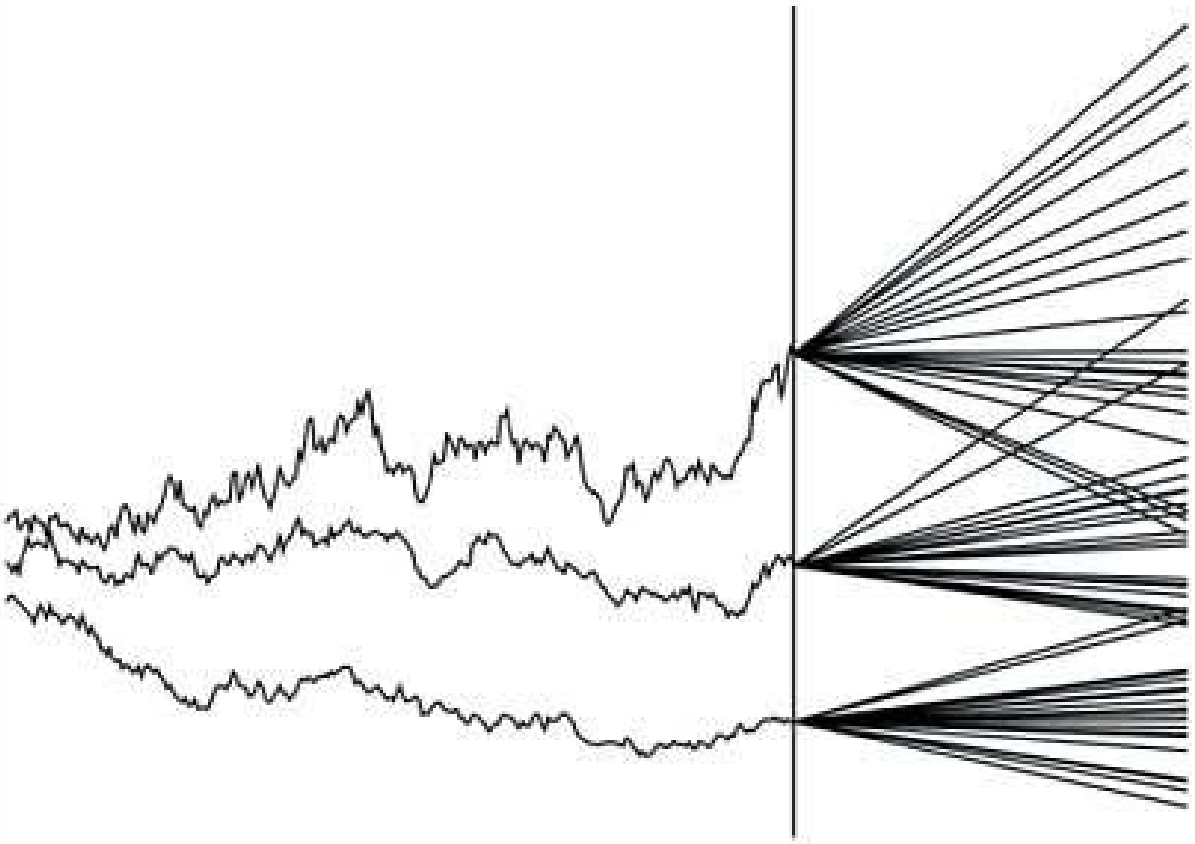
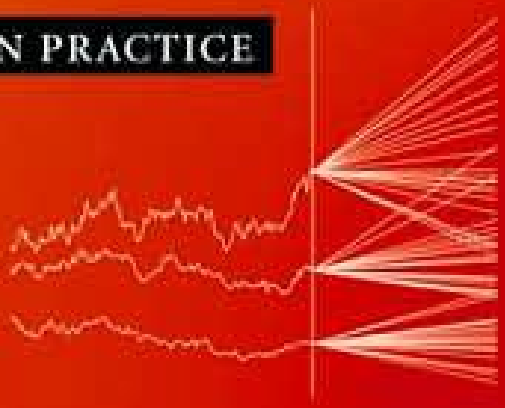
NGAI HANG CHAN
HOI YING WONG

Simulation Techniques in Financial Risk Management



WILEY

STATISTICS IN PRACTICE



Simulation Techniques in Financial Risk Management

STATISTICS IN PRACTICE

Advisory Editor

Peter Bloomfield
North Carolina State University, USA

Founding Editor

Vic Barnett
Nottingham Trent University, UK

Statistics in Practice is an important international series of texts which provide detailed coverage of statistical concepts, methods and worked case studies in specific fields of investigation and study.

With sound motivation and many worked practical examples, the books show in down-to-earth terms how to select and use an appropriate range of statistical techniques in a particular practical field within each title's special topic area.

The books provide statistical support for professionals and research workers across a range of employment fields and research environments. Subject areas covered include medicine and pharmaceutics; industry, finance and commerce; public services; the earth and environmental sciences, and so on.

The books also provide support to students studying statistical courses applied to the above areas. The demand for graduates to be equipped for the work environment has led to such courses becoming increasingly prevalent at universities and colleges.

It is our aim to present judiciously chosen and well-written workbooks to meet everyday practical needs. Feedback of views from readers will be most valuable to monitor the success of this aim.

A complete list of titles in this series appears at the end of the volume.

Simulation Techniques in Financial Risk Management

NGAI HANG CHAN

HOI YING WONG

The Chinese University of Hong Kong
Shatin, Hong Kong

A JOHN WILEY & SONS, INC., PUBLICATION

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.
Published simultaneously in Canada.

Library of Congress Cataloging-in-Publication Data:

Chan, Ngai Hang.
Simulation techniques in financial risk management / Ngai Hang Chan, Hoi Ying Wong.
p. cm.
Includes bibliographical references and index.
ISBN-13 978-0-471-46987-2 (cloth)
ISBN-10 0-471-46987-4 (cloth)
1. Finance—Simulation methods. 2. Risk management—Simulation methods. I. Wong, Hoi Ying. II. Title.

HG173.C47 2006
338.5—dc22 2005054992

Printed in the United States of America.

10 9 8 7 6 5 4 3 2 1

To Pat, Calvin, and Dennis

N.H. Chan

and

To Mei Choi

H.Y. Wong

Contents

List of Figures

List of Tables

Preface

Risk management is an important subject in finance. Despite its popularity, risk management has a broad and diverse definition that varies from individual to individual. One fact remains, however. Every modern risk management method comprises a significant amount of computations. To assess the success of a risk management procedure, one has to rely heavily on simulation methods. A typical example is the pricing and hedging of exotic options in the derivative market. These over-the-counter options experience very thin trading volume and yet their nonlinear features forbid the use of analytical techniques. As a result, one has to rely upon simulations in order to examine their properties. It is therefore not surprising that simulation has become an indispensable tool in the financial and risk management industry today.

Although simulation as a subject has a long history by itself, the same cannot be said about risk management. To fully appreciate the power and usefulness of risk management, one has to acquire a considerable amount of background knowledge across several disciplines: finance, statistics, mathematics, and computer science. It is the synergy of various concepts across these different fields that marks the success of modern risk management. Even though many excellent books have been written on the subject of simulation, none has been written from a risk management perspective. It is therefore timely and important to have a text that readily introduces the modern techniques of simulation and risk management to the financial world.

This text aims at introducing simulation techniques for practitioners in the financial and risk management industry at an intermediate level. The only

prerequisite is a standard undergraduate course in probability at the level of Hogg and Tanis (2006), say, and some rudimentary exposure to finance. The present volume stems from a set of lecture notes used at the Chinese University of Hong Kong. It aims at striking a balance between theory and applications of risk management and simulations, particularly along the financial sector. The book comprises three parts.

- Part one consists of the first three chapters. After introducing the motivations of simulation in Chapter 1, basic ideas of Wiener processes and Itô's calculus are introduced in chapters 2 and 3. The reason for this inclusion is that many students have experienced difficulties in this area because they lack the understanding of the theoretical underpinnings of these topics. We try to introduce these topics at an operational level so that readers can immediately appreciate the complexity and importance of stochastic calculus and its relationship with simulations. This will pave the way for a smooth transition to option pricing and Greeks in later chapters. For readers familiar with these topics, this part can be used as a review.

- Chapters 4 to 6 comprise the second part of the book. This part constitutes the main core of an introductory course in risk management. It covers standard topics in a traditional course in simulation, but at a much higher and succinct level. Technical details are left in the references, but important ideas are explained in a conceptual manner. Examples are also given throughout to illustrate the use of these techniques in risk management. By introducing simulations this way, both students with strong theoretical background and students with strong practical motivations get excited about the subject early on.

- The remaining chapters 7 to 10 constitute part three of the book. Here, more advanced and exotic topics of simulations in financial engineering and risk management are introduced. One distinctive feature in these chapters is the inclusion of case studies. Many of these cases have strong practical bearings such as pricing of exotic options, simulations of Greeks in hedging, and the use of Bayesian ideas to assess the impact of jumps. By means of these examples, it is hoped that readers can acquire a first-hand knowledge about the importance of simulations and apply them to their work.

Throughout the book, examples from finance and risk management have been incorporated as much as possible. This is done throughout the text, starting at the early chapter that discusses VaR of Dow to pricing of basket options in a multi-asset setting. Almost all of the examples and cases are illustrated with Splus and some with Visual Basics. Readers would be able to reproduce the analysis and learn about either Splus or Visual Basics by replicating some of the empirical work.

Many recent developments in both simulations and risk management, such as Gibbs sampling, the use of heavy-tailed distributions in VaR calculation, and principal components in multi-asset settings are discussed and illustrated in detail. Although many of these developments have found applications in the academic literature, they are less understood among practitioners. Inclusion of these topics narrows the gap between academic developments and practical applications.

In summary, this text fills a vacuum in the market of simulations and risk management. By giving both conceptual and practical illustrations, this text not only provides an efficient vehicle for practitioners to apply simulation techniques, but also demonstrates a synergy of these techniques. The examples and discussions in later chapters make recent developments in simulations and risk management more accessible to a larger audience.

Several versions of these lecture notes have been used in a simulation course given at the Chinese University of Hong Kong. We are grateful for many suggestions, comments, and questions from both students and colleagues. In particular, the first author is indebted to Professor John Lehoczky at Carnegie Mellon University, from whom he learned the essence of simulations in computational finance. Part two of this book reflects many of the ideas of John and is a reminiscence of his lecture notes at Carnegie Mellon. We would also like to thank Yu-Fung Lam and Ka-Yung Lau for their help in carrying out some of the computational tasks in the examples and for producing the figures in LaTeX, and to Mr. Steve Quigley and Ms. Susanne Steitz, both from Wiley, for their patience and professional assistance in guiding the preparation and production of this book. Financial support from the Research Grant Council of Hong Kong throughout this project is gratefully acknowledged. Last, but not least, we would like to thank our families for their understanding and encouragement in writing this book. Any remaining errors are, of course, our sole responsibility.

NGAI HANG CHAN AND HOI YING WONG

Shatin, Hong Kong

Simulation Techniques in Financial Risk Management
by Ngai Hang Chan and Hoi Ying Wong

1 Introduction

1.1 QUESTIONS

In this introductory chapter, we are faced with three basic questions:

- What is simulation?
- Why does one need to learn simulation?
- What has simulation to do with risk management and, in particular, financial risk management?

1.2 SIMULATION

When faced with uncertainties, one tries to build a probability model. In other words, risks and uncertainties can be handled (managed) by means of stochastic models. But in real life, building a full-blown stochastic model to account for every possible uncertainty is futile. One needs to compromise between choosing a model that is a realistic replica of the actual situation and choosing one whose mathematical (statistical) analysis is tractable.

But even equipped with the best insight and powerful mathematical knowledge, solving a model analytically is an exception rather than a rule. In most situations, one relies on an approximated model and learns about this model with approximated solutions. It is in this context that simulation comes into the picture. Loosely speaking, one can think of simulations as computer experiments. It plays the role of the experimental part in physics. When one

studies a physical phenomenon, one relies on physical theories and experimental verifications. When one tries to model a random phenomenon, one relies on building an approximated model (or an idealized model) and simulations (computer experiments).

Through simulations, one learns about different characteristics of the model, behaviors of the phenomenon, and features of the approximated solutions. Ultimately, simulations offer practitioners the ability to replicate the underlying scenario via computer experiments. It helps us to visualize the model, to study the model, and to improve the model.

In this book, we will learn some of the features of simulations. We will see that simulation is a powerful tool for analyzing complex situations. We will also study different techniques in simulations and their applications in risk management.

1.3 EXAMPLES

Practical implementation of risk management methods usually requires substantial computations. The computational requirement comes from calculating summaries, such as value-at-risk, hedging ratio, market β, and so on. In other words, summarizing data in complex situations is a routine job for a risk manager, but the same can be said for a statistician. Therefore, many of the simulation techniques developed by statisticians for summarizing data are equally applicable in the risk management context. In this section, we shall study some typical examples.

1.3.1 Quadrature

Numerical integration, also known as quadrature, is probably one of the earliest techniques that requires simulation. Consider a one-dimensional integral

$$I = \int_a^b f(x)\,dx, \tag{1.1}$$

where f is a given function. Quadrature approximates I by calculating f at a number of points $x_1, x_2, \ldots, x_n$ and applying some formula to the resulting values $f(x_1), \ldots, f(x_n)$. The simplest form is a weighted average

$$\hat{I} = \sum_{i=1}^{n} w_i\, f(x_i),$$

where $w_1, \ldots, w_n$ are some given weights. Different quadrature rules are distinguished by using different sets of design points $x_1, \ldots, x_n$ and different sets of weights $w_1, \ldots, w_n$. As an example, the simplest quadrature rule divides the interval $[a, b]$ into n equal parts, evaluates $f(x)$ at the midpoint of

each subinterval, and then applies equal weights. In this case

$$\hat{I} = \frac{b-a}{n} \sum_{i=1}^{n} f(a + (2i-1)(b-a)/(2n)).$$

This rule approximates the integral by the sum of the area of rectangles with base $(b-a)/n$ and height equal to the value of $f(x)$ at the midpoint of the base. For n large, we have a sum of many tiny rectangles whose area closely approximates I in exactly the same way that integrals are introduced in elementary calculus.

Why do we care about evaluating (1.1)? For one, we may want to calculate the expected value of a random quantity X with probability distribution function (p.d.f.) $f(x)$. In this case, we calculate

$$\mathrm{E}(X) = \int x f(x)\, dx,$$

and quadrature techniques may become handy if this integral cannot be solved analytically. Improvements over the simple quadrature have been developed, for example, Simpson's rule and the Gaussian rule. We will not pursue the details here, but interested readers may consult Conte and de Boor (1980). Clearly, generalizing this idea to higher dimensions is highly nontrivial. Many of the numerical integration techniques break down for evaluating high dimensional integrals. (Why?)

1.3.2 Monte Carlo

Monte Carlo integration is a different approach to evaluating an integral of f. It evaluates $f(x)$ at *random* points. Suppose that a series of points $x_1, \ldots, x_n$ are drawn independently from the distribution with density $g(x)$. Now

$$I = \int f(x)\, dx = \int [f(x)/g(x)] g(x)\, dx = \mathrm{E}_g\{\frac{f(x)}{g(x)}\}, \tag{1.2}$$

where E_g denotes expectation with respect to the distribution g. Now, the sample of points $x_1, \ldots, x_n$ drawn independently from g gives a sample of values $f(x_i)/g(x_i)$ of the function $f(x)/g(x)$. We estimate the integral (1.2) by the sample mean

$$\hat{I} = \frac{1}{n} \sum_{i=1}^{n} \frac{f(x_i)}{g(x_i)}.$$

According to classical statistics, $\hat{I}$ is an unbiased estimate of I with variance

$$\mathrm{Var}(\hat{I}) = \frac{1}{n} \mathrm{Var}_g \frac{f(x)}{g(x)}.$$

As n increases, $\hat{I}$ becomes a more and more accurate estimate of I. The variance (verify) can be estimated by its sample version, viz.,

$$\frac{1}{n^2}\sum_{i=1}^{n}\frac{f^2(x_i)}{g^2(x_i)} - \frac{\hat{I}^2}{n}. \tag{1.3}$$

Besides the Monte Carlo method, we should also mention that the idea of the quasi-Monte Carlo method has also enjoyed considerable attention recently. Further discussions on this method are beyond the scope of this book. Interested readers may consult the survey article by Hickernell, Lemieux, and Owen(2005).

1.4 STOCHASTIC SIMULATIONS

In risk management, one often encounters stochastic processes like Brownian motions, geometric Brownian motion, and lognormal distributions. While some of these entities may be understood analytically, quantities derived from them are often less tractable. For example, how can one evaluate integrals like $\int_0^1 W(t)\,dW(t)$ numerically? More importantly, can we use simulation techniques to help us understand features and behaviors of geometric Brownian motions or lognormal distributions? To illustrate the idea, we begin with the lognormal distribution.

Since the lognormal distribution plays such an important role in modeling the stock returns, we discuss some properties of the lognormal distribution in this section. First, recall that if $X \sim \mathrm{N}(\mu, \sigma^2)$, then the random variable $Y = e^X$ is lognormally distributed, i.e., $\log Y = X$ is normally distributed with mean μ and variance σ^2. Thus, the distribution of Y is given by

$$\begin{aligned} G(y) &= P(Y \le y) = P(X \le \log y) \\ &= P((X-\mu)/\sigma \le (\log y - \mu)/\sigma) \\ &= \Phi((\log y - \mu)/\sigma), \end{aligned}$$

where $\Phi(\cdot)$ denotes the distribution function of a standard normal random variable. Differentiating $G(y)$ with respect to y gives rise to the p.d.f of Y. To calculate $\mathrm{E}Y$, we can integrate it directly with respect to the p.d.f. of Y or we can make use of the normal distribution properties of X. Recall that the moment generating function of X is given by

$$M_X(t) = \mathrm{E}(e^{tX}) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}.$$

Thus,

$$\mathrm{E}Y = \mathrm{E}(e^X) = M_X(1) = e^{\mu + \frac{1}{2}\sigma^2}.$$

By a similar argument, we can calculate the second moment of Y and deduce that

$$\mathrm{Var}(Y) = e^{2\mu+\sigma^2}(e^{\sigma^2} - 1).$$

To generate 1,000 lognormal random variables in SPLUS with $\mu = 0$ and $\sigma^2 = 1$, i.e., $EY = e^{0.5}$ and $\text{Var}(Y) = e(e-1)$, type

```
>x_c(0.001,0.9)
>bounds_range(qlnorm(x),qnorm(x))
>points.x_seq(bounds[1],bounds[2],length=1000)
>points.qlnorm_dlnorm(points.x)
>points.qnorm_dnorm(points.x)
> plot(0,0,type='n',xlim=bounds,ylim=range(c(points.qlnorm,points.qnorm)),
+ xlab='', ylab='Density Value')
> lines(points.x,points.qlnorm,col=1,lty=1)
> lines(points.x,points.qnorm,col=1,lty=3)
```

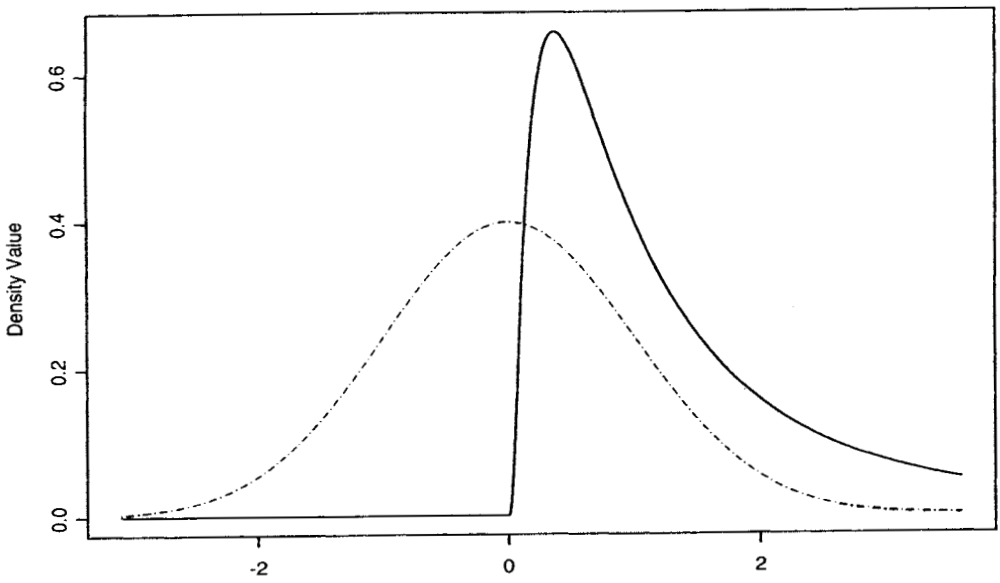


Fig. 1.1 Densities of a lognormal distribution with mean $e^{0.5}$ and variance $e(e-1)$, i.e., $\mu = 0$ and $\sigma^2 = 1$ and a standard normal distribution.

It can be seen from Fig. 1.1 that a lognormal density can never be negative. Further, it is skewed to the right and it has a much thicker tail than a normal random variable. Note that we have not tried to introduce SPLUS in detail here. We will only provide an operational discussion for readers to follow. For a comprehensive introduction to SPLUS, see Venables and Ripley (2002).

For readers who prefer to use Visual Basic, the corresponding codes are listed as follows:

```
Sub LogNormDist()

    Dim x(2) As Double
    x(1) = 0.001
    x(2) = 0.9

    Dim qlnorm(2) As Double
    Dim qnorm(2) As Double
    qlnorm(1) = Application.WorksheetFunction.LogInv(x(1), 0, 1)
    qlnorm(2) = Application.WorksheetFunction.LogInv(x(2), 0, 1)
    qnorm(1) = Application.WorksheetFunction.NormSInv(x(1))
    qnorm(2) = Application.WorksheetFunction.NormSInv(x(2))

    Dim bounds(2) As Double
    Dim range As Double
    bounds(1) = Application.WorksheetFunction.Min(qlnorm, qnorm)
    bounds(2) = Application.WorksheetFunction.Max(qlnorm, qnorm)
    range = bounds(2) - bounds(1)

    Dim points_x() As Double
    Dim n As Integer
    n = 1000
    ReDim points_x(n)
    points_x(1) = bounds(1)
    Cells(2, 1) = points_x(1)
    Dim i As Integer
    For i = 1 To n - 1
        points_x(i + 1) = points_x(i) + range / (n - 1)
        Cells(i + 2, 1) = points_x(i + 1)
    Next i

    Dim points_qlnorm() As Double
    ReDim points_qlnorm(n)
    Dim points_qnorm() As Double
    ReDim points_qnorm(n)
    Dim a, b As Double

    For i = 1 To n
        If points_x(i) < 0 Then
            points_qlnorm(i) = 0
        Else
            a = Application.WorksheetFunction.LogNormDist( _
```

```
            points_x(i), 0, 1)
        b = Application.WorksheetFunction.LogNormDist( _
            (points_x(i) + 0.00001), 0, 1)
        points_qlnorm(i) = (b - a) / 0.00001
    End If
    Cells(i + 1, 2) = points_qlnorm(i)
    a = Application.WorksheetFunction.NormSDist(points_x(i))
    b = Application.WorksheetFunction.NormSDist(points_x(i) _
        + 0.00001)
    points_qnorm(i) = (b - a) / 0.00001
    Cells(i + 1, 3) = points_qnorm(i)
Next i

Charts.Add
ActiveChart.ChartType = xlXYScatterSmoothNoMarkers
ActiveChart.SetSourceData Source:=Sheets("Sheet1").range( _
    "A2:C1001 "), PlotBy:=xlColumns
ActiveChart.Location Where:=xlLocationAsNewSheet
With ActiveChart
    .HasTitle = True
    .ChartTitle.Characters.Text = ""
    .Axes(xlCategory, xlPrimary).HasTitle = False
    .Axes(xlValue, xlPrimary).HasTitle = True
    .Axes(xlValue, xlPrimary).AxisTitle.Characters.Text = _
        "Density Value"
End With
With ActiveChart.Axes(xlCategory)
    .HasMajorGridlines = False
    .HasMinorGridlines = False
End With
With ActiveChart.Axes(xlValue)
    .HasMajorGridlines = False
    .HasMinorGridlines = False
End With
ActiveChart.HasLegend = False
ActiveChart.Axes(xlValue).Select
With ActiveChart.Axes(xlValue)
    .MinimumScale = -0.01
    .MaximumScale = 0.7
    .MinorUnit = 0.2
    .MajorUnit = 0.2
    .Crosses = xlAutomatic
    .ReversePlotOrder = False
    .ScaleType = xlLinear
    .DisplayUnit = xlNone
```

```
    End With
    Selection.TickLabels.NumberFormat = "0.0"
    ActiveChart.Axes(xlCategory).Select
    With Selection.Border
        .Weight = xlHairline
        .LineStyle = xlNone
    End With
    With Selection
        .MajorTickMark = xlOutside
        .MinorTickMark = xlNone
        .TickLabelPosition = xlNextToAxis
    End With
    With ActiveChart.Axes(xlCategory)
        .MinimumScaleIsAuto = True
        .MaximumScaleIsAuto = True
        .MinorUnit = 2
        .MajorUnit = 2
        .Crosses = xlCustom
        .CrossesAt = -4
        .ReversePlotOrder = False
        .ScaleType = xlLinear
        .DisplayUnit = xlNone
    End With
    Selection.TickLabels.NumberFormat = "0"

End Sub
```

Before ending this chapter, we would like to bring the readers' attentions to some existing books written on this subject. In the statistical community, many excellent texts have been written on this subject of simulations, see, for example, Ross (2002) and the references therein. These texts mainly discuss traditional simulation techniques without too much emphasis in finance and risk management. They are more suitable for a traditional audience in statistics.

In finance, there are several closely related texts. A comprehensive treatise on simulations in finance is given in the book by Glasserman (2004). A more succinct treatise on simulations in finance is given by Jaeckel (2002). Both of these books assume a considerable amount of financial background from the readers. They are intended for readers at a more advanced level. A book on simulation based on MATLAB is Brandimarte (2002). Another related book on Monte Carlo in finance is McLeish (2005). The survey article by Broadie and Glasserman (1998) offers a succinct account of the essence of simulations in finance. For readers interested in knowing more about the background of risk management, the two special volumes of Alexander (1998), the encyclopedic treatise of Crouchy, Galai and Mark (2001) and the special

volume of Dempster (2002) are excellent sources. The recent monograph of McNeil, Frey and Embrechts (2005) offers an up-to-date account on topics of quantitative risk management.

The current text can be considered as a synergy between Ross (2002) and Galsserman (2004), but at an intermediate level. We hope that readers with some (but not highly technical) background in either statistics or finance can benefit from reading this book.

1.5 EXERCISES

1. Verify equation (1.3).

2. Explain the possible difficulties in implementing quadrature methods to evaluate high dimensional numerical integrations.

3. Using either SPLUS or Visual Basic, simulate 1,000 observations from a lognormal distribution with a mean e^2 and variance $e^4(e^2 - 1)$. Calculate the sample mean and sample variance for these observations and compare their values with the theoretical values.

4. Let a stock have price S at time 0. At time 1, the stock price may rise to S_u with probability p or fall to S_d with probability $(1 - p)$. Let $R_S = (S_1 - S)/S$ denote the return of the stock at the end of period 1.

   (a) Calculate $m_S = \mathrm{E}(R_S)$.

   (b) Calculate $v_S = \sqrt{\mathrm{Var}(R_S)}$.

   (c) Let C be the price of a European call option of the stock at time 0 and C_1 be the price of this option at time 1. Suppose that $C_1 = C_u$ when the stock price rises to S_u and $C_1 = C_d$ when the stock price falls to S_d. Correspondingly, define the return of the call option at the end of period 1 as $R_C = (C_1 - C)/C$. Calculate $m_C = \mathrm{E}(R_C)$.

   (d) Show that $v_C = \sqrt{\mathrm{Var}(R_C)} = \sqrt{p(1-p)}(C_u - C_d)/C$.

   (e) Let $\Omega = \frac{(C_u - C_d)}{C} / \frac{(S_u - S_d)}{S}$, the so-called elasticity of the option. Show that $v_C = \Omega\, v_S$.

Simulation Techniques in Financial Risk Management
by Ngai Hang Chan and Hoi Ying Wong

2

Brownian Motions and Itô's Rule

2.1 INTRODUCTION

In this chapter, we will learn about the notion of Brownian motion and geometric Brownian motion, the latter being one of the most popular models in financial theory. In addition, the issue of Itô's calculus will also be introduced. The key element of this last concept is to develop an operational understanding of Itô's calculus so that readers will be able to do simple stochastic integration such as $\int_0^1 W^2(t)\,dW(t)$. Finally, we shall learn how to simulate these processes and study their corresponding features.

2.2 WIENER'S AND ITÔ'S PROCESSES

Consider the model defined by

$$W(t_{k+1}) = W(t_k) + \epsilon_{t_k}\sqrt{\Delta t}, \tag{2.1}$$

where $t_{k+1} - t_k = \Delta t$, and $k = 0, \ldots, N$ with $t_0 = 0$. In this equation, $\epsilon_{t_k} \sim \mathrm{N}(0,1)$ are identical independent distributed (i.i.d.) random variables. Further, assume that $W(t_0) = 0$. This is known as the random walk model (except for the factor $\sqrt{\Delta t}$, this equation matches with the familiar random walk model introduced in elementary courses). Note that from this model, for $j < k$,

$$W(t_k) - W(t_j) = \sum_{i=j}^{k-1} \epsilon_{t_i}\sqrt{\Delta t}.$$

There are a number of consequences:

1. As the right-hand side is a sum of normal random variables, it means that $W(t_k) - W(t_j)$ is also normally distributed.

2. By taking expectations, we have

$$\mathrm{E}(W(t_k) - W(t_j)) = 0,$$

$$\mathrm{Var}(W(t_k) - W(t_j)) = \mathrm{E}[\sum_{i=j}^{k-1} \epsilon_{t_i}\sqrt{\Delta t}]^2 = (k-j)\Delta t = t_k - t_j.$$

3. For $t_1 < t_2 \leq t_3 < t_4$,

$$W(t_4) - W(t_3) \text{ is uncorrelated with } W(t_2) - W(t_1).$$

Equation (2.1) provides a way to simulate a standard Brownian motion (Wiener process). To see how, consider partitioning $[0,1]$ into n subintervals each with length $\frac{1}{n}$. For each number t in $[0,1]$, let $[nt]$ denotes the greatest integer part of it. For example, if $n = 10$ and $t = \frac{1}{3}$, then $[nt] = [\frac{10}{3}] = 3$. Now define a stochastic process in $[0,1]$ as follows. For each t in $[0,1]$, define

$$S_{[nt]} = \frac{1}{\sqrt{n}} \sum_{i=1}^{[nt]} \epsilon_i, \tag{2.2}$$

where ϵ_i are i.i.d. standard normal random variables. Clearly,

$$S_{[nt]} = S_{[nt]-1} + \epsilon_{[nt]} \frac{1}{\sqrt{n}}, \tag{2.3}$$

which is a special form of (2.1) with $\Delta t = \frac{1}{n}$ and $W(t) = S_{[nt]}$. Furthermore, we know that at $t = 1$,

$$S_{[nt]} = S_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \epsilon_i,$$

has a standard normal distribution. Also by the Central Limit Theorem, we know that S_n tends to a standard normal random variable in distribution even if the ϵ_i are only i.i.d. but not necessarily normally distributed. The idea is that by taking the limit as n tends to ∞, the process $S_{[nt]}$ would tend to a Wiener process in distribution. Consequently, to simulate a sample path of a Wiener process, all we need to do is to iterate equation (2.3). Fig. 2.1 shows the simulations based on (2.3).

To generate Fig. 2.1 in SPLUS, type:

```
par(mfrow=c(1,1))
```

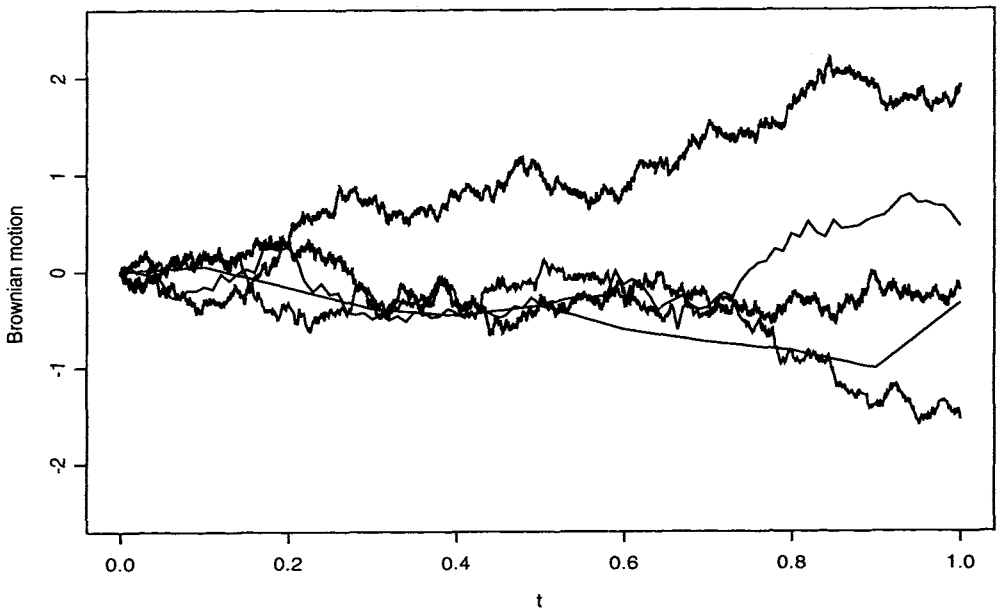


Fig. 2.1 Sample paths of the process $S_{[nt]}$ for different n and the same sequence of ϵ_i.

```
BMsim <- function(npaths,nSamples)
{
p <- npaths
N <- nSamples
y <- matrix(rep(0,(N+1)*p),nrow=N+1)
t <- (c(0:N))/N
for (j in 1:p)
{
z <- rnorm(N,0,1)
y[1,j] <- 0
for (i in 1:N)
{
y[i+1,j] <- (1/sqrt(N))*sum(z[1:i])
}
}
y
}
```

```
fig <- function(npath)
{
p <- npath
for(i in 1:p)
{
N <- 10^i
y <- BMsim(1, N)
t <- (c(0:N))/N
if(i == 1) {
matplot(t, y, type = "l", xlab = "t", ylab = "Brownian motion",
lty = 1, col = 1,ylim=range(-2.5,2.5))
}
else if(i > 1)
{
lines(t, y)
}
}
}
fig(5)
```

To generate Fig. 2.2 in SPLUS, type:

```
par(mfrow=c(2,2))
BMsim <- function(npaths,nSamples)
{
p <- npaths
N <- nSamples
y <- matrix(rep(0,(N+1)*p),nrow=N+1)
t <- (c(0:N))/N
for (j in 1:p)
{
z <- rnorm(N,0,1)
y[1,j] <- 0
for (i in 1:N)
{
y[i+1,j] <- (1/sqrt(N))*sum(z[1:i])
}
}
matplot(t,y,type="l",xlab="t",ylab="Brownian motion",lty=1,col=1)
}
BMsim(1,1000)
BMsim(5,1000)
BMsim(20,1000)
BMsim(100,1000)
```

Here are the corresponding Visual Basic codes for simulating Brownian motion:

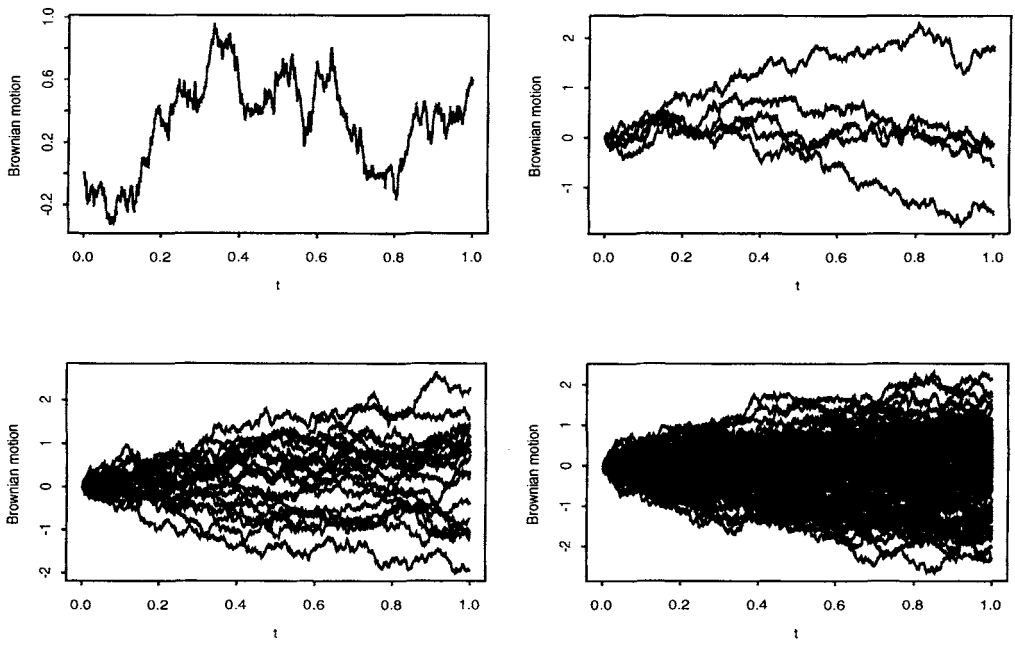


Fig. 2.2 Sample paths of Brownian motions on [0,1].

```
Sub BMsim()

    Dim npaths As Integer
    Dim nSamples As Integer

    npaths = 10          'no. of paths
    nSamples = 1000      'no. of samples in one path

    Dim t() As Double
    ReDim t(0 To nSamples)

    Dim j As Integer
    For j = 0 To nSamples
        t(j) = j / nSamples
    Next j

    Dim S() As Double
```

```
    ReDim S(0 To npaths, 0 To nSamples)

    Dim epsilon As Double
    Dim i As Integer
    For i = 0 To npaths
        epsilon = Application.WorksheetFunction.NormSInv(Rnd)
        S(i, 0) = 0
        For j = 1 To nSamples
            epsilon = Application.WorksheetFunction.NormSInv(Rnd)
            S(i, j) = S(i, j - 1) + (1 / Sqr(nSamples)) * epsilon
        Next j
    Next i

    Sheets("Sheet1").Select
    Cells.Select
    Selection.ClearContents
    If i <> 1 Then
        For i = 1 To npaths
            For j = 0 To nSamples
                Cells(j + 1, i) = S(i, j)
            Next j
        Next i
    Else
        For j = 0 To nSamples
            Cells(j + 1, 1) = S(1, j)
        Next j
    End If

    Cells(1, 1).Select

End Sub
```

In other words, by taking limit as Δt tends to zero, we get a Wiener process (Brownian motion), i.e.,

$$dW(t) = \epsilon(t)\sqrt{dt},$$

where $\epsilon(t)$ are uncorrelated standard normal random variables. We can interpret this equation as a continuous-time approximation of the random walk model (2.1), see Chan (2002). Of course, such an approximation can be dubious because we do not know if this limiting operation is well defined. In more advanced courses in probability, see Billingsley (1999), for example, it is shown that this limiting operation is well defined and, indeed, we obtain a Wiener process as a limit of the above operation. Formally, we define a Wiener process $W(t)$ as a stochastic process as follows.

Definition 2.1 *A Wiener process $W(t)$ is a stochastic process that satisfies the following properties:*

- *For $s < t$, $W(t) - W(s)$ is a normally distributed random variable with mean 0 and variance $t - s$.*
- *For $0 \leq t_1 < t_2 \leq t_3 < t_4$, $W(t_4) - W(t_3)$ is uncorrelated with $W(t_2) - W(t_1)$. This is known as the independent increment property.*
- *$W(t_0) = 0$ with probability one.*

From this definition, we can deduce a number of properties.

1. For $t < s$, $\mathrm{E}(W(s)|W(t)) = \mathrm{E}(W(s) - W(t) + W(t)|W(t)) = W(t)$. This is known as the *martingale* property of the Brownian motion.
2. The process $W(t)$ is nowhere differentiable. Consider
$$\mathrm{E}\left((\frac{W(s) - W(t)}{s - t})^2\right) = \frac{1}{s - t}.$$
This term tends to ∞ as $s - t$ tends to 0. Hence, the process cannot be differentiable and we cannot give a precise mathematical meaning to the process $dW(t)/dt$.
3. If we formally represent $\xi(t) = \frac{dW(t)}{dt}$ and call it the white noise process, we can only use it as a symbol and its mathematical meaning has to be interpreted in terms of an integration in the context of a stochastic differential equation.

The idea of Wiener process can be generalized as follows. Consider a process $X(t)$ satisfying the following equation:

$$dX(t) = \mu\, dt + \sigma\, dW(t), \tag{2.4}$$

where μ and σ are constants and $W(t)$ is a Wiener process defined previously. If we integrate (2.4) over $[0, t]$, we get

$$X(t) = X(0) + \mu t + \sigma W(t),$$

i.e., the process $X(t)$ satisfies the integral equation

$$\int dX(t) = \mu \int dt + \sigma \int dW(t).$$

The process $X(t)$ is also known as a diffusion process or a generalized Wiener process. In this case, the solution $X(t)$ can be written down analytically in terms of the parameters μ and σ and the Wiener process $W(t)$. To extend this idea further, we can let the parameters μ and σ depend on the process $X(t)$ as well. In that case, we have what is known as a general diffusion process or an Itô's process.

Definition 2.2 *An Itô's process is a stochastic process that is the solution to the following stochastic differential equation (SDE):*

$$dX(t) = \mu(x,t)\,dt + \sigma(x,t)\,dW(t). \tag{2.5}$$

In this equation, $\mu(x,t)$ is known as the drift function and $\sigma(x,t)$ is known as the volatility function of the underlying process. Of course, we need conditions for $\mu(x,t)$ and $\sigma(x,t)$ to ensure (2.5) has a solution. We will not discuss these technical details here; further details can be found in Karatzas and Shreve (1997) or Dana and Jeanblanc (2002). We will just assume that the drift and the volatility are "nice" enough functions so that the existence of a stochastic process $\{X(t)\}$ that satisfies (2.5) is guaranteed. Again, this equation has to be interpreted through integration.

