t,X) + \tfrac{1}{2}f''(X)\sigma^2(t,X)\,dt$$
$$+ \int_a^b f'(X)\sigma(t,X)\,dW. \qquad (\text{B.1})$$

Introduce the PDF $p(t,x)$ by taking expectations: recall that

$$E\{g(t,X)\} = \int g(t,x)p(t,x)\,dx,$$

119

where the limits of the x integration are implicitly over all x. The expectation of the left-hand side of (B.1) is

$$
\begin{aligned}
E\{f(X(b,\omega)) - f(X(a,\omega))\} &= \int f(x)p(x,b)\,dx - \int f(x)p(x,a)\,dx \\
&= \int f(x)[p(x,b) - p(x,a)]dx \\
&= \int f(x)\int_a^b \frac{\partial p}{\partial t}\,dt\,dx \\
&= \int_a^b \int f(x)\frac{\partial p}{\partial t}\,dx\,dt
\end{aligned}
$$

by the fundamental theorem of calculus that $p(x,b) - p(x,a) = \int_a^b \frac{\partial p}{\partial t}\,dt$. The factor $\frac{\partial p}{\partial t}$ appearing in the integrand becomes the $\frac{\partial p}{\partial t}$ term in the Fokker–Planck equation (3.2).

We derive the x derivative terms in the Fokker–Planck equation from the right-hand side of (B.1). Consider its expectation,

$$
E\left\{ \int_a^b f'(X)\mu(t,X) + \tfrac{1}{2}f''(X)\sigma^2(t,X)\,dt \right\} + E\left\{ \int_a^b f'(X)\sigma(t,X)\,dW \right\}.
$$

By the martingale property of Ito integrals (see Theorem 4.5) that $E\left\{ \int_a^b \cdot dW \right\} = 0$, the second expectation above is zero. Hence,

$$
\begin{aligned}
&E\{f(X(b,\omega)) - f(X(a,\omega))\} \\
&= E\left\{ \int_a^b f'(X)\mu(t,X) + \tfrac{1}{2}f''(X)\sigma^2(t,X)\,dt \right\} \\
&= \int_a^b E\{f'(X)\mu(t,X)\} + E\left\{ \tfrac{1}{2}f''(X)\sigma^2(t,X) \right\}\,dt \\
&= \int_a^b \int f'(x)\mu(t,x)p(t,x)\,dx + \int \tfrac{1}{2}f''(x)\sigma^2(t,x)p(t,x)\,dx\,dt.
\end{aligned}
$$

By integration by parts,

- $\int f'\mu p\,dx = [f\mu p]_{x=-\infty}^{\infty} - \int f\frac{\partial}{\partial x}(\mu p)\,dx = \int -f\frac{\partial}{\partial x}(\mu p)\,dx$ as the PDF $p(t,x)$ must go to zero as $x \to \pm\infty$; if it did not, then there would be no way that the area under the PDF could be one; whereas

- integrating by parts twice gives

$$
\int \tfrac{1}{2}f''\sigma^2 p\,dx = \tfrac{1}{2}\big[f'\sigma^2 p - f(\sigma^2 p)'\big]_{-\infty}^{\infty} + \int \tfrac{1}{2}f\frac{\partial^2}{\partial x^2}(\sigma^2 p)\,dx = \int \tfrac{1}{2}f\frac{\partial^2}{\partial x^2}(\sigma^2 p)\,dx
$$

as again the PDF and its derivative must go to zero as $x \to \pm\infty$.

Consequently the right-hand side becomes

$$E\{f(X(b,\omega)) - f(X(a,\omega))\} = \int_a^b \int -f\frac{\partial}{\partial x}(\mu p)\,dx + \int \tfrac{1}{2}f\frac{\partial^2}{\partial x^2}(\sigma^2 p)\,dx\,dt$$

$$= \int_a^b \int f\left[-\frac{\partial}{\partial x}(\mu p) + \tfrac{1}{2}\frac{\partial^2}{\partial x^2}(\sigma^2 p)\right]dx\,dt.$$

See the right-hand side of the Fokker–Planck equation (3.2) in this integrand.