2.3 STOCK PRICE

Recall the multiplicative model

$$\log S(k+1) = \log S(k) + w(k).$$

The continuous-time version of this equation is

$$d\log S(t) = \nu\,dt + \sigma\,dW(t).$$

The right-hand side of this equation is normally distributed with mean $\nu\,dt$ and variance $\sigma^2\,dt$. Solving this equation by integration,

$$\log S(t) = \log S(0) + \nu t + \sigma W(t).$$

Then, $\mathrm{E}\log S(t) = \log S(0) + \nu t$. Since the expected log price grows linearly with t, just as in a continuous compound interest formula, the process $S(t)$ is known as a geometric Brownian motion (GBM). Formally, we define

Definition 2.3 *Let $X(t)$ be a Brownian motion with drift ν and variance σ^2, i.e.,*

$$dX(t) = \nu\,dt + \sigma\,dW(t).$$

The process $S(t) = e^{X(t)}$ is called a geometric Brownian motion with drift parameter μ, where $\mu = \nu - \frac{1}{2}\sigma^2$. In particular, $S(t)$ satisfies

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

and

$$d\log S(t) = \left(\mu - \frac{1}{2}\sigma^2\right)dt + \sigma\,dW(t). \tag{2.6}$$

To simulate 1,000 geometric Brownian motions in SPLUS with $\mu = 0.03$ and $\sigma^2 = 0.04$, type

```
par(mfrow=c(1,1))
p <- 1 #no. of paths
N <- 1000 #no. of samples in one path
S0 <- 1 #current stock price
mu <- 0.03 #mean value
sigma <- 0.2 #standard deviation
nu <- mu - sigma^2/2
x <- matrix(rep(0,(N+1)*p),nrow=(N+1))
y <- matrix(rep(0,(N+1)*p),nrow=(N+1))
t <- (c(0:N))/N
for (j in 1:p)
{
z <- rnorm(N,0,1)
x[1,j] <- 0
y[1,j] <- S0
for (i in 1:N)
{
x[i+1,j] <- (1/sqrt(N))*sum(z[1:i])
y[i+1,j] <- y[1,j]*exp(nu*t[i+1]+sigma*x[i+1,j])
}
}
matplot(t,y,type="l",xlab="t",ylab="Geometric Brownian motion")
```

A sample path is plotted in Fig. 2.3. The corresponding Visual Basic codes for simulating the geometric Brownian motion are:

```
Sub GBMsim()

    Dim npaths As Integer
    Dim nSamples As Integer
    Dim S0 As Double
    Dim mu As Double
    Dim sigma As Double

    npaths = 1  'no. of paths
    nSamples = 1000  'no. of samples in one path
    S0 = 1  'current stock price
    mu = 0.03  'mean value
    sigma = 0.2  'standard deviation
    nu = mu - sigam^2/2
    Dim t() As Double
    ReDim t(0 To nSamples)
    Dim j As Integer
```

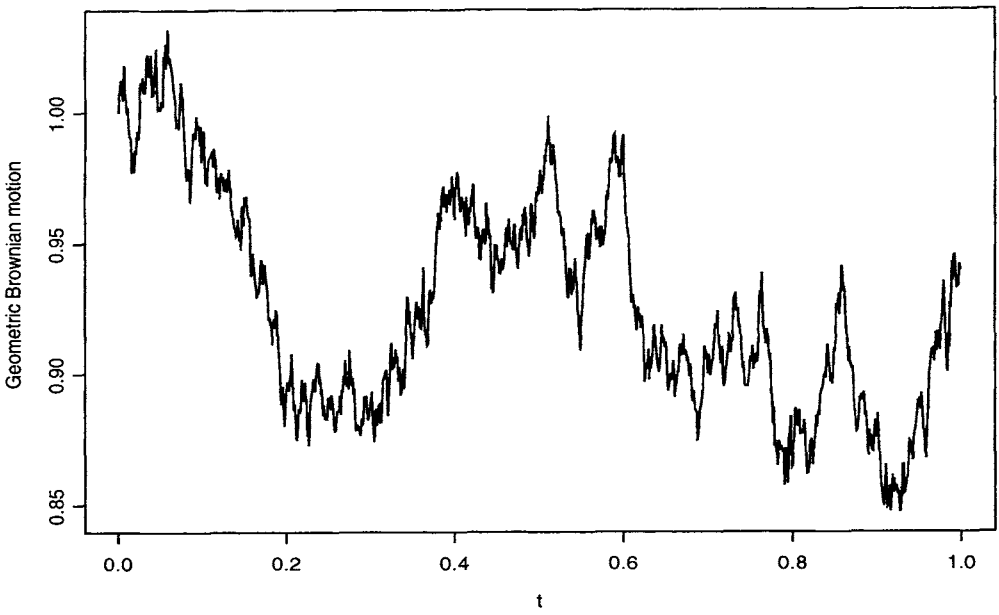


Fig. 2.3 Geometric Brownian motion.

```
For j = 0 To nSamples
    t(j) = j / nSamples
Next j

Dim X() As Double
Dim S() As Double
ReDim X(0 To npaths, 0 To nSamples)
ReDim S(0 To npaths, 0 To nSamples)
Dim epsilon As Double
Dim i As Integer
For i = 0 To npaths
    epsilon = Application.WorksheetFunction.NormSInv(Rnd)
    X(i, 0) = 0
    S(i, 0) = S0
    For j = 1 To nSamples
        epsilon = Application.WorksheetFunction.NormSInv(Rnd)
        X(i, j) = X(i, j - 1) + (1 / Sqr(nSamples)) * epsilon
```

```
            S(i, j) = S0 * Exp(nu * t(j) + sigma * X(i, j))
        Next j
    Next i

    Sheets("Sheet1").Select
    Cells.Select
    Selection.ClearContents
    If i <> 1 Then
        For i = 1 To npaths
            For j = 0 To nSamples
                Cells(j + 1, i) = S(i, j)
            Next j
        Next i
    Else
        For j = 0 To nSamples
            Cells(j + 1, 1) = S(1, j)
        Next j
    End If

End Sub
```

Equivalently, $S(t)$ is a geometric Brownian motion starting at $S(0) = z$ if

$$S(t) = ze^{X(t)} = ze^{\nu t + \sigma W(t)} = ze^{(\mu - \frac{1}{2}\sigma^2)t + \sigma W(t)}.$$

Using this definition, we see that for $t_0 < t_1 < \cdots < t_n$, the successive ratios

$$\frac{S(t_1)}{S(t_0)}, \frac{S(t_2)}{S(t_1)}, \cdots, \frac{S(t_n)}{S(t_{n-1})}$$

are independent random variables by virtue of the independent increment property of the Wiener process. The mean and variance of a geometric Brownian motion can be computed as in the lognormal distribution. Notice that since a Brownian motion is normally distributed, we conclude:

1. $\log S(t) = X(t) \sim \mathrm{N}(\log S(0) + \nu t, \sigma^2 t)$.
2. As $S(t) = S(0)e^{X(t)}$,

$$\begin{aligned}
\mathrm{E}(S(t)) &= \mathrm{E}(\mathrm{E}(S(t)|S(0) = z)) = \mathrm{E}(\mathrm{E}(ze^{(\nu t + \sigma W(t))}|S(0) = z)) \\
&= ze^{(\mu - \frac{1}{2}\sigma^2)t}\mathrm{E}(e^{\sigma W(t)}) \\
&= ze^{(\mu - \frac{1}{2}\sigma^2)t}\mathrm{E}(e^{\sigma\sqrt{t}\xi}) \qquad (\xi = W(t)/\sqrt{t} \sim \mathrm{N}(0,1)) \\
&= ze^{(\mu - \frac{1}{2}\sigma^2)t}e^{\frac{1}{2}\sigma^2 t} \\
&= ze^{\mu t} = S(0)e^{\mu t}.
\end{aligned}$$

This equation has an interesting economic implication in the case where μ is positive but small relative to σ^2. On one hand, if $\mu > 0$, then the

mean value $\mathrm{E}(S(t))$ tends to ∞ as t tends to ∞. On the other hand, if $0 < \mu < \frac{1}{2}\sigma^2$, then the process $X(t) = X(0) + (\mu - \frac{1}{2}\sigma^2)t + \sigma W(t)$ has a negative drift, i.e., it is drifting in the negative direction as t tends to ∞. It is intuitively clear that (which can be shown mathematically) $X(t)$ tends to $-\infty$. As a consequence, the original price $S(t) = S(0)e^{X(t)}$ tends to 0. The geometric Brownian motion $S(t)$ is drifting closer to zero as time goes on, yet its mean value $\mathrm{E}(S(t))$ is continuously increasing. This example demonstrates the fact that the mean function sometimes can be misleading in describing the process.

3. Similarly, we can show that

$$\mathrm{Var}(S(t)) = S(0)^2 e^{2\nu t + \sigma^2 t}(e^{\sigma^2 t} - 1) = S(0)^2 e^{2\mu t}(e^{\sigma^2 t} - 1).$$

2.4 ITÔ'S FORMULA

In the preceding section, we define $S(t)$ in terms of $\log S(t)$ as a Brownian motion. Although such a definition facilitates many of the calculations, it may sometimes be desirable to examine the behavior of the original price process $S(t)$ directly. To see how this can be done, first recall from calculus that

$$d \log S(t) = \frac{dS(t)}{S(t)}.$$

We might be tempted to substitute this elementary fact into (2.6) to get

$$\frac{dS(t)}{S(t)} = \nu\, dt + \sigma\, dW(t). \tag{2.7}$$

However, this computation is **NOT** exactly correct since it involves the differential $dW(t)$. A rule of thumb is that whenever we need to substitute quantities regarding $dW(t)$, there is a correction term that needs to be accounted for. We shall provide an argument of this correction term later. For the time being, the correct expression of the previous equation should be

$$\begin{aligned} \frac{dS(t)}{S(t)} &= (\nu + \frac{1}{2}\sigma^2)\, dt + \sigma\, dW(t) \\ &= \mu\, dt + \sigma\, dW(t), \end{aligned}$$

as $\nu = \mu - \frac{1}{2}\sigma^2$. The correction term required when transforming $\log S(t)$ to $S(t)$ is known as the **Itô's lemma**. We shall talk about this in the next theorem. Before doing that, there are a number of remarks.

Remarks

1. The term $dS(t)/S(t)$ can be thought of as the differential return of a stock and equation (2.7) says that the differential return possesses a simple form $\mu\, dt + \sigma\, dW(t)$.

2. Note that in (2.7), it is an equation about the ratio $dS(t)/S(t)$. This term can also be thought of as the instantaneous return of the stock. Hence equation (2.7) is describing the dynamics of the instantaneous return process.

3. In the case of a deterministic dynamics, i.e., without the stochastic component $dW(t)$ in (2.7), this equation reduces to the familiar form of a compound return. For example, let $P(t)$ denote the price of a bond that pays \$1 at time $t = T$. Assume the interest rate r is constant over time and there are no other payments before maturity, the price of the bond satisfies
$$\frac{dP(t)}{P(t)} = r\,dt.$$
In other words, $P(t) = P(0)e^{rt} = e^{r(t-T)}$, after taking the boundary condition $P(T) = 1$ into account.

4. Note that equation (2.7) provides a way to simulate the price process $S(t)$. Suppose we start at t_0 and let $t_k = t_0 + k\Delta t$. According to (2.7), the simulation equation is
$$S(t_{k+1}) - S(t_k) = \mu S(t_k)\,\Delta t + \sigma S(t_k)\epsilon(t_k)\sqrt{\Delta t},$$
where $\epsilon(t_k)$ are i.i.d. standard normal random variables. Iterating this equation we get
$$S(t_{k+1}) = [1 + \mu\,\Delta t + \sigma\epsilon t_k\sqrt{\Delta t}]S(t_k), \tag{2.8}$$
which is a multiplicative model, but the coefficient is normal rather than lognormal. So this equation does not generate the lognormal price distribution. However, when Δt is sufficiently small, the differences may be negligible.

5. Instead of using (2.7), we can use equation (2.6) for the log prices and get
$$\log S(t_{k+1}) - \log S(t_k) = \nu\,\Delta t + \sigma\epsilon(t_k)\sqrt{\Delta t}.$$
This equation leads to
$$S(t_{k+1}) = e^{\nu\,\Delta t + \sigma\epsilon(t_k)\sqrt{\Delta t}}S(t_k), \tag{2.9}$$
which is also a multiplicative model, but now the random coefficient is lognormal. In general, we can use either (2.8) or (2.9) to simulate stock prices.

With these backgrounds, we are now ready to state the celebrated Itô's lemma, which accounts for the correction term.

Theorem 2.1 *Suppose the random process $x(t)$ satisfies the diffusion equation*

$$dx(t) = a(x,t)\,dt + b(x,t)\,dW(t),$$

where $W(t)$ is a standard Brownian motion. Let the process $y(t) = F(x,t)$ for some function F. Then the process $y(t)$ satisfies the Itô's equation

$$dy(t) = (\frac{\partial F}{\partial x}a + \frac{\partial F}{\partial t} + \frac{1}{2}\frac{\partial^2 F}{\partial x^2}b^2)\,dt + \frac{\partial F}{\partial x}b\,dW(t). \tag{2.10}$$

Proof. Observe that if the process is deterministic, ordinary calculus shows that for a function of two variables like $y(t) = F(x,t)$, the total differential dy is given by

$$\begin{aligned} dy &= \frac{\partial F}{\partial x}\,dx + \frac{\partial F}{\partial t}\,dt \\ &= \frac{\partial F}{\partial x}(a\,dt + b\,dW) + \frac{\partial F}{\partial t}\,dt. \end{aligned}$$

Comparing this expression with (2.10), we see that there is an extra correction term $\frac{1}{2}\frac{\partial^2 F}{\partial x^2}b^2$ in front of dt. To see how this term arises, consider expanding the function F in a Taylor's expansion up to terms of first order in Δt. Note that since ΔW and hence Δx are of order $\sqrt{\Delta t}$, such an expansion would lead to terms with the second order in Δx. In this case,

$$\begin{aligned} y + \Delta y &= F(x,t) + \frac{\partial F}{\partial x}\Delta x + \frac{\partial F}{\partial t}\Delta t + \frac{1}{2}\frac{\partial^2 F}{\partial x^2}(\Delta x)^2 \\ &= F(x,t) + \frac{\partial F}{\partial x}(a\,\Delta t + b\,\Delta W) + \frac{\partial F}{\partial t}\Delta t + \frac{1}{2}\frac{\partial^2 F}{\partial x^2}(a\,\Delta t + b\,\Delta W)^2. \end{aligned}$$

Now focus at the quadratic expression of the last term. When expanded, it becomes

$$a^2(\Delta t)^2 + 2ab(\Delta t)(\Delta W) + b^2(\Delta W)^2.$$

The first two terms of the above expression are of orders higher than Δt, so they can be dropped as we only want terms up to the order of Δt. The last term $b^2(\Delta W)^2$ is all that remains. Recall that $\Delta W \sim \mathrm{N}(0, \Delta t)$ (recall the earlier fact that $dW(t) = \epsilon(t)\sqrt{dt}$), it can be shown that $(\Delta W)^2 \to \Delta t$. In other words, we have the following approximation

$$dW(t)^2 \cong dt \quad \text{or} \quad dW(t) \cong \sqrt{dt}.$$

Substituting this into the expansion, we have

$$y + \Delta y = F(x,t) + (\frac{\partial F}{\partial x}a + \frac{\partial F}{\partial t} + \frac{1}{2}\frac{\partial^2 F}{\partial x^2}b^2)\,\Delta t + \frac{\partial F}{\partial x}b\,\Delta W.$$

Taking limit as $\Delta t \to 0$ and noting $y(t) = F(x,t)$ complete the proof. □

Example 2.1 *Suppose $S(t)$ satisfies the geometric Brownian motion equation*

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

Now use Itô's formula to find the equation governing the process $F(S(t)) = \log S(t)$. Using (2.10), we identify $a = \mu S$ and $b = \sigma S$. Further, we know that $\partial F/\partial S = 1/S$ and $\partial^2 F/\partial S^2 = -1/S^2$. According to (2.10), we get

$$d\log S = (\frac{a}{S} - \frac{1}{2}\frac{b^2}{S^2})\,dt + \frac{b}{S}\,dW = (\mu - \frac{1}{2}\sigma^2)\,dt + \sigma\,dW,$$

which agrees with the earlier discussion. □

Example 2.2 *Evaluate*

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

To evaluate this integral, let us first guess the answer to be the one given by the classical integration by parts formula. That is, we might guess $tW(t) - \int_0^t W(s)\,ds$ to be the answer. To verify it, we need to differentiate this quantity to see if it matches the answer. To do this, use the following steps:

1. Let $X(t) = W(t)$, then $dX(t) = dW(t)$ and we identify $a = 0$ and $b = 1$ in (2.10).
2. Let $Y(t) = F(W(t)) = tW(t)$. Then $\partial F/\partial W = t$, $\partial^2 F/\partial W^2 = 0$, and $\partial F/\partial t = W(t)$.
3. Substitute these expressions into Itô's Lemma, we have $dY(t) = t\,dW(t) + W(t)\,dt$.
4. Integrating the preceding equation, we have

$$Y(t) = \int_0^t s\,dW(s) + \int_0^t W(s)\,ds,$$

that is,

$$\int_0^t s\,dW(s) = tW(t) - \int_0^t W(s)\,ds,$$

as required.

□

Example 2.3 *Evaluate*

$$\int_0^t W(s)\,dW(s).$$

First guess an answer, $W^2(t)/2$, say. Is this answer correct? To check, we differentiate again and apply Itô's Lemma. Using the recipe,

1. Let $X(t) = W(t)$, then $dX(t) = dW(t)$ and we identify $a = 0$ and $b = 1$ in (2.10).

2. Let $Y(t) = F(W(t)) = W^2(t)/2$. Then $\partial F/\partial W = W$, $\partial^2 F/\partial W^2 = 1$, and $\partial F/\partial t = 0$.

3. Recite Itô's Lemma:

$$dY(t) = [\frac{\partial F}{\partial X}a + \frac{\partial F}{\partial t} + \frac{1}{2}\frac{\partial^2 F}{\partial X^2}b^2]\,dt + \frac{\partial F}{\partial X}b\,dW(t),$$

so that

$$dY(t) = \frac{1}{2}\,dt + W(t)\,dW(t).$$

4. Integrating the preceding equation, we get

$$W^2(t)/2 = Y(t) = \frac{t}{2} + \int_0^t W(s)\,dW(s).$$

In other words,

$$\int_0^t W(s)\,dW(s) = \frac{W^2(t)}{2} - \frac{t}{2}!!!$$

5. This time, our initial guess was not correct. We need the extra correction term $\frac{t}{2}$ from Itô's Lemma.

□

Example 2.4 *Let W_t be a standard Brownian motion and let $Y_t = W_t^3$. Evaluate dY_t.*

Let $X_t = W_t$ and $F(X,t) = X_t^3$. Then the diffusion is $dX_t = dW_t$ with $a = 0$ and $b = 1$. Further

$$\frac{\partial F}{\partial X} = 3X^2, \quad \frac{\partial^2 F}{\partial X^2} = 6X, \quad \frac{\partial F}{\partial t} = 0.$$

Using Itô's lemma, we have

$$dY_t = 3W_t\,dt + 3W_t^2\,dW_t.$$

Integrating both sides of this equation, we get

$$\begin{aligned} \int_0^t dY_s &= \int_0^t 3W_s\,ds + \int_0^t 3W_s^2\,dW_s, \\ Y_t = W_t^3 &= 3\int_0^t W_s\,ds + 3\int_0^t W_s^2\,dW_s, \end{aligned}$$

In other words,

$$\int_0^t W_s^2\, dW_s = \frac{W_t^3}{3} - \int_0^t W_s\, ds.$$

In general, one gets

$$\int_0^t W_s^m\, dW_s = \frac{W_t^{m+1}}{m+1} - \frac{m}{2}\int_0^t W_s^{m-1}\, ds, \quad m = 0, 1, 2, \ldots . \tag{2.11}$$

□

Example 2.5 *Let*

$$dX_t = \frac{1}{2}X_t\, dt + X_t\, dW_t. \tag{2.12}$$

Evaluate $d\log X_t$.

From the given diffusion, we have $a = \frac{X_t}{2}$ and $b = X_t$. Let $Y_t = F(X,t) = \log X_t$. Then

$$\frac{\partial F}{\partial X} = \frac{1}{X}, \quad \frac{\partial^2 F}{\partial X^2} = -\frac{1}{X^2}, \quad \frac{\partial F}{\partial t} = 0.$$

Using Itô's lemma, we get $dY_t = d\log X_t = dW_t$. That is, $Y_t = W_t$ Therefore, $X_t = e^{W_t}$ is a solution to (2.12). □

Example 2.6 *Let the diffusion be*

$$dX_t = \frac{1}{2}\, dt + dW_t. \tag{2.13}$$

Evaluate de^{X_t}.

From the given diffusion, we have again $a = \frac{X_t}{2}$ and $b = X_t$. Let $Y_t = F(X,t) = e^{X_t}$. Then

$$\frac{\partial F}{\partial X} = e^{X_t}, \quad \frac{\partial^2 F}{\partial X^2} = e^{X_t}, \quad \frac{\partial F}{\partial t} = 0.$$

Using Itô's lemma, we get $dY_t = e^{X_t}\, dt + e^{X_t}\, dW_t$ so that

$$dY_t = Y_t\, dt + Y_t\, dW_t.$$

□

Example 2.7 *Find the solution to the stochastic differential equation*

$$dX_t = X_t\, dt + dW_t, \quad X_0 = 0.$$

Multiplying the integrating factor e^{-t} to both sides of the SDE, we have

$$e^{-t}\, dX_t = e^{-t}X_t\, dt + e^{-t}\, dW_t.$$

Let $Y_t = e^{-t}X_t$. Then $Y_0 = 0$ and by means of Itô's lemma, we have

$$dY_t = e^{-t}\,dW_t.$$

Integrating both sides of this equation,

$$Y_t - Y_0 = \int_0^t e^{-s}\,dW_s,$$

so that

$$X_t = e^t Y_t = \int_0^t e^{(t-s)}\,dW_s.$$

More generally, if we are given the SDE

$$dX_t = \mu X_t\,dt + \sigma\,dW_t,$$

then using the same method by considering the process $Y_t = e^{\mu t}X_t$, it can be easily shown that the solution to this SDE is given by the process

$$X_t = \sigma \int_0^t e^{\mu(t-s)}\,dW_s + X_0.$$

Such a process is known as the Ornstein-Uhlenbeck process, which is often used in modeling bond prices. □

2.5 EXERCISES

1. Let W_t be a Wiener process. Now is at time t_0. Find the mean and variance of X_t if

   (a) $X_t = \sigma_1 (W_t - W_{t_2}) - \sigma_2 (W_{t_1} - W_{t_0})$, $t > t_2 > t_1 > t_0$.

   (b) $X_t = \sigma_1 (W_t - W_{t_2}) - \sigma_2 (W_{t_1} - W_{t_0})$, $t > t_1 > t_2 > t_0$.

   (c) $X_t = \sum_{j=1}^{n} f(W_{t_{j-1}}) \left(W_{t_j} - W_{t_{j-1}}\right)$, $t_0 < t_1 < \cdots t_n = t$.

   (d) Use (c) to show that

   $$\mathrm{E}\left[\int_0^t f(W_\tau, \tau)\,dW_\tau\right] = 0$$

   and

   $$\mathrm{E}\left[\int_0^t f(W_\tau, \tau)\,dW_\tau\right]^2 = \int_0^t \mathrm{E} f(W_\tau, \tau)^2\,d\tau.$$

   Notice that the above two identities are known as Itô's identities.

2. Let X_t satisfy the stochastic differential equation

$$dX_t = -\frac{1}{3}\,dt + \frac{1}{2}\,dW_t,$$

where $X_0 = 0$ and W_t is a standard Brownian motion process. Define $S_t = e^{X_t}$ so that $S_0 = 1$.

(a) Find the stochastic differential equation that governs S_t.

(b) Simulate 10 independent paths of S_t for $t = 1, \ldots, 30$. Call these paths $S_t^i, i = 1, \ldots, 10$ and plot them on the same graph.

(c) What can you conclude about S_t for t large?

(d) With $n = 10$, evaluate

$$\bar{S}_{30} = \frac{1}{n} \sum_{i=1}^{n} S_{30}^i \tag{2.14}$$

at $t = 30$.

(e) Simulate 100 independent paths and calculate (2.14) with $n = 100$. What can you conclude about $\bar{S}_{1000}$ when n tends to infinity?

3. A stock price is governed by

$$dS(t) = \alpha S(t)\, dt + \beta S(t)\, dW(t),$$

where α and β are given constants and $W(t)$ is a standard Brownian motion process. Find the stochastic differential equation that governs

$$G(t) = \sqrt{S(t)}.$$

4. Consider a stock price S governed by the geometric Brownian motion process

$$\frac{dS(t)}{S(t)} = 0.10\, dt + 0.30\, dW(t),$$

where $W(t)$ is a standard Brownian motion process.

(a) Using $\Delta t = 1/12$ and $S(0) = 1$, simulate 5,000 years of the process $\log S(t)$ and evaluate

$$\frac{1}{t} \log S(t) \tag{2.15}$$

as a function of t. Note that (2.15) tends to a limit p. What is the theoretical value of p? Does your simulation match with this value?

(b) Evaluate

$$\frac{1}{t} \{\log S(t) - pt\}^2 \tag{2.16}$$

as a function of t. Does this tend to a limit?

5. Simulate a standard Brownian motion process $W(t)$ at grids $0 < \frac{1}{n} < \frac{2}{n} < \cdots < \frac{n-1}{n} < 1$ with $n = 10,000$. Let $W_i = W(\frac{i}{n})$ for $i = 0, \ldots, n$ with $W(0) = 0$. Suppose you want to evaluate the integral

$$\int_0^1 W(s)\, dW(s) \tag{2.17}$$

via the approximating sum

$$S_\epsilon = \sum_{i=0}^{n-1} \{(1-\epsilon)W_i + \epsilon W_{i+1}\}\{W_{i+1} - W_i\}. \tag{2.18}$$

   (a) Based on simulated values of W_i, use (2.18) to evaluate (2.17) with $\epsilon = 0$. Does your result match with the one obtained from Itô's formula?

   (b) Based on simulated values of W_i, use (2.18) to evaluate (2.17) with $\epsilon = \frac{1}{2}$. This is known as the *Stratonovich integral*. Using your calculated results, can you guess the difference between Itô's integral and the Stratonovich integral?

6. Let W_t denote a standard Brownian motion process.

   (a) Let $Y_t = F(W_t) = e^{W_t}$. Write down the diffusion equation that governs Y_t.

   (b) Evaluate $\int_0^t e^{W_s}\, dW_s$.

7. Denote X_t as the Brownian motion with drift μ and volatility σ.

   (a) Find df and dg where $f(t, X) = tX_t$ and $g(t, X) = tX_t^2$.

   (b) Financial market practitioners usually consider the time average of the underlying asset price when making investment decision. If the asset evolves as a Brownian motion X_t, then the time average line can be viewed as a stochastic variable

$$A_t = \frac{1}{t}\int_0^t X_\tau\, d\tau.$$

   What is the distribution for A_t?

   (c) Suppose $X_0 = 70$, $\mu = 0.05$, and $\sigma = 0.4$. Simulate X_1 and A_1 with $\Delta t = 0.01$. What are the sample means and variances for X_1 and A_1 for 1,000 simulations? What is the covariance between the two random variables, X_1 and A_1?

   (d) Comment on your simulation result.

Simulation Techniques in Financial Risk Management
by Ngai Hang Chan and Hoi Ying Wong

3

Black-Scholes Model and Option Pricing

3.1 INTRODUCTION

In this chapter, we will apply Itô's Lemma to derive the celebrated option pricing formula by Black and Scholes in the early 1970s. This formula has far reaching consequences and plays a fundamental role in modern option pricing theory. Immediately after Black and Scholes, Merton strengthened and improved the option pricing theory in several ways. To recognize their contributions, Merton and Scholes were awarded the Nobel prize in economics in 1997.

What is an option? An option is a financial derivative (contingent claim) that gives the holder the right (but not the obligation) to buy or to sell an asset for a certain price by a certain date. The option gives the holder a purchasing right is termed a *call option* whereas the *put option* gives the holder the selling right. The price in the contract is known as the *exercise price* or *strike price* (K); the date is known as the expiration or maturity (T). American options can be exercised at any time up to expiration. European options can be exercised only on the expiration date. As option holders are given a right, they have to pay an option premium to enter the contract. This premium is usually known as the option price.

Four basic option positions are possible:

1. A long position in a call option. Payoff $= \max(S_T - K, 0)$.
2. A long position in a put option. Payoff $= \max(K - S_T, 0)$.
3. A short position in a call option. Payoff $= -\max(S_T - K, 0)$.

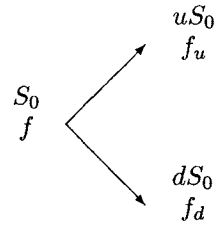


Fig. 3.1 One period binomial tree.

4. A short position in a put option. Payoff $= -\max(K - S_T, 0)$.

Notice that the long position in a put option is different from the short position of a call option. A long position in an option always has a non-negative payoff whereas a short position in an option always has a non-positive payoff, but the option premium is collected up front. Option pricing means determining the correct option premium.

To illustrate the Black-Scholes formula, we shall first discuss some fundamental concepts in a one period binomial model from which a risk-neutral argument is introduced.

3.2 ONE PERIOD BINOMIAL MODEL

Consider a binomial model in one period. Let S_0 and f denote the initial price of one share of a stock and an option on the stock. After one period, the price of the stock can either be uS_0 or dS_0 where $u > 1$ designates an upward movement of the stock price and $d < 1$ designates a downward movement of the stock price. Correspondingly, the payoff of the option after one period can either be f_u or f_d depending on whether the stock moves up or down. For instance, $f_u = \max(Su - K, 0)$ and $f_d = \max(Sd - K, 0)$ for a call option. Schematically, the one period outcome can be represented by Fig. 3.1.

Now consider constructing a hedging portfolio as follows. Suppose we long (buy and hold) Δ shares of the stock and short (sell) one call option (European). Suppose that the option lasts for one period T and, during the life of the option, the stock can either move up from S_0 to uS_0 or down from S_0 to dS_0. Further, suppose that the risk-free rate in this period is denoted by r. The value of this hedging portfolio in the next period is

$$\begin{aligned} &\Delta u S_0 - f_u, && \text{if stock moves up,} \\ &\Delta d S_0 - f_d, && \text{if stock moves down.} \end{aligned}$$

This portfolio will be risk-free if Δ is chosen so that the value of this portfolio is the same at the end of one period regardless of the stock going up or down, i.e.,

$$\Delta u S_0 - f_u = \Delta d S_0 - f_d.$$

Solving for Δ, we get

$$\Delta = \frac{f_u - f_d}{uS_0 - dS_0}.$$

Since this portfolio is risk-free in the sense that it attains the same value regardless of the outcome of the stock, it must earn the risk-free rate. Otherwise, one could take advantage of an arbitrage opportunity. For example, if the return of this hedging portfolio is larger than the risk-free rate, one could borrow money from the bank to purchase this portfolio and lock in the fixed return. After one period, the proceeds from the portfolio can be used to repay the loan and the arbitrageur pockets the difference. Consequently, the present value of this portfolio must equal to $(\Delta u S_0 - f_u)e^{-rT}$. If we let f denote the value of the option at present, then the present value of the portfolio is $S_0\Delta - f$ and according to the no arbitrage assumption,

$$S_0\Delta - f = (\Delta u S_0 - f_u)e^{-rT}.$$

Consequently,

$$\begin{aligned}
f &= S_0\Delta - (\Delta u S_0 - f_u)e^{-rT} \\
&= S_0\Delta(1 - ue^{-rT}) + f_u e^{-rT} \\
&= \frac{f_u - f_d}{u - d}(1 - ue^{-rT}) + f_u e^{-rT} \\
&= e^{-rT}[e^{rT}\frac{f_u - f_d}{u - d}(1 - ue^{-rT}) + f_u] \\
&= e^{-rT}(e^{rT}\frac{f_u - f_d}{u - d} - u\frac{f_u - f_d}{u - d} + f_u\frac{u - d}{u - d}) \\
&= e^{-rT}(f_u\frac{e^{rT} - d}{u - d} + f_d\frac{u - e^{rT}}{u - d}) \\
&= e^{-rT}[p f_u + (1 - p)f_d],
\end{aligned}$$

where $p = \frac{e^{rT} - d}{u - d}$. This identity has a very natural interpretation. If we let the value p, just defined as the probability of the stock, move up in a risk-neutral world, then the above formula simply states the fact that, in the risk-neutral world,

$$f = e^{-rT}\hat{\mathrm{E}}(f) = e^{-rT}(p f_u + (1 - p)f_d),$$

i.e., the expected value of the option in one period discounted by the risk-free rate equals to the present value of the option. Note that the expected value in this case is denoted by $\hat{\mathrm{E}}$, which is the expectation taken under the new probability measure p. For this reason, p is known as the risk-neutral probability. The same reasoning can be used to evaluate the stock itself. Note

that

$$\begin{aligned}\hat{\mathrm{E}}(S_1) &= puS_0 + (1-p)dS_0 \\ &= pS_0(u-d) + dS_0 \\ &= \frac{e^{rT}-d}{u-d}S_0(u-d) + dS_0 \\ &= e^{rT}S_0.\end{aligned}$$

In other words, the stock grows like a risk-free rate under the risk-neutral probability (in the risk-neutral world). Therefore, setting the probability of the stock price moving up to be p is tantamount to assuming that the return of the stock grows like the risk-free rate in a risk-neutral world. In a risk-neutral world, all individuals are indifferent to risk and require no compensation for risk. The expected return of all securities is the risk-free interest rate. It is for this reason that such a computation is usually known as the risk-neutral valuation and it is equivalent to the no arbitrage assumption in general.

Example 3.1 *Suppose the current price of one share of a stock is \$20 and in a period of three months, the price will either be \$22 or \$18. Suppose we sold a European call option with a strike price of \$21 in three months. Let the annual risk-free rate be 12% and let p denote the probability that the stock moves up in 3 months in the risk-neutral world. Note that the payoff of the option is either* $f_u = \$1$ *if the stock moves up or* $f_d = \$0$ *if the stock moves down. How much is the option, f, worth today? To find f, we can use the risk-neutral valuation method. Recall that from the above discussion,*

$$22p + 18(1-p) = 20e^{0.12/4},$$

so that $p = 0.6523$. *Using the expected payoff of the option, we get*

$$\hat{\mathrm{E}}(f) = pf_u + (1-p)f_d = p + (1-p)0 = p = 0.6523.$$

Therefore, the value of the option for today is

$$f = e^{-rT}\hat{\mathrm{E}}(f) = e^{-0.12/4}(pf_u + (1-p)f_d) = 0.633.$$

□

Alternatively, we can try to solve the same problem using the arbitrage-free argument.

Example 3.2 *With the same parameters as in the preceding example, consider solving for* Δ. *First, since we want a risk-free profit for the hedging portfolio, we want to purchase* Δ *shares of the stock and short one European call option expiring in three months. After three months, the value of the portfolio can either be*

$$\begin{aligned} 22\Delta - 1, &\quad \textit{if the stock price moves to } \$22, \\ \textit{or} & \\ 18\Delta, &\quad \textit{if the stock price moves to } \$18. \end{aligned}$$

This portfolio is risk-free if Δ *is chosen so that the value of the portfolio remains the same for both alternatives, i.e.,*

$$22\Delta - 1 = 18\Delta \quad \textit{which means} \quad \Delta = 0.25.$$

The value of the portfolio in three months becomes

$$22 \times 0.25 - 1 = 4.5 = 18 \times 0.25.$$

By the no arbitrage consideration, this risk-free profit must earn the risk-free interest rate. In other words, the value of the portfolio today must equal to the present value of \$4.5*, i.e.,* $4.5e^{-0.12/4} = 4.367$*. If the value of the option today is denoted by* f*, then the present value of the portfolio equals to*

$$20 \times 0.25 - f = 4.5e^{-0.12/4} = 4.367.$$

Solving for f *gives*

$$f = 0.633,$$

which matches with the answer of the preceding example. □.

In general, this principle can be applied to a multi-period binomial tree. We will not go into the analysis of a multi-period model and refer the readers to Chapter 11 of Hull (2006) for further details. For a comprehensive discussion on the discrete-time approach, see Pliska (1997). Although these two examples are illustrated with a call option, by the same token, the same principle can be used to price a put option, again details can be found in Hull (2006).

3.3 THE BLACK-SCHOLES-MERTON EQUATION

The Black-Scholes option pricing equation has initiated modern theory of finance. Its development has triggered an enormous amount of research and revolutionized the practice of finance. The equation was developed under the assumption that the price fluctuation of the underlying security can be described by a diffusion process studied earlier. The logic behind the equation is conceptually identical to the binomial lattice: at each moment two available securities are combined to construct a portfolio that reproduces the local behavior of a contingent claim. Historically, the Black-Scholes theory predates the binomial lattice.

To begin, let S denote the price of an underlying security (stock) governed by a geometric Brownian motion over a time interval $[0, T]$ by

$$dS = \mu S\, dt + \sigma S\, dW, \tag{3.1}$$

where W is a standard Brownian motion process. Assume further that there is also a risk-free asset (bond) carrying an interest rate r over the time interval $[0, T]$ such that

$$dB = rB\, dt. \tag{3.2}$$

Consider a contingent claim that is a derivative (call option) of S. The price of this derivative is a function of S and t, i.e., let $f(S,t)$ be the price of the claim at time t when the stock price is S. Our goal is to find an equation that models the behavior of $f(S,t)$. This goal is attained by the celebrated Black-Scholes-Merton equation.

Theorem 3.1 *Using the notation just defined, and assuming that the price and the bond are described by the geometric Brownian motion* (3.1) *and the compound interest rate model* (3.2) *respectively, then the price of the derivative of this security satisfies*

$$\frac{\partial f}{\partial t} + \frac{\partial f}{\partial S} rS + \frac{1}{2}\frac{\partial^2 f}{\partial S^2}\sigma^2 S^2 = rf. \tag{3.3}$$

Proof. The idea of this proof is the same as the binomial lattice. In deriving the binomial model, we form a portfolio with portions of the stock and the bond so that the portfolio exactly matches the return characteristics of the derivative in a period-by-period manner. In the continuous-time framework, the matching is done at each instant. Specifically, by Itô's lemma, recall that

$$df = (\frac{\partial f}{\partial S}\mu S + \frac{\partial f}{\partial t} + \frac{1}{2}\frac{\partial^2 f}{\partial S^2}\sigma^2 S^2)\,dt + \frac{\partial f}{\partial S}\sigma S\,dW. \tag{3.4}$$

This is also a diffusion process for f with drift $(\frac{\partial f}{\partial S}\mu S + \frac{\partial f}{\partial t} + \frac{1}{2}\frac{\partial^2 f}{\partial S^2}\sigma^2 S^2)$ and diffusion coefficient $\frac{\partial f}{\partial S}\sigma S$.

Construct a portfolio of S and B that replicates the behavior of the derivative. At each time t, we select an amount x_t of the stock and an amount y_t of the bond, giving a total portfolio value of $G(t) = x_t S(t) + y_t B(t)$. We wish to select x_t and y_t so that $G(t)$ replicates the derivative value $f(S,t)$. The instantaneous gain in value of this portfolio due to changes in security prices is

$$\begin{aligned} dG &= x_t\,dS + y_t\,dB \\ &= x_t(\mu S\,dS + \sigma S\,dW) + y_t rB\,dt \\ &= (x_t\mu S + y_t rB)\,dt + x_t\sigma S\,dW. \end{aligned} \tag{3.5}$$

Since we want the portfolio gain of $G(t)$ to behave like the gain of f, we match the coefficients of dt and dW in (3.5) to those of (3.4). First, we match the coefficient of dW in these two equations and we get

$$x_t = \frac{\partial f}{\partial S}.$$

Second, since $G(t) = x_t S(t) + y_t B(t)$, we get

$$y_t = \frac{1}{B(t)}(G(t) - x_t S(t)).$$

Third, remember we want $G = f$, therefore,

$$y_t = \frac{1}{B(t)} \left(f(S,t) - \frac{\partial f}{\partial S} S(t) \right).$$

Substituting this expression into (3.5) and matching the coefficient of dt in (3.4), we have

$$\frac{\partial f}{\partial S} \mu S + \frac{1}{B(t)} \left(f(S,t) - \frac{\partial f}{\partial S} S(t) \right) rB(t) = \frac{\partial f}{\partial S} \mu S + \frac{\partial f}{\partial t} + \frac{1}{2} \frac{\partial^2 f}{\partial S^2} \sigma^2 S^2.$$

Consequently,

$$\frac{\partial f}{\partial t} + \frac{\partial f}{\partial S} rS + \frac{1}{2} \frac{\partial^2 f}{\partial S^2} \sigma^2 S^2 = rf.$$

□

Remarks

1. If $f(S,t) = S$, then $\frac{\partial f}{\partial t} = 0, \frac{\partial f}{\partial S} = 1$, and $\frac{\partial^2 f}{\partial S^2} = 0$ and equation (3.3) reduces to $rS = rS$ so that $f(S,t) = S$ is a solution to (3.3).

2. As another simple example, consider a bond where $f(S,t) = e^{rt}$. This is a trivial derivative of S and it can be easily shown that this f satisfies (3.3).

3. In general, (3.3) provides a way to price a derivative by using the appropriate boundary conditions. Consider a European call option with strike price K and maturity T. Let the price be $C(S,t)$. Clearly, this derivative must satisfy

$$\begin{aligned} C(0,t) &= 0, \\ C(S,T) &= \max(S-K, 0). \end{aligned}$$

   For a European put option, the boundary conditions are

$$\begin{aligned} P(\infty,t) &= 0, \\ P(S,T) &= \max(K-S, 0). \end{aligned}$$

   Other derivatives may have different boundary conditions. For a knock-out option that will be canceled if the underlying asset breaches a pre-specified barrier level (H), in addition to the above conditions, we have an extra boundary condition

$$f(S = H, t) = 0.$$

4. With these boundary conditions, one can try to solve for the function f from the Black-Scholes equation. One problem is that this is a partial differential equation and there is no guarantee that an analytical solution

exists. Except in the simple case of a European option, one cannot find an analytic formula for the function f. In practice, either simulation or numerical methods have to be used to find an approximate solution.

5. Alternatively, we can derive equation (3.3) as follows. Construct a portfolio that consists of shorting one derivative and longing $\frac{\partial f}{\partial S}$ shares of the stock. Let the value of this portfolio be Π and let the value of the derivative be $f(S,t)$. Then

$$\Pi = -f + \frac{\partial f}{\partial S} S. \tag{3.6}$$

The change $\Delta\Pi$ in the value of this portfolio in the time interval Δt is given by

$$\Delta\Pi = -\Delta f + \frac{\partial f}{\partial S} \Delta S. \tag{3.7}$$

Recall that S follows a geometric Brownian motion so that

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

Also, from (3.4), the discrete version of df is

$$\Delta f = (\frac{\partial f}{\partial S} \mu S + \frac{\partial f}{\partial t} + \frac{1}{2} \frac{\partial^2 f}{\partial S^2} \sigma^2 S^2) \, \Delta t + \frac{\partial f}{\partial S} \sigma S \, \Delta W.$$

Substituting these two expressions into equation (3.7), we get

$$\Delta\Pi = (-\frac{\partial f}{\partial t} - \frac{1}{2} \frac{\partial^2 f}{\partial S^2} \sigma^2 S^2) \, \Delta t. \tag{3.8}$$

Note that by holding such a portfolio, the random component ΔW has been eliminated completely. Because this equation does not involve ΔW, this portfolio must equal to the risk-free rate during the time Δt. Consequently,

$$\Delta\Pi = r\Pi \, \Delta t,$$

where r is the risk-free rate. In other words, using (3.8) and (3.6), we obtain

$$(\frac{\partial f}{\partial t} + \frac{1}{2} \frac{\partial^2 f}{\partial S^2} \sigma^2 S^2) \, \Delta t = r(f - \frac{\partial f}{\partial S} S) \, \Delta t.$$

Therefore,

$$\frac{\partial f}{\partial t} + \frac{\partial f}{\partial S} rS + \frac{1}{2} \frac{\partial^2 f}{\partial S^2} \sigma^2 S^2 = rf.$$

It should be noted that the portfolio used in deriving (3.3) is not permanently risk-free. It is risk-free only for an infinitesimally short period of time. As S and t change, $\frac{\partial f}{\partial S}$ also changes. To keep the portfolio risk-free, we have to

change the relative proportions of the derivative and the stock in the portfolio continuously.

Example 3.3 *Let f denote the price of a forward contract on a non-dividend-paying stock with delivery price K and delivery date T. Its price at time t is given by*

$$f(S,t) = S - Ke^{-r(T-t)}. \tag{3.9}$$

Hence,

$$\frac{\partial f}{\partial t} = -rKe^{-r(T-t)}, \quad \frac{\partial f}{\partial S} = 1, \quad \textit{and} \quad \frac{\partial^2 f}{\partial S^2} = 0.$$

Substituting these into (3.3), *we get*

$$-rKe^{-r(T-t)} + rS = rf.$$

Thus, the price formula of f given by (3.9) *is a solution of the Black-Scholes equation, indicating that* (3.9) *is the correct formula.* □

The Black-Scholes equation generates two important insights. The first one is the concept of risk-neutral pricing. As the Black-Scholes equation does not involve the drift, μ, of the underlying asset price, the option pricing formula should be independent of the drift. Therefore, individual preferences toward the performance or the trend of a particular asset price does not affect the current price of the option on that asset. The second insight is that one would be able to derive a price representation of a European option with any payoff function from the equation. It is summarized in the following theorem.

Theorem 3.2 *Consider a European option with payoff $F(S)$ and expiration time T. Suppose the continuous compounding interest rate is r. Then, the current European option price is determined by*

$$f(S,0) = e^{-rT}\widehat{\mathrm{E}}[F(S_T)], \tag{3.10}$$

where $\widehat{\mathrm{E}}$ denotes the expectation under the risk-neutral probability that is derived from the risk-neutral process

$$\frac{dS}{S} = r\,dt + \sigma\,dW(t). \tag{3.11}$$

Proof. Notice that the current price of the option $f(S,0)$ is a deterministic function of time $t = 0$ and the current asset price S. Consider a stochastic process $\{X_t\}$ that satisfies

$$X_0 = S \quad \text{and} \quad \frac{dX_t}{X_t} = r\,dt + \sigma\,dW(t).$$

Then, $f(S,0) = f(X,0)$. Consider the process $f(X,t)$ derived from the stochastic process of $\{X_t\}$. By Itô's lemma, the differential form of f is

$$df = \left(\frac{\partial f}{\partial t} + rX\frac{\partial f}{\partial X} + \frac{1}{2}\sigma^2 X^2 \frac{\partial^2 f}{\partial X^2}\right) dt + \sigma X \frac{\partial f}{\partial X}\, dW.$$

The Black-Scholes equation says that the coefficient of dt is identical to the term rf, see Theorem 3.1. The total differential for the pricing function is simplified as

$$df = rf\, dt + \sigma X \frac{\partial f}{\partial X}\, dW,$$

which implies

$$df - rf\, dt = \sigma X \frac{\partial f}{\partial X}\, dW.$$

The left-hand side of the above equation can be combined with the product rule of differentiation to yield

$$e^{rt}\, d\left[e^{-rt} f(X,t)\right] = \sigma X \frac{\partial f}{\partial X}\, dW.$$

This expression has an equivalent integration form,

$$e^{-rT} f(X_T, T) - f(X,0) = \sigma \int_0^T e^{-rt} X \frac{\partial f}{\partial X}\, dW.$$

The right-hand side is a sum of Gaussian processes so that it has an expected value of zero. A fter taking expectation on both sides,

$$\widehat{\mathrm{E}}\left[e^{-rT} f(X_T, T) - f(X,0)\right] = 0.$$

This implies

$$f(X,0) = e^{-rT}\, \widehat{\mathrm{E}}[f(X_T, T)].$$

By the terminal condition specified in the Black-Scholes equation, $f(X_T, T) = F(X_T)$, the payoff of the option contract. Hence, we have

$$f(S,0) = e^{-rT}\, \widehat{\mathrm{E}}[F(X_T)],$$

where the expectation with respect to the random variable X_T is called the risk-neutral expectation and the process $\{X_t\}$ is called the risk-neutral asset dynamics. To avoid confusion, financial economists always use the term "asset price process in the risk-neutral world (S_t)" to represent the X_t in this proof. It establishes (3.10) and (3.11) and completes the proof.

□

3.4 BLACK-SCHOLES FORMULA

We are now ready to state the pricing formula of a European call option. A corresponding formula can also be deduced for a European put option. We first establish a key fact about lognormal random variables.

Lemma 3.1 *Let S be a lognormally distributed random variable such that $\log S \sim \mathrm{N}(m, \nu^2)$ and let $K > 0$ be a given constant. Then*

$$\mathrm{E}(\max\{S - K, 0\}) = \mathrm{E}(S)\Phi(d_1) - K\Phi(d_2), \tag{3.12}$$

where $\Phi(\cdot)$ denotes the distribution function of a standard normal random variable and

$$\begin{aligned} d_1 &= \frac{1}{\nu}(-\log K + m + \nu^2) = \frac{1}{\nu}\log\left(\mathrm{E}(\frac{S}{K}) + \frac{\nu^2}{2}\right), \\ d_2 &= \frac{-\log K + m}{\nu} = \frac{1}{\nu}\log\left(\mathrm{E}(\frac{S}{K}) - \frac{\nu^2}{2}\right). \end{aligned}$$

Proof. Let $g(s)$ denote the p.d.f. of the random variable S. Then

$$\mathrm{E}(\max(S - K, 0)) = \int_0^\infty \max(s - K, 0)g(s)\,ds = \int_K^\infty (s - K)g(s)\,ds.$$

By definition, since $\log S \sim \mathrm{N}(m, \nu^2)$,

$$\mathrm{E}(S) = e^{(m + \frac{1}{2}\nu^2)} \quad \text{so that} \quad \log \mathrm{E}(S) = m + \frac{1}{2}\nu^2.$$

Define the variable Q as

$$Q = \frac{\log S - m}{\nu} \quad \text{so that} \quad Q \sim \mathrm{N}(0, 1).$$

The p.d.f. of Q is given by $\phi(q) = \frac{1}{\sqrt{2\pi}}e^{-\frac{q^2}{2}}$, the p.d.f. of a standard normal random variable. Since $q = \frac{\log s - m}{\nu}$, $s = e^{m+q\nu}$ so that $dq = \frac{ds}{s\nu}$. Therefore,

$$\begin{aligned} \mathrm{E}(\max(S - K, 0)) &= \int_K^\infty \max(s - K, 0)g(s)\,ds \\ &= \int_{\frac{1}{\nu}(\log K - m)}^\infty (e^{m+q\nu} - K)g(e^{m+q\nu})s\nu\,dq \\ &= \int_{\frac{1}{\nu}(\log K - m)}^\infty (e^{m+q\nu} - K)\phi(q)\,dq \\ &= \int_{\frac{1}{\nu}(\log K - m)}^\infty e^{m+q\nu}\phi(q)\,dq - K\int_{\frac{1}{\nu}(\log K - m)}^\infty \phi(q)\,dq \\ &= I - II. \end{aligned} \tag{3.13}$$

Note that the third equality follows from the fact that the p.d.f. g of a lognormal random variable S has the form

$$g(s) = \phi(\frac{\log s - m}{\nu})\frac{1}{s\nu}, \quad \text{so that } g(e^{m+q\nu})s\nu = \phi(q). \tag{3.14}$$

We now analyze each of the terms I and II in (3.13). Consider the first term,

$$\begin{aligned} I &= e^{m+\frac{\nu^2}{2}} \int_{\frac{1}{\nu}(\log K - m)-\nu}^{\infty} \phi(q-\nu)\, d(q-\nu) \\ &= e^{m+\frac{\nu^2}{2}}(1 - \Phi(\frac{\log K - m}{\nu} - \nu)) \\ &= e^{m+\frac{\nu^2}{2}}\Phi(\frac{-\log K + m}{\nu} + \nu). \end{aligned}$$

For the second term, we have

$$II = K\int_{\frac{1}{\nu}(\log K - m)}^{\infty} \phi(q)\, dq = K\Phi(\frac{-\log K + m}{\nu}).$$

Substituting these two expressions into (3.13), we have

$$\mathrm{E}(\max(S-K,0)) = e^{m+\frac{\nu^2}{2}}\Phi(\frac{-\log K + m}{\nu} + \nu) - K\Phi(\frac{-\log K + m}{\nu}).$$

Observe that since $\log \mathrm{E}(S/K) = -\log K + m + \frac{\nu^2}{2}$,

$$\begin{aligned} \frac{-\log K + m}{\nu} + \nu &= \frac{-\log K + m + \nu^2}{\nu} \\ &= \frac{1}{\nu}(\log \mathrm{E}(S/K) + \frac{\nu^2}{2}) \\ &= d_1. \end{aligned}$$

Similarly, it can be easily shown that

$$d_2 = \frac{-\log K + m}{\nu}.$$

This completes the proof of the lemma. □

Using this lemma, we are now ready to state the Black-Scholes pricing formula.

Theorem 3.3 *Consider a European call option with strike price K and expiration time T. If the underlying stock pays no dividends during the time $[0,T]$ and if there is a continuously compounded risk-free rate r, then the price of this contract at time 0, $f(S,0) = C(S,0)$, is given by*

$$C(S,0) = S\,\Phi(d_1) - Ke^{-rT}\Phi(d_2), \tag{3.15}$$

where $\Phi(x)$ denotes the cumulative distribution function of a standard normal random variable evaluated at the point x,

$$\begin{aligned} d_1 &= [\log(S/K) + (r + \sigma^2/2)T]\frac{1}{\sigma\sqrt{T}}, \\ d_2 &= [\log(S/K) + (r - \sigma^2/2)T]\frac{1}{\sigma\sqrt{T}} \\ &= d_1 - \sigma\sqrt{T}. \end{aligned}$$

Proof. The proof of this result relies on the risk-neutral valuation. By Theorem 3.2, we have

$$C(S) = e^{-rT}\hat{\mathrm{E}}(\max\{S_T - K, 0\}), \tag{3.16}$$

where S_T denotes the stock price at time T, $\hat{\mathrm{E}}$ denotes the risk-neutral expectation, and

$$dS = rS\,dt + \sigma S\,d\hat{W}, \tag{3.17}$$

In this case, we have

$$\mathrm{E}S_T = S_0 e^{rT}. \tag{3.18}$$

From the preceding lemma, we get

$$\hat{\mathrm{E}}(\max\{S_T - K, 0\}) = \hat{\mathrm{E}}(S_T)\Phi(d_1) - K\Phi(d_2).$$

The remaining job is to identify d_1, d_2, and $\hat{\mathrm{E}}(S_T)$. By construction, $\hat{\mathrm{E}}S_T = S_0 e^{rT}$. Recall from (3.17), we can easily deduce from Itô's lemma that

$$d\log S_t = \gamma\,dt + \sigma\,d\hat{W}_t, \quad \text{with} \quad \gamma = r - \frac{1}{2}\sigma^2. \tag{3.19}$$

Consequently,

$$\begin{aligned} m &= \hat{\mathrm{E}}(\log S_T) = \log S_0 + rT - \frac{1}{2}\sigma^2 T, \\ \nu^2 &= \hat{\mathrm{Var}}(\log S_T) = \sigma^2 T. \end{aligned}$$

According to the lemma,

$$\begin{aligned} d_1 &= \frac{-\log K + m + \nu^2}{\nu} \\ &= \frac{1}{\sigma\sqrt{T}}[-\log K + \log S_0 + (r - \frac{1}{2}\sigma^2)T + \sigma^2 T] \\ &= \frac{1}{\sigma\sqrt{T}}[\log(\frac{S_0}{K}) + (r + \frac{1}{2}\sigma^2)T]. \end{aligned}$$

By similar substitutions, it can be easily shown that

$$d_2 = [\log(S/K) + (r - \sigma^2/2)(T)]\frac{1}{\sigma\sqrt{T}} = d_1 - \sigma\sqrt{T}.$$

This completes the proof of the Black-Scholes formula (3.15). □

Example 3.4 *Consider a five-month European call option on an underlying stock with a current price of* \$62, *strike price* \$60, *annual risk-free rate* 10%, *and the volatility of this stock is* 20% *per year. In this case,* $S = 62, K = 60, r = 0.1, \sigma = 0.2$, *and* $T = \frac{5}{12}$. *Applying* (3.15), *we get*

$$\begin{aligned} d_1 &= \frac{1}{0.2\sqrt{5/12}}\left[\log(\frac{62}{60}) + (0.1 + \frac{0.2^2}{2})\frac{5}{12}\right] \\ &= 0.641287, \\ d_2 &= d_1 - 0.2\sqrt{5/12} = 0.512188. \end{aligned}$$

From the normal table, we get $\Phi(d_1) = 0.739332$ *and* $\Phi(d_1) = 0.695740$. *Consequently,*

$$C = (62)(0.739332) - (60)e^{-(0.1)(5/12)}(0.695740) = 5.798.$$

□

Remarks

1. Note that the Black-Scholes pricing formula is derived using a risk-neutral valuation argument here. Alternatively, for a given derivative such as a European call option, we can try to solve the partial differential equation(PDE) given by the Black-Scholes equation (3.3) subject to the explicit boundary conditions given in Remark 3 in Section 6.3. This was the original idea of Black and Scholes and it is commonly known as the PDE approach. Although feasible, due to the complexity of the PDE of the Black-Scholes equation, the risk-neutral valuation argument offers a more intuitive approach based on the arbitrage-free argument.

2. For a European put option, the corresponding pricing formula is given by
$$P = Ke^{-rT}\Phi(-d_2) - S_0\Phi(-d_1),$$
where r, K, d_1, and d_2 are defined as in (3.15).

3. To interpret the Black-Scholes formula, look at what happens to d_1 and d_2 as $T \to 0$. If $S_0 > K$, they both tend to ∞ so that $\Phi(d_1) = \Phi(d_2) = 1$ and $\Phi(-d_1) = \Phi(-d_2) = 0$. This means that
$$C = S_0 - K \quad \text{and} \quad P = 0.$$

On the other hand, if $S_0 < K$, the reverse argument shows d_1 and d_2 tend to $-\infty$ as $T \to 0$ so that

$$C = 0 \text{ and } P = K - S_0.$$

Is this reasonable? When $S_0 > K$, and when $T = 0$, the call option should be worth $S_0 - K$ and the put option is of course worthless. On the other hand, if $S_0 < K$ and $T = 0$, the put option should be worth $K - S_0$ and the call option becomes worthless. Thus, the Black-Scholes formula offers the price that is consistent with the boundary condition.

4. What happens when $T \to \infty$? In this case, $d_1 = d_2 = \infty$ and $C = S_0, P = 0$. This is known as the perpetual call. If we own the call for a long time, the stock value will almost certainly increase to a very large value so that the strike price K is irrelevant. Hence, if we own the call we could obtain the stock later for essentially nothing, duplicating the position we would have if we initially bought the stock. Thus, $C = S_0$.

5. The Black-Scholes formula is derived for a European call option under the situation where the stock pays no dividends. When the underlying stock does pay dividends at a specific time during the life of the option, a similar formula to price the option can also be deduced. Again, we refer the interested readers to Hull (2006) for further details.

6. For an American option where early exercise is allowed, one can no longer find an exact analytic formula such as (3.15) for the price of a call. Instead, a range of possible values can be deduced and details are given in Hull (2006).

7. In using the Black-Scholes formula, one important quantity required is the value of σ, the volatility or the risk of the underlying stock. To use the formula, we can estimate σ from the historical data and put this estimate into the Black-Scholes equation. Such an approach is known as the historical volatility approach. On the other hand, one can also use the Black-Scholes formula to imply the value of σ, known as the implied volatility. In this latter approach, we substitute the observed price of the derivative as the real price into the Black-Scholes formula to solve for σ, giving it the name of implied volatility. This quantity can be used to monitor the market's opinion about the volatility of a particular stock. Analysts often calculate implied volatilities from actively traded options on a certain stock and use them to calculate the price of a less actively traded option on the same stock.

3.5 EXERCISES

1. A company's share price is now \$60. Six month from now, it will be either \$75 with risk-neutral probability 0.7 or \$50 with risk-neutral prob-

ability 0.3. A call option exists on the stock that can be exercised only at the end of six months with exercise price of \$65.

(a) If you wish to establish a perfectly hedged position, what would you do?

(b) Under each of the two possibilities, what will be the value of your hedged position?

(c) What is the expected value of option price at the end of the period?

(d) What is the reasonable option price today?

2. Consider the binomial model of Section 3.2.

(a) Show that the European call option price of the two period model is given by

$$c_2 = \left[p^2 c_{uu} + 2p(1-p)c_{ud} + (1-p)^2 c_{dd}\right] e^{-rT},$$

where T is the option maturity and

$$\begin{aligned} c_{uu} &= \max(Su^2 - K, 0) \\ c_{ud} &= \max(Sud - K, 0) \\ c_{dd} &= \max(Sd^2 - K, 0). \end{aligned}$$

(b) Show by induction that the n-period call price is given by

$$c_n = e^{-rT} \sum_{j=0}^{n} \left\{ {}_nC_j q^j (1-q)^{n-j} \max\left(Su^j d^{n-j} - K, 0\right) \right\}.$$

(c) Cox, Ross, and Rubinstein [CRR, 1979] propose that $u = e^{\sigma\sqrt{\Delta t}}$ and $d = e^{-\sigma\sqrt{\Delta t}}$, where σ is the annualized asset volatility, are respectively appropriate choices for the upward and downward factors in implementing the binomial model. Show that

$$\lim_{n\to\infty} c_n = S\,\Phi(d_1) - Ke^{-rT}\Phi(d_2),$$

the Black-Scholes call price, if the CRR proposal is adopted.

3. By Theorem 3.2, show the put-call parity relation

$$p + S = c + Ke^{-rT}.$$

4. A fixed strike geometric Asian call option has the payoff function $\max(G_T - K, 0)$ where

$$G_T = \exp\left(\frac{1}{T}\int_0^T \log S_t\, dt\right).$$

By Theorem 3.2 and Lemma 3.1, determine the analytical solution for the fixed strike geometric Asian call option. (Hints: 1. Apply the result of question 7(b) of Chapter 2, 2. You can find the answer in Chapter 7.)

5. Consider the partial differential equation:

$$\begin{aligned} \frac{\partial f}{\partial t} + \frac{1}{2}\sigma^2(t,x)\frac{\partial^2 f}{\partial x^2} + \mu(t,x)\frac{\partial f}{\partial x} + a(t,x)f &= 0 \\ f(T,x) &= F(x). \end{aligned}$$

By modifying the proof of Theorem 3.2, show that

$$f(t,x) = \mathrm{E}\left[e^{\int_t^T a(\tau)\,d\tau} F(X_T)\right],$$

where X_T is the solution to the SDE:

$$dX = \mu(\tau, X)\,d\tau + \sigma(\tau, X)\,dW_\tau, \;\; X_t = x.$$

This result is called the Feyman-Kac formula.

6. Suppose the risk-free interest rate and the volatility of an asset are deterministic functions of time. That means,

$$r = r(t) \qquad \text{and} \qquad \sigma = \sigma(t).$$

(a) Show that the Black-Scholes equation governing European option prices, $f(t,S)$, is given by

$$\frac{\partial f}{\partial t} + \frac{1}{2}\sigma^2(t)S^2\frac{\partial^2 f}{\partial S^2} + r(t)S\frac{\partial f}{\partial S} - r(t)f = 0.$$

(b) Show that the European call option price satisfies:

$$f(t,S) = e^{-\int_t^T r(\tau)\,d\tau}\widehat{\mathrm{E}}[\max(S_T - K, 0)],$$

where

$$dS_\tau = r(\tau)S_\tau\,d\tau + \sigma(\tau)S_\tau\,dW_\tau, \;\; \tau > t, \text{ and } S_t = S.$$

Hint: Use the result of question 5.

(c) Hence, show that

$$f(t,S) = C_{BS}(t,S; r = \bar{r}, \sigma = \bar{\sigma}),$$

where C_{BS} is the Black-Scholes formula for call option with constant parameters,

$$\bar{r} = \frac{1}{T-t}\int_t^T r(\tau)\,d\tau \qquad \text{and} \qquad \bar{\sigma} = \sqrt{\frac{1}{T-t}\int_t^T \sigma^2(\tau)\,d\tau}.$$

7. A stochastic process $X(t)$ is said to be a martingale under a probability measure $\mathcal{P}$ if $\mathrm{E}^{\mathcal{P}}[X(t)|X(s), s < \tau] = X(\tau)$, with probability one.

   (a) Consider the asset price dynamics under the risk-neutral measure:

   $$dS = rS\,dt + \sigma S\,dW.$$

   Show that $X(t) = S(t)\,e^{-rt}$ is a martingale.

   (b) Denote $C(t, S; T)$ as the Black-Scholes formula for a European call option with maturity T. Show that $Ce^{r(T-t)}$ is a martingale.

Simulation Techniques in Financial Risk Management
by Ngai Hang Chan and Hoi Ying Wong

4 Generating Random Variables

4.1 INTRODUCTION

The first stage of simulation is generation of random numbers. Random numbers serve as the building block of simulation. The second stage of simulation is generation of random variables based on random numbers. This includes generating both discrete and continuous random variables of known distributions. In this chapter, we shall study techniques for generating random variables.

4.2 RANDOM NUMBERS

Random numbers can be generated in a number of ways. For example, they were generated manually or mechanically by spinning wheels or rolling dice in the old days. Of course, the notion of randomness may be a subjective judgment. Things that look apparently random may not be random according to the strict definition. The modern approach is to use a computer to generate pseudo-random numbers successively. These pseudo-random numbers, although deterministically generated, constitute a sequence of values having the appearance of uniformly (0, 1) distributed random variables.

One of the most popular devices to generate uniform random numbers is the congruential generator. Starting with an initial value x_0, called the seed, the computer successively calculates the values $x_n, n \geq 1$ via

$$x_n = ax_{n-1} + c \text{ modulo } m, \tag{4.1}$$

where a, c, and m are given positive integers, and the equality means that the value ax_{n-1} is divided by m and the remainder is taken as the value of x_n. Each x_n is either $0, 1, \ldots, m-1$ and the quantity $\frac{x_n}{m}$ is taken as an approximation to the values of a uniform (0, 1) random variable. Since each of the numbers x_n assumes one of the values of $0, 1, \ldots, m-1$, it follows that after some finite number of generated values a value must repeat itself. For example, if we take $a = c = 1$ and $m = 16$, then

$$x_n = x_{n-1} + 1 \text{ modulo } 16.$$

With $x_0 = 1$, then the range of x_n is the set

$$\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 0, \ldots\}.$$

When $a = 5$, $c = 1$, and $m = 16$, then the range of x_n becomes

$$\{0, 1, 6, 15, 12, 13, 2, 11, 8, 9, 14, 7, 4, 5, 10, 3, 0, \ldots\}.$$

We usually want to choose a and m such that for any given seed x_0, the number of variables that can be generated before repetition occurs is large. In practice, one may choose $m = 2^{31} - 1$ and $a = 7^5$, where the number 31 corresponds to the bit size of the machine.

Any set of pseudo-random numbers will by definition fail on some problems. It is therefore desirable to have a second generator available for comparison. In this case, it may be useful to compare results for a fundamentally different generator.

From now on, we will assume that we can generate a sequence of random numbers that can be taken as an approximation to the values of a sequence of independent uniform (0, 1) random variables. We will not explore the technical details about the construction of good generators, interested reader may consult L'Ecuyer (1994) for a survey of random number generators.

4.3 DISCRETE RANDOM VARIABLES

A discrete random variable X is specified by its probability mass function given by

$$P(X = x_j) = p_j, \quad j = 0, 1, \ldots, \quad \sum_j p_j = 1. \tag{4.2}$$

To generate X, generate a random number U, which is uniformly distributed in (0, 1) and set

$$X = \begin{cases} x_0 & \text{if} \quad U < p_0, \\ x_1 & \text{if} \quad p_0 \le U < p_0 + p_1, \\ \vdots & \\ x_j & \text{if} \quad \sum_{i=1}^{j-1} p_i \le U < \sum_{i=1}^{j} p_i, \\ \vdots & \end{cases}$$

Recall that for $0 < a < b < 1$, $P(a < U < b) = b - a$. Thus,

$$P(X = x_j) = P(\sum_{i=1}^{j-1} p_i \leq U < \sum_{i=1}^{j} p_i) = p_j, \tag{4.3}$$

so that X has the desired distribution. Note that if the x_i are ordered so that $x_0 < x_1 < \cdots$ and if F denotes the distribution function of X, then $F(x_k) = \sum_{i=0}^{k} p_i$ and so

$$X \text{ equals to } x_j \text{ if } F(x_{j-1}) \leq U < F(x_j).$$

That is, after generating U, we determine the value of X by finding the interval $[F(x_{j-1}), F(x_j))$ in which U lies. This also means that we want to find the inverse of $F(U)$ and thus the name of inverse transform.

Example 4.1 *Suppose we want to generate a binomial random variable X with parameters n and p.*

The probability mass function of X is given by

$$p_i = P(X = i) = \frac{n!}{i!(n-i)!} p^i (1-p)^{n-i}, \quad i = 0, 1, \ldots, n.$$

From this probability mass function, we see that

$$p_{i+1} = \frac{n-i}{i+1} \frac{p}{1-p} p_i.$$

The algorithm goes as follows:

1. Generate U
2. If $U < p_0$, set $X = 0$ and stop
3. If $p_0 < U < p_0 + p_1$, set $X = 1$ and stop

   $\vdots$

4. If $p_0 + \cdots + p_{n-1} < U < p_0 + \cdots + p_n$, set $X = n$ and stop

Recursively, by letting i be the current value of X, $pr = p_i = P(X = i)$, and $F = F(i) = P(X \leq i)$, the probability that X is less than or equal to i, the above algorithm can be succinctly written as:

STEP 1: Generate U

STEP 2: $c = p/(1-p)$, $i = 0$, $pr = (1-p)^n$, $F = pr$

STEP 3: If $U < F$, set $X = i$ and stop

STEP 4: $pr = [c(n-i)/(i+1)]pr$, $F = F + pr$, $i = i+1$

STEP 5: Go to Step 3 □

To generate a a binomial random variable X with parameters $n = 10$ and $p = 0.7$ in SPLUS, type:

```
n <- 10
p <- 0.7
U <- runif(1,0,1)
c <- p/(1-p)
i <- 0
pr <- (1-p)^n
f <- pr
for (i in 0:n)
{
if (U < f)
{
X <- i
break
}
else
{
pr <- c*(n-i)/(i+1)*pr
f <- f+pr
}
}
X
```

4.4 ACCEPTANCE-REJECTION METHOD

In the preceding example, we see how the inverse transform can be used to generate a known discrete distribution. For most of the standard distributions, we can simulate their values easily by means of standard built-in routines available in standard packages. But when we move away from standard distributions, simulating values become more involved. One of the most useful methods is the acceptance-rejection algorithm.

Suppose we have an efficient method, e.g., a computer package, to simulate a random variable Y having probability mass function $\{q_j, j \geq 0\}$. We can use this as a basis for simulating a distribution X having probability mass function $\{p_j, j \geq 0\}$ by first simulating Y and then accepting this simulated value with a probability proportional to p_Y/q_Y. Specifically, let c be a constant such that

$$\frac{p_j}{q_j} \leq c \text{ for all } j \text{ such that } p_j > 0.$$

Then we can simulate the values of X having probability mass function $p_j = P(X = j)$ as follows:

STEP 1: Simulate the value of Y from q_j

STEP 2: Generate a uniform random number U

STEP 3: If $U < p_Y/(cq_Y)$, set $X = y$ and stop. Otherwise, go to Step 1.

Theorem 4.1 *The acceptance-rejection algorithm generates a random variable X such that*

$$P(X = j) = p_j, \quad j = 0, 1, \ldots .$$

In addition, the number of iterations of the algorithm needed to obtain X is a geometric random variable with mean c.

Proof. First consider the probability that a single iteration produces the accepted value j. Note that

$$\begin{aligned} P(Y = j, \text{ it is accepted}) &= P(Y = j)P(\text{accepted}|Y = j) \\ &= q_j P(U \le p_j/(cq_j)) \\ &= q_j p_j/(cq_j) \\ &= p_j/c. \end{aligned}$$

Summing over j, we get the probability that a generated random variable is accepted as

$$P(\text{accepted}) = \sum_j p_j/c = 1/c.$$

Since each iteration independently results in an accepted value with probability $1/c$, the number of iterations needed is geometric with mean c. Finally,

$$\begin{aligned} P(X = j) &= \sum_n P(j \text{ accepted on iteration } n) \\ &= \sum_n (1 - 1/c)^{n-1} p_j/c = p_j. \end{aligned}$$

□

Example 4.2 *Suppose we want to simulate a random variable X taking values in $\{1, 2, \ldots, 10\}$ with probabilities as follows:*

i	1	2	3	4	5	6	7	8	9	10
$P(X = i)$	0.11	0.12	0.09	0.08	0.12	0.1	0.09	0.11	0.07	0.11

Using the acceptance-rejection method, first generate discrete uniform random variables over the integers $\{1, \ldots, 10\}$. That is, $P(Y = j) = q_j = 1/10$

for $j = 1, \ldots, 10$. First, compute the number c by setting $c = \max \frac{p_j}{q_j} = 1.2$. Now generate a discrete uniform random variable Y by letting $Y = [10U_1]+1$, where $U_1 \sim U(0,1)$. Then generate another $U_2 \sim U(0,1)$ and compare if $U_2 \leq p_Y/(cq_Y)$. If this condition is satisfied, then $X = Y$ is the simulated value. Otherwise, repeat the steps again. □

The code in SPLUS is as follows:

```
k <- 1000
x <- c(rep(0,k))
u1 <- c(rep(0,k))
u2 <- c(rep(0,k))
y <- c(rep(NA,k))
c1 <- 0
n <- 10
p <- c(0.11,0.12,0.09,0.08,0.12,0.1,0.09,0.11,0.07,0.11)
q <- c(rep(1/n,10))
b <- (max(p/q))/n
for (i in 1:k)
{
  u1[i]<- trunc(runif(1,0,1)*n)+1
  u2[i]<- runif(1,0,1)

  if (u2[i]<=p[u1[i]]/b) x[i]<-u1[i]
  if (x[i]==0) c1 <- c1+1
     else  y[i] <- x[i]
}
c1
hist(na.omit(y),prob=T,breaks=((0:10)+0.5))
lines(density(na.omit(y)))
```

4.5 CONTINUOUS RANDOM VARIABLES

Generating continuous random variables is very similar to generating discrete random variables. It again relies on two main approaches using uniform random numbers: the inverse transform and the acceptance-rejection method.

4.5.1 Inverse Transform

Theorem 4.2 *Let U be a uniform (0, 1) random variable. For any continuous distribution function F, the random variable X defined by $X = F^{-1}(U)$ has distribution F. Here*

$$F^{-1}(u) = \inf\{x : F(x) \geq u\}.$$

Proof. Let F_X denote the distribution of $X = F^{-1}(U)$. Then

$$\begin{aligned} F_X(x) &= P(X \leq x) \\ &= P(F^{-1}(U) \leq x) \\ &= P(U \leq F(x)) \\ &= F(x). \end{aligned}$$

□

Example 4.3 *Let X be an exponential distribution with rate 1. Then its distribution function is given by $F(x) = 1 - e^{-x}$. Let $x = F^{-1}(u)$, then $u = F(x) = 1 - e^{-x}$, so that $x = -\log(1-u)$. Thus, we can generate X by generating U and setting $X = -\log(1-U)$. Moreover, since $(1-U)$ has the same distribution as U, which is uniform (0, 1), we can simply set $X = -\log U$. Finally, it can be seen easily that if $Y \sim \exp(\lambda)$, then $\mathrm{E}(Y) = 1/\lambda$ and $Y = X/\lambda$, where $X \sim \exp(1)$. In this case, we can simulate Y by first simulating U and setting $Y = -\frac{1}{\lambda}\log U$.* □

The previous example illustrates how to apply the inverse transform method when the inverse of F can be written down easily. The next example demonstrates the case when the inverse of F is not readily available.

Example 4.4 *Let $X \sim \Gamma(n, \lambda)$. Then it has distribution function*

$$F_X(x) = \int_0^x \frac{\lambda e^{-\lambda y}(\lambda y)^{n-1}}{(n-1)!}\, dy.$$

Clearly, finding the inverse of F_X is not feasible. But recall that $X = \sum_{i=1}^n Y_i$, where $Y_i \sim \Gamma(1, \lambda)$ i.i.d. Furthermore, each Y_i has distribution function

$$F_Y(y) = \int_0^y \lambda e^{-\lambda s}\, ds,$$

which is the distribution function of an exponential distribution with rate λ. Therefore, we can generate X via

$$X = -\frac{1}{\lambda}\log U_1 - \cdots - \frac{1}{\lambda}\log U_n = -\frac{1}{\lambda}\log(U_1 \cdots U_n).$$

□

To generate a random variable X that follows a gamma distribution with parameters $n = 5$ and $\lambda = 10$ in SPLUS, type:

```
n <- 5
lambda <- 10
U <- runif(n,0,1)
X <- -(1/lambda)*sum(log(U[1:n]))
X
```

The message from these two examples is that, although the inverse transform method is simple, we may need to conduct certain simplifications before applying the method.

4.5.2 The Rejection Method

Suppose that we can simulate from a density g easily. We can use this as a basis to simulate from a density $f(x)$ by first generating Y from g and then accepting the generated value with probability proportional to $f(Y)/g(Y)$. Specifically, let c be such that

$$\frac{f(y)}{g(y)} \leq c \text{ for all } y.$$

Then we generate from f via the following algorithm:

STEP 1: Generate Y from a density g

STEP 2: Generate a uniform random number U

STEP 3: If $U \leq \frac{f(y)}{cg(y)} := h(y)$ set $X = Y$, else go to Step 1

This is exactly the same acceptance-rejection method in the discrete case. Correspondingly, we have the following result whose proof is almost the same as in the discrete case.

Theorem 4.3 *The random variable X generated by the rejection method has density f. Moreover, the number of iterations that this algorithm needs is a geometric random variable with mean c.*

Proof. Let $f(x) = cg(x)h(x)$, where $c \geq 1$ is a constant, $g(x)$ is also a p.d.f. and $0 < h(x) \leq 1$. Let Y have p.d.f. g and $U \sim U(0,1)$. Consider

$$f_Y(x|U \leq h(Y)) = \frac{P(U \leq h(Y)|Y = x)g(x)}{P(U \leq h(Y))}.$$

For the first part in the numerator, we have

$$P(U \leq h(Y)|Y = x) = P(U \leq h(x)) = h(x).$$

For the denominator, consider

$$\begin{aligned} P(U \leq h(Y)) &= \int P(U \leq h(Y|Y = x))g(x)\,dx \\ &= \int h(x)g(x)\,dx \\ &= \int \frac{1}{c}f(x)\,dx \\ &= \frac{1}{c}. \end{aligned}$$

Therefore, $f_Y(x|U \leq h(Y)) = h(x)g(x)c = f(x)$. □

One of the difficulties in using the rejection method is determining the constant c. Our goal is to find the function $cg(x)$ so that $cg(x) \geq f(x)$ and sample easily from the density $g(x)$. This can be achieved using trial-by-error, or in certain circumstances, can be achieved by simple analysis, as illustrated in the following example.

Example 4.5 *Suppose we want to simulate from the density*

$$f(x) = 20x(1-x)^3, \ 0 < x < 1.$$

First note that f is defined only on the interval (0,1). We may try g that can be simulated easily over the same interval, uniform (0, 1), say, that is, $g(x) = 1, 0 < x < 1$. To determine the smallest number c such that $f(x)/g(x) \leq c$ for all $0 < x < 1$, we first find the maximum value of the ratio $f(x)/g(x) = 20x(1-x)^3$. Using calculus, differentiating and setting to zero,

$$\frac{d}{dx}\left(\frac{f(x)}{g(x)}\right) = 0,$$

we solve $x = 1/4$ to be the maximum of f/g. Thus

$$\frac{f(x)}{g(x)} \leq 20(1/4)(3/4)^3 = 135/64 = c.$$

Therefore,

$$\frac{f(x)}{cg(x)} = 20(64)/(135)x(1-x)^3.$$

The algorithm becomes:

STEP 1: Generate random numbers U_1 and U_2

STEP 2: If $U_2 \leq \frac{256}{27}U_1(1-U_1)^3$, stop and set $X = U_1$,

else go to Step 1 □

To simulate from this distribution, use the following code:

```
k <- 1000
x <- c(rep(0,k))
u1 <- c(rep(0,k))
u2 <- c(rep(0,k))
c1 <- 0
y <- c(rep(NA,k))
for (i in 1:k)
{
  u1[i] <- runif(1,0,1)
  u2[i] <- runif(1,0,1)
  if (u2[i] <= (256/27)*u1[i]*(1-u1[i])^3) x[i]<-u1[i]
```

```
  if (x[i]==0) c1<-c1+1
}
c1
#c1 counts the number of rejected values
for (i in 1:k)
{
  if (x[i]!=0) y[i] <- x[i]
}
y[1:20]
#only the values of y with y not equaling to NA are plotted
hist(na.omit(y),prob=T)
lines(density(na.omit(y)))
for (i in 1:k)
{
  f[i] <- 20*(i/k)*(1-i/k)^3
}
plot(f,type="l")
```

4.5.3 Multivariate Normal

An important application of simulation is to handle high dimensional problems. High dimensional problems are usually related to multivariate normal distributions (Gaussian distribution). However, most software packages do not provide algorithms for generating multivariate normal random variables. This section studies algorithms for generating multivariate normal random variables.

A random vector $\boldsymbol{X}$ is said to follow a multivariate normal distribution if all of its elements are normal random variables. The distribution of $\boldsymbol{X}$ is then described as

$$\boldsymbol{X} \sim \mathrm{N}(\boldsymbol{m}, \Sigma), \tag{4.4}$$

where $\boldsymbol{m} = \mathrm{E}[\boldsymbol{X}]$ is the mean vector and $\Sigma = \mathrm{Var}[\boldsymbol{X}]$ is the variance-covariance matrix. Consider a vector $\boldsymbol{X} = (X_1, \ldots, X_n)^T$ with $X_i \sim \mathrm{N}(\mu_i, \sigma_i^2)$. In this case, the mean vector $\boldsymbol{m} = (\mu_1, \ldots, \mu_n)^T$ and the $n \times n$ matrix $\Sigma = [\mathrm{Cov}(X_i, X_j)]$, $i, j = 1, \ldots, n$.

There is a convenient way to generate a normal random vector $\boldsymbol{X}$ when $\Sigma = I$. $\Sigma = I$ indicates that the elements of $\boldsymbol{X}$ are independent random variables. Therefore, we can generate X_i independently and then stack them up to form the vector $\boldsymbol{X}$. For a normal random vector with dependent components, i.e., $\Sigma \neq I$, decomposition methods are useful.

4.5.3.1 Cholesky Decomposition The first method is the Cholesky decomposition. Consider two correlated standard normal random variables X_1 and X_2

with correlation coefficient ρ, written as,

$$\boldsymbol{X} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim \mathrm{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right).$$

Theorem 4.4 *Correlated random variables X_1 and X_2 can be decomposed into two uncorrelated random variables Z_1 and Z_2 through the linear transformation:*

$$\begin{aligned} Z_1 &= X_1 \\ Z_2 &= \frac{X_2 - \rho X_1}{\sqrt{1-\rho^2}}. \end{aligned}$$

In other words,

$$\boldsymbol{X} = \begin{pmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{pmatrix} \boldsymbol{Z}, \tag{4.5}$$

where

$$\boldsymbol{Z} \sim \mathrm{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right).$$

Proof. As X_1 and X_2 are linear combinations of normal random variables, they are also normally distributed. Furthermore,

$$\begin{aligned} \mathrm{E}(X_1) &= \mathrm{E}(X_2) = 0 \\ \mathrm{Var}(X_1) &= 1, \quad \mathrm{Var}(X_2) = (1-\rho^2)\,\mathrm{Var}(Z_2) + \rho^2\,\mathrm{Var}(Z_1) = 1 \\ \mathrm{Cov}(X_1, X_2) &= \mathrm{Cov}(Z_1, Z_2\sqrt{1-\rho^2} + \rho Z_1) = \rho. \end{aligned}$$

Thus, X_1 and X_2 have the desired distribution. □

The linear transformation of (4.5) is called the Cholesky decomposition. It enables us to generate (X_1, X_2) by the following procedures.

STEP 1: Generate $Z_1, Z_2 \sim \mathrm{N}(0,1)$ i.i.d.

STEP 2: Set $X_1 = Z_1$ and $X_2 = Z_2\sqrt{1-\rho^2} + \rho Z_1$.

In fact, there is a Cholesky decomposition for $\mathrm{N}(\boldsymbol{m}, \Sigma)$. Since Σ is a semi-positive definite matrix, i.e., $\boldsymbol{v}^T \Sigma \boldsymbol{v} > 0$ for all vector $\boldsymbol{v}$, there exists a lower triangular matrix L such that $\Sigma = LL^T$. The Cholesky decomposition is an algorithm to obtain this lower triangular matrix L.

For $n \times n$ matrices $\Sigma = [a_{ij}]$ and $L = [l_{ij}]$, the Cholesky decomposition algorithm works as follows.

STEP 1: Set $l_{11} = \sqrt{a_{11}}$.

STEP 2: For $j = 2, \ldots, n$ set $l_{j1} = a_{j1}/l_{11}$.

STEP 3: For $i = 2, \ldots, n-1$ conduct STEP 4 and STEP 5.

STEP 4: Set $l_{ii} = \left[a_{ii} - \sum_{k=1}^{i-1} l_{ik}^2\right]^{1/2}$.

STEP 5: For $j = i+1, \ldots, n$, set $l_{ji} = \frac{1}{l_{ii}}\left[a_{ji} - \sum_{k=1}^{i-1} l_{jk} l_{ik}\right]$.

STEP 6: Set $l_{nn} = \left[a_{nn} - \sum_{k=1}^{n-1} l_{nk}^2\right]^{1/2}$.

The algorithm can be implemented with a VBA code:

```
Function nCholesky(Correlation)
'Creating the Cholesky Decomposition Factors
Dim mRs As Single
mRs = nCountRC(Correlation, True) 'Correlation.Rows.Count

Dim aCholesky() As Double 'Cholesky Decomposition Matrix
ReDim aCholesky(1 To mRs, 1 To mRs) As Double

For i = 1 To mRs
aCholesky(i, 1) = Correlation(i, 1)
Next i
For i = 2 To mRs
For j = 2 To i
If i = j Then
aCholesky2 = 0
For k = 1 To j - 1
aCholesky2 = aCholesky2 + aCholesky(i, k) ^ 2
Next k
aCholesky(i, j) = Sqr(1 - aCholesky2)
Else
aCholeskyA = 0
For k = 1 To j - 1
aCholeskyA = aCholeskyA + aCholesky(i, k) * aCholesky(j, k)
Next k
aCholesky(i, j) = (Correlation(i, j) - aCholeskyA) / aCholesky(j, k)
End If
Next j
Next i
nCholesky = aCholesky
End Function
```

In SPLUS, Cholesky decomposition is solved with the subroutine 'chol()'.

Given the matrix L, a random vector $\boldsymbol{X} \sim \mathrm{N}(\boldsymbol{m}, \Sigma)$ is generated by

$$\boldsymbol{X} = \boldsymbol{m} + L\boldsymbol{Z}, \quad \boldsymbol{Z} \sim \mathrm{N}(\boldsymbol{0}, I) \tag{4.6}$$

The following SPLUS code generates an n-dimensional multivariate normal random vector with mean $\boldsymbol{m}$ and variance-covariance matrix Σ.

```
for (i in 1:n){
  Z[i] <- rnorm(1,0,1)
}
X <- m + t(chol(Sigma))%*%Z
```

Theorem 4.5 *The $\boldsymbol{X}$ obtained in (4.6) follows* $\mathrm{N}(\boldsymbol{m}, \Sigma)$.

Proof. The random vector $\boldsymbol{X}$ has a Gaussian distribution as it is a linear combination of Gaussian random variables. Therefore, it suffices to check the mean and variance of $\boldsymbol{X}$. For the mean,

$$\mathrm{E}[\boldsymbol{X}] = \boldsymbol{m} + \mathrm{E}[L\boldsymbol{Z}] = \boldsymbol{m}.$$

For the variance,

$$\mathrm{Var}[\boldsymbol{X}] = \mathrm{Var}[L\boldsymbol{Z}] = L\left(\mathrm{Var}[\boldsymbol{Z}]\right)L^T = LL^T = \Sigma.$$

□

Example 4.6 *Consider a portfolio of three assets:* $P(t) = S_1(t) + 2S_2(t) + 3S_3(t)$. *The current assets values are* $S_1(0) = 100, S_2(0) = 60$ *and* $S_3(0) = 30$. *Suppose rate of returns of three assets follow a multivariate normal distribution. Specifically, we let*

$$R_i(t) = \frac{S_i(t+\Delta t) - S_i(t)}{S_i(t)} \quad \textit{and} \quad \boldsymbol{R}(t) = (R_1(t), R_2(t), R_3(t))^T,$$

where

$$\boldsymbol{R}(t) = \begin{bmatrix} 0.1\Delta t + 0.2\sqrt{\Delta t}X_1 \\ -0.03\Delta t + 0.4\sqrt{\Delta t}X_2 \\ 0.2\Delta t + 0.25\sqrt{\Delta t}X_3 \end{bmatrix},$$

$$\begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} \sim \mathrm{N}\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & -0.1 & 0.2 \\ -0.1 & 1 & 0.1 \\ 0.2 & 0.1 & 1 \end{bmatrix}\right).$$

Simulate 10 sample paths of the portfolio with $\Delta t = 1/100$.

The SPLUS code for the problem is given as follows:

```
n <- 10
N <- 100
dt<- 1/N
#input the variance-covariance matrix Sigma
Sigma<-matrix(c(1,-0.1,0.2,-0.1,1,0.1,0.2,0.1,1), nrow=3, ncol=3)
L<- t(chol(Sigma))
```

```
stock<-matrix(0,3,N+1)
stock[,1]<-c(100,60,30) #initial stock price
P<-matrix(0,n,N+1)

for(k in 1:n){
for(i in 1:N){
X <- L%*%rnorm(3,0,1)
R<- c(0.1,-0.03,0.2)*dt + c(0.2,0.4,0.25)*sqrt(dt)*X
stock[,i+1]<- stock[,i]*(1+R)  #Generate asset prices paths
}
P[k,]<-stock[1,]+2*stock[2,]+3*stock[3,]
}

t<-c(0:N)/N

#plot 10 sample paths for the portfolio
plot(t,P[1,],type="l",ylim=c(240,420),pty=0,xlab="t",ylab="P(t)")
for(k in 2:n)  {lines(t, P[k,])}

#plot the last sample paths of individual assets and the portfolio
plot(t,P[n,],type="l",ylim=c(0,400),pty=0,xlab="t",ylab="prices")
for(k in 1:3)  {lines(t, stock[k,], lty=2) }
```

Two graphs are produced by the programme. Fig. 4.1 plots 10 portfolio sample paths against time. Fig. 4.2 plots one sample path for each individual assets and one sample path of the portfolio. Asset and the portfolio can be identified by their initial values.

4.5.3.2 Eigenvalue Decomposition The second method is the eigenvalue decomposition. Given an $n \times n$ matrix Σ, if a constant value λ and a *non-zero* vector $\boldsymbol{v}$ satisfy:

$$\Sigma \boldsymbol{v} = \lambda \boldsymbol{v}, \tag{4.7}$$

then λ is called an eigenvalue of the matrix Σ and $\boldsymbol{v}$ is the corresponding eigenvector. In principle, there are n eigenvalues for an $n \times n$ matrix. For the variance-covariance matrix Σ, we know that all eigenvalues are non-negative and eigenvectors are orthogonal because Σ is semi-positive definite.

In multivariate analysis, eigenvalues of a variance-covariance matrix Σ are arranged in descending order as $\lambda_1 > \lambda_2 > \cdots > \lambda_n$ and the corresponding eigenvectors are chosen to have unit length. This means $||\boldsymbol{v}_i|| = 1$ for $i = 1, 2, \ldots, n$. Under these specifications, $\boldsymbol{v}_1$ is called the first principle component, $\boldsymbol{v}_2$ is the second principle component and so on. More importantly, the matrix Σ can be decomposed into a product of three square matrices:

$$\Sigma = PDP^T, \tag{4.8}$$

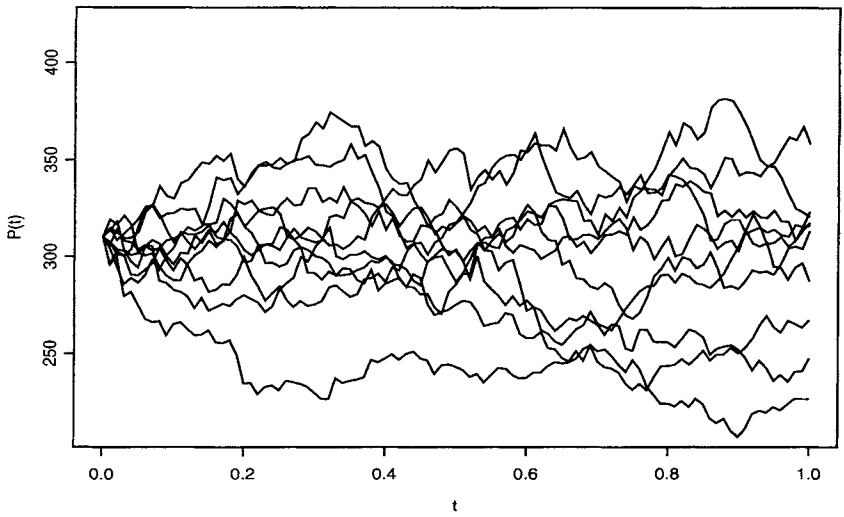


Fig. 4.1 Sample paths of the portfolio.

where $P = [\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n]$ and $D = diag(\lambda_1, \lambda_2, \ldots, \lambda_n)$ is a diagonal matrix. In SPLUS, eigenvalues and eigenvectors are easily obtained with the subroutine 'eigen()'.

Theorem 4.6 *If* $\boldsymbol{Z} \sim \mathrm{N}(\boldsymbol{0}, I)$, *then* $\boldsymbol{X} = \boldsymbol{m} + P\sqrt{D}\boldsymbol{Z} \sim \mathrm{N}(\boldsymbol{m}, \Sigma)$.