Now put the two parts together. Equate the left and right-hand sides:

$$\int_a^b \int f(x)\frac{\partial p}{\partial t}\,dx\,dt = \int_a^b \int f(x)\left[-\frac{\partial}{\partial x}(\mu p) + \tfrac{1}{2}\frac{\partial^2}{\partial x^2}(\sigma^2 p)\right]dx\,dt.$$

Then put all terms under one pair of integrals on the left:

$$\int_a^b \int f(x)\left[\frac{\partial p}{\partial t} + \frac{\partial}{\partial x}(\mu p) - \tfrac{1}{2}\frac{\partial^2}{\partial x^2}(\sigma^2 p)\right]dx\,dt = 0.$$

All the parts of the Fokker–Planck equation (3.2) appear in this integrand—that is, all the parts except the "$= 0$" bit.

Use proof by contradiction to get the $= 0$. Recall that $f(x)$ is arbitrary, as is the time interval $[a, b]$. The only way that the continuous factor in the integrand in square brackets can be zero for all f, a, and b is if it is always zero itself. To see this, *suppose* the factor inside the brackets $[\,]$ is nonzero, say positive, at some $x = \xi$ and $t = \tau$. Since the expression inside the brackets $[\,]$ is continuous, then it must be still positive in some small time interval around τ, $a < \tau < b$ say, and in some small interval around ξ, $c < \xi < d$ say. Choose $f(x)$ to be any smooth function that is positive inside the interval $[c, d]$ and zero outside. Then the integral $\int_a^b \int f(x)[\,]\,dx\,dt$ must be > 0, as the integrand is ≥ 0 and is > 0 over a finite part of the domain. This is the contradiction, as we know the integral is always zero. Hence the supposition must be false: the factor in the brackets $[\,]$ must be everywhere zero. Consequently,

$$\frac{\partial p}{\partial t} + \frac{\partial}{\partial x}(\mu p) - \tfrac{1}{2}\frac{\partial^2}{\partial x^2}(\sigma^2 p) = 0,$$

which when rearranged slightly is the Fokker–Planck equation (3.2). $\qquad\qquad\square$

B.2 Kolmogorov backward equation

This section provides an alternate proof of Theorem 3.15 that the Kolmogorov backward equation (3.11) governs the evolution of the conditional PDF $p(t, x|s, y)$ of a stochastic Ito process.

Proof. Consider the conditional PDF $p(\tau, \xi|s, y)$ for any ξ and y and for any times $s < \tau$. Recall that $p\,d\xi$ gives the probability of passing through $x = \xi$ at time $t = \tau$ given the Ito process $X(t, \omega)$ starts at $x = y$ at time $t = s$, that is, $X(s, \omega) = y$.

Define a new Ito process $Z(t,\omega) = p(\tau,\xi|t,X(t,\omega))$ for $s \leq t \leq \tau$. This is fine because the conditional PDF p is some smooth function—we may or may not know what it is, but it does exist. The Ito process Z also varies with ξ and τ, but we focus on its t dependence. Apply Ito's formula, remembering $dX = \mu\,dt + \sigma\,dW$:

$$
dZ = \left.\frac{\partial p}{\partial s}\right|_{(\tau,\xi|t,X)} dt + \left.\frac{\partial p}{\partial y}\right|_{(\tau,\xi|t,X)} dX + \frac{1}{2}\left.\frac{\partial^2 p}{\partial y^2}\right|_{(\tau,\xi|t,X)} dX^2
$$

$$
= \left[\frac{\partial p}{\partial s} + \mu(t,X)\frac{\partial p}{\partial y} + \frac{1}{2}\sigma^2(t,X)\frac{\partial^2 p}{\partial y^2}\right]\Bigg|_{(\tau,\xi|t,X)} dt + \left.\frac{\partial p}{\partial y}\right|_{(\tau,\xi|t,X)} \sigma(t,X)\,dW.
$$

See that the terms in the Kolmogorov backward equation (3.11) appear in the dt term; our task is to disentangle them somehow. For simplicity define the new Ito process

$$
K(t,\omega) = \left[\frac{\partial p}{\partial s} + \mu(t,X)\frac{\partial p}{\partial y} + \frac{1}{2}\sigma^2(t,X)\frac{\partial^2 p}{\partial y^2}\right]\Bigg|_{(\tau,\xi|t,X)},
$$

which contains the terms of the Kolmogorov backward equation (3.11). Our goal is then to extract K from all the other terms.