Proof. Again, it suffices to check the mean and variance of $\boldsymbol{X}$. For the mean,

$$\mathrm{E}[\boldsymbol{X}] = \boldsymbol{m} + \mathrm{E}[P\sqrt{D}\boldsymbol{Z}] = \boldsymbol{m}.$$

For the variance,

$$\mathrm{Var}[\boldsymbol{X}] = \mathrm{Var}[P\sqrt{D}\boldsymbol{Z}] = P\sqrt{D}\,(\mathrm{Var}[\boldsymbol{Z}])\,[P\sqrt{D}]^T = PDP^T = \Sigma.$$

□

The following is the SPLUS code for generating Gaussian normal random vectors by the eigenvalue decomposition.

```
for (i in 1:n){
Z[i] <- rnorm(1,0,1)
}
A <- (eigen(Sigma)$vectors)%*%diag(sqrt(eigen(Sigma)$values))
X <- m + A%*%Z
```

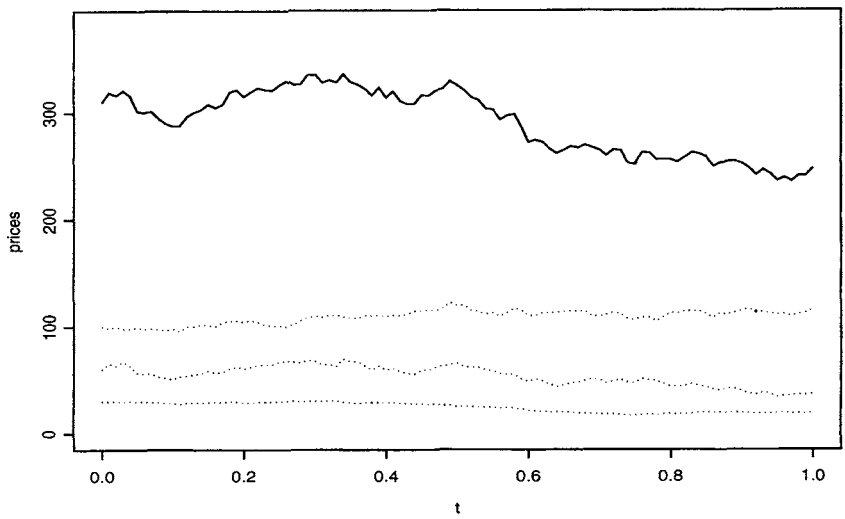


Fig. 4.2 Sample paths of the assets and the portfolio.

Remarks: VBA users may worry about matrix operations used in the above algorithms. Fortunately, there are free downloads available on the Web that provide necessary subroutines under the platform of EXCEL. For instance, the PoPTools, from http://www.cse.csiro.au/poptools/index.htm, includes routines of Cholesky and eigenvalue decompositions.

4.6 EXERCISES

1. Using the inverse transform method to generate a random variable X with the probability mass function.
   (a) $P(X = j) = \frac{1}{j(j+1)}, \quad j = 1, 2, ...$
   (b) $P(X = j) = {}_{(n+j-1)}C_j(1-p)^j p^n, \quad j = 0, 1, 2..., $ where n and p are given parameters.
2. We simulate X, Y, Z from an inverse transform algorithm. Suppose $U \sim$ U$(0, 1)$. Determine the distributions of the following random variables:
   (a) $X =$ int$(10U(1-U))$
   (b) $Y =$ int$(1/U)$
   (c) $Z = (B-3)^2, \quad B \sim$ Bin$(5, 0.5)$.

3. Determine the p.d.f. of

   (a) $X = -10 \log U + 5$

   (b) $X = 2 \tan(\pi U) + 10$

   (c) $W = nU - \text{int}(nU)$. Show that it is independent of $I = \text{int}(nU)$. (Hint: Show that $P(W \leq w, I = i) = w/n$.)

4. Let X have probability mass function

i	1	2	3	4	5	6	7
$P(X = i)$	0.3	0.12	0.09	0.12	0.1	0.17	0.1

   (a) Use the acceptance-rejection algorithm, simulate 1,000 data points from this distribution. You may use a discrete uniform as your g.

   (b) Plot out the histogram of your simulation.

   (c) What is the expected number of acceptance for this distribution? Does that match your simulation results?

5. Suppose we want to simulate from the density

$$f(x) = x + 1/2, \ 0 < x < 1.$$

   (a) Using the inverse transformation method, simulate 1,000 values from f.

   (b) Using the acceptance-rejection method, simulate another 1,000 values from f. Which algorithm is more efficient?

6. Suppose we want to simulate $|Z|$, where $Z \sim \text{N}(0, 1)$. That is, the absolute value of a standard normal random variable. First note that the p.d.f. of $|Z|$ is given by

$$f(x) = \frac{2}{\sqrt{2\pi}} e^{-x^2/2}, \ 0 < x < \infty.$$

   Suppose you want to use the acceptance-rejection algorithm to simulate $|Z|$. Take g to be the exponential distribution,

$$g(x) = e^{-x}, \ 0 < x < \infty.$$

   (a) Determine the value c such that $c = \max \frac{f(x)}{g(x)}$.

   (b) Use the acceptance-rejection method, simulate 1,000 values of $|Z|$.

   (c) Suppose you want to recover Z from the simulated values of $|Z|$. One way to do it is to generate a random number U and set

$$Z = \begin{cases} |Z| & \text{if } U > 1/2, \\ -|Z| & \text{if } U \leq 1/2. \end{cases}$$

   Using this method, obtain 1,000 values of Z and plot its density.

Simulation Techniques in Financial Risk Management
by Ngai Hang Chan and Hoi Ying Wong

5

Standard Simulations in Risk Management

5.1 INTRODUCTION

Risk management applications require simulation experiments. In this chapter, we introduce some standard simulation techniques and discuss their applications in risk management.

5.2 SCENARIO ANALYSIS

Scenario analysis of risk management refers to simulating possible scenarios to analyze the risk of a decision and consequences. The ultimate goal of a scenario analysis may be to reach a decision, to verify a model or to validate a certain conjecture.

Suppose a newspaper boy buys a newspaper from an agent for \$4 each and sells it for \$6. His problem is to decide how many newspapers to buy each morning. In other words, what would be a prudent purchasing strategy?

To analyze the situation, he examines the sales record for the past 100 days given in Table 5.1. After reviewing the data in Table 5.1, he comes up with the following strategies:

1. Each day, purchase the same number of papers sold the day before.

2. Each day, purchase a fixed number of papers, say 23.

Number of newspapers	Days occurring
21	15
22	20
23	30
24	21
25	14

Table 5.1 Sales record.

To test each of these two strategies, one could simulate the scenarios using inverse transform. First convert the information in Table 5.1 into the empirical probability mass function (p.m.f.):

Number of newspapers	p.m.f.	Cumulative distribution
21	0.15	0.15
22	0.20	0.35
23	0.30	0.65
24	0.21	0.86
25	0.14	1.00

Table 5.2 Probability mass function.

Now simulate 10 future days and compare the two policies following the p.m.f. given in Table 5.2. The simulation draws a standard uniform random variables u. The demands of newspaper are generated according to where the random variables fall. For instance, if $u = 0.17$, which belongs to the range of 0.15 to 0.35, then the corresponding demand is 22. To have a fair comparison, assume that the newspaper boy orders 23 papers on Day 1. Table 5.3 lists the results of the simulation. The interval $[0, 1]$ is partitioned according to the cumulative frequency in Table 5.2. According to Table 5.3, policy 2 is better than policy 1. One can repeat the simulation for many times to see if this phenomenon is consistent.

The newspaper boy problem illustrates several important elements in scenario analysis. Decision makers identify possible scenarios based on empirical data or experience. In this example, scenarios correspond to the daily demand of newspapers. Simulation is then developed to replicate future possibilities. We use the inverse transform with the empirical density function in this ex-

	$u \sim U(0,1)$	Number of newspapers	Profit of 1	Profit of 2
Day 1	0.5828	23	\$46	\$46
Day 2	0.0235	21	\$34	\$34
Day 3	0.5155	23	\$42	\$46
Day 4	0.3340	22	\$40	\$40
Day 5	0.4329	23	\$44	\$46
Day 6	0.2259	22	\$40	\$40
Day 7	0.5798	23	\$44	\$46
Day 8	0.7604	24	\$46	\$46
Day 9	0.8298	24	\$48	\$46
Day 10	0.6405	23	\$42	\$46
		Total Profit =	\$426	\$436

Table 5.3 Policy simulation and evaluation.

ample. After generating scenarios, a risk manager analyzes consequences corresponding to each scenario. If the first policy is adopted, then the number of newspapers purchased equals the number sold yesterday; otherwise, 23 papers are purchased. Finally, evaluation and comparison can be conducted using the simulated results.

5.2.1 Value at Risk

In finance, risk scenario analysis is usually conducted for evaluating value-at-risk (VaR), a widely adopted risk measure

Definition 5.1 *VaR summarizes the worst loss of a portfolio over a target horizon with a given level of confidence.*

Statistically speaking, VaR describes the specified *quantile* or percentile of the projected distribution of profits and losses over the target horizon. Let R_t be the return of a portfolio for a horizon t. Then, the $c\%$ confidence VaR of the portfolio is measured through the expression:

$$P(R_t < -\text{VaR}) = (1-c)\% := \alpha. \tag{5.1}$$

Hence, VaR is the negative of the α-th percentile of the probability distribution of profits and losses. The larger the VaR, the higher the risk of the portfolio. An advantage of VaR is that it allows the user to specify the confidence level to reflect individual risk-averseness. For more details, see Jorion (2000).

VaR is indispensable for market risk analysis because it is the number that splits future possible asset returns into two scenarios: risky and nonrisky.

Returns less than the negative of VaR belong to the class of risky scenario. Decision makers can evaluate their policies by examining consequences under the risky scenario. For instance, a bank may check if it maintains enough money for an extremely risky situation.

A conventional way to measure VaR often assumes portfolio returns to follow a normal distribution. VaR obtained in this way is called normal VaR. A typical model is

$$R_t = \mu + \sigma Z, \ Z \sim \mathrm{N}(0,1). \tag{5.2}$$

In such a parametric model, it is easy to derive that

$$\mathrm{VaR}_\alpha(t) = -z_\alpha \sigma - \mu, \tag{5.3}$$

where z_α is the α-quantile of the standard normal distribution, μ is the drift and σ is the standard deviation of the return R_t over the horizon t.

Though one can prove (5.3) mathematically, we would like to verify it by simulation. The algorithm is given as follows.

STEP 1: Generate n independent standard normal random variables, namely $Z_j \sim \mathrm{N}(0,1)$ i.i.d., $j = 1, 2, \ldots, n$.

STEP 2: Set $R_j = \mu + \sigma Z_j$.

STEP 3: Rank $\{R_1, R_2, \cdots, R_n\}$ in ascending order as $\{R_1^*, R_2^*, \cdots, R_n^*\}$.

STEP 4: Set $\mathrm{VaR} = -R_k^*$, where $k = int(\alpha \times n)$.

Example 5.1 *Let $\mu = 0.003$, $\sigma = 0.23$, $\alpha = 5\%$, and $n = 10,000$. Then, the 95% VaR corresponds to the 500th smallest return generated from the simulation. Our simulation shows that the VaR = 0.3783, which is close to the value, 0.3753, obtained by (5.3). The* SPLUS *code is as follows:*

```
n <- 10000
alp <- 0.05
k <- round(n*alp)
mu <- 0.003
sigma <- 0.23
R <- rnorm(n,mu,sigma)
SR <- sort(R)
VaR <- -SR[k]
```

5.2.2 Heavy-Tailed Distribution

In reality, returns of market prices may not follow a normal distribution but a heavy-tailed distribution. This means the two tails of the empirical density decay less rapidly than the normal density. Since closed-form solution for the

VaR of a heavy-tailed distribution is not readily available, a feasible alternative is to generate random variables according to a heavy-tailed distribution.

One commonly used form for heavy-tailed distribution is the generalized error distribution (GED). The p.d.f. of GED with parameter ξ is given by

$$\begin{aligned} f(z) &= \frac{\xi \exp\left(-\frac{1}{2}|z/\lambda|^{\xi}\right)}{\lambda 2^{1+1/\xi}\Gamma(1/\xi)}, \\ \lambda &= \left[\frac{2^{-2/\xi}\Gamma(1/\xi)}{\Gamma(3/\xi)}\right]^{1/2}, \end{aligned} \tag{5.4}$$

where $\Gamma(\cdot)$ denotes the Gamma function. Fig. 5.1 plots the p.d.f. of GED and Fig. 5.2 zooms in at the left-tail of the density function. It is seen that the smaller the ξ is, the heavier the left-tail of the density function is.

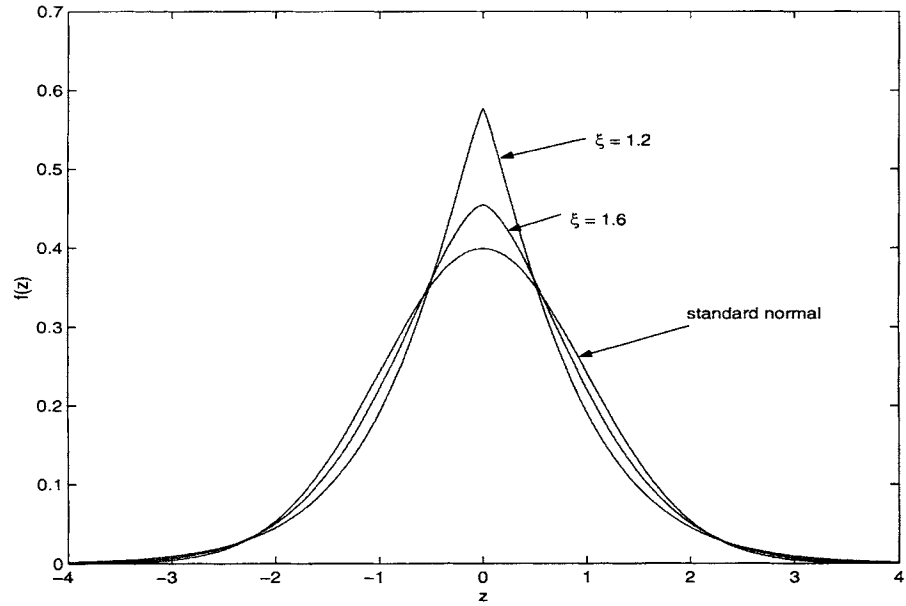


Fig. 5.1 The shape of GED density function.

The key to simulate VaR is to generate random variables following the desired distribution. Here, we apply the rejection method introduced in Chapter 4 using an exponential distribution for g. The algorithm goes as follows.

STEP 1: Generate $Y \sim \text{Exp}(1)$.

STEP 2: Generate $U \sim U(0,1)$.

STEP 3: If $U \leq 2f(Y)e^{Y}/a$, then $Z = Y$; else go to STEP 1.

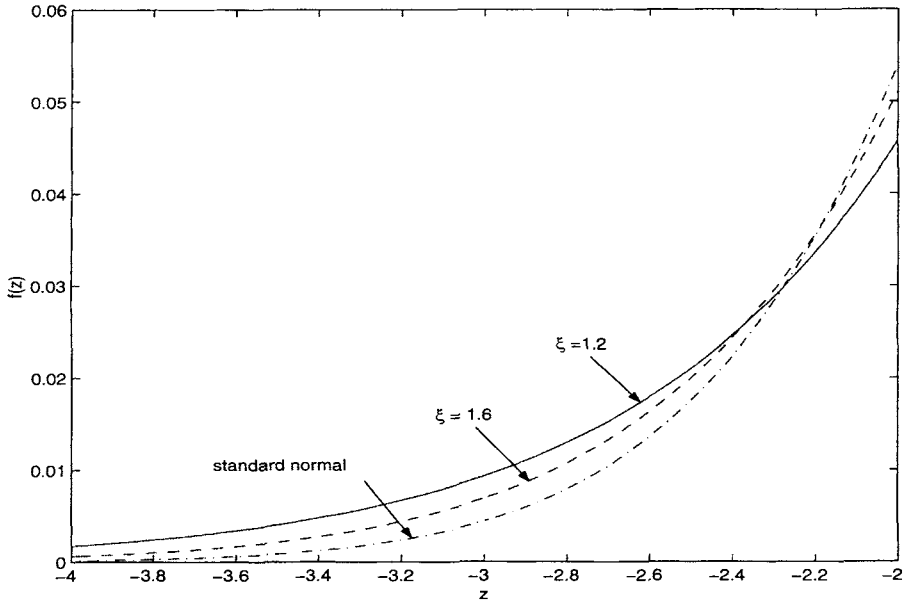


Fig. 5.2 Left tail of GED.

STEP 4: Generate $V \sim U(0,1)$. If $V < -1/2$, then $Z = -Y$.

STEP 5: Repeat STEP 1 - 4 for n times to get $\{Z_1, Z_2, \ldots, Z_n\}$.

STEP 6: Set $R_i = \mu + \sigma Z_i$.

STEP 7: Sort the returns in ascending order as $\{R_1^*, R_2^*, \ldots, R_n^*\}$.

STEP 8: Set VaR $= -R_k^*$ where $k = \text{int}(\alpha \times n)$.

Remarks:

1. In Step 3, a is a constant no less than $\max_y\{2f(y)e^y\}$.
2. As the exponential distribution is defined with a domain of positive real numbers, Steps 1 to 3 of the algorithm generate positive GED. Step 4 converts a positive GED random variable into a GED random variable.

5.2.3 Case Study: VaR of Dow Jones

We demonstrate the use of GED-VaR by considering 10-year daily closing prices of Dow Jones Industrial Index (DJI) in the period of August 8, 1995, to August 7, 2004. Data downloaded from http://finance.yahoo.com consists of 2,265 prices. The prices are converted into 2,264 daily returns by the formula:

$$R_t = \frac{S_t - S_{t-1}}{S_{t-1}}.$$

Sample mean and standard deviation of the returns are 0.04% and 1.16% in a daily scale, respectively. From (5.3), the 95% and 99% normal VaR from the sample are 1.87% and 2.66%, respectively.

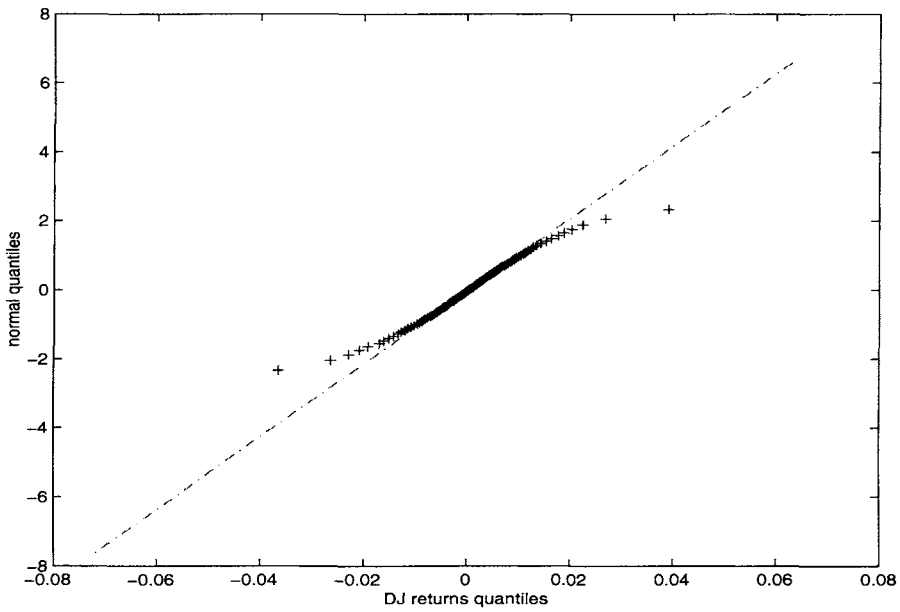


Fig. 5.3 QQ plot of normal quantiles against daily Dow Jones returns.

To access the quality of normal VaR, one has to test the normality assumption or, more precisely, the distributional assumption used in the VaR computation. Here, we introduce a simple but valuable tool, known as the quantile-quantile (QQ) plot. The idea is to plot the quantiles of the sample returns against the quantiles of the distribution used. If the returns truly follow the target distribution, then the graph should look like a straight line. For testing normality, the target distribution is the normal distribution. Systematic deviations from the line signal that the returns are not well described by the normal distribution.

Fig. 5.3 shows a QQ plot of our sample against the normal distribution. Large deviations are observed by the two tails of the empirical data. Specifically, the empirical quantile is less than the normal quantile in the left tail but larger than the normal quantile in the right tail. The deviations strongly suggest heavy-tailed distribution from the empirical data.

We use GED to reduce the deviation from the QQ plot. Returns are first standardized by the sample mean and standard deviation as

$$SR_t = \frac{R_t - 0.04\%}{1.16\%},$$

where SR_t denotes the standardized return at time t. We conjecture that $SR_t \sim GED(\xi)$, identically and independently. The parameter ξ is estimated from the SR using maximum likelihood estimation. Our estimation shows that $\xi = 1.21$ (see Appendix). Then, GED-VaR is estimated from the eight-step algorithm in Section 5.2.1, where the constant a is required in Step 3. The value of a can be deduced from the plot of $2f(y)e^y$ against y, where $f(y)$ is the p.d.f. of $GED(1.21)$. Fig. 5.4 shows that the maximum function value is bounded above by 1.2 so that we set $a = 1.2$.

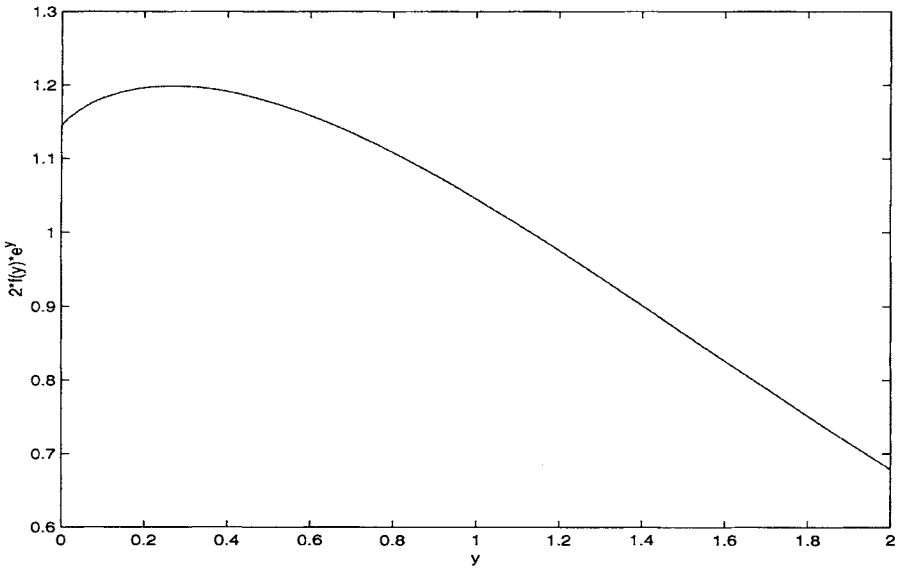


Fig. 5.4 Determine the maximum of $2f(y)e^y$ graphically.

We write the SPLUS code for simulating GED-VaR.

```
########## positive GEDpdf(v) ##########
funGED<-function(x,v){
lamda<-((2^(-2/v)*gamma(1/v))/gamma(3/v))^0.5
positiveGED<-2*(v*exp(-0.5*(x/lamda)^v))/(gamma(1/v)*lamda*2^(1+1/v))
}
funEXP<-function(x){
EXP<-exp(-x)
}

N<-10000
VaR95<-c(rep(0,1000))
VaR99<-c(rep(0,1000))
```

```
for (j in 1:1000){
#A program for generating GED by
#using rejection method (Proposal Density is exp(1))
# First generate positive GED distribution
N<-10000 #number of random number we want to generate
e<-c(rep(0,N))
for (i in 1:N){
u2<- 1 #initializing u2 for STEP 2
Y <- 2
while (u2>funGED(Y,1.21)/(funEXP(Y)*1.2)){ #Check conditions for STEP 3
  u1<-runif(1) #STEP 1
  Y<- -log(u1) #STEP 1: Generate the exp(1)
  u2<-runif(1) #STEP 2
  e[i]<-Y #STEP 3
}
}

#Then generate GED by assign half of the
#positive GED with negative sign
for (k in 1:N){
u3<-runif(1)
if(u3<0.5){
e[k]<- -e[k]    #STEP 4
}
}
R<-(0.04+1.16*e)/100        #STEP 6
sR<-sort(R)                 #STEP 7
VaR95[j]<- - sR[N*0.05]     #STEP 8 for 95% VaR
VaR99[j]<- - sR[N*0.01]     #STEP 8 for 99% VaR
}
###Repeat the whole algorithm for 10,000 times to get the C.I.###

sVaR95<-sort(VaR95)
sVaR99<-sort(VaR99)
E95VaR<-mean(VaR95)
UCIVaR95<-sVaR95[1000*0.975] #95% upper CI
LCIVaR95<-sVaR95[1000*0.025] #95% lower CI
E99VaR<-mean(VaR99)
UCIVaR99<-sVaR99[1000*0.975] #99% upper CI
LCIVaR99<-sVaR99[1000*0.025] #99% lower CI
E95VaR
UCIVaR95
LCIVaR95
E99VaR
```

```
UCIVaR99
LCIVaR99
```

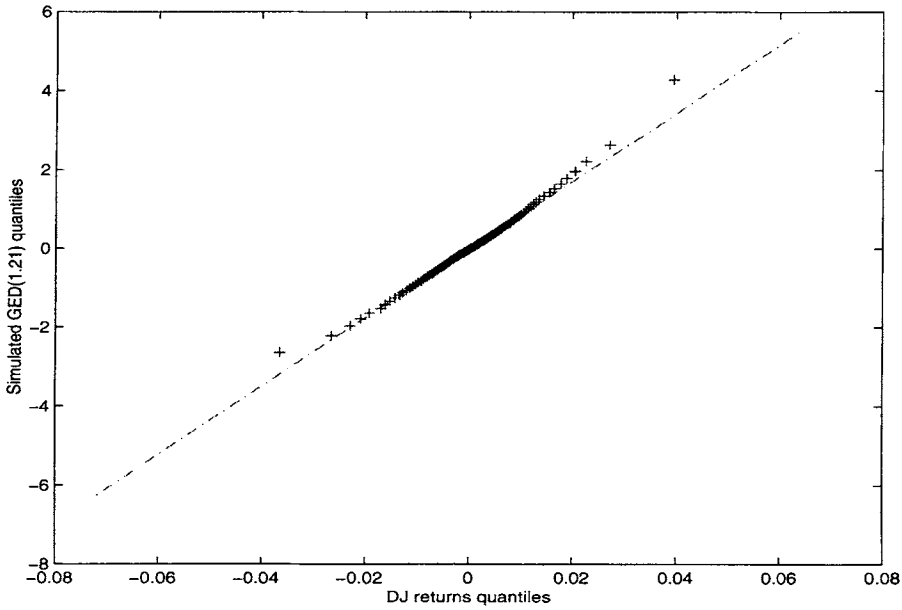


Fig. 5.5 QQ plot GED(1.21) quantiles against Dow Jones return quantiles.

The program estimates 95% and 99% VaR by generating 10,000 GED(1.21) random variables. For the confidence intervals, it repeats the process 1,000 times to get 1,000 VaR estimates. After arranging the simulated VaRs in ascending order, the 95% two-tailed confidence interval (CI) is the range between the 25th VaR and the 975th VaR.

To check the performance of GED-VaR, we use the QQ-plot of Fig. 5.5 based on one simulation. It is seen that deviations from the straight line have been substantially reduced. From this exercise, we see that GED(1.21) is appropriate for modeling the sample of Dow Jones returns. The average 95% VaR and 99% VaR from the 1,000 simulation are 1.87% and 3.02%, respectively. Therefore, 95% GED-VaR and 95% normal VaR give similar values whereas 99% GED-VaR is 10% more than the 99% normal VaR.

These findings may be useful for a risk manager. As normal VaR is commonly used in the financial industry, it is essential for a risk manager to understand the limitation of the normal VaR. The rationale of this empirical study is that normal VaR is a good estimate for potential losses of a portfolio under "normal, nonextreme" scenarios. However, it underestimates potential losses when "extreme events" happen, especially for those happening with probability less than 1%. To measure VaR with higher confidence level, e.g.,

99% VaR, the risk manager may consider GED-VaR. For further discussion about extreme values, see Embrechts, Klüppelberg, and Mikosch(1997) and the themed volume of Finkenstädt and Rootzén (2004).

5.3 STANDARD MONTE CARLO

In the preceding chapters, we studied the idea of simulating random variables. One of the main reasons to simulate random variables is to estimate quantities like $\mathrm{E}(X)$, which is related to the evaluation of definite integrals. Suppose we have already generated n values of a random variable X, it would be very natural to estimate the quantity $\theta = \mathrm{E}(X)$ by $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. We shall study some standard statistical techniques to assess the accuracy of such an estimate, which are based on the law of large numbers and the central limit theorem. Whenever we estimate quantities like $\mathrm{E}(X)$ based on standard applications of simulations, we refer these methods as standard Monte Carlo simulations. We shall study other more sophisticated simulation methods in later chapters.

5.3.1 Mean, Variance, and Interval Estimation

Suppose that X is a given random variable with mean θ and variance σ^2. A natural way to evaluate $\theta = \mathrm{E}(X)$ using simulations is to generate random values $X_1, \ldots, X_n$ and calculate the quantity

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i,$$

which is called the sample mean of $\{X_1, \ldots, X_n\}$. It is easy to see that

$$\mathrm{E}(\bar{X}_n) = \mathrm{E}(X) = \theta, \text{ unbiasedness property,} \tag{5.5}$$

$$\mathrm{Var}(\bar{X}_n) = \frac{\sigma^2}{n}. \tag{5.6}$$

To assess the accuracy of $\bar{X}_n$ as an estimate of θ, we rely on two important results. The first one is the law of large numbers, which asserts that as the number of simulations n gets bigger, the closer is $\bar{X}_n$ to θ, see, for example, Casella and Berger (2001). Specifically,

Theorem 5.1 *Let $X_1, \ldots, X_n$ be i.i.d. random variables with mean θ and variance σ^2. Then for any given $\epsilon > 0$,*

$$P(|\bar{X}_n - \theta| > \epsilon) \to 0 \text{ as } n \to \infty.$$

This result is sometimes written as $\bar{X}_n \to \theta$ in probability.

The second one is the central limit theorem, which asserts that as n tends to infinity, the distribution of the random variable $\bar{X}_n$ behaves like a normal distribution approximately.

Theorem 5.2 *Let $X_1, \ldots, X_n$ be i.i.d. random variables with mean θ and variance $\sigma^2 > 0$. Then as n tends to infinity*

$$P(\sqrt{n}\frac{(\bar{X}_n - \theta)}{\sigma} \leq z) \to \Phi(z),$$

where $\Phi(z)$ denotes the c.d.f. of a standard normal distribution evaluated at the point z.

A equivalent definition of this result is that the random variable $\sqrt{n}(\bar{X}_n - \theta)/\sigma$ converges in distribution to Z, written as

$$\sqrt{n}\frac{(\bar{X}_n - \theta)}{\sigma} \to_{\mathcal{L}} Z,$$

where $Z \sim \mathrm{N}(0,1)$. The proof of these two results can be found in standard text books in probability, see Billingsley (2001) for example. One immediate application of the central limit theorem is to construct approximate confidence intervals for θ. According to Theorem 5.2,

$$P(\,|\bar{X}_n - \theta| > \frac{\sigma}{\sqrt{n}}c) \sim P(\,|Z| \geq c) = 2(1 - \Phi(c)).$$

As a result, if we let $c = 1.96$, then the probability of $\bar{X}_n$ differs from θ by more than $1.96\sigma/\sqrt{n}$ would be approximately equal to 0.05. In other words, we are relatively confident (95%) that our estimate is within two standard errors ($1.96\sigma/\sqrt{n}$) from θ. To make use of this result, we have to have knowledge about the value σ, which is usually unavailable. A simple fix is to estimate it from the simulated values. The sample variance, which is defined as

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2,$$

constitutes an estimate of σ^2. It can be easily shown that

$$\mathrm{E}(S^2) = \sigma^2, \text{ unbiasedness property,} \tag{5.7}$$

$$S^2_{j+1} = (1 - 1/j)S^2_j + (j+1)(\bar{X}_{j+1} - \bar{X}_j)^2. \tag{5.8}$$

One frequently asked question in simulations is that after simulating X and evaluating $\bar{X}_n$, when should we stop? The answer to this question is given by the following scheme:

1. Choose an appropriate value d for the standard deviation of the estimation. That is, d represents the margin of error we can tolerate using simulations.

2. Generate at least 100 values of X.

3. Continue generating X and stopping when we have k values of X such that $S/\sqrt{k} < d$.

4. The desired estimate is given by $\bar{X}_k$.

Finally, we can form an interval estimation for θ by using the notion of confidence intervals.

Definition 5.2 *If $\bar{X}_n = \bar{x}, S = s$, then the interval*

$$(\bar{x} - z_{\alpha/2}\frac{s}{\sqrt{n}}, \bar{x} + z_{\alpha/2}\frac{s}{\sqrt{n}})$$

is an approximate $100(1-\alpha)\%$ confidence interval for θ.

In particular, when $\alpha = 0.05, z_{\alpha/2} = 1.96$ and $(\bar{x} \pm 1.96s/\sqrt{n})$ is an approximate 95% confidence interval for θ and thus giving rise to the rule of "two sigma."

5.3.2 Simulating Option Prices

To illustrate the ideas of standard simulations in risk management, consider first simulating stock prices. Let S denote the price of a stock. Recall that we usually assume that S follows a geometric Brownian motion

$$dS = \mu S\, dt + \sigma S\, dW.$$

Equivalently,

$$d\log S = \nu\, dt + \sigma\, dW,$$

where $\nu = \mu - \sigma^2/2$. Using the last equation and letting ϵ to denote a standard normal random variable, we can generate S according to the formula

$$S(t+dt) = S(t)\exp(\nu\, dt + \sigma\epsilon\sqrt{dt}).$$

In particular

$$S(T) = S(0)\exp(\nu T + \sigma\epsilon\sqrt{T}). \tag{5.9}$$

Notice that according to the risk neutral valuation principle, we usually take $\mu = r$, the risk free rate.

Example 5.2 *Let $S_0 = 10, \mu = r = 0.03, \sigma = 0.4$, and $dt = 1/52$. We want to simulate weekly prices of the stock $S_i, i = 1, \ldots, 52$ for a one-year period. Then $\nu = \mu - \sigma^2/2 = -0.05$ and the results are given in* Table 5.4. *The* SPLUS *code is as follows:*

```
N <- 52
S0 <- 10
mu <- 0.03
sigma <- 0.4
nu <- mu - sigma^2/2
t <- (0:N)/N
dt <- 1/N
x <- rep(0,N+1)
y <- rep(0,N+1)
z <- rnorm(N,0,1)
y[1] <- S0
for (i in 1:N)
{
x[i+1] <- sqrt(dt)*sum(z[1:i])
y[i+1] <- y[1]*exp(nu*t[i+1]+sigma*x[i+1])
}
plot(t,y,type="l",xlab="t",ylab="Stock Price")
```

Suppose we want to calculate the price of a European call option maturing in one year with strike price $K = 12$. We can use the Black-Scholes formula to obtain the call price C as

$$C(S,t) = S\Phi(d_1) - Ke^{-r(T-t)}\Phi(d_2),$$

where $d_1 = \frac{1}{\sigma\sqrt{T-t}}(\log(S/K) + (r + \sigma^2/2)(T-t)$ and $d_2 = d_1 - \sigma\sqrt{T-t}$. Substituting the values of $r = 0.03, K = 12, T = 1, t = 0, \sigma = 0.4$, and $S_0 = 10$, we get

$$d_1 = \frac{1}{0.4\sqrt{1}}(\log(10/12)+(0.03+0.08)(1)) = -0.1808, d_2 = d_1 - 0.4 = -0.5808.$$

Using the SPLUS command `pnorm(z)` to evaluate $\Phi(z)$, we get $\Phi(d_1) = 0.4283$ and $\Phi(d_2) = 0.2807$. Hence,

$$C = 10(0.4283) - 12e^{-0.03}(0.2807) = 1.013918.$$

On the other hand, we can evaluate $C = e^{-rT}\hat{\mathrm{E}}(S_T - K)^+$.

Example 5.3 *The price of the European call option can now be computed using simulations.*

1. First generate n independent values of $S_1(T), \ldots, S_n(T)$ according to (5.9).
2. Compute simulated discounted call prices $C_i = e^{-rT}\max\{(S_i(T) - K), 0\}, i = 1, \ldots, n$.
3. Compute $\bar{C} = \frac{1}{n}\sum_{i=1}^{n} C_i$. $\bar{C}$ is an estimate of the discounted payoff $\hat{\mathrm{E}}(S_T - K)^+$.

Week	Price
0	10.000000
1	10.38419
2	10.37402
3	10.67406
4	11.65342
5	11.89871
6	11.28296
7	11.15327
8	10.33483
9	11.16090
10	12.14546
⋮	⋮
43	14.39009
44	13.78038
45	14.01125
46	12.72393
47	13.44627
48	13.05377
49	12.00424
50	12.74416
51	12.16204
52	12.15517

Table 5.4 Simulated prices of the first and the last 10 weeks.

4. Construct a 95% confidence interval for C from

$$\bar{C} \pm 1.96S/\sqrt{n},$$

where

$$S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(C_i - \bar{C})^2},$$

is the sample standard deviation of the simulated call prices C_i.

To simulate 100 paths

```
p <- 100
N <- 52
S0 <- 10
K <- 12
mu <- 0.03
sigma <- 0.4
```

```
nu <- mu - sigma^2/2
t <- (0:N)/N
dt <- 1/N
x <- matrix(rep(0,(N+1)*p),nrow=(N+1))
y <- matrix(rep(0,(N+1)*p),nrow=(N+1))
for (j in 1:p)
{
z <- rnorm(N,0,1)
for (i in 1:N)
{
x[i+1,j] <- sqrt(dt)*sum(z[1:i])
y[i+1,j] <- S0*exp(nu*t[i+1]+sigma*x[i+1,j])
}
}
ST <- y[N+1,]
C <- rep(0,p)
for (i in 1:p)
{
C[i] <- exp(-mu)*max(ST[i]-K,0)
}
C.bar <- mean(C)
S <- sqrt(var(C)*p/(p-1))
CI <- mean(C) - 1.96*S/sqrt(p)
CI[2] <- mean(C) + 1.96*S/sqrt(p)
```

Outputs of the simulated C_is are given in Table 5.5. The result of a 100-path simulation shows that the 95% confidence interval for C is $(0.37, 1.83)$. Fig. 5.6 shows that when the number of runs increases, the value of C converges to the limit of 1.01.

5.3.3 Simulating Option Delta

In risk management, hedging an option is sometimes more important than valuing the option. When a bank issues structured financial products to enhance sales, the embedded option risk would be of great concerns. Hedging is a useful device to manage such a risk. For a standard call option, the hedge ratio refers to the delta of the option, the partial derivative of the option price with respect to the underlying asset price. Under the Black-Scholes assumption, the delta of a call is defined as

$$delta = \frac{\partial c}{\partial S} = \Phi(d_1). \tag{5.10}$$

We use simulation to calculate the hedge ratio, delta, for general European options.

Path	C_i
1	0.0000000
2	0.0000000
3	0.0000000
4	5.9331955
5	1.1971242
6	0.0000000
7	2.2395878
8	0.0000000
9	0.0000000
10	4.0065595
11	0.0000000
12	1.3006804
13	0.0000000
14	0.0000000
15	0.0000000
16	0.0000000
17	6.0970236
18	0.0000000
19	0.0000000
20	0.1768191

Table 5.5 The discounted call prices for the first 20 paths.

The risk-neutral valuation asserts that an option with payoff $F(S_T)$ can be valued as $e^{-rT}\widehat{\mathrm{E}}[F(S_T)|S_0 = S]$. Therefore, delta equals

$$delta = e^{-rT}\frac{\partial}{\partial S}\widehat{\mathrm{E}}[F(S_T)|S_0 = S]. \tag{5.11}$$

In order to compute delta under the Black-Scholes dynamics, the following theorem is established.

Theorem 5.3 *The delta of a European option with payoff $F(S_T)$ is given by*

$$delta = e^{-rT}\,\widehat{\mathrm{E}}\left[F(S_T)\frac{W_T}{S\sigma T}\right], \tag{5.12}$$

where W_T is the standard Brownian motion driving S_T.

Proof: Ignoring the discount factor, the definition of delta in (5.11) is

$$\frac{\partial}{\partial S}\int_{-\infty}^{\infty} F(e^x)\phi(x|\log S)\,dx = \int_{-\infty}^{\infty} F(e^x)\frac{1}{S}\frac{\partial\phi(x|\log S)}{\partial \log S}\,dx,$$

where

$$\phi(x|y) = \frac{1}{\sigma\sqrt{2\pi T}}\exp\left[-\frac{(x-y-\nu T)^2}{2\sigma^2 T}\right].$$

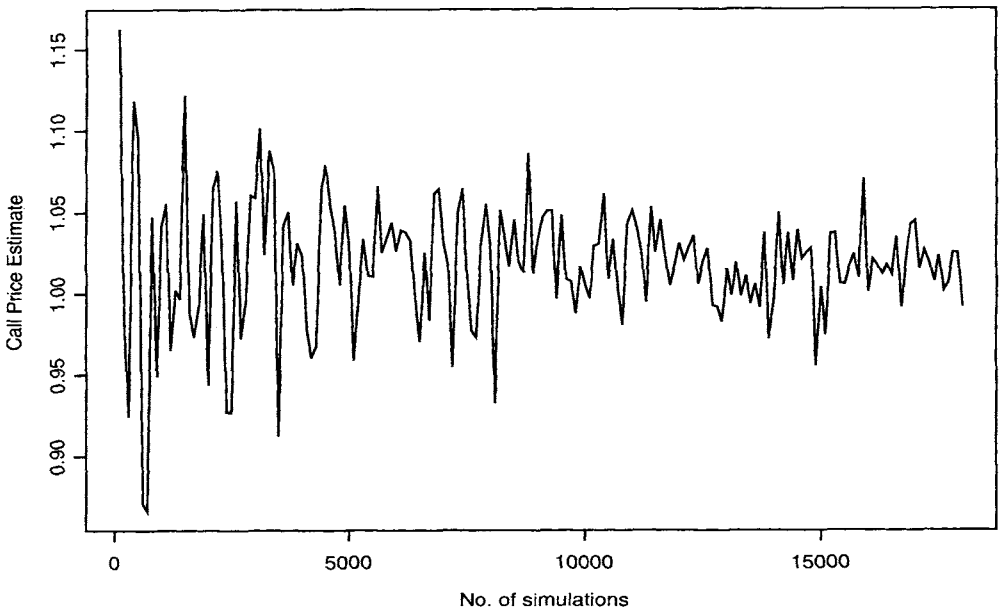


Fig. 5.6 Simulations of the call price against the size.

Standard differentiation shows that

$$\frac{\partial \phi(x|y)}{\partial y} = \phi(x|y)\frac{x - y - \nu T}{\sigma^2 T}.$$

Hence, we have

$$delta = e^{-rT}\int_{-\infty}^{\infty} F(e^x)\frac{x - \log S - \nu T}{S\sigma^2 T}\phi(x|\log S)\,dx.$$

Recall that $x = \log S_T$,

$$x - \log S - \nu T = \log S_T - \log S - \nu T = \sigma W_T.$$

This completes the proof. □

Theorem 5.3 enables us to simulate option delta (or even gamma) as follows.

STEP 1: Generate $Z_1, Z_2, \ldots, Z_n \sim \mathrm{N}(0,1)$ i.i.d.

STEP 2: Set $Y_j = F\left(Se^{(r-\sigma^2/2)T+\sigma Z_j\sqrt{T}}\right)\frac{Z_j}{S\sigma\sqrt{T}}$.

STEP 3: Set $delta = \frac{1}{n}\sum_{j=1}^{n} Y_j$.

The theorem can be extended to the case of path dependent options. However, the derivation requires some knowledge of Malliavin calculus, which is beyond of the scope of the book. For details of this generalization, we refer to the paper of Fournie et al. (1999).

Example 5.4 *The current price is \$10, interest rate 5%, volatility 40%. Simulate the price and delta of a call option with strike price \$12 and maturity 1 year by generating 10,000 terminal asset prices.*

An algorithm can be constructed as follows:

STEP 1: Generate 10,000 terminal asset prices by the formula

$$S_T^j = S_0 \exp\left[(r - \sigma^2/2)T + \sigma\sqrt{T}Z_j\right], \ Z_j \sim \mathrm{N}(0,1).$$

STEP 2: For $j = 1$ to 10,000, Compute

$$C_j = \max(S_T^j - K) * \exp(-rT) \text{ and } Del_j = C_j * Z_j/(\sigma\sqrt{T}S_0).$$

STEP 3: Compute call price = mean(C_j) and delta = mean(Del_j).

The corresponding SPLUS code is given as follows.

```
###Define parameter#######
check1<-proc.time() #Set the initial time count
n <-10000
S0 <- 10
maturity <-1
K <-12
r <- 0.05
sigma <-0.4
nu <- r -sigma^2/2
S <- rep(0,n)
C <- rep(0,n)
C0 <-rep(0,n)
Del <- rep(0,n)

####Simulation##############
z <- rnorm(n,0,1)
S <- S0*exp(nu*maturity+sigma*sqrt(maturity)*z)
C <- pmax(S-K,0)*exp(-r*maturity)
Del<-C*z/S0/sigma/sqrt(maturity)

####Compute Call and Delta values#######
Call <- mean(C)
Delta <- mean(Del)
```

```
check2 <- proc.time() # checkpoint for method without Ito
Call
Delta

####The CPU time############
check2-check1
```

With a CPU time of 0.2 seconds, our simulation finds that the call price is 1.075 and the delta is 0.441. The Black-Scholes call price and the delta are 1.08 and 0.448, respectively. This demonstrates the efficiency and accuracy of the simulation algorithm.

One thing we have to stress is that Theorem 5.3 is very useful for simulating deltas of single asset European options, with arbitrary payoff $F(S_T)$. However, it may not be applicable for path dependent options and multi-asset options. Therefore, we shall introduce alternative methods in later chapters.

5.4 EXERCISES

1. Write the SPLUS code for the newsboy problem of Section 5.2.

2. Suppose the asset return follows the t-distribution with 2 degrees of freedom. Write a SPLUS code to simulate the 95% confidence VaR with parameters given in Example 5.1. Compare your result with the one obtained by normal VaR.

3. Implement the rejection method for generating GED when $\nu = 1.4$.

4. Verify (5.5) and (5.6).

5. Verify (5.7) and (5.8).

6. Let $S_0 = 100, \mu = r = 0.05, \sigma = 0.3$. Use the geometric Brownian motion method to simulate 20 daily prices of the stock $S_i, i = 1, \ldots, 20$.

   (a) Suppose you want to determine the price of a European put option maturing in 20 days with a strike price $K = 100$. Use simulation techniques to estimate this price.

   (b) Compare your result with the one obtained from the Black-Scholes formula. Are they similar?

7. The gamma of an option is defined as

$$\frac{\partial(delta)}{\partial S}.$$

(a) What is the financial interpretation of the gamma?

(b) By modifying the proof of Theorem 5.3, show that

$$gamma = e^{-rT}\widehat{\mathrm{E}}\left[F(S_T)\frac{1}{S^2\sigma T}\left(\frac{W_T^2}{\sigma T} - W_T - \frac{1}{\sigma}\right)\right].$$

(c) Construct and implement a simulation algorithm to compute the call option gamma with a SPLUS code.

(d) Suppose $S = 10, K = 12, r = 0.1, \sigma = 0.3$, and $T = 0.8$. Compare your simulation result with the closed-form solution:

$$gamma = \frac{1}{S\sigma^2 T\sqrt{2\pi}} \exp\left[-\frac{\left(\ln\frac{S}{K} + (r + \sigma^2/2)T\right)^2}{2\sigma^2 T}\right].$$

5.5 APPENDIX

The data comprise 2,264 daily rates of returns. These data are transformed into standardized returns by using the sample mean and standard derivation. We assume that standardized returns follow a GED distribution with parameter ξ. Our goal is to estimate ξ. The density function of the GED distribution is given in (5.4). Hence, the likelihood function is

$$\begin{aligned} L(\xi) &= \prod_{i=1}^{2,264} \frac{\xi \exp\left(-\frac{1}{2}\left|Z_i/\lambda\right|^{\xi}\right)}{\lambda 2^{1+1/\xi}\Gamma(1/\xi)}, \\ \lambda &= \left[\frac{2^{-2/\xi}\Gamma(1/\xi)}{\Gamma(3/\xi)}\right]^{1/2}, \end{aligned}$$

where $Z_1, \ldots, Z_{2,264}$ are standardized returns. Instead of deriving the maximum likelihood estimation (MLE) theoretically, we search the maximum point of the likelihood function with a numerical method. To confine the target point in a small interval, we plot the likelihood function against the parameter ξ. In Fig. 5.7, we recognize that a unique maximum appears for $\xi \in (1, 1.3)$. The SPLUS code of the likelihood function and the plot is given after this paragraph. We then use the bisection method to search for the solution. Specifically, we compare $L(1)$ and $L(1.3)$ and discard the smaller one. The next step compares the remaining one with $L(1.15)$, the functional value at the mid-point of 1 and 1.3. We discard the point with a smaller value in L and repeat the procedure until a sufficiently accurate solution is obtained. Ultimately, $\xi = 1.21$, which has been input to generate GED-VaR in Section 5.2.3.

```
loglikelihood<-function(Z,xi) {
L <- 0
lambda <- sqrt( 2^(-2/xi)*gamma(1/xi)/gamma(3/xi) )
for (i in 1:length(Z)){
temp1 <- log(xi)-abs(Z[i]/lambda)^xi/2
temp2 <- log(lambda)+log(2)*(1+1/xi)+lgamma(1/xi)
L <- L + temp1 - temp2
}
L
}

### input returns from a text file ###
Z <- read.table("table.txt")
Z <- Z[,1]
Z <- (Z-mean(Z))/sqrt(var(Z))
print(Z)
a<-0
XI <- (50:250)/100
for (i in 1:length(XI)){
a[i] <- loglikelihood(Z,XI[i])
}

plot(XI,a, type='l',ylab='log likelihood')
print(XI[which(max(a)==a)])
```

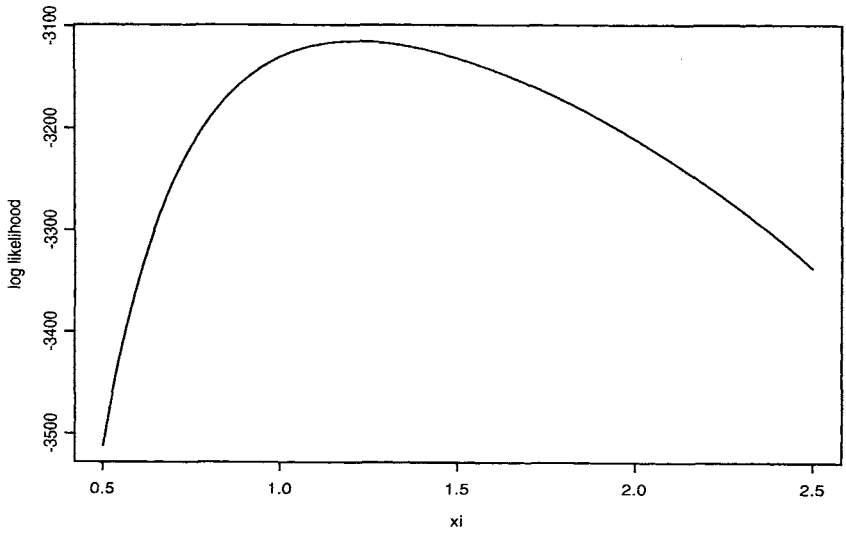


Fig. 5.7 The log likelihood against ξ.

Simulation Techniques in Financial Risk Management
by Ngai Hang Chan and Hoi Ying Wong

6

Variance Reduction Techniques

6.1 INTRODUCTION

In standard Monte Carlo, we estimate the unknown quantity $\theta = EX$ by generating random numbers $X_1, \ldots, X_n$ and use $\bar{X}_n$ to estimate θ. Recall that in the preceding chapter, the standard error for $\bar{X}_n$ is $\sigma/\sqrt{n}$, where σ^2 is the variance of X. There are two sources of contributions to the standard error of estimation. One is the factor $1/\sqrt{n}$, which is intrinsic to the Monte Carlo method, and not much can be done about it. The other one is the standard error σ of the output X, which by some techniques, can be improved upon. There are usually four standard methods to reduce σ:

1. Antithetic Variables
2. Control Variates
3. Stratification
4. Importance Sampling

We shall discuss each of these methods in the subsequent sections.

6.2 ANTITHETIC VARIABLES

The idea of antithetic variables can best be illustrated by considering a special example. Suppose we want to estimate $\theta = EX$ by generating two outputs

X_1 and X_2 such that $EX_1 = EX_2 = \theta$ and $\mathrm{Var}X_1 = \mathrm{Var}X_2 = \sigma^2$. Then

$$\begin{aligned}\mathrm{Var}(\frac{1}{2}(X_1 + X_2)) &= \frac{1}{4}(\mathrm{Var}X_1 + \mathrm{Var}X_2 + 2\mathrm{Cov}(X_1, X_2)) \\ &= \frac{\sigma^2}{2} + \frac{1}{2}\mathrm{Cov}(X_1, X_2) \\ &\leq \frac{\sigma^2}{2}, \quad \text{if } \mathrm{Cov}(X_1, X_2) \leq 0.\end{aligned}$$

Note that when X_1 and X_2 are independent, then $\mathrm{Var}((X_1 + X_2)/2) = \sigma^2/2$. Thus, the above inequality asserts that if X_1 and X_2 are negatively correlated, then the variance of the mean of the two would be less than the case when X_1 and X_2 were independent.

How do we generate negatively correlated random numbers? Suppose we simulate $U_1, \ldots, U_m$, which are uniform random numbers. Then $V_1 = 1 - U_1, \ldots, V_m = 1 - U_m$ would also be uniform random numbers with the property that (U_i, V_i) being negatively correlated (exercise). If $X_1 = h(U_1, \ldots, U_m)$, then $X_2 = h(V_1, \ldots, V_m)$ must have the same distribution as X_1. It turns out that if h is a monotone function (either increasing or decreasing) in each of its arguments, then X_1 and X_2 are negatively correlated. This result will be proved later at the end of this section. Thus, after generating $U_1, \ldots, U_m$ to compute X_1, instead of generating another new independent set of Us to compute X_2, we compute X_2 by

$$X_2 = h(V_1, \ldots, V_m) = h(1 - U_1, \ldots, 1 - U_m).$$

Accordingly, $(X_1 + X_2)/2$ should have smaller variance.

In general, we may generate $X_i = F^{-1}(U_i)$ using the inverse transform method. Let $Y_i = F^{-1}(V_i)$. Since F is monotone, so is F^{-1} and, hence, X_i and Y_i will be negatively correlated. Both $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$ generated in this way are i.i.d. sequences with c.d.f. F, but negatively correlated.

Definition 6.1 *The Y_i sequence is called the sequence of antithetic variables.*

For normal distributions, generating antithetic variables is straightforward. Suppose that $X_i \sim \mathrm{N}(\mu, \sigma^2)$, then $Y_i = 2\mu - X_i$ also has a normal distribution with mean μ and variance σ^2 and X_i and Y_i are negatively correlated.

More generally, if we want to compute $E(H(X))$ for some function H, standard Monte Carlo suggests using $\frac{1}{n}\sum_{i=1}^{n} H(X_i)$. Then an antithetic estimator of $E(H(X))$ is

$$\hat{H}_{AN} = \frac{1}{2n}\sum_{i=1}^{n}(H(X_i) + H(Y_i)),$$

where Y_i is a sequence of antithetic variables. To see how variance reduction is achieved by using this antithetic estimator, let $\mathrm{Var}(H(X)) = \sigma^2$ and

$\text{Corr}(H(X), H(Y)) = \rho$. Consider

$$\begin{aligned}\text{Var}(\hat{H}_{AN}) &= \frac{1}{4n^2}\sum_{i=1}^{n}\{\text{Var}H(X_i) + \text{Var}H(Y_i) + 2\text{Cov}(H(X_i), H(Y_i))\} \\ &= \frac{1}{4n^2}(2n\sigma^2 + 2n\rho\sigma^2) \\ &= \frac{\sigma^2}{2n}(1+\rho).\end{aligned}$$

Note that when $H(X)$ and $H(Y)$ are uncorrelated ($\rho = 0$), then the variance would be reduced by a factor of 2, which is equivalent to doubling the simulation size. On the other hand, if $\rho = -1$, then the variance would be reduced to zero. As long as ρ is negative, some form of variance reduction can be achieved. An obvious question is that in view of this observation, why not choose Y so that $\rho = -1$? Such Ys may be difficult to construct since ρ represents the correlation between $H(X)$ and $H(Y)$. In the case $H(X) = X$, then $\hat{H}_{AN}$ reduces to a constant, which is the perfect scenario. In view of these caveats, we usually choose the antithetic variables Y so that ρ is negative, not necessarily -1. When H is linear, such as the case $H(X) = X$, the antithetic variable works best. In general, the more linear the H is, the more effective the antithetic variable is.

Example 6.1 *Let* $\theta = \text{E}(e^U) = \int_0^1 e^x\,dx$.

We know that $\theta = e - 1$. Consider the antithetic variable $V = 1 - U$. Recall that the moment generating function of U equals $\text{E}(e^{tU}) = (e^t - 1)/t$. Now

$$\begin{aligned}\text{Cov}(e^U, e^V) &= \text{E}(e^U e^V) - \text{E}(e^U)\text{E}(e^V) \\ &= \text{E}(e^U e^{1-U}) - \text{E}(e^U)\text{E}(e^{1-U}) \\ &= e - (e-1)^2 = -0.2342.\end{aligned}$$

Furthermore,

$$\text{Var}(e^U) = \text{E}(e^{2U}) - (\text{E}(e^U))^2 = (e^2 - 1)/2 - (e-1)^2 = 0.242.$$

Thus, for U_1 and U_2 to be independent uniform (0,1) random variables,

$$\text{Var}[(e^{U_1} + e^{U_2})/2] = \text{Var}(e^U)/2 = 0.121.$$

But

$$\text{Var}[(e^U + e^V)/2] = \text{Var}(e^U)/2 + \text{Cov}(e^U, e^V)/2 = 0.121 - 0.2342/2 = 0.0039,$$

achieving a substantial variance reduction of 96.7%. □

We are now ready to justify the argument used in advocating antithetic variables.

Theorem 6.1 *Let $X_1, \ldots, X_n$ be independent, then for any increasing functions f and g of n variables,*

$$\mathrm{E}(f(\boldsymbol{X})\, g(\boldsymbol{X})) \geq \mathrm{E}f(\boldsymbol{X})\, \mathrm{E}g(\boldsymbol{X}),$$

where $\boldsymbol{X} = (X_1, \ldots, X_n)$.

Proof. By mathematical induction. Consider $n = 1$, then

$$(f(x) - f(y))\,(g(x) - g(y)) \geq 0, \text{ for all } x \text{ and } y,$$

as both factors are either nonnegative ($x \geq y$) or non-positive ($x \leq y$). Thus, for any random variables X and Y,

$$(f(X)-f(Y))\,(g(X)-g(Y)) \geq 0 \quad \text{implying} \quad \mathrm{E}((\mathrm{f(X)}-\mathrm{f(Y)})\,(\mathrm{g(X)}-\mathrm{g(Y)})) \geq 0.$$

In other words,

$$\mathrm{E}(f(X)\, g(X)) + \mathrm{E}(f(Y)\, g(Y)) \geq \mathrm{E}(f(X)\, g(Y)) + \mathrm{E}(f(Y)\, g(X)).$$

If X and Y are independent and identically distributed, then

$$\mathrm{E}(f(X)\, g(X)) = \mathrm{E}(f(Y)\, g(Y))$$

and

$$\mathrm{E}(f(X)\, g(Y)) = \mathrm{E}(f(Y)\, g(X)) = \mathrm{E}(f(Y))\, \mathrm{E}(g(X)) = \mathrm{E}(f(X))\, \mathrm{E}(g(X))$$

so that

$$\mathrm{E}(f(X)\, g(X)) \geq \mathrm{E}(f(X))\, \mathrm{E}(g(X)),$$

proving the result for the case $n = 1$. Assume the result for $n - 1$. Suppose $X_1, \ldots, X_n$ are independent and let f and g be increasing functions. Then

$$\begin{aligned}
\mathrm{E}(f(\boldsymbol{X})\, g(\boldsymbol{X}) | X_n = x_n) &= \mathrm{E}(f(X_1, \ldots, X_{n-1}, x_n)\, g(X_1, \ldots, X_{n-1}, x_n) | X_n = x_n) \\
&= \mathrm{E}(f(X_1, \ldots, X_{n-1}, x_n)\, g(X_1, \ldots, X_{n-1}, x_n)) \\
&\quad \text{(because of independence)} \\
&\geq \mathrm{E}(f(X_1, \ldots, X_{n-1}, x_n))\, \mathrm{E}(g(X_1, \ldots, X_{n-1}, x_n)) \\
&\quad \text{(by induction hypothesis)} \\
&= \mathrm{E}(f(\boldsymbol{X}) | X_n = x_n)\, \mathrm{E}(g(\boldsymbol{X}) | X_n = x_n).
\end{aligned}$$

Hence,

$$\mathrm{E}(f(\boldsymbol{X})\, g(\boldsymbol{X}) | X_n) \geq \mathrm{E}(f(\boldsymbol{X}) | X_n)\, \mathrm{E}(g(\boldsymbol{X}) | X_n).$$

Upon taking expectation on both sides of this equation, we have

$$\mathrm{E}(f(\boldsymbol{X})\, g(\boldsymbol{X})) \geq \mathrm{E}[\mathrm{E}(f(\boldsymbol{X}) | X_n)\, \mathrm{E}(g(\boldsymbol{X}) | X_n)].$$

Observe that $\mathrm{E}(f(\boldsymbol{X})|X_n)$ and $\mathrm{E}(g(\boldsymbol{X})|X_n)$ are increasing functions of X_n so that by the result of $n = 1$, we have

$$\begin{aligned}\mathrm{E}[\mathrm{E}(f(\boldsymbol{X})|X_n)\,\mathrm{E}(g(\boldsymbol{X})|X_n)] &\geq \mathrm{E}[\mathrm{E}(f(\boldsymbol{X}|X_n))]\,\mathrm{E}[\mathrm{E}(g(\boldsymbol{X}|X_n))] \\ &= \mathrm{E}(f(\boldsymbol{X}))\,\mathrm{E}(g(\boldsymbol{X})).\end{aligned}$$

This completes the proof for the case of n. □

Corollary 6.1 *If $h(X_1, \ldots, X_n)$ is a monotone function of each of its arguments, then for a set $U_1, \ldots, U_n$ of independent random numbers,*

$$\mathrm{Cov}[h(U_1, \ldots, U_n), h(1 - U_1, \ldots, 1 - U_n)] \leq 0.$$

Proof. Without loss of generality, by redefining h, we may assume that h is increasing in its first r arguments and decreasing in its remaining $n - r$ arguments. Let

$$\begin{aligned}f(x_1, \ldots, x_n) &= h(x_1, \ldots, x_r, 1 - x_{r+1}, \ldots, 1 - x_n), \\ g(x_1, \ldots, x_n) &= -h(1 - x_1, \ldots, 1 - x_r, x_{r+1}, \ldots, x_n).\end{aligned}$$

It follows that both f and g are increasing functions. By the preceding theorem,

$$\mathrm{Cov}[f(U_1, \ldots, U_n), g(U_1, \ldots, U_n)] \geq 0.$$

That is,

$$\mathrm{Cov}[h(U_1, \ldots, U_r, V_{r+1}, \ldots, V_n), h(V_1, \ldots, V_r, U_{r+1}, \ldots, U_n)] \leq 0, \tag{6.1}$$

where $V_i = 1 - U_i$. Observe that since $(h(U_1, \ldots, U_n), h(V_1, \ldots, V_n))$ has the same joint distribution as $(h(U_1, \ldots, U_r, V_{r+1}, \ldots, V_n), h(V_1, \ldots, V_r, U_{r+1}, \ldots, U_n))$, it follows from (6.1) that

$$\mathrm{Cov}[h(U_1, \ldots, U_n), h(V_1, \ldots, V_n)] \leq 0,$$

proving the corollary. □

When is antithetic variable effective? Here are some guidelines:

- Antithetic variables will result in a lower variance estimate than independent simulations only if the values computed from a path and its antithetic variables are *negatively correlated.*
- If H is monotone in each of its arguments, then antithetic variables reduce variance in estimating $\mathrm{E}(H(Z_1, \ldots, Z_n))$.
- If H is linear, then an antithetic estimate of $\mathrm{E}(H(Z_1, \ldots, Z_n))$ has zero variance.

- If H is symmetric, that is, $H(-Z) = H(Z)$, then an antithetic estimate of sample size $2n$ has the same variance as an independent sample of size n.

Example 6.2 *To illustrate some of these points, consider the simulations of payoff of options using antithetic variables. The function H here maps*

$$z \to \max\{0, S_0 \exp([r - \sigma^2/2]T + \sigma\sqrt{T}z) - K\}.$$

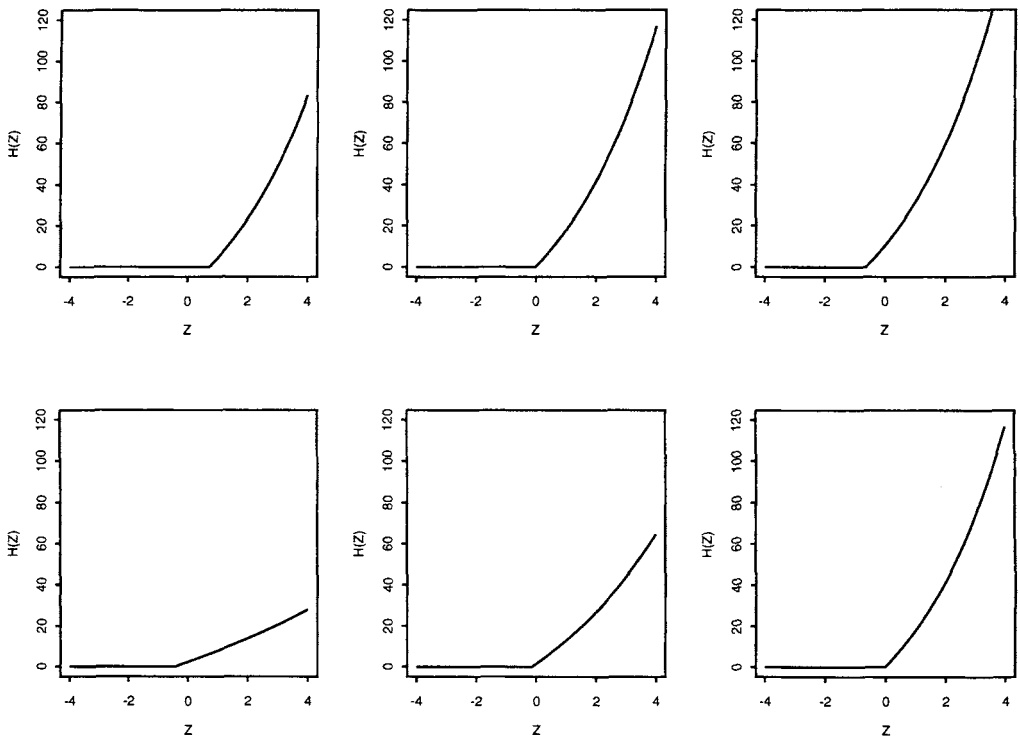


Fig. 6.1 Illustration of payoffs for antithetic comparisons.

In Fig. 6.1, the vertical axis is the payoff and the horizontal axis is the value of z, the input standard normal. All cases have $r = 0.05\%$, $K = 50$, *and* $T = 0.5$. *The top three cases have* $\sigma = 0.3$ *and* $S_0 = 40$, 50, *and* 60; *the second three cases have* $S_0 = 50$ *and* $\sigma = 0.10, 0.20, 0.30$. *The top three graphs correspond to the function H for options that are out-of-money* ($S_0 = 40$), *at-the-money* ($S_0 = 50$), *and in-the-money* ($S_0 = 60$), *respectively; the bottom three graphs correspond to low, intermediate, and high volatility for an at-the-money option. (The precise parameter values are given in the caption of the*

figure.) As one would expect, increasing moneyness and decreasing volatility both increase the degree of linearity. For the values indicated in the figure, we find numerically that antithetics reduce variance by 14%, 42%, *and* 80% *in the top three cases and by* 65%, 49%, *and* 42% *in the bottom three. Clearly, the more linear the function H is, the more effective the antithetic variable technique is.* □

Example 6.3 *Fig. 6.2 plots the payoff of $|S_T - K|$ on a straddle as a function of z. The parameter values are given in the caption. The graph shows a high degree of symmetry around zero, suggesting that antithetic variables may not be as effective as in the other cases. Numerical results here indicate that an antithetic estimate based on m pairs of antithetic variables has higher variance than an estimate based on $2m$ independent samples.* □

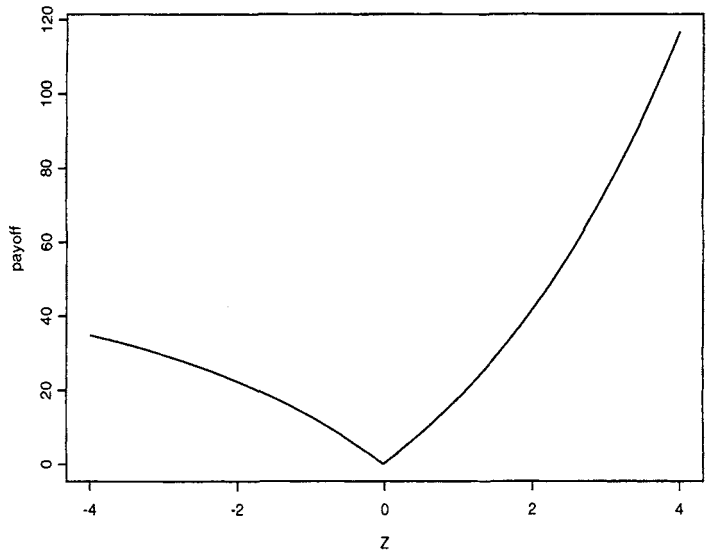


Fig. 6.2 Payoff on a straddle as a function of input normal Z based on the parameters $S_0 = K = 50$, $\sigma = 0.30$, $T = 1$, and $r = 0.05$.

```
######## Part(1): Standard method ########
p <- 10000
S0 <- 50
K <- 50
t <- 0.5
mu <- 0.05
sigma <- 0.3
nu <- mu - sigma^2/2
```

```
ST <- rep(0,2*p)
C <- rep(0,2*p)
P <- rep(0,2*p)
for (i in 1:(2*p))
{
z <- rnorm(1)
ST[i] <- S0*exp(nu*t+sigma*sqrt(t)*z)
C[i] <- max(ST[i]-K,0)
P[i] <- max(K-ST[i],0)
}
C.bar <- mean(C)
a1 <- var(C)/(2*p)
straddle <- C+P
b1 <- var(straddle)/(2*p)

######## Part(2): Antithetics ########
ST1 <- rep(0,p)
ST2 <- rep(0,p)
ST <- rep(0,p)
C1 <- rep(0,p)
C2 <- rep(0,p)
C <- rep(0,p)
P1 <- rep(0,p)
P2 <- rep(0,p)
P <- rep(0,p)
for (i in 1:p)
{
z1 <- rnorm(1)
z2 <- -z1
ST1[i] <- S0*exp(nu*t+sigma*sqrt(t)*z1)
ST2[i] <- S0*exp(nu*t+sigma*sqrt(t)*z2)
C1[i] <- max(ST1[i]-K,0)
C2[i] <- max(ST2[i]-K,0)
P1[i] <- max(ST1[i]-K,0)
P2[i] <- max(ST2[i]-K,0)
}
C <- (C1+C2)/2
P <- (P1+P2)/2
C.bar <- mean(C)
a2 <- var(C)/p
straddle <- C+P
b2 <- var(straddle)/p
a2/a1
b2/b1
```

6.3 STRATIFIED SAMPLING

The idea of stratification is often used in sample surveys (Barnett, 1991). The idea lies in the observation that the population may be heterogeneous and consists of various homogeneous subgroups (such as gender, race, social-economic status). If we wish to learn about the whole population (such as whether people in Hong Kong would like to have universal suffrage in 2007), we can take a random sample from the whole population to estimate that quantity. On the other hand, it would be more efficient to take small samples from each subgroup and combine the estimates in each subgroup according to the fraction of the population that subgroup represents. Since we can learn about the opinion of a homogeneous subgroup with a relatively small sample size, this stratified sampling procedure would be more efficient.

In general, if we want to estimate EX, where X depends on a random variable S that takes on one of the values in $\{1, \ldots, k\}$ with known probabilities, then the technique of stratification runs into k groups, with the ith group having $S = i$, letting $\bar{X}_i$ be the average values of X in those runs with $S = i$, and then estimating by $EX = \sum_{i=1}^{k} E(X|S = i)P(S = i)$ by

$$\sum_{i=1}^{k} \bar{X}_i P(S = i).$$

This is known as stratified sampling.

To illustrate this idea, suppose we want to estimate $E(g(U)) = \int_0^1 g(x)\,dx$. Consider two estimators based on a sample of $2n$ runs. The first one is the standard method,

$$\hat{g} = \frac{1}{2n} \sum_{i=1}^{2n} g(U_i).$$

Note that $E(\hat{g}) = E(g(U))$ and

$$\mathrm{Var}(\hat{g}) = \frac{1}{4n^2} \sum_{i=1}^{2n} \mathrm{Var}(g(U_i)) = \frac{1}{2n}[\int_0^1 g^2(x)\,dx - (\int_0^1 g(x)\,dx)^2].$$

On the other hand, we can write

$$E(g(U)) = \int_0^{1/2} g(x)\,dx + \int_{1/2}^1 g(x)\,dx.$$

Instead of selecting Us from [0,1], we can select the first n Us from [0,1/2] and the remaining n Us from [1/2,1] to construct a new estimator

$$\hat{g}_s = \frac{1}{2n}[\sum_{i=1}^{n} g(U_i/2) + \sum_{i=n+1}^{2n} g((U_i + 1)/2)].$$

It can be easily seen that if $U \sim U(0,1)$, then $V = a + (b-a)U$ is distributed as uniform (a,b). In particular, $U/2 \sim U(0,1/2)$ and $(U+1)/2 \sim U(1/2,1)$. To compute the variance of the new estimator, consider

$$\mathrm{Var}(\hat{g}_s) = \frac{1}{4n^2}\left\{\sum_{i=1}^{n}\mathrm{Var}(g(U_i/2)) + \sum_{i=n+1}^{2n}\mathrm{Var}(g((U_i+1)/2))\right\}.$$

Direct computations show that if $U_i \sim U(0,1)$, then

$$\begin{aligned}\mathrm{Var}(g(\frac{U_i}{2})) &= 2\int_0^{1/2} g^2(x)\,dx - 4m_1^2,\\ \mathrm{Var}(g(\frac{U_i+1}{2})) &= 2\int_{1/2}^{1} g^2(x)\,dx - 4m_2^2,\end{aligned}$$

where $m_1 = \int_0^{1/2} g(x)\,dx$ and $m_2 = \int_{1/2}^{1} g(x)\,dx$. Now

$$\mathrm{Var}(g(\frac{U_i}{2})) + \mathrm{Var}(g(\frac{U_i+1}{2})) = 2\int_0^1 g^2(x)\,dx - 4(m_1^2+m_2^2).$$

Consequently,

$$\mathrm{Var}(\hat{g}_s) = \frac{1}{2n}\left\{\int_0^1 g^2(x)\,dx - 2(m_1^2+m_2^2)\right\}.$$

Note that

$$(m_1+m_2)^2 + (m_1-m_2)^2 = 2(m_1^2+m_2^2).$$

Therefore,

$$\begin{aligned}\mathrm{Var}(\hat{g}_s) &= \frac{1}{2n}\left\{\int_0^1 g^2(x)\,dx - (m_1+m_2)^2 - (m_1-m_2)^2\right\}\\ &= \mathrm{Var}(\hat{g}) - \frac{1}{2n}(m_1-m_2)^2.\end{aligned}$$

Since this second term is always non-negative, stratification reduces the variance by an amount of this second term. The bigger the difference in m_1 and m_2, the greater the reduction in variance. In general, if more strata are introduced, more reduction will be achieved. One can generalize this result to the multi-strata case, but we will omit the mathematical details here.

Example 6.4 *Consider again* $\theta = \mathrm{E}(e^U) = \int_0^1 e^x\,dx$.

Recall that by standard Monte Carlo with $n = 2$,

$$\hat{g} = \frac{1}{2}(e^{U_1} + e^{U_2}),$$

and $\text{Var}(\hat{g}) = 0.121$. On the other hand, using stratification, we have

$$\hat{g}_s = \frac{1}{2}(e^{U_1/2} + e^{(U_2+1)/2}),$$

and $\text{Var}(\hat{g}_s) = \text{Var}(\hat{g}) - (m_1 - m_2)^2/2$, where $m_1 = \int_0^{1/2} e^x\,dx = e^{1/2} - 1$ and $m_2 = \int_{1/2}^1 e^x\,dx = e - e^{1/2}$. Thus,

$$\text{Var}(\hat{g}_s) = 0.121 - (2e^{1/2} - e - 1)^2/2 = 0.0325,$$

resulting a variance reduction of 73.13%. □

Stratified sampling is also very useful to draw random samples from designated ranges. For example, if we want to sample $Z_1, \ldots, Z_{100}$ from a standard normal distribution, the standard technique would partition the whole real line $(-\infty, \infty)$ into a number of bins and sample Zs from these bins randomly. In such a case, it is inevitable that some bins may have more samples while other bins, particularly those near the tails, may have no sample at all. Therefore, a random sample drawn this way would under represent the tails. While this may not be a serious issue in general, it may have severe effect when the tail is the quantity of interest, such as the case in the simulation of VaR. To ensure that the bins are regularly represented, we may generate the Zs as follows. Let

$$V_i = \frac{1}{100}(U_i + (i-1)), \quad i = 1, \ldots, 100,$$

where $U_i \sim U(0,1)$ i.i.d. By the property of uniform distribution, $V_i \sim U(\frac{i-1}{100}, \frac{i}{100})$. Now let $Z_i = \Phi^{-1}(V_i)$. Then Z_i falls between the $i-1$ and i percentiles of the standard normal distribution. For example, if $i = 1$, then $V = U/100 \sim U(0, 1/100)$ so that $Z = \Phi^{-1}(V)$ falls between $\Phi^{-1}(0) = -\infty$ and $\Phi^{-1}(0.01)$, i.e., the 0th and the 1st percentile of a standard normal distribution.

This method gives equal weight to each of the 100 equiprobable strata. Of course, the number 100 can be replaced by any number that is desirable. The price we pay in stratification is the loss of independence of the Zs. This complicates statistical inference for simulation results.

Example 6.5 *As an illustration of stratification, consider simulating standard normal random numbers via standard method and stratification method, respectively. As can be clearly seen from Fig. 6.3 and Fig. 6.4, stratified sampling generates samples much more uniformly over the range than standard Monte Carlo. The* SPLUS *codes for these simulations are given as follows.* □

```
######## Standard method ########
N <- 500
U1 <- runif(N)
X1 <- qnorm(U1)
hist(X1,freq=F,xlim=range(-3.5,3.5),nclass=50)
```

```
######## Stratified Sampling ########
U2 <- runif(N)
i <- 0:(N-1)
V <- (U2+i)/N
X2 <- qnorm(V)
hist(X2,freq=F,xlim=range(-3.5,3.5),nclass=50)
```

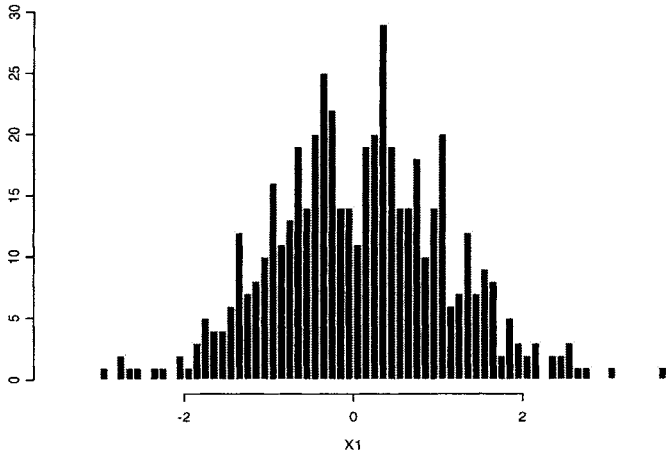


Fig. 6.3 Simulations of 500 standard normal random numbers by standard Monte Carlo.

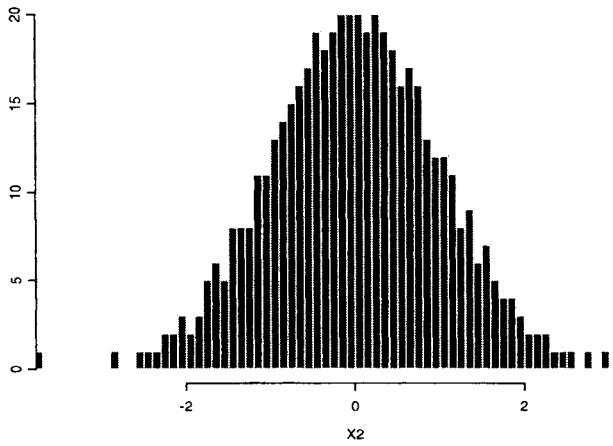


Fig. 6.4 Simulations of 500 standard normal random numbers by stratified sampling.

Example 6.6 *As a second illustration of stratification, consider the simulation of a European call option of Example 5.2 again.*

In Example 5.2, we simulate the terminal prices $S_1(T), \ldots, S_n(T)$ according to (5.9) and then compute the estimate as

$$\bar{C} = \frac{e^{-rT}}{n} \sum_{i=1}^{n} \max\{S_i(T) - K, 0\}.$$

In this standard simulation, the random normals are samples arbitrarily over the whole real line. We can improve the efficiency by introducing stratification.

1. Partition $(-\infty, \infty)$ into B strata or bins.
2. Set $V_i = \frac{1}{B}(U_i + (i-1))$, $i = 0, \ldots, B$ and generate the desired number of random samples (N_B, say) of Vs in the ith bin.
3. Apply $\Phi^{-1}(V_i)$ to get the desired normal random numbers from each bin and calculate $\bar{C}_i$ from each bin.
4. Average the $\bar{C}_i$ over the total number of bins to get an overall estimate $\bar{C}$.
5. Calculate the standard error as in the previous cases.

This numerical example uses $S_0 = 10$, $K = 12$, $r = 0.03$, $\sigma = 0.40$, and $T = 1$. The theoretical Black-Scholes price is 1.0139. We simulate the European option price for different bin sizes with $N_B \times B = 1,000$ in all cases. The effect of stratification increases as we increase the number of bins. The SPLUS code and the results (Table 6.1) are as follows. □

```
n <- 1000 # total sample size
B <- 100 #no. of bins
NB <- n/B
S0 <- 10
K <- 12
mu <- 0.03
sigma <- 0.4
nu <- mu-sigma^2/2
t <- 1
u <- 0
z <- 0
ST <- rep(0,NB)
Ci <- rep(0,NB)
Ci.bar <- 0
varr <- 0
for (i in 0:(B-1))
{
u <- runif(NB)
z <- qnorm((u+i)/B)
```

```
for (j in 1:NB)
{
ST[j] <- S0*exp(nu*t+sigma*sqrt(t)*z[j])
Ci[j] <- exp(-mu*t)*max(ST[j]-K,0)
}
Ci.bar <- Ci.bar + mean(Ci)
varr <- varr + var(Ci)
}
C <- Ci.bar/B
SE <- sqrt(varr/NB)/B
C
SE
```

Bins (B)	N_B	Mean ($\bar{C}$)	Std. Err.
1	1000	0.9744	0.0758
2	500	1.0503	0.0736
5	200	1.0375	0.0505
10	100	0.9960	0.0389
20	50	0.9874	0.0229
50	20	1.0168	0.0146
100	10	0.9957	0.0092
200	5	1.0208	0.0094
500	2	1.0151	0.0062
1000	1	1.0091	NA

Table 6.1 Effects of stratification for simulated option prices with different bin sizes.

Regular stratification puts equal weight on each of the B bins. Such an allocation may not be ideal as one would like to have sample sizes directly related to the variability of the target function over that bin. To illustrate this point, consider the payoff of a European call option again.

Example 6.7 *Stratified sampling for a European call with the same parameter values as in Example 6.5.*

We know that if $S_T < K$, then the payoff of the call is zero. Recall

$$S_T = S_0 e^{[r-\sigma^2/2]T+\sigma\sqrt{T}Z}.$$

Therefore, $S_T < K$ iff $S_0 e^{[r-\sigma^2/2]T+\sigma\sqrt{T}Z} < K$. That is,

$$Z < [\log(K/S_0) - (r - \sigma^2/2)T]/(\sigma\sqrt{T}) := L.$$

Every simulated $Z < L$ is being wasted as it just returns the value 0. We should only be concentrating on the interval $[L, \infty)$. How can we achieve this goal?

1. Find out the c.d.f. of a normal distribution Y restricted on $[L, \infty)$. It can be shown that Y has c.d.f.

$$F(y) = \frac{\Phi(y) - \Phi(L)}{1 - \Phi(L)}.$$

2. Use the inverse transform method to generate Y. Consider the inverse transformation of F, i.e., solve for y such that $y = F^{-1}(x)$. Writing it out, we have $x = F(y) = \frac{\Phi(y)-\Phi(L)}{1-\Phi(L)}$ so that

$$y = \Phi^{-1}(x(1 - \Phi(L)) + \Phi(L)).$$

   Now generate U from uniform (0,1) and evaluate

$$Y = \Phi^{-1}(U(1 - \Phi(L)) + \Phi(L)).$$

3. Plug in the generated Y into the simulation step of the payoff of the call and complete the analysis. Note that when evaluating the new estimator for the payoff, we need to multiply the factor $1 - \Phi(L)$. That is,

$$C^* = (1 - \Phi(L))\bar{C},$$

   where $\bar{C}$ is the average of the simulated payoffs using the truncated normal random variables.

In general, we would like to apply the stratification technique to bins in which the variability of the integrand is largest. Here, we just focus the entire sample on the case $S_T > K$. □

The SPLUS code and the output (Table 6.2) are given as follows:

```
n <- 1000 # total sample size
B <- 100 #no. of bins
NB <- n/B
S0 <- 10
K <- 12
mu <- 0.03
sigma <- 0.4
nu <- mu-sigma^2/2
t <- 1
```

```
u <- 0
z <- 0
ST <- rep(0,NB)
Ci <- rep(0,NB)
Ci.bar <- 0
varr <- 0
L <- {log(K/S0)-(mu-sigma^2/2)*t}/(sigma*t)
for (i in 0:(B-1))
{
u <- runif(NB)
v <- (u+i)/B
z <- qnorm(v*(1-pnorm(L))+pnorm(L))
for (j in 1:NB)
{
ST[j] <- S0*exp(nu*t+sigma*sqrt(t)*z[j])
Ci[j] <- exp(-mu*t)*max(ST[j]-K,0)
}
Ci.bar <- Ci.bar + mean(Ci)
varr <- varr + var(Ci)
}
C <- (1-pnorm(L))*Ci.bar/B
SE <- sqrt(varr/NB)/B
C
SE
##  The theoretical Black-Scholes value is 1.0139.
```

6.4 CONTROL VARIATES

The idea of control variates is very simple. Suppose we want to estimate $\theta = \mathrm{E}X$ from the simulated data. Suppose that for some other variable Y, the mean $\mu_Y = \mathrm{E}Y$ is known. Then for any given constant c, the quantity

$$X_{CV} = X + c(Y - \mu_Y)$$

is also an unbiased estimate of θ since $\mathrm{E}(X_{CV}) = \theta$. Presumably, if we choose the constant c cleverly, some form of variance reduction can be achieved. How can we do this? In other words, what would be a good choice of c? To answer this question, first consider the variance of the new estimator X_{CV}, call it σ^2_{CV}.

$$\sigma^2_{CV} = \mathrm{Var}(X + c(Y - \mu_Y)) = \mathrm{Var}X + c^2\mathrm{Var}Y + 2c\mathrm{Cov}(X, Y).$$

Bins (B)	N_B	Mean ($\bar{C}$)	Std. Err.	Adj. Mean	SE
1	1000	0.9744	0.0758	0.9842	0.1102
2	500	1.0503	0.0736	1.0303	0.0823
5	200	1.0375	0.0505	1.0235	0.0524
10	100	0.9960	0.0389	1.0101	0.0404
20	50	0.9874	0.0229	1.0058	0.0238
50	20	1.0168	0.0146	1.0147	0.0153
100	10	0.9957	0.0092	1.0089	0.0095
200	5	1.0208	0.0094	1.0160	0.0099
500	2	1.0151	0.0062	1.0143	0.0066
1000	1	1.0091	NA	1.0125	NA

Table 6.2 Effects of stratification for simulated option prices with restricted normal.

We would like to find c such that σ^2_{CV} is minimized. Differentiate the preceding expression with respect to c and set it equal to zero, we have

$$2c\text{Var}Y + 2\text{Cov}(X, Y) = 0.$$

Solving for such a c, we get, $c^* = -\text{Cov}(X, Y)/\text{Var}Y$ as the value of c that minimizes σ^2_{CV}. For such a c^*,

$$\sigma^2_{c^*} = \text{Var}X - \frac{\text{Cov}^2(X, Y)}{\text{Var}Y}.$$

The variable Y used in this way is known as a control variate for the simulation estimator X. Recall that $\text{Corr}(X, Y) = \text{Cov}(X, Y)/(\text{Var}X\text{Var}Y)^{1/2}$. Therefore,

$$\sigma^2_{c^*} = \text{Var}X(1 - \text{Corr}^2(X, Y)).$$

Hence, as long as $\text{Corr}(X, Y) \neq 0$, some form of variance reduction is achieved. In practice, quantities like $\sigma^2_Y = \text{Var}Y$ and $\text{Cov}(X, Y)$ are usually not available, they have to be estimated from the simulations based on sample values.

For example, let $\bar{X} = \sum_{i=1}^{n} X_i/n$ and $\bar{Y} = \sum_{i=1}^{n} Y_i/n$. Then

$$\begin{aligned}
\hat{\text{Cov}}(X,Y) &= \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}), \\
\hat{\sigma}_Y^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2, \\
\hat{c}^* &= -\frac{\hat{\text{Cov}}(X,Y)}{\hat{\sigma}_Y^2}.
\end{aligned}$$

Suppose we use $\bar{X}$ from simulation to estimate θ. Then the control variate would be $\bar{Y}$ and the control variate estimator is

$$\bar{X} + c^*(\bar{Y} - \mu_Y),$$

with variance equaling to

$$\frac{1}{n}(\text{Var}X - \frac{\text{Cov}^2(X,Y)}{\text{Var}Y}) = \frac{\sigma_X^2}{n}(1-\rho^2).$$

Equivalently, one can use the simple linear regression equation

$$X = a + bY + e, \quad e \sim i.i.d.\ (0, \sigma^2), \tag{6.2}$$

to estimate c^*. In fact, it can be easily shown that the least squares estimates of b, $\hat{b} = -\hat{c}^*$, see Weisberg (1985). In such a case, the control variate estimator is given by

$$\bar{X} + c^*(\bar{Y} - \mu_Y) = \bar{X} - \hat{b}(\bar{Y} - \mu_Y) = \hat{a} + \hat{b}\mu_Y, \tag{6.3}$$

where $\hat{a} = \bar{X} - \hat{b}\bar{Y}$ is the least squares estimate of a in (6.2). That is, the control variate estimate is equal to the estimated regression equation evaluated at the point μ_Y.