Integrate over *any* time interval $[a,b]$ such that $s \leq a < b \leq \tau$ to deduce

$$
Z(b,\omega) - Z(a,\omega) = \int_a^b K\,dt + \int_a^b \left.\frac{\partial p}{\partial y}\right|_{(\tau,\xi|t,X)} \sigma(t,X)\,dW.
$$

Disentangling K is done by taking the expectation of this equation:

$$
E\{Z(b,\omega)\} - E\{Z(a,\omega)\} = \int_a^b E\{K\}\,dt + E\left\{\int_a^b \left.\frac{\partial p}{\partial y}\right|_{(\tau,\xi|t,X)} \sigma(t,X)\,dW\right\}. \qquad \text{(B.2)}
$$

First, by the martingale property of Ito integrals (see Theorem 4.5) that $E\{\int_a^b \cdot dW\} = 0$, the last integral above vanishes. Second, consider $E\{Z(t,\omega)\}$: from the definition of Z and an expectation,

$$
E\{Z(t,\omega)\} = \int p(\tau,\xi|t,x)p(t,x|s,y)\,dx = p(\tau,\xi|s,y),
$$

as the integral is the probability of going from (s,y) to (t,x), and then from (t,x) to (τ,ξ), integrated over all possible x values. Thus the integral must be just the probability of going from (s,y) to (τ,ξ). But the right-hand side $p(\tau,\xi|s,y)$ is independent of the intermediate time t, and hence the expectation $E\{Z(t,\omega)\}$ is constant in t. Consequently, its change over the interval $E\{Z(b,\omega)\} - E\{Z(a,\omega)\} = 0$. Thus, equation (B.2) becomes

$$
0 = \int_a^b E\{K\}\,dt.
$$

But this integral is zero for *all* intervals $[a,b]$, $s \leq a < b < \tau$, and so, by the same argument as used to prove the Fokker–Planck equation, the integrand

$$
E\{K(t,\omega)\} = 0 \quad \text{for all } s < t < \tau.
$$

We have extracted $K(t, \omega)$ at the expense of the expectation.

But $K(t, \omega)$ is a continuous Ito process, so the expectation $E\{K(t, \omega)\}$ must also be continuous; hence $E\{K(t, \omega)\} = 0$ for time $t = s$. Take the limit as $t \to s$; then $X(t, \omega) \to y$ and

$$E\{K(s, \omega)\} = E\left\{ \left[\frac{\partial p}{\partial s} + \mu(s, y) \frac{\partial p}{\partial y} + \frac{1}{2} \sigma^2(s, y) \frac{\partial^2 p}{\partial y^2} \right] \Bigg|_{(\tau, \xi | s, y)} \right\} = 0.$$

At $t = s$ there is nothing stochastic remaining inside the expectation. So the expectation is irrelevant and we deduce

$$\frac{\partial p}{\partial s} + \mu(s, y) \frac{\partial p}{\partial y} + \frac{1}{2} \sigma^2(s, y) \frac{\partial^2 p}{\partial y^2} = 0,$$

which must be satisfied for all τ, ξ, s, and y, as they are arbitrary, and hence proves the Kolmogorov backward equation (3.11). $\square$

Bibliography

Abramowitz, M. and Stegun, I. A., eds. (1966), *Handbook of Mathematical Functions*, Dover, New York.

Higham., D. J. (2001), An algorithmic introduction to numerical simulation of stochastic differential equations, *SIAM Review* **43**(3), 525–546. Available online at `http://link.aip.org/link/?SIR/43/525/1`

Higham, D. J. (2008), Modeling and simulating chemical reactions, *SIAM Review* **50**(2), 347–368. Available online at `http://link.aip.org/link/?SIR/50/347/1`

Kao, E. P. C. (1997), *An Introduction to Stochastic Processes*, Duxbury Press, Pacific Grove, CA.

Kloeden, P. E. and Platen, E. (1992), *Numerical Solution of Stochastic Differential Equations*, Vol. 23 of *Applications of Mathematics*, Springer-Verlag, Berlin.

Kreyszig, E. (1999), *Advanced Engineering Mathematics*, 8th ed., Wiley, New York.

Øksendal, B. K. (1998), *Stochastic Eifferential Equations: An Introduction with Applications*, Springer-Verlag, Berlin.

Roberts, A. J. (1994), *A One-Dimensional Introduction to Continuum Mechanics*, World Scientific, River Edge, NJ.

Stampfli, J. and Goodman, V. (2001), *The Mathematics of Finance: Modeling and Hedging*, Brooks/Cole, Pacific Grove, CA.

Index

Page numbers in italics denote the page of definition of the term.