Notice that there is a very simple geometric interpretation using (6.2). First observe that the estimated regression line

$$\begin{aligned}
\hat{X} &= \hat{a} + \hat{b}Y \\
&= \bar{X} + \hat{b}(Y - \bar{Y}).
\end{aligned}$$

Thus, this line passes through the point $(\bar{Y}, \bar{X})$. Second, from (6.3),

$$\hat{X}_{CV} = \hat{a} + \hat{b}\mu_Y = \bar{X} - \hat{b}(\bar{Y} - \mu_Y).$$

Suppose that $\bar{Y} < \mu_Y$, that is, the simulation run underestimates μ_Y and suppose that X and Y are positively correlated. Then it is likely that $\bar{X}$ would underestimate $\text{E}(X) = \theta$. We therefore need to adjust the estimator upward and this is indicated by the fact that $\hat{b} = -\hat{c}^* > 0$. The extra amount

that needs to be adjusted upward equals $-\hat{b}(\bar{Y} - \mu_Y)$, which is governed by the linear equation (6.3).

Finally, $\hat{\sigma}^2$, the regression estimate of σ^2 is the estimate of $\mathrm{Var}(X - \hat{b}Y) = \mathrm{Var}(X + \hat{c}^*Y)$. To see this, recall from regression that

$$\begin{aligned}
\hat{\sigma}^2 &= \frac{1}{n}\sum_{i=1}^{n} \hat{e}_i^2 \\
&= \frac{1}{n}\sum_{i=1}^{n} (X_i - \hat{a} - \hat{b}Y_i)^2 \\
&= \frac{1}{n}\sum_{i=1}^{n} (X_i - (\bar{X} - \hat{b}\bar{Y}) - \hat{b}Y_I)^2 \\
&= \frac{1}{n}\sum_{i=1}^{n} ((X_i - \bar{X}) + \hat{b}(Y_i - \bar{Y}))^2 \\
&= \frac{1}{n}\sum_{i=1}^{n} ((X_i - \bar{X})^2 - \hat{b}^2(Y_i - barY)^2) \\
&= \hat{\mathrm{Var}}(X) - \hat{b}^2\hat{\mathrm{Var}}(Y) \\
&= \hat{\mathrm{Var}}(X - \hat{b}Y).
\end{aligned}$$

The last equality follows from a standard expansion of the variance estimate (see exercise 6.2). It follows that the estimated variance of the control variate estimator $\bar{X} + \hat{c}^*(\bar{Y} - \mu_Y)$ is $\hat{\sigma^2}/n$.

Example 6.8 *Consider the problem* $\theta = \mathrm{E}(e^U)$ *again.*

Clearly, the control variate is U itself. Now

$$\begin{aligned}
\mathrm{Cov}(e^U, U) &= \mathrm{E}(Ue^U) - \mathrm{E}(U)\mathrm{E}(e^U) \\
&= \int_0^1 xe^x\,dx - (e-1)/2 \\
&= 1 - (e-1)/2 = 0.14086.
\end{aligned}$$

The second last equality makes use of the facts from the previous examples that $\mathrm{E}(U) = 1/2, \mathrm{Var}U = 1/12$, and $\mathrm{Var}(e^U) = 0.242$. It follows that the control variate estimate has variance

$$\mathrm{Var}(e^U + c^*(U - 1/2)) = \mathrm{Var}(e^U)(1 - 12(0.14086)^2/0.242) = 0.0039,$$

resulting a variance reduction of $(.242 - .0039)/.242 \times 100\% = 98.4\%$. □

In general, if we want to have more than one control variate, we can make use of outputs from the multiple linear regression model given by

$$X = a + \sum_{i=1}^{k} b_i Y_i + e, \quad e \sim i.i.d.\ (0, \sigma^2).$$

In this case, the least squares estimates of a and b_is, $\hat{a}$ and $\hat{b}_i$s can be easily shown to satisfy $\hat{c}_i^* = -\hat{b}_i$, $i = 1, \ldots, k$. Furthermore, the control variate estimate is given by

$$\bar{X} + \sum_{i=1}^{k} (\bar{Y}_i - \mu_i) = \hat{a} + \sum_{i=1}^{k} \hat{b}_i \mu_i,$$

where $\mathrm{E}(Y_i) = \mu_i$, $i = 1, \ldots, k$. In other words, the control variate estimate is equal to the estimated multiple regression line evaluated at the point $(\mu_1, \ldots, \mu_k)$. By the same token, the variance of the control variate estimate is given by $\hat{\sigma^2}/n$, where $\hat{\sigma}^2$ is the regression estimate of σ^2.

Example 6.9 *Plunging along the same line, consider simulating the vanilla European call option as in Example 6.5, using the terminal value S_T as the control variate.*

The control variate estimator is given by

$$C_{CV} = C + c^*(S_T - \mathrm{E}(S_T)).$$

Recall $S_T = S_0 e^{(\nu T + \sigma\sqrt{T}Z)}$, it can be easily deduced that

$$\mathrm{E}(S_T) = S_0 e^{rT}, \tag{6.4}$$

$$\mathrm{Var}(S_T) = S_0^2 e^{2rT}(e^{\sigma^2 T} - 1). \tag{6.5}$$

The algorithm goes as follows:

1. For $i = 1, \ldots, N_1$, simulate a pilot of N_1 independent paths to get

$$\begin{aligned} S_T(i) &= S_0 e^{\nu T + \sigma\sqrt{T}Z_i}, \\ C(i) &= e^{-rT} \max\{0, S_T(i) - K\}. \end{aligned}$$

2. Compute $\mathrm{E}(S_T)$ as $S_0 e^{rT}$ or estimate it by $\sum_{i=1}^{N_1} S_T(i)/N_1$. Compute $\mathrm{Var}(S_T)$ as $S_0^2 e^{2rT}(e^{\sigma^2 T} - 1)$ or estimate it by $\frac{1}{N_1 - 1} \sum_{i=1}^{N_1} (S_T(i) - \bar{S}_T)^2$. Now estimate covariance by

$$\hat{\mathrm{Cov}}(S_T, C) = \frac{1}{N_1 - 1} \sum_{i=1}^{N_1} (S_T(i) - \bar{S}_T)(C(i) - \bar{C}),$$

where $\bar{C} = \sum_{i=1}^{N_1} C(i)/N_1$ and $\bar{S}_T = \sum_{i=1}^{N_1} S_T(i)/N_1$.

3. Repeat the simulations of S_T and C by means of control variate. For $i = 1, \ldots, N_2$, independently simulate

$$\begin{aligned} S_T(i) &= S_0 e^{\nu T + \sigma\sqrt{T}Z_i}, \\ C(i) &= e^{-rT} \max\{0, S_T(i) - K\}, \\ C_{CV}(i) &= C(i) + c^*(S_T(i) - \mathrm{E}(S_T(i))), \end{aligned}$$

where $c^* = -\hat{\text{Cov}}(S_T, C)/\hat{\text{Var}}S_T$ is computed from the preceding step.

4. Calculate the control variate estimator by

$$\bar{C}_{CV} = \frac{1}{N_2}\sum_{i=1}^{N_2} C_{CV}(i).$$

Complete the simulation by evaluating the standard error of $\bar{C}_{CV}$ and construct confidence intervals.

Here is the SPLUS code and output. □

```
N1 <- 500
N2 <- 50000
S0 <- 10
K <- 12
r <- 0.03
sigma <- 0.4
nu <- r-sigma^2/2
t <- 1
ST <- rep(0,N1)
C <- rep(0,N1)
ST2 <- rep(0,N2)
C2 <- rep(0,N2)
CCV <- rep(0,N2)

for (i in 1:N1)
{
z <- rnorm(1)
ST[i] <- S0*exp(nu*t+sigma*sqrt(t)*z)
C[i] <- exp(-r*t)*max(ST[i]-K,0)
}
ST.bar <- S0*exp(r*t)
VarST.hat <- S0^2*exp(2*r*t)*(exp(sigma^2*t)-1)
#ST.bar <- mean(ST)
C.bar <- mean(C)
Cov.hat <- sum((ST-ST.bar)*(C-C.bar))/(N1-1)
#VarST.hat <- sum((ST-ST.bar)^2)/(N1-1)
c <- -Cov.hat/VarST.hat

for (i in 1:N2)
{
z <- rnorm(1)
ST2[i] <- S0*exp(nu*t+sigma*sqrt(t)*z)
C2[i] <- exp(-r*t)*max(ST2[i]-K,0)
CCV[i] <- C2[i]+c*(ST2[i]-ST.bar)
```

```
}
CCV.bar <- mean(CCV)
Var.CCV <- sum((CCV-CCV.bar)^2)/(N2-1)
SE <- sqrt(Var.CCV)
CI <- CCV.bar-1.96*SE/sqrt(N2)
CI[2] <- CCV.bar+1.96*SE/sqrt(N2)
CCV.bar
CI
```

For $N_1 = 500$ and $N_2 = 50,000$, we have a 95% confidence interval for C_{CV} of [1.0023 1.0247]. In this case, the estimated call price is 1.0135 with standard error 0.0057.

In using control variates, there are a number of features that should be kept in mind.

- What should constitute the appropriate control? We have seen that in simple cases, the underlying asset prices may be appropriate. In more complicated situation, we may use some easily computed quantities that are highly correlated with the object of interest as control variates. For example, standard calls and puts frequently provide convenient source of control variates for pricing exotic options, and so does the underlying asset itself.

- The control variate estimator is usually unbiased by construction. Also, we can separate the estimation of the coefficients ($\hat{c}_i^*$) from the estimation of prices.

- The flexibility of choosing the c_is suggests that we can sometimes make optimal use of information. In any event, we should exploit the specific feature of the problem under consideration, rather than generic applications of routine methods.

- Because of its close relationship with linear regression, control variates are easily computed and explained.

- We have only covered linear control. In practice, one can consider using nonlinear control variates, for example, $\bar{X}\bar{Y}/\mu_Y$. Statistical inference for nonlinear control may be tricky though.

6.5 IMPORTANCE SAMPLING

After studying three variance reduction methods, we will pursue one last method, namely, importance sampling. This method is similar in idea to the acceptance-rejection method that was discussed in Chapter 4. Its main idea lies in approximating at places where the quantity of interest carries the most information, hence the name of importance sampling. This chapter

then concludes with examples illustrating the different methods of variance reduction in risk management.

Suppose that we are interested in estimating

$$\theta = \mathrm{E}[h(\boldsymbol{X})] = \int h(\boldsymbol{x}) f(\boldsymbol{x})\, d\boldsymbol{x},$$

where $\boldsymbol{X} = (X_1, \ldots, X_n)$ denotes an n-dimensional random vector having a joint p.d.f. $f(\boldsymbol{x}) = f(x_1, \ldots, x_n)$. Suppose that a direct simulation of the random vector $\boldsymbol{X}$ is inefficient so that computing $h(\boldsymbol{x})$ is infeasible. This inefficiency may be due to difficulties encountered in simulating $\boldsymbol{X}$, or the variance of $h(\boldsymbol{x})$ being too large, or a combination of both.

Suppose there exists another density $g(\boldsymbol{x})$, which is easy to simulate and satisfies the condition that $f(\boldsymbol{x}) = 0$ whenever $g(\boldsymbol{x}) = 0$. Then θ can be estimated by

$$\begin{aligned} \theta &= \mathrm{E}[h(\boldsymbol{x})] \\ &= \int \frac{h(\boldsymbol{x}) f(\boldsymbol{x})}{g(\boldsymbol{x})} g(\boldsymbol{x})\, d\boldsymbol{x} \\ &= \mathrm{E}_g \left[\frac{h(\boldsymbol{x}) f(\boldsymbol{x})}{g(\boldsymbol{x})} \right], \end{aligned}$$

where the notation E_g denotes the expectation of the random vector $\boldsymbol{X}$ taken under the density g, i.e., $\boldsymbol{X}$ has joint p.d.f. $g(\boldsymbol{x})$. It follows from this identity that θ can be estimated by generating $\boldsymbol{X}$ with density g and then using as the estimator the average of the values of $h(\boldsymbol{X}) f(\boldsymbol{X}) / g(\boldsymbol{X})$. In other words, we could construct a Monte Carlo estimator of $\theta = \mathrm{E}(h(\boldsymbol{X}))$ by first computing i.i.d. random vectors $\boldsymbol{X}_i$ with p.d.f. $g(\boldsymbol{X})$, then using the estimator

$$\hat{\theta} = \frac{1}{n} \sum_{i=1}^{n} \frac{h(\boldsymbol{X}_i) f(\boldsymbol{X}_i)}{g(\boldsymbol{X}_i)}.$$

If a density $g(\boldsymbol{x})$ can be chosen so that the random variable $h(\boldsymbol{X}) f(\boldsymbol{X}) / g(\boldsymbol{X})$ has a small variance, then this approach is known as the importance sampling approach and can result in an efficient estimator of θ.

To see how it works, note that the ratio $f(\boldsymbol{X})/g(\boldsymbol{X})$ represents the likelihood ratio of obtaining $\boldsymbol{X}$ with respective densities f and g. If $\boldsymbol{X}$ is distributed according to g, then $f(\boldsymbol{X})$ would be small relative to $g(\boldsymbol{X})$ and therefore when $\boldsymbol{X}$ is simulated according to g, the likelihood ratio $f(\boldsymbol{X})/g(\boldsymbol{X})$ will usually be small in comparison to 1. On the other hand, it can be seen that

$$\mathrm{E}_g \left[\frac{f(\boldsymbol{X})}{g(\boldsymbol{X})} \right] = \int \frac{f(\boldsymbol{x})}{g(\boldsymbol{x})} g(\boldsymbol{x})\, d\boldsymbol{x} = \int f(\boldsymbol{x})\, d\boldsymbol{x} = 1.$$

Thus, even though the likelihood ratio $f(\boldsymbol{X})/g(\boldsymbol{X})$ is smaller than 1, its mean is equal to 1, suggesting that it occasionally takes large values and results in a large variance.

To make the variance of $h(\boldsymbol{X})f(\boldsymbol{X})/g(\boldsymbol{X})$ small, we arrange for a density g such that those values of $\boldsymbol{X}$ for which $f(\boldsymbol{X})/g(\boldsymbol{X})$ is large are precisely the values for which $h(\boldsymbol{X})$ is small, thus making the ratio $h(\boldsymbol{X})f(\boldsymbol{X})/g(\boldsymbol{X})$ stays small. Since importance sampling requires h to be small sometimes, it works best when estimating a small probability. Further discussions on importance sampling and likelihood method are given in Glasserman (2003).

Example 6.10 *Consider the problem* $\theta = \mathrm{E}(U^5)$.

Suppose that we use the standard method $\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} U_i^5$, then we oversample the data near the origin and undersample the data near 1. It is easy to compute that

$$\mathrm{Var}(\hat{\theta}) = \frac{1}{n}\{\mathrm{E}U^{10} - (\mathrm{E}U^5)^2\} = \frac{1}{n}(\frac{1}{11} - \frac{1}{36}) = \frac{0.0631}{n}.$$

Now, suppose we use the importance sampling, putting more weights near 1. Let $g(x) = 5x^4$ for $0 < x < 1$. Then

$$\theta_I = \mathrm{E}_g(\frac{X^5 \cdot 1}{5X^4}) = \frac{\mathrm{E}_g X}{5}.$$

The variance of this method is

$$\begin{aligned}
\mathrm{Var}(\hat{\theta_I}) &= \frac{1}{25n}\{\mathrm{E}_g X^2 - (\mathrm{E}_g X)^2\} \\
&= \frac{1}{25n}\{\int_0^1 x^2(5x^4)\,dx - (\int_0^1 x(5x^4)\,dx)^2\} \\
&= \frac{1}{25n}\{5\int_0^1 x^6\,dx - (5\int_0^1 x^5\,dx)^2\} \\
&= \frac{1}{25n}\{\frac{5}{7} - (\frac{5}{6})^2\} \\
&= \frac{0.00794}{n},
\end{aligned}$$

resulting a variance reduction of 98.74%. □

How do we choose g in general? This requires the notion of the so-called tilted density. Recall the notation that $M(t) = \mathrm{E}(e^{tX})$ represents the moment generating function (m.g.f.) of the random variable X with density f.

Definition 6.2 *A density function*

$$f_t(x) = \frac{e^{tx}f(x)}{M(t)}$$

is called a tilted density of a given f, $-\infty < t < \infty$.

Note that from this definition, a random variable with density f_t tends to be larger than the one with density f when $t > 0$, and tends to be smaller when $t < 0$.

Example 6.11 *Let f be a Bernoulli density with parameter p. Then $f(x) = p^x(1-p)^{1-x}$, $x = 0, 1$. In this case, the m.g.f. is $M(t) = \mathrm{E}(e^{tX}) = pe^t + (1-p)$ so that*

$$\begin{aligned} f_t(x) &= \frac{1}{M(t)} e^{tx} f(x) \\ &= \frac{1}{M(t)} (pe^t)^x (1-p)^{1-x} \\ &= \left(\frac{pe^t}{pe^t + 1 - p}\right)^x \left(\frac{1-p}{pe^t + 1 - p}\right)^{1-x}. \end{aligned}$$

Thus, the tilted density f_t is a Bernoulli density with parameter $p_t = pe^t/(pe^t + 1 - p)$. □

In many instances, we are interested in sums of independent random variables. In these cases, the joint density $f(\boldsymbol{x})$ of $\boldsymbol{x} = (x_1, \ldots, x_n)$ can be written as the product of the marginals f_i of x_i so that

$$f(\boldsymbol{x}) = f_1(x_1) \cdots f_n(x_n).$$

In this situation, it is often useful to generate the X_i according to their tilted densities with a common t.

Example 6.12 *Let $X_1, \ldots, X_n$ be independent with marginal densities f_i. Suppose we are interested in estimating the quantity*

$$\theta = P(S \geq a),$$

where $S = \sum_{i=1}^n X_i$ and $a > \sum_{i=1}^n \mathrm{E}(X_i)$ is a given constant. We can apply tilted densities to estimate θ. Let $I\{S \geq a\}$ equal 1 if $S \geq a$ and 0 otherwise. Then

$$\theta = \mathrm{E}(I\{S \geq a\}),$$

where the expectation is taken with respect to the joint density. Suppose we simulate X_i according to the tilted density function $f_{t,i}$, where the value of $t > 0$ is to be specified. To construct the importance sampling estimator, note that $h(\boldsymbol{X}) = I\{S \geq a\}$, $f(\boldsymbol{X}) = \prod f_i(X_i)$, and $g(\boldsymbol{X}) = \prod f_{t,i}(X_i)$. The importance sampling estimator would be

$$\hat{\theta} = I\{S \geq a\} \prod_i \frac{f_i(X_i)}{f_{t,i}(X_i)}.$$

Now $f_i(X_i)/f_{t,i}(X_i) = M_i(t)e^{-tX_i}$, therefore,

$$\begin{aligned} \hat{\theta} &= I\{S \geq a\} \prod_i M_i(t) e^{-tX_i} \\ &= I\{S \geq a\} M(t) e^{-tS}, \quad (M(t) = \prod_i M_i(t)). \end{aligned}$$

Since it is assumed that $t > 0$, $S \geq a$ *iff* $e^{-tS} \leq e^{-ta}$ *and*

$$I\{S \geq a\}e^{-tS} \leq e^{-ta},$$

so that

$$\hat{\theta} \leq M(t)e^{-ta}.$$

We now find $t > 0$ *such that the right-hand side of the above inequality is minimized. In that case, we obtain an estimator that lies between 0 and* $\min_t M(t)e^{-ta}$. *It can be shown that such* t *can be found by solving the equation*

$$\mathrm{E}_t(S) = a.$$

After solving for t, *it can be utilized in the simulation. To be specific, suppose* $X_1, \ldots, X_n$ *are i.i.d. Bernoulli trials with* $p = p_i = 0.4$. *Let* $n = 20$ *and* $a = 16$. *Then*

$$\hat{\theta} = I\{S \geq a\}e^{-tS} \prod_i (pe^t + 1 - p).$$

Recall from the preceding example that the tilted density $f_{t,i}$ *is the p.d.f. of a Bernoulli trial with parameter* $p^* = pe^t/(pe^t + 1 - p)$. *It follows that*

$$\mathrm{E}_t(S) = 20p^* = \sum_{i=1}^{20} \frac{pe^t}{pe^t + 1 - p}.$$

Plugging in $n = 20, p = 0.4, a = 16$, *we have*

$$20\frac{0.4e^t}{0.4e^t + 0.6} = 16,$$

which leads to $e^{t^*} = 6$. *Therefore, we should generate Bernoulli trials with parameter* $0.4e^{t^*}/(0.4e^{t^*} + 0.6) = 0.8$ *as the g and evaluate* $M(t^*) = (0.4e^{t^*} + 0.6)^{20}$ *and* $e^{-t^*S} = (1/6)^S$. *The importance sampling estimator is now*

$$\hat{\theta} = I\{S \geq 16\}M(t^*)e^{-t^*S} = I\{S \geq 16\}3^{20}(1/6)^S.$$

Furthermore, we know that

$$\hat{\theta} \leq M(t^*)e^{-t^*a} = 3^{20}(1/6)^{16} = 0.001236.$$

Thus, in each iteration, the value of the importance sampling estimator lies between 0 *and* 0.001236.

On the other hand, we can also evaluate $\theta = P(S \geq 16)$ *exactly, which equals to the probability that a Binomial random variable with parameters* 20 *and* 0.4 *be at least as big as* 16. *This value turns out to be* 0.000317. *Recall the function* $h(\boldsymbol{X}) = I\{S \geq 16\}$. *This is a Bernoulli trial with parameter* $\theta =$ 0.000317. *Therefore, if we simulate directly from* X*s, the standard estimator* $\hat{\theta}_S$ *has variance*

$$\mathrm{Var}(\hat{\theta}_S) = \theta(1 - \theta) = 3.169 \times 10^{-4}.$$

As $0 \leq \hat{\theta} \leq 0.001236$, *it can be shown that*

$$\mathrm{Var}(\hat{\theta}) \leq (0.001236)^2/4 = 3.819 \times 10^{-7},$$

which is much smaller than the variance of the standard estimator $\hat{\theta}_S$. □

Another application of importance sampling is to estimate tail probabilities (recall at the beginning we mentioned that importance sampling works best in small probability). Suppose we are interested in estimating $P(X > a)$, where X has p.d.f. f and a is a given constant. Let $I(X > a) = 1$ if $X > a$ and 0 otherwise. Then

$$\begin{aligned} P(X > a) &= \mathrm{E}_f(I(X > a)) \\ &= \mathrm{E}_g[I(X > a)\frac{f(X)}{g(X)}] \\ &= \mathrm{E}_g[I(X > a)\frac{f(X)}{g(X)}|X > a]P(X > a) \\ &\quad +\mathrm{E}_g[I(X > a)\frac{f(X)}{g(X)}|X \leq a]P(X \leq a) \\ &= \mathrm{E}_g[\frac{f(X)}{g(X)}|X > a]P(X > a). \end{aligned}$$

Take $g(x) = \lambda e^{-\lambda x}$, $x > 0$, an exponential density with parameter λ. Then the above derivation shows

$$P(X > a) = \mathrm{E}_g[e^{\lambda X} f(X)|X > a]e^{-\lambda a}/\lambda.$$

Using the so-called "memoryless property", i.e., $P(X > s+t|X > s) = P(X > t)$, of an exponential distribution, it can be easily seen that the conditional distribution of an exponential distribution conditioned on $\{X > a\}$ has the same distribution as $a + X$. Therefore,

$$\begin{aligned} P(X > a) &= \frac{e^{-\lambda a}}{\lambda}\mathrm{E}_g[e^{\lambda(X+a)} f(X + a)] \\ &= \frac{1}{\lambda}\mathrm{E}_g[e^{\lambda X} f(X + a)]. \end{aligned}$$

We can now estimate θ by generating $X_1, \ldots, X_n$ according to an exponential distribution with parameter λ and using

$$\hat{\theta} = \frac{1}{\lambda}\frac{1}{n}\sum_{i=1}^{n} e^{\lambda X_i} f(X_i + a).$$

Example 6.13 *Suppose we are interested in* $\theta = P(X > a)$, *where* X *is standard normal. Then* f *is the normal density. Let* g *be an exponential*

density with $\lambda = a$. *Then*

$$\begin{aligned} P(X > a) &= \frac{1}{a}\mathrm{E}_g[e^{aX} f(X+a)] \\ &= \frac{1}{a\sqrt{2\pi}}\mathrm{E}_g[e^{aX-(X+a)^2/2}]. \end{aligned}$$

We can therefore estimate θ *by generating* X, *an exponential distribution with rate* a, *and then using*

$$\hat{\theta} = \frac{e^{-a^2/2}}{a\sqrt{2\pi}} \frac{1}{n} \sum_{i=1}^{n} e^{-X_i^2/2}$$

to estimate θ. *To compute the variance of* $\hat{\theta}$, *we need to compute quantities* $\mathrm{E}_g[e^{-X^2/2}]$ *and* $\mathrm{E}_g[e^{-X^2}]$. *These can be computed numerically and can be shown to be*

$$\mathrm{E}_g[e^{-X^2/2}] = ae^{a^2/2}\sqrt{2\pi}(1-\Phi(a)),\ \mathrm{E}_g[e^{-X^2}] = ae^{a^2/4}\sqrt{\pi}(1-\Phi(a/\sqrt{2})).$$

For example, if $a = 3$ *and* $n = 1$, *then* $\mathrm{Var}(e^{-X^2/2}) = 0.0201$ *and* $\mathrm{Var}(\hat{\theta}) = (\frac{e^{-4.5}}{3\sqrt{2\pi}})^2 \times 0.0201 \sim 4.38 \times 10^{-8}$. *On the other hand, a standard estimator has variance* $\theta(1-\theta) = 0.00134$. □

Consider simulating a vanilla European call option price again, using the importance sampling technique. Suppose that we evaluate the value of a deep out-of-money ($S_0 << K$) European call option with a short maturity T. Many sampling paths result $S_T \leq K$ and give zero-values. Thus, these samples are wasted. One possible way to deal with this problem is to increase the values of Z_is by sampling them from a distribution with large mean and large variance. Sample $\tilde{Z}_i$ from $\mathrm{N}(\frac{m}{\sigma\sqrt{T}}, s^2)$ so that

$$\sigma\sqrt{T}\tilde{Z}_i \sim \mathrm{N}(m, \sigma^2 T s^2).$$

Note that $\tilde{Z}_i$ can be written as

$$\tilde{Z}_i = \frac{m}{\sigma\sqrt{T}} + sZ_i, \quad Z_i \sim \mathrm{N}(0,1).$$

The importance sampling estimator is then given by

$$C_I = e^{-rT}\frac{1}{N}\sum_{i=1}^{N} \max\{S_0 e^{(r-\sigma^2/2)T+\sigma\sqrt{T}\tilde{Z}_i} - K, 0\} R(\tilde{Z}_i),$$

where

$$R(\tilde{Z}_i) = \frac{\frac{1}{\sqrt{2\pi}}\exp(-\tilde{\mathrm{Z}}_\mathrm{i}^2/2)}{\frac{1}{\sqrt{2\pi}\mathrm{s}}\exp(-\frac{1}{2\mathrm{s}^2}(\tilde{\mathrm{Z}}_\mathrm{i} - \frac{\mathrm{m}}{\sigma\sqrt{\mathrm{T}}})^2)} = s\exp\left(\frac{Z_i^2}{2} - \frac{\tilde{Z}_i^2}{2}\right).$$

Thus, C_I can be expressed as

$$C_I = s\,e^{-rT}\frac{1}{N}\sum_{i=1}^{N}\max\{S_0 e^{(r-\sigma^2/2)T+m+s\sigma\sqrt{T}Z_i}-K,0\}\exp(\frac{Z_i^2}{2}-\frac{(\frac{m}{\sigma\sqrt{T}}+sZ_i)^2}{2}).$$

Example 6.14 *Let* $S_0 = 100$, $K = 140$, $r = 0.05$, $\sigma = 0.3$, *and* $T = 1$. *We simulate the value of this deep out-of-money European call option, using the importance sampling technique and compare it with the result of standard method.*

The SPLUS code is as follows:

```
######## Part(1): Standard method ########
N <- 10000
S0 <- 100
K <- 140
t <- 1
r <- 0.05
sigma <- 0.3
nu <- r - sigma^2/2

Ci <- rep(0,N)
for (i in 1:N)
{
z <- rnorm(1)
ST <- S0*exp(nu*t+sigma*sqrt(t)*z)
Ci[i] <- exp(-r*t)*max(ST-K,0)
}
C.bar <- mean(Ci)
SE <- sqrt(var(Ci)/(N-1))
C.bar
SE

######## Part(2): Importance Sampling ########
N <- 10000
S0 <- 100
K <- 140
t <- 1
r <- 0.05
sigma <- 0.3
nu <- r - sigma^2/2
m <- 0.5
s <- 1.1

Ci <- rep(0,N)
for (i in 1:N)
```

```
{
z1 <- rnorm(1)
z2 <- m/(sigma*sqrt(t))+s*z1
ST <- S0*exp(nu*t+m+s*sigma*sqrt(t)*z1)
Ci[i] <- s*exp(-r*t)*max(ST-K,0)*exp(z1^2/2-z2^2/2)
}
C.bar <- mean(Ci)
SE <- sqrt(var(Ci)/(N-1))
C.bar
SE
```

For $N = 10,000$, we have $\bar{C}_I = 3.1202$ with standard error 0.0264 using importance sampling while getting $\bar{C} = 3.0166$ with standard error 0.1090 using standard method. The result shows that the importance sampling technique gives a more precise estimate of the price of the option, which has a theoretical Black-Scholes price 3.1187. □

6.6 EXERCISES

1. Let $U \sim U(0,1)$ and let a and b be two given constants with $a < b$. Show that $Y = a + (b-a)U$ is distributed as a $U(a,b)$ random variable.

2. Let $\hat{b}$ be the least squares estimate of b in the simple linear regression model $X = a + bY + e$, $e \sim (0, \sigma^2)$ *i.i.d.* Show that

$$\text{Var}(X - \hat{b}Y) = \text{Var}(X) - \hat{b}^2\text{Var}(Y).$$

3. Suppose you want to estimate $\theta = \int_0^1 e^{x^2}\,dx$. Show that generating a random number U and then using the antithetic estimator $(e^{U^2}(1 + e^{1-2U}))/2$ is better than generating two random numbers U_1 and U_2 and using the standard estimator $(e^{U_1^2} + e^{U_2^2})/2$.

4. Consider estimating $\theta = \int_0^1 4x^3\,dx$.
   (a) Using standard simulation technique, estimate θ.
   (b) Using antithetic variable technique, construct an improved estimate of θ.
   (c) Using stratification, construct another estimate of θ.
   (d) Construct a control variate estimate of θ.
   (e) Compare the performance of these different estimates.
   (f) Can you combine the above methods to improve the result?

5. Consider $\theta = \int_2^\infty (x-2)e^{-x}\,dx$.

(a) It is known that $\theta = \mathrm{E}[f(X)]$ where $X \sim$Exp(1). What is $f(X)$?

(b) Provide an algorithm to sample X from the interval $[2, \infty)$.

(c) Provide an algorithm to stratify X in the interval $[2, \infty)$ with equal probability $1/4$ for each stratified interval.

(d) Provide a Monte Carlo algorithm using $(X - 2)$ as the control variate.

6. Redo Examples 6.6, 6.7, and 6.9 using $S_0 = K = 100, r = 0.05, \sigma = 0.1$, and $T = 1$. Calculate the theoretical Black-Scholes price also.

7. Verify equations (6.4) and (6.5).

8. Consider a *truncated payoff* vanilla call option with maturity T and strike price K. The payoff function is given by

$$h(S_T) = \begin{cases} S_T - K & \text{if } \mathrm{K} \leq \mathrm{S_T} \leq \mathrm{S_b}, \\ 0 & \text{otherwise.} \end{cases}$$

The given constant S_b acts as a barrier, canceling the option whenever $S_T > S_b$. Assuming that the stock price follows a geometric Brownian motion with $\nu = r - \sigma^2/2$, where the risk free rate r and the volatility σ are known. Using the idea of antithetic variables, write a variance reduction algorithm to estimate the payoff function.

Simulation Techniques in Financial Risk Management
by Ngai Hang Chan and Hoi Ying Wong

7

Path-Dependent Options

7.1 INTRODUCTION

Contingent claims other than standard call and put options are known as exotic options. The most common type of exotic options is path-dependent options. As indicated by the name, the payoff of a path-dependent option depends on the entire path of the underlying asset prices, not just the terminal asset price alone. According to this definition, American options are path-dependent options because the option holder has to determine whether the options are worth exercising at each time point. The path-dependent feature of an option usually complicates the analytical tractability of valuation. Simulation would be the most useful alternative.

Owing to the need to value exotic options, this chapter studies simulation techniques for European- and American-style path-dependent options. Some of the options considered in this chapter have no analytical solutions.

7.2 BARRIER OPTION

Barrier options have become increasingly popular nowadays. A barrier option is very much like a "vanilla" option, which becomes alive when the barrier is crossed. Let K be the strike price, T be the time to maturity and V be the value of the barrier. A **down-and-in** barrier option becomes **alive** only if the stock price (usually counting only closing prices) goes below V before T. A **down-and-out** barrier option is **killed** if the stock price goes below V

before T. A down-and-in barrier call option is a cheaper tool to hedge against the upside risk. From the definition, it can be easily seen that holding both a down-and-in and down-and-out options with the same strike price K and maturity T is the same as holding a "vanilla" option. Let C_{di} and C_{do} be the option value of the down-and-in call and the down-and-out call, respectively. Then

$$C_{di} + C_{do} = C,$$

where C is the vanilla call price. Let

$$S_{min} = \min_{0<t\leq T} S(t) \quad \text{and} \quad I\{S_{min} < V\} = \begin{cases} 1 & S_{min} < V, \\ 0 & S_{min} \geq V, \end{cases}$$

be the realized minimum asset price and the indicator of the down-and-in option, respectively. Then, the value of the option can be written as

$$C_{di} = e^{-rT}\hat{\mathrm{E}}\{I\{S_{min} < V\}(S(T) - K)^+\},$$

where $\hat{\mathrm{E}}$ denotes the risk-neutral expectation. The other types of barrier options can be evaluated analogously.

To simulate the value of a down-and-in call option, the algorithm goes as follows:

1. Generate the daily stock price $S(t_1), S(t_2), \ldots, S(t_n = T)$. If $\min_i S(t_i) < V$, then set
$$C = e^{-rT}\max(S(T) - K, 0),$$
else set $C = 0$.

2. Repeat step 1 N times to obtain $C_1, \ldots, C_N$. The value of the down-and-in call option is given by
$$\bar{C} = \frac{1}{N}\sum_{i=1}^{N} C_i,$$
and the standard error of the estimator is given by
$$\sqrt{\frac{1}{N(N-1)}\sum_{i=1}^{N}(C_i - \bar{C})^2}.$$

Example 7.1 *Let $S_0 = 10$, $r = 0.23$, $\sigma = 0.4$, and $dt = 1/250$. Compute the value of a down-and-in call option with strike price $K = 12$, maturity $T = 1$, and barrier $V = 9$.*

The SPLUS code is as follows:

```
N <- 10000 #no. of path
```

```
S0 <- 10
K <- 12
V <- 9
Time <- 1
r <- 0.03
sigma <- 0.4
dt <- 1/250
nu <- r - sigma^2/2

t <- (1:(1/dt))*dt
ST <- rep(0,1/dt)
Ci <- rep(0,N)

for (j in 1:N)
{
z <- rnorm(1/dt)
for (i in 1:(1/dt))
{
ST[i] <- S0*exp(nu*t[i]+sigma*sqrt(dt)*sum(z[1:i]))
}
if (min(ST)<V) Ci[j] <- exp(-r*T)*max(ST-K,0)
else Ci[j] <- 0
}
C.bar <- mean(Ci)
SE <- sqrt(var(Ci)/(N-1))
CI <- C.bar - 1.96*SE
CI[2] <- C.bar + 1.96*SE
C.bar
SE
CI
```

For $N = 10,000$, we get $\bar{C} = 1.0273$ and the standard error of $\bar{C}$ is 0.02048. The 95% confidence interval for C is [0.9872,1.0675]. □

7.3 LOOKBACK OPTION

The payoffs of lookback options depend on the maximum or the minimum stock price during the life of the option. Denote the maximum (minimum) of the stock price over the time period $[0, T]$ by $S_{max}(T)$ ($S_{min}(T)$). Four popular lookback options are:

1. Floating strike lookback call (c_{fl}): payoff $= S_T - S_{min}(T)$;
2. Floating strike lookback put (p_{fl}): payoff $= S_{max}(T) - S_T$;
3. Fixed strike lookback call (c_{fix}): payoff $= \max(S_{max}(T) - K, 0)$;

4. Fixed strike lookback put (p_{fix}): payoff $= \max(K - S_{min}(T), 0)$.

There are lookback put-call parities connecting the floating strike lookback call (put) to the fixed strike lookback put (call). Specifically, four put-call parities of lookback options are:

1. $c_{fl}(t, S, S_{min}(t)) = S - e^{-r(T-t)} S_{min}(t) + p_{fix}(t, S, S_{min}(t); K = S_{min}(t))$;
2. $p_{fl}(t, S, S_{max}(t)) = e^{-r(T-t)} S_{max}(t) - S + c_{fix}(t, S, S_{max}(t); K = S_{max}(t))$;
3. $c_{fix}(t, S, S_{max}(t); K) = S - e^{-r(T-t)} K + p_{fl}(t, S, \max(S_{max}(t), K))$;
4. $p_{fix}(t, S, S_{min}(t); K) = e^{-r(T-t)} K - S + c_{fl}(t, S, \min(S_{min}(t), K))$.

These four put-call parities are model-independent, meaning that they are applicable to any asset dynamics. For a proof, we refer to the paper of Wong and Kwok (2003).

Pricing lookback options with simulation is very similar to that of the barrier option. Consider the floating strike lookback call option. The SPLUS code of Example 7.1 can be modified to obtain the lookback option price. We just compute

$$\frac{e^{-rT}}{N} \sum_{i=1}^{N} \left[S_i(T) - \min_j S_i(t_j) \right].$$

Other lookback options are valued in the same manner.

It is interesting to notice that simulating fixed strike lookback options requires less storage than simulating the floating ones. The reason is that payoffs of fixed strike lookback options do not depend on the terminal asset price, S_T. Therefore, after generating a sample path, only the maximum or minimum price of the path is required. With this observation and the lookback put-call parities, a storage saving approach to simulating floating strike lookback options can be developed. For valuing a floating strike lookback call, a fixed strike lookback put with strike price equaling to the realized minimum asset value is simulated. Then, the floating strike call price is extracted from the first put-call parity.

7.4 ASIAN OPTION

Asian options payoffs depend on the average of the underlying asset prices during the option life. Asian options are popular in the financial industry because they cost less than their vanilla counterparts and are less sensitive to the change in underlying asset prices. The common forms of averaging in option contracts can either be geometric average or arithmetic average of the underlying variables. Denote the geometric average and arithmetic average of

the underlying asset in the period $[0,T]$ by G_T and A_T, respectively. Then,

$$\begin{aligned} G_T &= \lim_{n\to\infty}\left[\prod_{i=1}^{n} S(t_i)\right]^{\frac{1}{n}} = \exp\left[\frac{1}{T}\int_0^T \log S(t)\,dt\right], \qquad (7.1)\\ A_T &= \lim_{m\to\infty}\frac{1}{n}\sum_{i=1}^{n} S(t_i) = \frac{1}{T}\int_0^T S(t)\,dt. \end{aligned}$$

For geometric Asian options, analytical pricing formulas are available in the literature, see, for example, Wong and Cheung (2004). However, almost all Asian options are traded with arithmetic average. For instance, two frequently traded Asian options are:

1. Floating strike Asian call. Payoff $= \max(S_T - A_T, 0)$;
2. Fixed strike Asian call. Payoff $= \max(A_T - K, 0)$.

In practice, the geometric Asian option prices are used as a control variate in simulating their arithmetic counterparts.

Let us illustrate the procedure by considering a fixed strike Asian call. The geometric version of the option has the payoff $\max(G_T - K, 0)$. Denote X_T by $\log G_T$, i.e.,

$$X_T = \frac{1}{T}\int_0^T \log S(\tau)\,d\tau.$$

By Itô's lemma,

$$\log S_\tau = \log S_t + \nu(\tau - t) + \sigma(W(\tau) - W(t)) \text{ (Recall: } \nu = r - \sigma^2/2),$$

which implies

$$\begin{aligned} X_T &= X_t\frac{t}{T} + \frac{1}{T}\int_t^T \log S(\tau)\,d\tau\\ &= X_t\frac{t}{T} + \frac{T-t}{T}\log S_t + \nu\frac{(T-t)^2}{2T} + \frac{\sigma}{T}\left[\int_t^T W(\tau)\,d\tau - (T-t)W_t\right]\\ &= X_t\frac{t}{T} + \frac{T-t}{T}\log S_t + \nu\frac{(T-t)^2}{2T} + \frac{\sigma}{T}\left[T(W_T - W_t) - \int_t^T \tau\,dW_\tau\right]\\ &= X_t\frac{t}{T} + \frac{T-t}{T}\log S_t + \nu\frac{(T-t)^2}{2T} + \frac{\sigma}{T}\int_t^T (T-\tau)\,dW_\tau, \end{aligned}$$

where the second to last line uses the integration by parts formula, see Example (2.2). By Itô's identities, see Exercise 1(d) of Chapter 2, we have

$$\mathrm{E}\int_t^T (T-\tau)\,dW(\tau) = 0 \quad \text{and} \quad \mathrm{Var}\left[\int_t^T (T-\tau)\,dW(\tau)\right] = \int_t^T (T-\tau)^2\,d\tau = \frac{(T-t)^3}{3}.$$

Therefore,

$$X_T \sim \mathrm{N}\left(\frac{t}{T}X_t + \frac{T-t}{T}\log S_t + \nu\frac{(T-t)^2}{2T}, \sigma^2\frac{(T-t)^3}{3T^2}\right). \tag{7.2}$$

Risk-neutral valuation asserts that

$$C_G^{fix}(t,S,G_t) = e^{-r(T-t)}\widehat{\mathrm{E}}\left[\max(e^{X_T}-K,0)\right].$$

Applying Lemma 3.1, we obtain the closed form solution as

$$C_G^{fix}(t,S,G_t) = S\left(\frac{G_t}{S}\right)^{\frac{t}{T}} e^{R(t,T)}\Phi(\hat{d}_1) - Ke^{-r(T-t)}\Phi(\hat{d}_2), \tag{7.3}$$

where

$$\begin{aligned}
\hat{d}_1 &= \frac{T\log\frac{S}{K} + t\log\frac{G_t}{S} + (r-\frac{\sigma^2}{2})\frac{(T-t)^2}{2} + \sigma^2\frac{(T-t)^3}{3T}}{\sqrt{\sigma^2\frac{(T-t)^3}{3}}}, \qquad (7.4)\\
\hat{d}_2 &= \hat{d}_1 - \sqrt{\sigma^2\frac{(T-t)^3}{3T^2}},\\
R(t;T) &= \left(r-\frac{\sigma^2}{2}\right)\frac{(T-t)^2}{2T} + \sigma^2\frac{(T-t)^3}{6T^2} - r(T-t).
\end{aligned}$$

With the analytical solution of the geometric Asian call, we simulate the arithmetic Asian price via control variate. The algorithm is presented as follows.

STEP 1: Generate daily stock prices $S(t_1), S(t_2), \ldots, S(t_n)$.

STEP 2: Set

$$\begin{aligned}
G_j &= \left[\prod_{i=1}^{n} S(t_i)\right]^{\frac{1}{n}},\ C_G^j = e^{-r(T-t)}\max(G_j - K, 0),\\
A_j &= \frac{1}{n}\sum_{i=1}^{n} S(t_i),\ C_A^j = e^{-r(T-t)}\max(A_j - K, 0).
\end{aligned}$$

STEP 3: Repeat Steps 1 and 2 N times.

STEP 4: Compute the regression coefficients a and b by fitting

$$C_A^j = a + bC_G^j,\ j = 1,2,\ldots,N.$$

STEP 5: $C_A^{fix} = a + bC_G^{fix}(t,S,G_t)$ with formula (7.3) applied.

Example 7.2 *Consider the parameter values:* $S_t = 10, r = 0.03, \sigma = 0.4, t = 0.2, T = 1$ *and the realized arithmetic average* $A_t = 10.5$. *Simulate the arithmetic Asian call option with a fixed strike price of \$12.*

We implement the preceding algorithm by the following SPLUS code:

```
##Define parameters###
N  <- 1000
TIME <- 1
time <- 0.2
r <- 0.03
sigma <- 0.4
St <- 10
K  <- 12
At <- 10.5
Gt <- 10.5
nu <- r - (sigma^2/2)
dT <- TIME-time
dt <- TIME/100
n  <- floor(dT*100)
S <- c(rep(St,N))
A <- c(rep(At*(100-n)+St,N))
lnG <- c(rep(log(Gt)*(100-n)+log(St),N))
CA <- c(rep(0,N))
CG <- c(rep(0,N))

#Generate asset price paths and averages
for (j in 1:n){
S <- S * exp ( nu*dt + sigma*sqrt(dt)*rnorm(N,0,1) )
A <- A + S
lnG <- lnG + log(S)
}
A <- A/101
lnG <- lnG/101
CA <- exp(-r*dT)*pmax(A-K,0)
CG <- exp(-r*dT)*pmax(exp(lnG)-K,0)

#Compute the analytical solution of the geometric Asian option
SIGMA <- sigma^2*dT^3/3/TIME^2
d1 <- (log(St/K)+time/TIME*log(Gt/St)+nu*dT^2/2/TIME+SIGMA)/sqrt(SIGMA)
d2 <- d1 - sqrt(SIGMA)
RtT <- nu*dT^2/2/TIME + SIGMA/2 - r*dT
CGfix <- St*(Gt/St)^(time/TIME)*exp(RtT)*pnorm(d1) - K*exp(-r*dT)*pnorm(d2)

#Compute the regression coefficients
a <- which ( CA *CG !=0)
aa <- coef( lm( CA ~ CG ) )
price<- aa[1]+CGfix*aa[2]
```

This simulation gives the arithmetic Asian call (AAC) price to be 0.1698. The analytical price for the geometric Asian call (GAC) is computed as 0.1318. The AAC is a bit more expensive than the GAC because the arithmetic mean always dominates the geometric mean. The computational time is about 10 seconds. □

7.5 AMERICAN OPTION

American options allow the holder to exercise prior to maturity. This early exercise feature exists in major financial markets. The valuation and optimal exercise of American options is one of the most challenging problems in derivatives finance, especially when more than one factor is involved in the option contract.

Although simulation techniques can be used to generate future scenarios, the forward looking feature of simulation complicates the valuation of American option, where optimal exercising policy have to be constructed via backward reduction. When an American put option is valued with binomial tree, one has to determine if it is optimal to exercise the option at each node in a backward manner. A practical approach to valuing American options with simulation is proposed by Longstaff and Schwartz (2001). This section presents the idea of American option pricing using this approach.

7.5.1 Simulation: Least Squares Approach

The best way to illustrate the least squares approach of Longstaff and Schwartz (2001) is by means of a concrete example. In the following numerical example, we shall introduce the algorithm in detail first and explain the concepts later.

Example 7.3 *Let* $S(0) = 10$, $r = 0.03$, $\sigma = 0.4$. *Compute the value of an American put option with strike price* $K = 12$ *and maturity* $T = 1$. *For simplicity, assume that the option can be exercised at* $t = 1/3$, $2/3$ *and* 1.

We use the formula $S(t+\Delta t) = S(t)\exp[(r-\sigma^2/2)\,\Delta t + \sigma\,\Delta W_t]$ to generate asset prices at exercise time points: $t = 1/3, 2/3$ and 1. Table 7.1 gives eight sample paths. Terminal payoffs corresponding to each path, Y_3, are given by the last column of the table. Discounting the sample mean of the terminal payoffs estimates the European put price to be \$2.4343. This is a lower bound for the American put option.

At time $t = 2/3$, the option holder must decide whether to exercise the option immediately or to continue the option when the option is in-the-money. To make the decision, the holder should compare the cash flows of immediate exercise with the expected payoff of continuation given the asset price at time 2/3. Therefore, it is essential to estimate the conditional expected payoff. To do this, we collect the response variable $Y_3 e^{-r\Delta t}$ and the explanatory variable

Path	$t = 1/3$	$t = 2/3$	$t = 1$	$Y_3 = \max(K - S(1), 0)$
1	8.3826	9.9528	6.7581	5.2419
2	11.9899	13.8988	14.5060	0
3	13.1381	17.4061	13.4123	0
4	6.8064	7.8115	10.6520	1.3480
5	7.0508	9.1293	7.4551	4.5449
6	11.2214	8.3600	9.2896	2.7104
7	8.9672	8.7787	9.0822	2.9178
8	11.5336	10.9398	8.6958	3.3042

Table 7.1 Sample paths.

$S(2/3)$ for in-the-money paths in Table 7.2, where $\Delta t = 1/3$. We model the expected payoff from continuation at time $t = 2/3$ as a quadratic polynomials, $f_2(S_t)$, of asset values at time $t = 2/3$. Coefficients of the polynomials are estimated from the data in Table 7.2 by the least squares method. Therefore, we estimate $\hat{a}_0, \hat{a}_1$ and $\hat{a}_2$ from the regression line:

$$Y_3 e^{-r\Delta t} = \hat{a}_0 + \hat{a}_1[S(2/3)] + \hat{a}_2[S(2/3)]^2 + \epsilon.$$

The resulting formula is

$$\mathrm{E}[Y_3 e^{-r\Delta t} | S(2/3)] = -82.5347 + 17.7788[S(2/3)] - 0.9063[S(2/3)]^2 := f_2(S).$$

Path	$Y_3 e^{-r\Delta t}$	$S(2/3)$	Exercise in-the-money?
1	5.1898	9.9528	Yes
2	——	13.8988	No
3	——	17.4061	No
4	1.3346	7.8115	Yes
5	4.4997	9.1293	Yes
6	2.6834	8.3600	Yes
7	2.8888	8.7787	Yes
8	3.2714	10.9398	Yes

Table 7.2 Regression at $t = 2/3$.

With this conditional expectation function, $f_2(S)$, we are able to compare the value of immediate exercise, $K - S(2/3)$, and compute payoffs, Y_2, for each path at $t = 2/3$. The value of Y_2 is obtained by the formula,

$$Y_2 = \begin{cases} K - S(2/3), & \text{if } K - S(2/3) \geq f_2(S(2/3)), \\ e^{-\Delta t} Y_3, & \text{otherwise.} \end{cases}$$

This formula asserts that the payoff at time $t = 2/3$ is $K - S$ if exercising the option is worth more than the expected payoff from holding it; otherwise, the payoff at time 2/3 becomes the discounted cash flow in the next exercise time. The last column of Table 7.3 gives the expected payoffs, Y_2, for each sample path.

Path	Exercise $K - S(2/3)$	Continuation $f_2(S(2/3))$	$e^{-r\Delta t}Y_3$	Y_2
1	2.0472	4.6380	5.1898	5.1898
2	——	——	0	0
3	——	——	0	0
4	4.1885	1.0428	1.3346	4.1885
5	2.8707	4.2388	4.4997	4.4997
6	3.6400	2.7554	2.6834	3.6400
7	3.2213	3.6959	2.8888	2.8888
8	1.0602	3.4968	3.2714	3.2714

Table 7.3 Optimal decision at $t = 2/3$.

Next, we repeat the procedure for $t = 1/3$. In Table 7.4, all sample paths are in-the-money except path three. Then, the least squares estimation corresponding to in-the-money paths gives

$$\mathrm{E}[Y_2 e^{-r\Delta t}|S(1/3)] = -8.9488 + 3.3104S(1/3) - 0.2036[S(1/3)]^2 := f_1(S).$$

This regression function determines the exercising policy at $t = 1/3$.

Path	$Y_2e^{-r\Delta t}$	$S(1/3)$	Exercise in-the-money?
1	5.1381	8.3826	Yes
2	0	11.9899	Yes
3	0	13.1381	No
4	4.1468	6.8064	Yes
5	4.4549	7.0508	Yes
6	3.6038	11.2214	Yes
7	2.8600	8.9672	Yes
8	3.2388	11.5336	Yes

Table 7.4 Regression at $t = 1/3$.

Once again, the Y_1 in Table 7.5 is computed according to the optimal decision by the rule,

$$Y_1 = \begin{cases} K - S(1/3), & \text{if } K - S(1/3) \geq f_1(S(1/3)), \\ e^{-\Delta t} Y_2, & \text{otherwise.} \end{cases}$$

Finally, the current price of the American option is estimated by the average of $e^{-r\Delta t} Y_1$, i.e., \$3.0919, which is higher than the European option price \$2.4343. □

Path	Exercise $K - S(1/3)$	Continuation $f_1(S(1/3))$	$e^{-r\Delta t} Y_2$	Y_1
1	3.6174	4.4921	5.1381	5.1381
2	0.0101	1.4689	0	0
3	——	——	0	0
4	5.1936	4.1494	4.1468	5.1936
5	4.9492	4.2688	4.4549	4.9492
6	0.7786	2.5572	3.6038	3.6038
7	3.0328	4.3620	2.8600	2.8600
8	0.4664	2.1440	3.2388	3.2388

Table 7.5 Optimal decision at $t = 1/3$.

7.5.2 Analyzing the Least Squares Approach

Consider an American put option with exercise rights at $t_1 < \cdots < t_n = T$. To simplify matters, we assume $t_{j+1} - t_j = \Delta t$ for $j = 1, 2, \ldots, n-1$. Given a sample path of the underlying asset price, $\{S(t_1), S(t_2), \ldots, S(t_n)\}$, we study possible payoffs received by the option holder at each of the exercise time points. Clearly, if the option is not exercised prematurely, then the holder receives the terminal payoff, denoted as $Y_n = \max(K - S(t_n), 0)$. At time $t = t_{n-1}$, the corresponding payoff, Y_{n-1}, depends on the holder's decision of exercising the option. Therefore,

$$Y_{n-1} = \begin{cases} K - S(t_{n-1}), & \text{exercise,} \\ e^{-r\Delta t} Y_n, & \text{continue.} \end{cases}$$

This formula indicates that the option holder receives $K - S(t_{n-1})$ if the optimal decision is to exercise the option. Otherwise, the holder will receive a cash flow of Y_n at the next time step. The present value of this cash flow is obtained through multiplying a discounted factor $e^{-r\Delta t}$. Inductively, the payoff Y_j at time t_j can be described as

$$Y_j = \begin{cases} K - S(t_j), & \text{exercise,} \\ e^{-r\Delta t} Y_{j+1}, & \text{continue.} \end{cases} \tag{7.5}$$

This iterative process stops until Y_1 is obtained. Since the option holder has no exercise right in the time period $[0, t_1)$, the American put option can be viewed as a European option that expires at t_1 with payoff Y_1. Risk-neutral valuation allows us to value the American put, $P_A(0, S)$, as

$$P_A(0, S) = \widehat{\mathrm{E}}\left[e^{-rt_1} Y_1 | S_0 = S\right].$$

Therefore, a typical simulation algorithm generates N sample paths, each follows the algorithm to obtain $\{Y_1^{(1)}, \ldots, Y_1^{(N)}\}$. The American put is estimated by

$$P_A(0, S) = \frac{1}{N} \sum_{i=1}^{N} e^{-rt_1} Y_1^{(i)}. \tag{7.6}$$

The above simulation is incomplete, however. To simulate the American put, the payoff, Y_1, at time t_1 should be obtained via simulation. This requires the simulation algorithm to detect optimal exercise at each time point successively. In other words, we have to clarify the condition of exercising the option in (7.5). It is crucial that the optimal decision should not be made by simply comparing the values of $K - S(t_j)$ and $e^{-r\Delta t} Y_{j+1}$ in (7.5). The reason is that the decision at time t_j should be based on the information up to t_j. However, the value Y_{j+1} depends on the asset value at t_{j+1}. The correct approach is to compare the immediate exercise cash flow $K - S(t_j)$ with the expectation on the discounted cash flow conditional on the asset price $S(t_j)$. This leads (7.5) to

$$Y_j = \begin{cases} K - S(t_j), & \text{if } K - S(t_j) \geq f_j(S(t_j)), \\ e^{-r\Delta t} Y_{j+1}, & \text{if } K - S(t_j) < f_j(S(t_j)), \end{cases} \tag{7.7}$$

where $f_j(S(t_j))$ is the conditional expectation function at t_j, that is,

$$f_j(S(t_j)) = \widehat{\mathrm{E}}\left[e^{-r\Delta t} Y_{j+1} | S(t_j)\right]. \tag{7.8}$$

The key to the Longstaff and Schwartz (2001) approach is the use of least squares to estimate the function, $f_j(S)$. Under certain technical conditions, it can be shown that the function $f_j(S(t_j))$ can be approximated by a polynomial of $S(t_j)$. In other words,

$$f_j(S(t_j)) = \sum_{k=0}^{\infty} a_k [S(t_j)]^k,$$

where $\{a_k\}$ converges to zero rapidly. Therefore, one way to approximate $f_j(S)$ is by truncating the polynomial of infinite order to a finite order polynomial. Coefficients of the finite order polynomials are estimated through the least squares method.

In Example 7.3, we use a polynomials of degree 2 to approximate $f_j(S)$. The simulation starts by generating N asset price paths, $\{S_i(t_1), \ldots, S_i(t_n)\}$ for $i = 1, 2, \ldots, N$. When $t = t_n$, it is clear that $Y_n^{(i)} = \max[K - S_i(t_n), 0]$ for the path i. We go one step back to the time point $t = t_{n-1}$, where N possible asset prices have been generated. Then, the coefficients a_0, a_1, and a_2 are obtained by taking least squares estimation to the regression line:

$$f_{n-1}(S) = \widehat{\mathrm{E}}\left[e^{-r\Delta t} Y_n | S\right] = a_0 + a_1[S(t_{n-1})] + a_2[S(t_{n-1})]^2. \tag{7.9}$$

The estimation is based on the sample $\{(S_i(t_{n-1}), Y_n^{(i)}) | K > S_i(t_{n-1}), i = 1, \ldots, N\}$, i.e., in-the-money paths. Then, payoffs at t_{n-1} are calculated via the rule (7.7). Having a sample of payoffs $\{Y_{n-1}^{(i)} | i = 1, 2, \ldots, N\}$ at t_{n-1}, we go one step back to the time point t_{n-2} and repeat the process. Eventually, we obtain N possible payoffs,$\{Y_1^i | i = 1, 2, \ldots, N\}$, at t_1. Monte Carlo simulation estimates the current option price by the average in (7.6).

Remarks:

1. In the regression equation (7.9), only in-the-money paths are used in the least squares estimation as these paths are sensitive to immediate exercise. Remember that the option holder will exercise the option only when it is in-the-money.

2. An obvious way to improve the accuracy is to increase the number of terms in (7.9). However, one has to strike a balance between increasing the number of terms and the quality of estimates. Numerical experiments show that polynomials of degree 3 is a reasonable choice.

3. Instead of using ordinary monomials as basis functions in (7.9), one may consider other basis functions, like Hermite, Laguerre, Legendre, Chebyshev, Gegenbauer, and Jacobi polynomials. Numerical tests of Moreno and Navas (2003) show that the least squares approach is quite robust to the choice of basis functions. For more complex derivatives, this choice can slightly affect option prices.

4. The recent analysis of Stentoft (2004) indicates that a modified specification using ordinary monomials is preferred over the specification based on Laguerre polynomials used in Longstaff and Schwartz (2001). Furthermore, the least squares method is computationally more efficient than other numerical methods, such as finite difference, especially when high dimensional problems are concerned.

5. The paper by Longstaff and Schwartz (2001) points out that the R^2 values of the regressions are often low. This means that the volatility of unexpected cash flows is large relative to the expected cash flows. However, since the least squares simulation is based on conditional first

moments rather than higher moments, the R^2's of the regression should have little impact on estimated American option price.

6. If the user is really concerned about the R^2, it may be more efficient to use other techniques such as weight least squares and GMM in estimating the conditional expectation function.

Example 7.4 *Using the parameters in the preceding example, simulate the American put price with continuous exercise rights and hence determine the optimal exercise policy. The simulation is based on 10,000 sample paths with* $\Delta t = 1/100$.

The SPLUS code is as follows:

```
N <- 10000
n <- 100
dt<- 1/n
r <- 0.03
sigma <- 0.4
S0 <- 10
K  <- 12
nu <- r-sigma^2/2
stock <- matrix(0,N,n+1)
y <- c(rep(0,N))
put <- c(rep(0,N))
boundary <- c(rep(0,n))
stock[,1] <- S0
check1 <- proc.time()  #the first check point

# generate asset price paths
for (i in 1:n){
stock[,i+1] <- stock[,i]* exp(nu*dt+sigma*sqrt(dt)*rnorm(N,0,1))
}
y <- pmax( (K-stock[,n+1]), 0 )
for (j in n:2){
a <- which( stock[,j]<K ) # identify in-the-money paths
if ( length(a) >= 3 ) { # ensure there's a solution for the regression
# Compute the conditional expectation function
S <- stock[a,j]
A <- coef( lm( y[a] ~ S + S^2, singular.ok="T" ) )
if ( is.na(A[3]) ) { A[3]<-0 }
X <- matrix( c(rep(1,N),stock[,j],stock[,j]^2), ncol=3 )
put <- X %*% A
put <- exp(-r*dt)*pmax(put,0)

# determinate Y & find boundary of K-S(t) <f(S(t))
```

```
b <- which( (K-stock[,j]) > put )
y <- exp(-r*dt)*y
y[b] <- K - stock[b,j] # assign y as K-S when K-S > P
if ( length(b)==0 ) {boundary[j]<-NA} # boundary cannot be estimated
else { boundary[j] <- max(stock[b,j]) }
}
else {
y <- exp(-r*dt)*y
boundary[j] <- NA
}
}
check2 <- proc.time()  #the second check point
boundary[n+1] <- K
boundary[1:20] <- NA #give up the estimate for t<0.2
time<-c(0:n)/n
plot(time,boundary,type="h",ylim=c(0,K),xlab="time",ylab="asset price")

price <- exp(-r*dt)* mean(y)
check2-check1  #Check the CPU time for the valuation
price #price of American Put Option
```

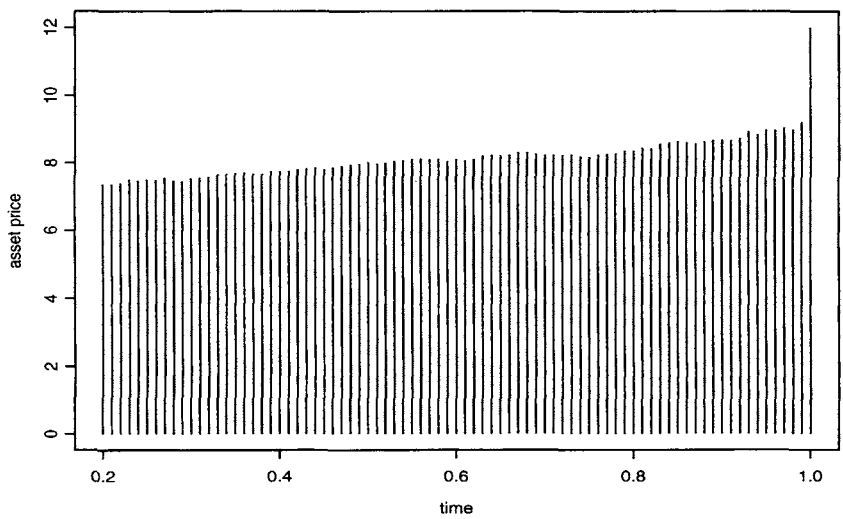


Fig. 7.1 The exercising region of the American put option.

By using quadratic conditional expectation functions, our simulation estimates the American put price as 2.739 within 15 seconds, which is consistent

with the binomial model of Hull (2006). For the early exercise policy, we collect the maximum asset value that belongs to the exercising region at each time. For $t \geq 0.2$, Fig. 7.1 plots the exercise policy against time. The option is optimal to exercise if the stock price falls into the shaded region. It is seen that the early exercise boundary looks like an increasing function of calendar time and hence a decreasing function of option maturity. For $t < 0.2$, our simulation has no path in the exercising region so that we are unable to graph the exercising boundary. □

7.5.3 American-Style Path-Dependent Options

The examples considered so far are relevant to pricing American put options; the least squares approach is applicable to any early exercisable contingent claims. Denote the terminal payoff function of a path-dependent option by $F(S_T, \xi_T)$ where ξ is an exogenous variable. For instance, $\xi_T = S_{min}(T)$ for a barrier option or a lookback option and $\xi_T = A_T$ for an Asian option. The American style path-dependent option with payoff $F(S_T, \xi_T)$ can be simulated as follows.

STEP 1: Generate asset price paths $\{S_i(t_1), S_i(t_2), \ldots, S_i(t_n)\}$ for $i = 1, 2, \ldots, N$. Set $j = n - 1$ and $Y_n^i = F(S_i(t_n), \xi_i(t_n))$.

STEP 2: Use least squares to estimate coefficients of a polynomials of degree m, $\mathcal{P}_m(S_i(t_j), \xi_i(t_j))$ from:

$$e^{-r\Delta t} Y_{j+1}^i = \mathcal{P}_m(S_i(t_j), \xi_i(t_j)),$$

for in-the-money paths.

STEP 3: If $F(S_i(t_j), \xi_i(t_j)) \geq \mathcal{P}_m(S_i(t_j), \xi_i(t_j))$, then set $Y_j^i = F(S_i(t_j), \xi_i(t_j))$; otherwise, set $Y_j^i = e^{-r\Delta t} Y_{j+1}^i$.

STEP 4: If $j > 1$, then set $j = j - 1$ and go to STEP 2.

STEP 5: The American option price $= \frac{1}{N} \sum_{i=1}^{N} e^{-r\Delta t} Y_1^i$.

Example 7.5 *Suppose* $S_0 = 100, r = 0.03, \sigma = 0.4, T = 7/12$ *(7 months). Simulate the American style floating strike arithmetic Asian put option and plot the optimal exercise regions for* $t = 0.2, 0.4, 0.6$ *and* 0.8. *The simulation is based on 10,000 sample paths with* $\Delta t = 1/100$.

We approximate the conditional expectation function, $f_j(S, A)$, by a two-variable quadratic polynomials, i.e.,

$$f_j(S, A) = a_{00} + a_{10} S + a_{20} S^2 + a_{11} SA + a_{01} A + a_{02} A^2.$$

The SPLUS code is as follows.

```
N <- 10000
n <- 100
dt<- 1/n
r <- 0.03
sigma <- 0.4
S0 <- 10
nu <- r-sigma^2/2
stock <- matrix(0,N,n+1)
K <- matrix(0,N,n+1)
y <- c(rep(0,N))
put <- c(rep(0,N))
stock[,1] <- S0
K[,1] <- S0
A <- matrix(0,6,n)
check1 <- proc.time()  #the first check point

#Generate asset price paths and realized averages
for (j in 1:n){
stock[,j+1] <- stock[,j]*exp( nu*dt + sigma*sqrt(dt)*rnorm(N,0,1) )
K[,j+1] <- K[,j]+stock[,j+1]
}
for (j in 1:(n+1)) { K[,j] <- K[,j]/j }
y <- pmax( (K[,n+1]-stock[,n+1]), 0 )

#Compute conditional expectation function
for (j in n:3){
# Collect in-the-money paths
a <- which( stock[,j] < K[,j] )
if ( length(a) >= 6 ) {

S <- stock[a,j]
zeta <- K[a,j]
A[,j] <- coef( lm( y[a] ~ S + S^2 + zeta + zeta^2 + S*zeta ) )
xx<- c(rep(1,N),stock[,j],stock[,j]^2,K[,j],K[,j]^2,stock[,j]*K[,j])
X <- matrix( xx, ncol=6 )
put <- X%*%A[,j]
put <- exp(-r*dt)*pmax(put,0)
# Find S that is in the exercising region
b <- which( (K[,j]-stock[,j]) > put )
y <- exp(-r*dt)*y
y[b] <- K[b,j]-stock[b,j]  # assign y as K-S when K-S > P
}
else { y <- exp(-r*dt)*y }
}
price <- exp(-2*r*dt)* mean(y)
```

```
check2 <- proc.time()  #the second check point
check2-check1  #Check the CPU time for the valuation

#########################################################
## Plot the exercising region at t=0.2,0.4,0.6,0.8 ##
## The continuation region is the shaded region    ##
#########################################################

par(mfrow=c(2,2))
for (j in c(20,40,60,80)) {
xx<-c(rep(1,N),stock[,j],stock[,j]^2,K[,j],K[,j]^2,stock[,j]*K[,j])
X <- matrix( xx, ncol=6 )
put <- exp(-r*dt) *X%*%A[,j]
b <- which((K[,j] - stock[,j]) > put)
if (j==80) {plot( K[b,j], stock[b,j], type="h", xlab="average",
   ylab="asset price", xlim=c(5,13), ylim=c(5,13), main="t = 0.8")}
if (j==60) {plot( K[b,j], stock[b,j], type="h", xlab="average",
   ylab="asset price", xlim=c(5,13), ylim=c(5,13), main="t = 0.6")}
if (j==40) {plot( K[b,j], stock[b,j], type="h", xlab="average",
   ylab="asset price", xlim=c(5,13), ylim=c(5,13), main="t = 0.4")}
if (j==20) {plot( K[b,j], stock[b,j], type="h", xlab="average",
   ylab="asset price", xlim=c(5,13), ylim=c(5,13), main="t = 0.2")}
}
```

Our simulation estimates the option price to be 9.783. This number is consistent with the one obtained by the finite difference method in Hansen and Jorgensen (2000). The CPU time is about 17 seconds for the computation. Fig. 7.2 plots the exercise boundaries at time 0.2, 0.4, 0.6 and 0.8. The boundaries are the interfaces between shaded and nonshaded regions. The shaded regions are those of the continuation regions. For $t = 0.2$, there are less points falling into the exercising region. Thus, the simulation is only able to graph the exercise boundary for underlying asset prices in the range of 7 to 11 at $t = 0.2$. □

7.6 GREEK LETTERS

As pointed out in Chapter 5, hedging is sometimes more important than pricing in risk management. Option hedging requires risk managers to compute option Greeks, like delta, gamma, vega, and theta. We refer interested readers to Hull (2006) for the application of Greeks in hedging and Joshi (2003) for discrete tree approximation. The Greek letters are actually representing partial differentiations of the option pricing formula with respect to different parameters. Since most options, especially path-dependent options, do not have closed form pricing formulas, Greeks should be obtained by means of

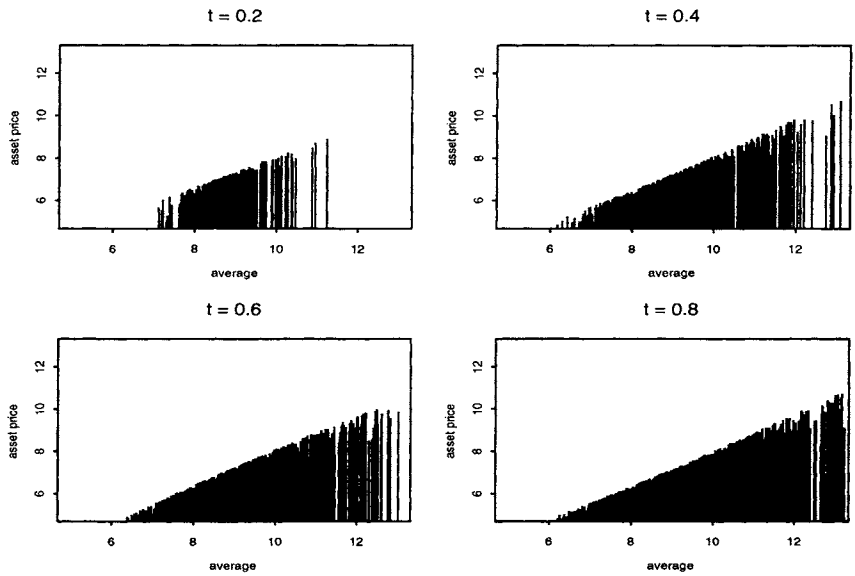


Fig. 7.2 Exercise regions of the American-style Asian option.

simulation. For single asset path independent options, the simulation can be constructed via Theorem 5.3. However, it is inapplicable for path-dependent options. Thus, we introduce an alternative practical approach to simulating Greeks.

Let V denote the pricing formula of an option. The option Greeks are defined as follows.

$$\begin{aligned} Delta &= \frac{\partial V}{\partial S}; \\ Gamma &= \frac{\partial^2 V}{\partial S^2}; \\ Vega &= \frac{\partial V}{\partial \sigma}; \\ Theta &= \frac{\partial V}{\partial t}; \\ Rho &= \frac{\partial V}{\partial r}, \end{aligned}$$

where S is the underlying asset price, σ is the volatility, t is the time variable and r is the spot interest rate. Hence, the Greeks can be obtained by standard differentiation techniques or approximated by the numerical finite difference method (FDM) if the option pricing formula is available. The FDM

computes numerical differentiation by approximating the first principle in differentiation. For instance, suppose we are interested in the Delta of an option. Then, the FMD approximates the value by

$$Delta \simeq \frac{V(S+h) - V(S)}{h}, \tag{7.10}$$

where h is an arbitrarily chosen small number and other parameters are fixed.

The approach introduced here combines simulation with FDM together. Suppose we need the Delta of an option. Then we proceed as follows. First, the option price is simulated as usual with the current realized asset price S. Second, we resimulate the option price again with a "perturbed" asset price $S+h$. Finally, the Delta is approximated by (7.10). However, the stability of this approach would be of great concerns since there are two sources of errors, simulation error and FDM error. The most critical one is the simulation error, which makes the numerator of (7.10) nonzero even when h tends to zero. To circumvent this difficulty, it is very common for market practitioners to use the same set of random numbers in the first and the second steps. We illustrate these ideas with the down-and-out call option in the following example.

Example 7.6 *Suppose* $S(0) = 100, r = 0.05, \sigma = 0.4, T = 1$ *(1 year). Estimate the delta of down-and-out call option with a strike price of 95 and provision on a downside barrier of 80.*

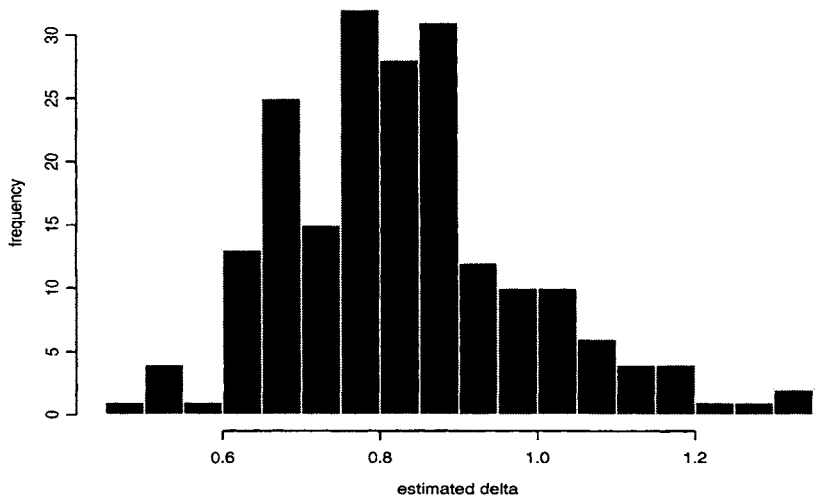


Fig. 7.3 The strike against the delta of a down-and-out call option.

We base our simulation on 100,000 sample paths, each of which is divided into 100 equally-spaced intervals. Therefore, this simulation requires 10 million independent normal random variables, namely ϵ_{ij} with $i = 1, 2, \ldots, 100$ and $j = 1, 2, \ldots, 100,000$. Using the set of $\{\epsilon_{ij}\}$, we produce the sample paths as $\{S_j(t_i), \ldots, S_j(t_{100})\}$ using the Black-Scholes dynamics of asset price with $S_j(0) = 100$ for all j. Therefore, we get the C_{do} price as in Section 7.2. To obtain delta, we repeat the above procedure by assuming $S_j(0) = 100 + h$, where $h = 0.01$, to estimate the option price again. It is important to recall that we must use the same set of ϵ_{ij}. After that, the delta is approximated by the finite difference method. Our simulation estimates the delta of the down-and-out call option to be 0.863. Fig. 7.3 shows the distribution of the delta estimates over 100 simulations. The corresponding programming code is given as follows. □

```
### compute the delta for European Down-and-Out Call Option ###
### parameters ###
N <- 100000
S0<- 100
L <- 80
r <- 0.05
sigma <- 0.4
nu <- (r-sigma^2/2)
t <- 1
n <- 100
dt<- t/n
h <- 0.05
K <- 93:110
returnS <- matrix(0,N,n)
deltaC  <- matrix(0,4,18)

### estimation of delta ####
for (i in 1:4){
returnS <-matrix(rnorm(N*n,nu*dt,sigma*sqrt(dt)),N,n)
for (k in 1:10){
S1 <- c(rep(S0,N))
S2 <- c(rep(S0+h,N))
out1 <- c(rep(1,N))
out2 <- c(rep(1,N))

##find the terminating price and whether it touches the barrier##
for (j in 1:n){
S1<-S1*exp(returnS[,j]);
S2<-S2*exp(returnS[,j]);

out1 <- ifelse (S1>L,1*out1,0);
```

```
out2 <- ifelse (S2>L,1*out2,0);
}

C1 <- mean(ifelse (out1==1, pmax(S1-K[k],0), 0))*exp(-r*t);
C2 <- mean(ifelse (out2==1, pmax(S2-K[k],0), 0))*exp(-r*t);
deltaC[i,k] <- mean(C2-C1)/h *exp(-r*t)
print(c(C1,C2,deltaC[i,k]))
}
}

### true value of delta (for K>boundary only) ###
eta2<- 2*r/(sigma^2) + 1;
a <-(log(S0/K)+(r+sigma^2/2)*t)/sigma/sqrt(t);
b <-(log(L^2/K/S0)+(r+sigma^2/2)*t)/sigma/sqrt(t);
DOCdelta <- pnorm(a)+(eta2-1)*(L/S0)^eta2*pnorm(b);
temp1<-(eta2-2)*exp(-r*t)*(L/S0)^(eta2-2);
temp2<-K/S0*pnorm(b-sigma*sqrt(t));
DOCdelta <-DOCdelta -temp1*temp2;

a<-c(0.7,1.3)
plot(K,deltaC[1,],type='l',xlab='strike',ylab='delta',ylim=a)
for (i in 2:4) { lines(K,deltaC[i,],type="l") }
lines(K,DOCdelta,type="o")
```

Other Greek letters can be obtained in a similar manner. For instance, the Gamma is the second-order partial differentiation of the option pricing formula with respect to the underlying asset price. To estimate its value, we can approximate the second order differentiation by central finite differencing such that

$$Gamma \simeq \frac{V(S+h) - 2V(S) + V(S-h)}{h^2}.$$

Therefore, we are required to compute $V(S-h)$ on top of $V(S)$ and $V(S+h)$.

Example 7.7 *Using the input parameters in Example 7.6, plot the gamma of down-and-out call option against strike price, where the strike price varies from 93 to 110.*

```
####compute the gamma for European Down-and-Out Call Option
########################################
### parameters as the same as before ###
########################################
returnS <- matrix(0,N,n)
deltaC  <- matrix(0,4,18)

for (i in 1:5) {
returnS <- matrix(rnorm(N*n,nu*dt,sigma*sqrt(dt)),N,n)
```

```
for (k in 1:18){
S1 <- c(rep(S0-h,N))
S2 <- c(rep(S0  ,N))
S3 <- c(rep(S0+h,N))
out1 <- c(rep(1,N))
out2 <- c(rep(1,N))
out3 <- c(rep(1,N))

#find the terminating price and whether it touches the barrier#
for (j in 1:n){
S1<-S1*exp(returnS[,j]);
S2<-S2*exp(returnS[,j]);
S3<-S3*exp(returnS[,j]);
out1 <- ifelse(S1>L,1*out1,0);
out2 <- ifelse(S2>L,1*out2,0);
out3 <- ifelse(S3>L,1*out3,0);
}

C1 <- mean(ifelse (out1==1, pmax(S1-K[k],0), 0))*exp(-r*t);
C2 <- mean(ifelse (out2==1, pmax(S2-K[k],0), 0))*exp(-r*t);
C3 <- mean(ifelse (out3==1, pmax(S3-K[k],0), 0))*exp(-r*t);
gammaC[i,k] <- mean(C1-C2*2+C3)*exp(-r*t)/h^2;
print(c(C1,C2,C3,gammaC[i,k]))
}
}

a<-c(min(gammaC),max(gammaC))
plot(K,gammaC[1,],type='l',xlab='strike',ylab='gamma',ylim=a)
for (i in 2:4) { lines(K,gammaC[i,],type="l") }
```

7.7 EXERCISES

1. Verify equations (7.1) and (7.3).

2. By modifying Example 7.1, simulate the price of down-and-out call, which will be knocked out if the underlying asset price goes below \$8.

3. By modifying Example 7.1, simulate prices of a fixed lookback put option and a floating lookback call if the fixed strike price and the realized minimum asset prices are both \$8. Verify the lookback put-call parities of these options.

4. Show that American call option price equals that of its European counterpart if the underlying asset pays no dividends. In other words, the

American call option is never optimal to exercise prior to maturity if the underlying asset pays no dividends.

5. By modifying Example 7.4, simulate the price of an American call option with a strike of \$12 and a dividend yield δ of 4%. Hint: The risk-neutral dynamics of an asset paying continuous dividend yield is given by

$$\frac{dS}{S} = (r - \delta)\, dt + \sigma\, dW.$$

What is the optimal exercising policy from your simulation? Plot the critical asset prices against time.

6. Forward start option is a path-dependent option that the strike price will be set as the underlying asset price in the future. For instance, the forward start call option payoff is

$$\max(S_T - S_{t_1}, 0),$$

where $0 < t_1 < T$.

(a) Suppose $S_0 = \$10$, $\sigma = 0.4$, $r = 0.1$, $\delta = 0.5$, $T = 0.5$ and $t_1 = 0.3$. Construct and implement an algorithm for the forward start call option with 1,000 sample paths.

(b) Denote $C_{BS}(S, t; K, T)$ by the Black-Scholes formula for the standard call option. Based on financial insights, a risk analyst speculates that the forward start call option is the discounted standard call price. That is

$$\text{Current forward start call price} = e^{-rt_1} C_{BS}(S_0, t_1; S_0, T).$$

Verify this conjecture by your simulation.

(c) Suppose that the option has a continuous early exercise right after $t = t_1$. Determine the option price by the least squares simulation with 10,000 sample paths.

Simulation Techniques in Financial Risk Management
by Ngai Hang Chan and Hoi Ying Wong

8

Multi-asset Options

8.1 INTRODUCTION

Multi-asset options are exotic options whose payoffs depend on values of multiple assets. Multi-asset options abound in the financial market. An obvious example is index options, where the underlying variable, the financial index, can be thought as a portfolio of multiple assets. Challenges of valuing multi-asset options are the curse of dimensionality and the lack of analytical tractability. Those problems can be circumvented by simulations.

Some examples of multi-asset options traded in the financial market are first introduced. Let $S_1, S_2, \ldots, S_n$ denote the prices of n different assets.

1. Exchange options: the right to exchange an asset for another. Thus, the option payoff is $\max(S_1 - cS_2, 0)$, where c is a constant multiplicative factor. This option is useful, for example, when a US investor wants to buy Japanese yen with eurodollars.

2. Quanto options: options on stocks in a foreign country, i.e., involving the exchange rate. If we treat S_1 as the exchange rate and S_2 as the underlying asset in the foreign country, then there are a number of possible quanto option payoffs, like $S_1 \max(S_2 - K, 0)$, $\max(S_1 S_2 - K, 0)$, $\max(S_1, C)$, $\max(S_2 - K, 0)$, and $C \max(S_2 - K, 0)$, where C is a fixed constant. The last payoff function appears to be of a single asset option. However, the volatility of the exchange rate, S_1, does contribute to the option price if S_1 and S_2 are correlated.

3. Basket options: options S on a portfolio. The payoff of a call on a portfolio is $\max(\Pi - K, 0)$, where $\Pi = \sum_{i=1}^{n} a_i S_i$.

4. Extreme options: options on the extrema of different assets. The maximum call option has the payoff: $\max[\max(S_1, S_2, \ldots, S_n) - K, 0]$.

All multi-asset options can be traded with European or American style. Complex multi-asset options, or structured products, may even involve path dependent features. In such cases, simulations are indispensable.

8.2 SIMULATING EUROPEAN MULTI-ASSET OPTIONS

Consider an option on two assets with payoff $F(S_1(T), S_2(T))$. In the risk-neutral world, assets are assumed to follow the dynamics of

$$\frac{dS_i}{S_i} = r\,dt + \sigma_i\,dW_i,\ i = 1, 2, \tag{8.1}$$

where

$$\widehat{\mathrm{E}}(dW_1 dW_2) = \rho\,dt, \tag{8.2}$$

and $\widehat{\mathrm{E}}$ denotes the risk-neutral expectation. Then, the option can be simulated via the Cholesky decomposition (see Theorem 4.4).

Example 8.1 *Suppose* $S_1(0) = S_2(0) = 10$, $\sigma_1 = 0.3$, $\sigma_2 = 0.4$, $\rho = 0.2$, *and* $r = 0.05$. *Simulate the price of an exchange option with maturity of six months.*

By Itô's lemma, we derive the terminal asset prices as

$$S_1(T) = S_1(0)e^{(r-\sigma_1^2/2)T+\sigma_1 X_1\sqrt{T}} \text{ and } S_2(T) = S_2(0)e^{(r-\sigma_2^2/2)T+\sigma_2 X_2\sqrt{T}}, \tag{8.3}$$

where

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim \mathrm{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right).$$

The option price, C_X, can be determined by evaluating the expectation:

$$C_X = e^{-rT}\widehat{\mathrm{E}}\left[\max(S_1(T) - S_2(T), 0)\right].$$

We estimate the option price by the following simulation algorithm.

STEP 1: For $i = 1$ to N, do Step 2 to Step 4 as follows:

STEP 2: Generate Z_1, Z_2 N(0,1) i.i.d.

STEP 3: Set $X_1 = Z_1$ and $X_2 = \rho Z_1 + \sqrt{1-\rho^2}Z_2$.

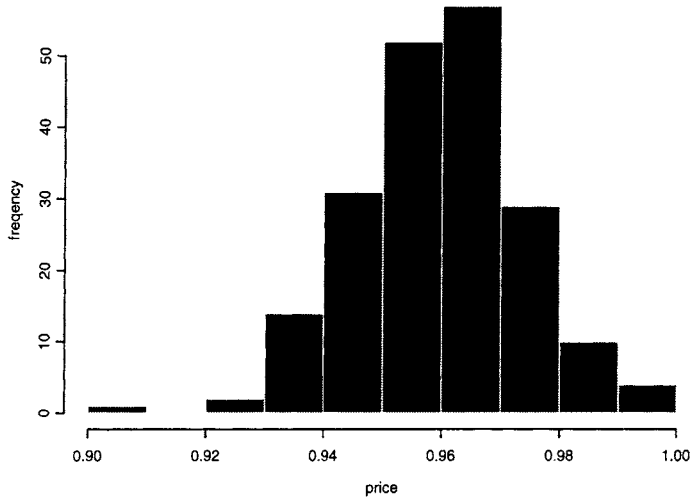


Fig. 8.1 The distribution of simulated price.

STEP 4: Compute $S_1^{(i)}(T)$ and $S_2^{(i)}(T)$ by (8.3).

STEP 5: Set $C_X = \frac{e^{-rT}}{N} \sum_{i=1}^{N} \max(S_1^{(i)}(T) - S_2^{(i)}(T), 0)$.

Fig. 8.1 plots the distribution of the estimated price over 100 simulations. We obtain the estimated option price to be 0.962. This algorithm is implemented with SPLUS and the programming code is given as follows: □

```
S0 <-c(10,10)
r <- 0.5
sigma1<-0.3
sigma2<-0.4
rho <-0.2
T <-0.5
N<-10000
Z1<-rnorm(N)
Z2<-rnorm(N)
X1<-Z1
X2<-Z1*rho+Z2*sqrt(1-rho^2)
S1<-S0[1]*exp((r-sigma1^2/2)*T+sigma1*sqrt(T)*X1)
S2<-S0[2]*exp((r-sigma2^2/2)*T+sigma2*sqrt(T)*X2)
C<-mean(pmax(S1-S2,0))*exp(-r*T)
C
```

8.3 CASE STUDY: ON ESTIMATING BASKET OPTIONS

In practice, basket options are often valued by assuming that the value of the portfolio of assets comprising the basket follows the Black-Scholes dynamics jointly rather than each asset follows the Black-Scholes dynamics individually. After estimating the portfolio volatility from the portfolio return, the basket call option is valued by substituting the portfolio volatility into the Black-Scholes formula. This approach offers a quick solution to traders. However, the risk manager needs to understand the risk of this simplifying assumption. We examine this approach by means of simulation.

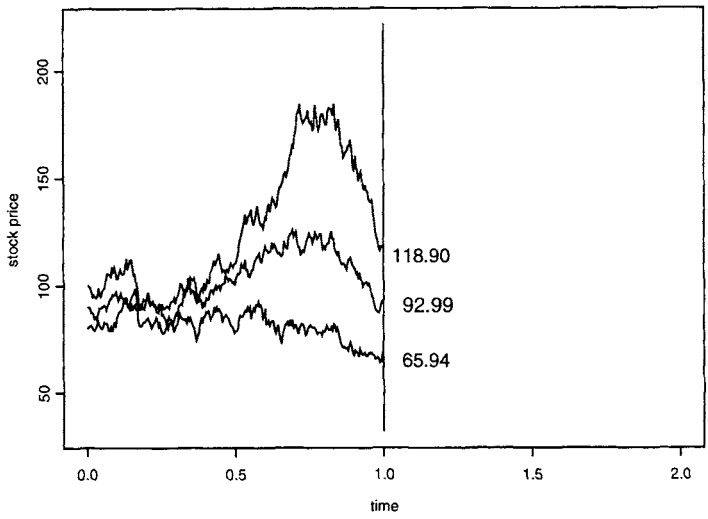


Fig. 8.2 The historical price of shocks.

Consider a basket call option with three underlying assets, S_1, S_2, and S_3. The payoff of this option is $\max(S_1 + S_2 + S_3 - K, 0)$. In other words, the holder of the option has the right to purchase the portfolio as a sum of the three assets for a fixed value of K. Suppose the current time is $t = 1$ and we observe the prices of three assets since $t = 0$. Fig. 8.2 depicts the paths of the three simulated asset prices.

```
#### simulate the historical stock price ####
n<-250;   r<-0.05;  dt<-1/n;
sigma <- c(0.4,0.3,0.35);
SIGMA <- matrix(c(1,0.3,0.3,0.3,1,0.3,0.3,0.3,1),3,3);
SIGMA <- SIGMA * ( sigma %*% t(sigma) )
```

```
S <- matrix(0,3,n+1);
S[,1] <- c(100,90,80);
for (i in 1:n) {
S[,i+1]<-S[,i]*exp(r*dt+rnorm(3)%*%chol(SIGMA)*sqrt(dt));
}
P <- S[1,]+S[2,]+S[3,];
print(P)
returnS <- diff(t(log(S)));
returnP <- diff(log(P));

cor(returnS)
A <- var(returnS)/dt
B <- var(returnP)/dt
C <- S[,n+1]

time<-(0:n)/n;
range<-c(min(S)*0.5,max(S)*1.2);
plot(time,S[1,],type='l',xlim=c(0,1.6), ylim=range, ylab='stock price')
lines(time,S[2,],type='l')
lines(time,S[3,],type='l')
lines(c(1,1),range,type='l')
#### cont'd  #####
```

At $t = 1$, the asset prices are $S_1 = 142.69$, $S_2 = 89.23$, and $S_3 = 49.73$. The current portfolio value is the sum of and equals 281.65. Based on the three asset price paths, the portfolio volatility is estimated to be 0.280. Consider the basket option with a strike price of 250, maturity of half a year and interest rate of 5%. The naive application of the Black-Scholes formula produces a value for the option as 44.81.

On the other hand, we can use MC simulation to estimate the option price by assuming individual asset follows the Black-Scholes dynamics. By examining the asset price paths, we estimate the variance-covariance matrix for assets returns as

$$\begin{pmatrix} 0.172 & 0.050 & 0.043 \\ 0.050 & 0.088 & 0.038 \\ 0.043 & 0.038 & 0.123 \end{pmatrix}.$$

Then, we simulate asset prices at $t = 1.5$ using the Cholesky decomposition for 10,000 times. Fig. 8.3 illustrates the idea of generating asset values at $t = 1.5$, the maturity of the option. Terminal values of individual assets are simulated based on an approach similar to (8.3) with three assets. The option price is then evaluated by discounting the sample mean of the option payoff using the interest rate of 5%. The simulated option price is 51.35, which is larger than the naive approach of 44.81.

We are also interested in the contribution of the error in estimating parameters of the option. We perform a control experiment assuming that the variance-covariance matrix can be estimated without error. Input the variance-covariance matrix as

$$\begin{pmatrix} 0.1600 & 0.0360 & 0.0420 \\ 0.0360 & 0.0900 & 0.0315 \\ 0.0420 & 0.0315 & 0.1225 \end{pmatrix}.$$

Using the same set of independent normal random numbers, we obtain the option price as 50.71. It seems that the error of estimating the variance-covariance matrix does not contribute too much to basket option values. Therefore, the MC and the naive approach to valuing basket options can produce significantly different results, irrespective to the estimation error.

In practice, banks and financial institutions usually have a lot of derivatives positions in their portfolio. Risk managers are responsible to check for the consistency of models that are used to value individual derivatives in the portfolio. Imagine a situation that a bank buys and sells options on individual assets and a basket of assets everyday. When individual assets are assumed to follow the Black-Scholes dynamics, it is crucial for the risk manager to realize what kind of assumptions have been imposed. The simulation shows that it is not appropriate to assume the portfolio constituting the basket to follow Black-Scholes dynamics jointly because this assumption is not consistent with the assumption on individual assets. In such a case, the value of basket options can be significantly underestimated.

```
#### cont'd  #####
range<-c(min(S)*0.5,max(S)*1.6);
plot(time,S[1,],type='l',xlim=c(0,1.6), ylim=range, ylab='stock price')
lines(time,S[2,],type='l')
lines(time,S[3,],type='l')
lines(c(1,1),range,type='l')
print(C)

#### simulate scenarios ###
K<-250; t<-0.5;
d1 <- (log(sum(C)/K)+(r+B^2/2)*t)/sqrt(B*t);
d2 <- d1-sqrt(B*t);
sum(C)*pnorm(d1)-K*exp(-r*t)*pnorm(d2);

N <- 20000;
S <- matrix(0,N,3);
S[,1]<-C[1]; S[,2]<-C[2]; S[,3]<-C[3];

for (i in 1:125){
S<-S*exp(r*dt+matrix(rnorm(3*N),N,3)%*%chol(A)*sqrt(dt));
```

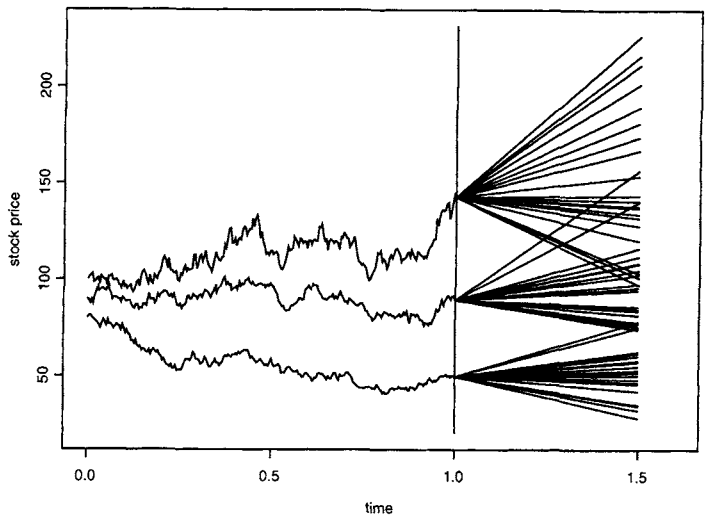


Fig. 8.3 Simulating terminal asset prices.

```
}

P<-S[,1]+S[,2]+S[,3];
price<-exp(-r*T)*pmax(P-K,0);
mean(price)

for (i in 1:20){
lines(c(1,1.5),c(C[1],S[i,1]),type='l')
lines(c(1,1.5),c(C[2],S[i,2]),type='l')
lines(c(1,1.5),c(C[3],S[i,3]),type='l')
}
```

8.4 DIMENSIONAL REDUCTION

For an n-asset option, simulation can be constructed by using the Cholesky decomposition (4.6). However, this requires generating n independent normal random variables for each scenario. To reduce the computational burden, we can use the principle component analysis (PCA) to approximate the n factors by a smaller number of factors, usually less than 10 in practice.

Suppose we have an n dimensional random vector $\boldsymbol{X} \sim \mathrm{N}(0, \Sigma)$ where Σ is an $n \times n$ variance-covariance matrix. PCA for normal random variables is to

approximate $\boldsymbol{X}$ by $\boldsymbol{Y}$ which follows a distribution similar to that of $\boldsymbol{X}$ but is easier to simulate.

PCA uses the eigenvalue decomposition in Chapter 4 to approximate the random vector $\boldsymbol{X}$. Let $\lambda_1, \lambda_2, \ldots, \lambda_n$ be eigenvalues of Σ and $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n$ be the corresponding eigenvectors. As variance-covariance matrices are positive definite, their eigenvalues are all positive real numbers so that the corresponding squared roots are positive real numbers. Theorem 4.6 asserts that the random vector

$$\boldsymbol{X} = \sqrt{\lambda_1}\boldsymbol{v}_1 Z_1 + \sqrt{\lambda_2}\boldsymbol{v}_2 Z_2 + \cdots + \sqrt{\lambda_n}\boldsymbol{v}_n Z_n, \tag{8.4}$$

where $Z_1, Z_2, \ldots, Z_n$ are i.i.d. standard normal random variables. The equality (8.4) is defined in the sense of distribution. In PCA, we arrange eigenvalues in descending order such that $\lambda_1 > \lambda_2 > \cdots > \lambda_n$. From (8.4), we see that the contribution of the term $\sqrt{\lambda_i}\boldsymbol{v}_i Z_i$ to the value of $\boldsymbol{X}$ decreases with the index i. The eigenvector $\boldsymbol{v}_i$ is called the i-th principle component (PC). To approximate $\boldsymbol{X}$, we truncate the sum in (8.4) such that

$$\boldsymbol{X} \simeq \sqrt{\lambda_1}\boldsymbol{v}_1 Z_1 + \sqrt{\lambda_2}\boldsymbol{v}_2 Z_2 + \cdots + \sqrt{\lambda_m}\boldsymbol{v}_m Z_m,$$

where $m < n$. If we are comfortable with this approximation, we then simulate m independent standard normal random variables Z_i and calculate everything based on this the approximation.

An important topic in PCA is to determine the value m. The number of terms used in the approximation depends on the accuracy of the outcome required by the modeler. If the user requires 100% accuracy besides simulation error, then he should use formula of (8.4). PCA is useful when the user requires an accuracy that is less than 100%. Suppose he requires an accuracy of at least 99%. Then, m is the minimum integer such that

$$\frac{\sum_{i=1}^{m} \lambda_i}{\sum_{i=1}^{n} \lambda_i} \geq 99\%.$$

A proof of this result can be found in standard texts in multivariate analysis, for example Anderson (2003).

Let us apply PCA in multi-asset option pricing. Consider an option with 10 underlying assets. Each asset follows the Black-Scholes dynamics such that

$$dS_i = \mu_i S_i\, dt + \sigma_i S_i\, dW_i, \quad i = 1, 2, \ldots 10,$$

where S_i is the value of the i-th asset. The $W_1, W_2, \ldots, W_{10}$ are correlated Brownian motions with correlation matrix:

$$\begin{pmatrix}
1.00 & 0.74 & 0.34 & -0.08 & 0.05 & -0.74 & 0.04 & -0.12 & 0.81 & 0.82 \\
0.74 & 1.00 & 0.81 & -0.04 & -0.57 & -0.25 & 0.06 & 0.47 & 0.89 & 0.92 \\
0.34 & 0.81 & 1.00 & -0.17 & -0.83 & 0.20 & -0.09 & 0.78 & 0.65 & 0.72 \\
-0.08 & -0.04 & -0.17 & 1.00 & 0.01 & -0.05 & 0.94 & -0.04 & -0.09 & -0.05 \\
0.05 & -0.57 & -0.83 & 0.01 & 1.00 & -0.55 & 0.00 & -0.94 & -0.41 & -0.45 \\
-0.74 & -0.25 & 0.20 & -0.05 & -0.55 & 1.00 & -0.16 & 0.65 & -0.40 & -0.40 \\
0.04 & 0.06 & -0.09 & 0.94 & 0.00 & -0.16 & 1.00 & -0.06 & 0.04 & 0.06 \\
-0.12 & 0.47 & 0.78 & -0.04 & -0.94 & 0.65 & -0.06 & 1.00 & 0.31 & 0.34 \\
0.81 & 0.89 & 0.65 & -0.09 & -0.41 & -0.40 & 0.04 & 0.31 & 1.00 & 0.91 \\
0.82 & 0.92 & 0.72 & -0.05 & -0.45 & -0.40 & 0.06 & 0.34 & 0.91 & 1.00
\end{pmatrix}$$

A discrete approximation to the asset price dynamics is

$$\Delta S_i = rS_i\,\Delta t + S_i\sqrt{\Delta t}\,\epsilon_i,$$

where ϵ_i are risk factors such that $[\epsilon_i] = \boldsymbol{X} \sim \mathrm{N}(0, \Sigma)$, and Σ is the correlation matrix given above.

For the given correlation matrix, eigenvalues are obtained as 4.719, 2.843, 1.931, 0.147, 0.104, 0.079, 0.062, 0.056, 0.038, 0.022. Summing up all the eigenvalues gives a value of 10. When we divide the sum of the first three eigenvalues by the total sum, the ratio is close to 95%. Therefore, if we accept an error of 5%, the first three PCs provide sufficient accuracy. Eigenvectors corresponding to the first three PCs are found to be:

$\boldsymbol{v}_1$:	0.31	0.44	0.41	−0.05	−0.31	−0.07	−0.00	0.27	0.42	0.43
$\boldsymbol{v}_2$:	0.41	0.08	−0.22	0.09	0.41	−0.57	0.14	−0.45	0.18	0.16
$\boldsymbol{v}_3$:	0.09	−0.03	0.01	−0.70	0.12	−0.05	−0.69	−0.10	0.02	−0.00

Based on the first three PCs, we generate three independent standard normal random variables, namely Z_1, Z_2, and Z_3, and approximate the n risk factors by

$$[\epsilon_i] \simeq Z_1\sqrt{\lambda_1}\,\boldsymbol{v}_1 + Z_2\sqrt{\lambda_2}\,\boldsymbol{v}_2 + Z_3\sqrt{\lambda_3}\,\boldsymbol{v}_3.$$

The 10 risk factors $\epsilon_1, \epsilon_2, \ldots, \epsilon_{10}$ are reduced to three independent factors only. Hence, we reduce a 10-dimensional problem to a 3-dimensional problem.

Example 8.2 *Value a maximum option on 10 assets with a strike price of \$95 and a maturity of half a year. All asset values are currently \$100 with volatilities of 30% for all assets. The correlation matrix of risk factors is given above. The interest rate is 4%. We accept a maximum error of 5%.*

The option payoff is $\max[\max(S_1, S_2, \ldots, S_{10}) - 95, 0]$. As the option is traded in European style, it is efficient to simulate terminal asset values directly. By Itô's lemma, we know that the terminal value of the i-th asset is given by

$$S_i(T) \quad = \quad S_i(0)\exp\left[(r - \sigma_i^2/2)T + \sigma_i W_i(T)\right]$$

$$= S_i(0) \exp\left[(r - \sigma_i^2/2)T + \sigma_i \sqrt{T}\,\epsilon_i\right], \tag{8.5}$$

where the vector $[\epsilon_i] \sim \mathrm{N}(0, \Sigma)$. Our simulation obtains the option price to be 49.12. The corresponding programming code is given as follows. □

```
#### input the correlation matrix ######
A<-matrix(0,10,10)
A[1,]<-c( 1.00, 0.74, 0.34,-0.08, 0.05,-0.74, 0.04,-0.12, 0.81, 0.82);
A[2,]<-c( 0.74, 1.00, 0.81,-0.04,-0.57,-0.25, 0.06, 0.47, 0.89, 0.92);
A[3,]<-c( 0.34, 0.81, 1.00,-0.17,-0.83, 0.20,-0.09, 0.78, 0.65, 0.72);
A[4,]<-c(-0.08,-0.04,-0.17, 1.00, 0.01,-0.05, 0.94,-0.04,-0.09,-0.05);
A[5,]<-c( 0.05,-0.57,-0.83, 0.01, 1.00,-0.55, 0.00,-0.94,-0.41,-0.45);
A[6,]<-c(-0.74,-0.25, 0.20,-0.05,-0.55, 1.00,-0.16, 0.65,-0.40,-0.40);
A[7,]<-c( 0.04, 0.06,-0.09, 0.94, 0.00,-0.16, 1.00,-0.06, 0.04, 0.06);
A[8,]<-c(-0.12, 0.47, 0.78,-0.04,-0.94, 0.65,-0.06, 1.00, 0.31, 0.34);
A[9,]<-c( 0.81, 0.89, 0.65,-0.09,-0.41,-0.40, 0.04, 0.31, 1.00, 0.91);
A[10,]<-c( 0.82, 0.92, 0.72,-0.05,-0.45,-0.40, 0.06, 0.34, 0.91, 1.00);

sigma <-0.3
A <- A*sigma^2
eigenvalues <- eigen(A)$values
eigenvectors<- eigen(A)$vectors
print(eigenvalues)   ##check that the first 3 values are the largest

### input parameters #####
S0<- 100;
K <- 95;
r <- 0.04;
n <- 5000;
t <- 1;
B <- matrix(0,n,3);
S <- matrix(S0,10,n);
price <- c(rep(0,n));

### generate scenarios #####
for (j in 1:3){ B[,j] <- rnorm(n,0,1)*sqrt(eigenvalues[j]) }
returnS <- eigenvectors[,1:3]%*%t(B);
S <- S*exp((r-sigma^2/2)*t +returnS*sqrt(t));

for (i in 1:n) {price[i] <- max(max(S[,i])-K,0)}
value<-mean(price)*exp(-r*t)
value
```

8.5 EXERCISES

1. Suppose $x(t)$ and $y(t)$ are two correlated Itô's processes such that

$$\begin{aligned} dx &= a(t,x)\,dt + b(t,x)\,dW_1 \\ dy &= \alpha(t,y)\,dt + \beta(t,y)\,dW_2 \\ \mathrm{E}(dW_1 dW_2) &= \rho\,dt. \end{aligned}$$

Consider a function, $f(t,x,y)$, which depends on both stochastic variables of $x(t)$ and $y(t)$. By modifying the proof of Theorem 2.1, show that the dynamic of $f(t,x,y)$ is

$$\begin{aligned} df &= \left(\frac{\partial f}{\partial t} + a\frac{\partial f}{\partial x} + \alpha\frac{\partial f}{\partial y} + \frac{b^2}{2}\frac{\partial^2 f}{\partial x^2} + \frac{\beta^2}{2}\frac{\partial^2 f}{\partial y^2} + \rho b\beta\frac{\partial^2 f}{\partial x \partial y}\right) dt \\ &\quad + b\frac{\partial f}{\partial x}\,dW_1 + \beta\frac{\partial f}{\partial y}\,dW_2. \end{aligned} \tag{8.6}$$

This formula is known as the Itô's lemma for two variables.

2. Answer the following questions by considering the property of martingales defined in Question 7 of Chapter 3.

   (a) Consider a pair of asset price dynamics under the risk-neutral measure:

$$\begin{aligned} dS_1 &= rS_1\,dt + \sigma_1 S_1\,dW_1 \\ dS_2 &= rS_2\,dt + \sigma_2 S_2\,dW_2 \\ \mathrm{E}(dW_1 dW_2) &= \rho\,dt. \end{aligned}$$

   Show that the stochastic process $X(t) = S_1(t)/S_2(t)$ is a martingale under the Brownian motions $W_1^*(t)$ and $W_2^*(t)$ where

$$W_1^*(t) = W_1(t) - \rho\sigma_2 t \quad \text{and} \quad W_2^*(t) = W_2(t) - \sigma_2 t.$$

   (b) Under (a), show that $X(t)$ has the dynamics:

$$\frac{dX}{X} = \sigma\,dW^*,$$

   where W^* is a Brownian motion and

$$\sigma^2 = \sigma_1^2 - 2\rho\sigma_1\sigma_2 + \sigma_2^2.$$

   (c) Consider a function of S_1 and S_2, $V(t,S_1,S_2)$, which has the property that

$$V(t,S_1,S_2) = S_2\,U(t,S_1/S_2) = S_2\,U(t,X).$$

   Show that $U(t,X)$ is a martingale under Brownian motions $W_1^*(t)$ and $W_2^*(t)$.

3. Consider the exchange option with payoff $\max(S_1(T) - S_2(T), 0)$. Denote the option pricing formula for this option as $V_{ex}(t, S_1, S_2)$. By using the no-arbitrage argument, one derives that the exchange option has the properties:

   - There exists a function U such that $V_{ex}(t, S_1, S_2) = S_2\, U(t, S_1/S_2)$.
   - There exists a probability measure $\mathcal{Q}$ such that $X(t) = S_1(t)/S_2(t)$ is a martingale.

   Based on these properties and the results obtained in Question 2, show that

   $$V_{ex} = S_1\Phi(d_1^*) - S_2\Phi(d_2^*),$$

   where

   $$\begin{aligned} d_1^* &= \frac{\ln(S_1/S_2) + \sigma^2(T-t)/2}{\sigma\sqrt{T-t}} \\ d_2^* &= d_1^* - \sigma\sqrt{T-t} \\ \sigma^2 &= \sigma_1^2 - \rho\sigma_1\sigma_2 + \sigma_2^2. \end{aligned}$$

   Herein, σ_1 and σ_2 are volatilities of S_1 and S_2, respectively, and ρ is the correlation coefficient between the returns of two assets.

4. Run the simulation program for pricing exchange option and compare the numerical result with the analytical one.

5. The so-called geometric basket option has the payoff function

   $$\max\left\{\left(\prod_{i=1}^{n} S_i(T)\right)^{1/n} - K, 0\right\}.$$

   (a) Show that this option has a value less than the usual basket option with payoff

   $$\max\left\{\frac{1}{n}\sum_{i=1}^{n} S_i(T) - K, 0\right\}.$$

   (b) Suppose individual assets follow the Black-Scholes dynamics. Derive the analytical pricing formula for the geometric basket option.

   (c) By regarding the price of the geometric basket option as a control variate, simulate the price of the usual basket option that depends on four assets with the following correlation matrix:

   $$\begin{pmatrix} 1.0000 & 0 & 0.3000 & 0.3000 \\ 0 & 1.0000 & 0.4000 & 0.2000 \\ 0.3000 & 0.4000 & 1.0000 & 0.3000 \\ 0.3000 & 0.2000 & 0.3000 & 1.0000 \end{pmatrix}.$$

We assume all assets sharing the same volatility of 30% and each asset individually follows the Black-Scholes dynamics.

6. Use simulation to determine the value and early exercise policy of American style exchange options. We assume the interest rate of 5%, $S_1 = 100$, $S_2 = 95$, $T = 1$ year and the variance-covariance matrix of asset returns:

$$\begin{pmatrix} 0.016 & 0.006 \\ 0.006 & 0.09 \end{pmatrix}.$$

Hint: you may use the LS model and a quadratic polynomial of S_1 and S_2 in your regression.

Simulation Techniques in Financial Risk Management
by Ngai Hang Chan and Hoi Ying Wong

9

Interest Rate Models

9.1 INTRODUCTION

Fixed income securities are concerned with the valuation of promised payments at a future date. For example, a zero coupon bond promises to pay a single payment on the maturity day. A straight US Treasury bond promises to make payments with the amount and date of the payments determined by the face value, maturity date, and coupon rate of the bond. Because cash flows are certain, we are not concerned with the risk of the volatility of the amount of cash. Instead we are interested in the following questions: *How much would a rational individual be willing to pay today for a promised payment in the future?* The answer to this question is related to the interest rate movement. This leads to the next question. *How to manage the interest rate risk?* Simulation can serve as a useful tool to gain some insights in answering these questions.

9.2 DISCOUNT FACTOR

Consider the simplest case: a zero coupon bond (zeros) paying \$100 a year from now. What is the maximum value one is willing to pay for this contract today? Purchasing this bond should be worth at least as much as putting the money into the bank. Let P be the payment at the current moment. Then,

$$P(1 + R) = 100,$$

where R is current annual interest paid by a bank (R is supposed to be a constant). That is

$$P = \frac{100}{1+R}.$$

P and $\frac{1}{1+R}$ are known as the zeros price and discount factor, respectively. One may regard the discount factor as the zeros price with unity payment on the maturity T. We usually write

$$P(0,T) = P \quad \text{and} \quad B(0,T) = \text{discount factor from } T \text{ to } 0.$$

For a zero coupon bond, we have

$$P(0,T) = P(T,T)B(0,T). \tag{9.1}$$

Suppose that the interest rate is paid in continuously compounding. Then, the discount factor becomes

$$B(t,T) = e^{-r(T-t)}. \tag{9.2}$$

Notice that the continuously compounding interest rate r and the annual interest rate R are connected by the formula

$$\frac{1}{1+R} = e^{-r}.$$

What about multiple dates at which payments will be made? A typical bond will pay coupons at semi-annual intervals and a principle payment at maturity. The key to evaluating such kinds of bonds is to view dollars promised at different future dates as separated zero coupon bonds. We then value each payment at each date using the discount factor for that date followed by summing up the values. For a series of cash flow $C(t_i)$ at time $t_1, t_2, ..., t_N$, the coupon bearing bond can be valued by the formula:

$$P(0,T) = \sum_{i=1}^{N} C(t_i)B(0,t_i). \tag{9.3}$$

For example, the value of a bond paying semi-annual coupon is given by

$$P(0,T) = \frac{C(1/2)}{1+R/2} + \frac{C(1)}{(1+R/2)^2} + \cdots + \frac{C(N)}{(1+R/2)^{2N}}.$$

9.2.1 Time-Varying Interest Rate

The previous examples value bonds when the interest rate is assumed to be a constant. A more realistic model considers the time-varying deterministic interest rate. For instance, we assume that the continuously compounding interest rate r is a function of time t, i.e. $r = r(t)$. We are concerned with the

evolution of discount factor $B(t,T)$, which can be viewed as the zeros price at t with unity payment at T. The rate of return of this zero coupon bond must be the instantaneous interest rate $r(t)$. Thus, we have

$$\frac{dB(t,T)}{B(t,T)} = r(t)\,dt \quad \text{and} \quad B(T,T) = 1.$$

Solving the above ordinary differential equation gives

$$B(t,T) = e^{-\int_t^T r(s)\,ds}. \tag{9.4}$$

If we use $y(t,T) = \frac{1}{T-t}\int_t^T r(s)\,ds$ to denote the mean of the interest rate in the interval $[t,T]$, then (9.4) can be written as

$$B(t,T) = e^{-y(t,T)(T-t)},$$

which agrees with (9.2). We can use (9.1) and (9.3) to evaluate zero prices and coupon-bearing bonds, respectively. As $y(t,T)$ is also defined as the yield of a bond, the discrete yield to maturity Y_t is given by

$$Y_t = 1 - e^{-y(t,t+1)}.$$

Y_t and $y(t,T)$ are the constant annual rate and constant instantaneous rate of a bond. Bond markets usually quote the yield Y_t in place of the interest rate $r(t)$. $y(t,T)$ and $r(t)$ are usually identical if and only if $r(t)$ is a constant value. Further discussions about yields and interest rate models can be found in Jarrow (2002).

9.3 STOCHASTIC INTEREST RATE MODELS AND THEIR SIMULATIONS

Deterministic interest rate models are inadequate to capture interest rate movements as no one knows future interest rates for certain. A better approach is to incorporate the stochastic feature of the interest rates. A stochastic interest rate model should match as (9.4) when the stochastic component is absent. A natural way is to consider

$$B(t,T) = \mathrm{E}\left(e^{-\int_t^T r(s,\boldsymbol{W}_s)\,ds}\right), \tag{9.5}$$

where $\boldsymbol{W}_s$ is a vector of stochastic factors. If stochastic factors are absent, the function inside the expectation becomes deterministic and the expectation equals to the function itself.

From a simulation perspective, expression (9.5) offers a means to conduct Monte Carlo simulations. Once an appropriate stochastic interest rate model,

such as the Vasicek model of Vasicek (1977), the CIR model of Cox, Ingersoll and Ross (1985), the Ho-Lee model of Ho and Lee (1986), and the Hull-White model of Hull and White (1988), is formulated, simulations can be conducted.

To illustrate the ideas, consider a short rate model that follows

$$dr = \alpha(t,r)\,dt + \beta(t,r)\,dW_t, \tag{9.6}$$

where r is the current continuous compounding interest rate and W_t is a Wiener process. For example, the Vasicek model assumes that $\alpha(t,r) = a(b-r)$ and $\beta(t,r) = \sigma$ whereas CIR model uses the same $\alpha(t,r)$ with $\beta(t,r) = \sigma\sqrt{r}$. Sample paths of short rate models of the form (9.6) can be generated by the following steps:

STEP 1: Set $r_i = r_0$ be the current market rate

STEP 2: Generate $\epsilon \sim \mathrm{N}(0,1)$

STEP 3: Set $r_{i+1} = r_i + \alpha(t_i, r_i)\,\Delta t + \beta(t_i, r_i)\,\epsilon\sqrt{\Delta t}$

STEP 4: Go to Step 2 □

Let $r^{(j)} = \{r^{(j)}(t) : t = 0, \frac{1}{n}, \frac{2}{n}, \ldots, T\}$ be the j-th interest rate path out of M sample paths generated by the preceding algorithm with $\Delta t = \frac{1}{n}$. By means of quadrature, we can make the following approximation

$$\int_{t_1}^{t_2} r^{(j)}(t)\,dt \simeq \frac{1}{n} \sum_{t \in [t_1, t_2]} r^{(j)}(t).$$

If we take $\Delta t = \frac{1}{280}$, the discount factor $B(0,1)$ can approximated by

$$\begin{aligned} B(0,1) &\simeq \mathrm{E}\left\{ \exp\left(-\frac{1}{280} \sum_{i=0}^{280} r^{(j)}\left(\frac{i}{280}\right)\right)\right\} \\ &\simeq \frac{1}{M}\sum_{j=1}^{M} \exp\left(-\frac{1}{280} \sum_{i=0}^{280} r^{(j)}\left(\frac{i}{280}\right)\right). \end{aligned}$$

In general, we write

$$\begin{aligned} B(t,T) &\simeq \frac{1}{M}\sum_{j=1}^{M} \exp\left(-\frac{1}{n} \sum_{t_i \in [t,T]} r^{(j)}(t_i)\right) \\ &= \frac{1}{M}\sum_{j=1}^{M} \exp\left(-Avg_{t_i \in [t,T]} r^{(j)}(t_i) \times (T-t)\right). \end{aligned} \tag{9.7}$$

Consider the Vasicek model proposed in Vasicek (1977),

$$dr(t) = a(b - r(t))\,dt + \sigma\,dW_t.$$

The discrete version is

$$r(t+\Delta t) = r(t)(1-a\Delta t) + ab\Delta t + \sigma(W_{t+\Delta t} - W_t).$$

A sample path can be generated by

$$\begin{aligned} r_1 &= r_0(1-a\Delta t) + ab\Delta t + \epsilon_1\sigma\sqrt{\Delta t}, \\ r_2 &= r_1(1-a\Delta t) + ab\Delta t + \epsilon_2\sigma\sqrt{\Delta t}, \\ &\vdots \\ r_n &= r_{n-1}(1-a\Delta t) + ab\Delta t + \epsilon_n\sigma\sqrt{\Delta t}, \end{aligned}$$

where $\Delta t = T/n$.

Using (9.5), one obtains a representation for the zero coupon bond price for the Vasicek model as

$$B(0,T) \simeq \mathrm{E}\left\{\exp\left(-\Delta t\sum_{i=1}^{n} r_i\right)\right\}.$$

Moreover, (9.7) can also be used to estimate a zero coupon bond price. To achieve an efficient simulation procedure, we need to scrutinize the preceding approximation more carefully. Express r_n in terms of $\epsilon_1, \ldots, \epsilon_n$ and r_0,

$$\begin{aligned} r_n &= (1-a\Delta t)^n r_0 + ab\Delta t\left[1 + (1-a\Delta t) + \cdots + (1-a\Delta t)^{n-1}\right] \\ &\quad +\sigma\sqrt{\Delta t}\left[(1-a\Delta t)^{n-1}\epsilon_1 + \cdots + (1-a\Delta t)\epsilon_{n-1} + \epsilon_n\right]. \end{aligned} \tag{9.8}$$

Since r_n equals to a constant plus a sum of normal variates, it is normally distributed. The coefficient of $ab\Delta t$ is a geometric progression $[1-(1-a\Delta t)^n]/(a\Delta t)$. This implies that the sum $X = -\Delta t\sum r_i$ is also a normal random variable. To obtain an efficient simulation, it is worth finding out the mean and the variance of X. Note that

$$\sum_{i=1}^{n} r_i = r_0\sum_{i=1}^{n}(1-a\Delta t)^i + ab\Delta t\sum_{i=1}^{n}\frac{1-(1-a\Delta t)^i}{a\Delta t} + \sigma\sqrt{\Delta t}\sum_{i=1}^{n}\epsilon_{n-i+1}\frac{1-(1-a\Delta t)^i}{a\Delta t}.$$

Simplifying,

$$(r_0-b)\frac{(1-a\Delta t)-(1-a\Delta t)^{n+1}}{a\Delta t} + bn + \sigma\sqrt{\Delta t}\sum_{i=1}^{n}\epsilon_{n-i+1}\frac{1-(1-a\Delta t)^i}{a\Delta t}.$$

Hence,

$$\begin{aligned} \mathrm{E}X &= -(r_0-b)\frac{(1-a\Delta t)-(1-a\Delta t)^{n+1}}{a} - bT, \\ \mathrm{Var}(X) &= \frac{\sigma^2}{a^2}\sum_{i=1}^{n}[1-(1-a\Delta t)^i]^2 \end{aligned}$$

$$= \frac{\sigma^2}{a^2}\left[n - 2(1-a\Delta t)\frac{1-(1-a\Delta t)^n}{a\Delta t} + (1-a\Delta t)^2\frac{1-(1-a\Delta t)^{2n}}{1-(1-a\Delta t)^2}\right].$$

Therefore, the Vasicek discount factor can be estimated by

$$B(0,T) \simeq \frac{1}{M}\sum_{j=1}^{M} e^{\mathrm{E}X + z_j\sqrt{\mathrm{Var}(X)}},$$

where $z_j \sim \mathrm{N}(0,1)$.

To improve the simulation further, we observe that $\mathrm{E}e^X$ is actually the moment generating function of X, which equals

$$\exp\left[\mathrm{E}X + \frac{1}{2}\mathrm{Var}(X)\right]. \tag{9.9}$$

The closed-form formula for the Vasicek discount factor can be obtained by allowing $n \to \infty$ (or $\Delta t \to 0$) in the mean and the variance expressions. The limit is

$$\begin{aligned} & B(0,T) \qquad (9.10)\\ = \; & \exp\left\{-bT - (r-b)\frac{1-e^{-aT}}{a} + \frac{\sigma^2}{4a^3}[4e^{-aT} - e^{-2aT} + 2aT - 3]\right\}. \end{aligned}$$

This closed-form formula allows us to check the accuracy of simulating the Vasicek model.

To define the yield in a stochastic interest model, we follow (9.4) and define the yield $y(t,T)$ as

$$B(t,T) = e^{-y(t,T)(T-t)}.$$

Equivalently,

$$y(t,T) = \frac{-1}{T-t}\log B(t,T) = \frac{-1}{T-t}\log\frac{P(t,T)}{P(T,T)}.$$

In practice, it is sometimes more relevant to know the yield of a bond rather than its value because market practitioners are used to extract bond price from the yield curve.

9.4 OPTIONS WITH STOCHASTIC INTEREST RATE

In Chapter 5, we simulate option prices for constant interest rate models. We now study simulating option price with stochastic interest rate. Recall that the call price can be written as

$$\text{Call} = \text{Discount factor} \times \mathrm{E}[\max(S_T - K, 0)].$$

Two features of stochastic interest rate are:

1. The discount factor should be calculated under a stochastic interest model.

2. The movement of the interest rate may affect the stock price.

The first feature can be tackled by calculating the discount factor using the method mentioned earlier. For the second feature, we consider the Black-Scholes asset dynamics under the Vasicek interest rate economy. The stock movement and interest rate movement are, respectively,

$$dS_t = r(t)S_t\,dt + \sigma_1 S_t\,dW_1 \quad \text{and} \quad dr(t) = a(b - r(t))\,dt + \sigma_s\,dW_2,$$

or, equivalently,

$$\begin{aligned} S_{t+1} &= S_t + r(t)S_t\,\Delta t + \sigma_1 S_t \epsilon_1 \sqrt{\Delta t}, \\ r(t+1) &= r(t) + a(b - r(t))\,\Delta t + \sigma_s \epsilon_2 \sqrt{\Delta t}. \end{aligned} \tag{9.11}$$

Suppose that the correlation coefficient of the stock and the interest rate is ρ. Then

$$\begin{bmatrix} \epsilon_1 \\ \epsilon_2 \end{bmatrix} \sim \mathrm{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right).$$

Computer software usually generates independent normal random variables by default. To generate correlated random variables, the following transformation of variables can be made:

$$\begin{aligned} z_1 &= \epsilon_1; \\ z_2 &= \frac{\epsilon_2 - \rho\epsilon_1}{\sqrt{1-\rho^2}}. \end{aligned}$$

It is easily seen that

$$\begin{aligned} \mathrm{Var}(z_1) &= 1, \\ \mathrm{Var}(z_2) &= 1, \\ \mathrm{Cov}(z_1, z_2) &= 0, \end{aligned}$$

and z_1 and z_2 are i.i.d. standard normal random variables. We generate (ϵ_1, ϵ_2) from the following procedures.

STEP 1: Generate $z_1, z_2 \sim \mathrm{N}(0,1)$ i.i.d.

STEP 2: Set $\epsilon_1 = z_1$ and $\epsilon_2 = z_2\sqrt{1-\rho^2} + \rho z_1$.

Substitute ϵ_1 and ϵ_2 into (9.11) to generate future possible asset prices and interest rates. The call option price can be obtained by the standard Monte Carlo method as

$$C(S,0) = B(0,T)\frac{1}{n}\sum_{j=1}^{n} \max(S_j(T) - K, 0), \tag{9.12}$$

where $B(0,T)$ is the Vasicek bond price obtained in (9.10) and $S_j(T)$ is the terminal asset price for the j-th path.

9.5 EXERCISES

1. Verify (9.10) by computing the limit.

2. Consider the CIR model:

$$dr = 0.1(0.05 - r)\,dt + 0.3\sqrt{r}\,dW \text{ and } r_0 = 0.052.$$

   (a) Construct and implement a standard Monte Carlo simulation to compute the discount factor.

   (b) Use the Vasicek discount factor of (9.10) as a control variate to improve the simulation in (a). (Hint: you may use $\sigma_{Vasicek} = \sigma_{CIR}\sqrt{r_0}$.)

   (c) Compare the difference between two prices based on 1,000 simulated prices.

3. Consider the Ho-Lee interest rate movement:

$$dr = \theta(t)\,dt + \sigma\,dW, \qquad (*)$$

   where $\theta(t) = a + e^{-bt}$, σ =constant and W is the standard Brownian motion.

   (a) Provide an algorithm to compute $B(0,t)$ by discretizing (*).

   (b) For pricing a five-year bond paying coupons semiannually, you adopted $\Delta t = 1/250$ to calculate the integration, and $M = 1,000$ to estimate each discount factor. What is the minimum size for the random sample used to compute the bond price with simulations in (a)?

   (c) Express r_t in terms of a, b, σ, h, and r_0 based on the algorithm in (b), where $h = \Delta t$. Hence, show that

$$r_t = r_0 + at + \frac{1 - e^{-bt}}{b} + \epsilon\sigma\sqrt{t}, \quad \text{when} \quad h \to 0,$$

   where $\epsilon \sim \mathrm{N}(0,1)$.

   (d) Modify the approach of Section 8.2.1 to derive closed form solution for the discount factor under the Ho-Lee model.

Simulation Techniques in Financial Risk Management
by Ngai Hang Chan and Hoi Ying Wong

10

Markov Chain Monte Carlo Methods

10.1 INTRODUCTION

Bayesian inference is an important area in statistics. It has found applications in a spectrum of disciplines. One of the main ingredients of Bayesian inference is to incorporate prior information via the specification of prior distributions. As information flows freely in financial markets, incorporating prior information with Bayesian ideas constitutes a natural approach. In this final chapter, we shall briefly introduce the essence of Bayesian statistics with reference to risk management. In particular, we shall discuss the celebrated Markov Chain Monte Carlo method in detail and illustrate its uses via a case study.

10.2 BAYESIAN INFERENCE

The essence of the Bayesian approach is to incorporate uncertainties for the unknown parameters. Predictive inference is conducted via the joint probability distribution of the parameters $\theta = (\theta_1, \theta_2, \ldots, \theta_r)$ conditional on the observable data $x = (x_1, \ldots, x_n)$. The joint distribution is deduced from the distribution of observable quantities via Baye's theorem. Many excellent texts have been written about the Bayesian paradigm; see, for example, DeGroot (1970), Box and Tiao (1973), Berger (1985), O'Hagan (1994), Bernardo and Smith (2000), Lee (2004), and Robert (2001), just to name a few. A succinct introduction to Bayesian inference for time series is given in Tsay (2006).

The observational (or sampling) distribution $f(x|\theta)$ is the likelihood function. Under the Bayesian framework, a *prior distribution* $p(\theta)$ is specified for the parameter θ. Inferences are conducted based on the posterior distribution $\pi(\theta|x)$ according to the following identity:

$$\pi(\theta|x) = \frac{f(x|\theta)p(\theta)}{f(x)},$$

where $f(x)$ is the marginal density such that

$$f(x) = \int f(x|\theta)p(\theta)\, d\theta. \tag{10.1}$$

The probability density function $\pi(\theta|x)$ is known as the *posterior* density function. Since x is observed, the marginal density in (10.1) is a constant. It is more convenient to express (10.1) as

$$\pi(\theta|x) \propto L(\theta)p(\theta), \tag{10.2}$$

where $L(\theta) = f(x|\theta)$ is the likelihood function. One way to estimate θ is to compute the posterior mean of θ, i.e.,

$$\hat{\theta} = \int \theta\pi(\theta|x)\, d\theta. \tag{10.3}$$

Prior and posterior are relative to the observables. A posterior distribution conditional on x can be used as a prior for a new observation y. This process can be iterated and eventually leads to a new posterior via the Bayes theorem. We illustrate this idea with a concrete example.

Example 10.1 *Suppose we observe $x_1, \ldots, x_n$ independent random variables each $N(\mu, \sigma^2)$ with μ unknown and σ^2 known. Estimate μ in a Bayesian setting.*

The likelihood function is

$$L(\mu) = \frac{1}{(2\pi\sigma)^{n/2}} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right] \propto \exp\left[-\frac{n}{2\sigma^2}(\overline{x} - \mu)^2\right],$$

where $\overline{x}$ is the sample mean of the observation. It seems natural to assume that μ follows a normal distribution by specifying the prior $p(\mu) \sim \mathrm{N}(m, \tau^2)$, where m and τ^2 are known as hyperparameters. Substituting this prior into (10.2), we have

$$\begin{aligned}\pi(\mu|x) &\propto \exp\left[-\frac{(\overline{x} - \mu)^2}{2\sigma^2/n}\right] \exp\left[-\frac{(\mu - m)^2}{2\tau^2/n}\right] \\ &\propto \exp\left[-\frac{(\mu - m_1)^2}{2\tau_1^2}\right],\end{aligned}$$

where

$$m_1 = \frac{\tau^2 \overline{x} + m\sigma^2/n}{\tau^2 + \sigma^2/n} \quad \text{and} \quad \tau_1^2 = \frac{\tau^2 \sigma^2}{n\tau^2 + \sigma^2},$$

equivalently,

$$\mu \sim \mathrm{N}(m_1, \tau_1^2).$$

The posterior mean $\hat{\mu} = \mathrm{E}(\mu) = m_1$ is an estimate of μ given x. Notice that m_1 tends to the sample mean $\overline{x}$ and τ_1^2 tends to zero as the number of observation increases. In most cases, the prior distribution plays a lesser role when the sample size is large. Another interesting observation is that the information contained in the prior becomes less when τ^2 increases. When when $\tau^2 \to \infty$, $p(\mu) \propto$ constant and $\pi(\mu|x) = \mathrm{N}(\overline{x}, \sigma^2/n)$. Such a prior is known as the noninformative prior as it provides no information about the distribution of μ.

There are many ways to specify a prior distribution in the Bayesian setting. Some prefer noninformative priors and others prefer priors that are analytically tractable. Conjugate priors are adopted to address the latter concern.

Given a likelihood function, the conjugate prior distribution is a prior distribution such that the posterior distribution belongs to the same class of distribution as the prior. Conjugate priors and posterior distributions are differed through hyperparameters. Example 10.1 serves as a good example. Conjugate priors facilitate statistical inferences because the posterior distributions belong to the same family as the prior distributions, which are usually of familiar forms. Moreover, updating posterior distributions with new information becomes straightforward as only hyperparameters have to be updated.

In the one-dimensional case, deriving conjugate priors is relatively simple when the likelihood belongs to the exponential family. Conjugacy within the exponential family is discussed in Lee (2004). Table 10.1 summarizes some of the commonly used conjugate families. Herein, Be denotes the Beta distribution, G the Gamma distribution, IG the Inverse Gamma distribution, and N the Normal distribution.

Likelihood $L(\theta)$	Conjugate prior $p(\theta)$
Poisson $\theta = \lambda$	$G(\alpha, \beta)$
Binomial $\theta = p$	$Be(\alpha, \beta)$
Normal $\theta = \mu$, σ^2 known	$\mathrm{N}(m, \tau^2)$
Normal $\theta = \sigma^2$, μ known	$IG(\alpha, \beta)$

Table 10.1 Conjugate priors.

10.3 SIMULATING POSTERIORS

Bayesian inference makes use of simulation techniques to estimate parameters naturally. As shown in (10.3), calculating a posterior mean is tantamount to numerically evaluating an integral. It is therefore not surprising that Monte Carlo simulation plays an important role. The integration in (10.3) is usually an improper integral (integration over an unbounded region), which renders standard numerical techniques useless. Although one may use numerical quadrature to bypass such a difficulty in the one-dimensional case, applying quadrature in higher dimensions are far from being simple. Financial modeling is usually of higher dimensions.

Monte Carlo simulation with importance sampling simplifies the computation of (10.3). Since it may be difficult to generate random variables from the posterior distribution $\pi(\theta|x)$ directly, we may take advantage of the fact that importance sampling enables us to compute integrations with a conveniently chosen density. Consider

$$\hat{\theta} = \int \theta\pi(\theta|x)\,d\theta = \int \frac{\theta\pi(\theta|x)}{q(\theta)} q(\theta)\,d\theta, \tag{10.4}$$

where $q(\theta)$ is an a priori specified density function that can be generated easily. Drawing n random samples θ_i from $q(\theta)$, we approximate the posterior mean by

$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} \frac{\theta_i\pi(\theta_i|x)}{q(\theta_i)}.$$

Note that the importance sampling is not used as a variance reduction device here. It is applied to facilitate the computation of the posterior mean. The variance of the computation can be large in some cases.

10.4 MARKOV CHAIN MONTE CARLO

One desirable feature of combining Markov chain simulation with Bayesian ideas is that the resulting method can handle high dimensional problems efficiently. Another desirable feature is to draw random samples from the posterior distribution directly. The Markov Chain Monte Carlo (MCMC) methods are developed with these two features in mind.

10.4.1 Gibbs Sampling

Gibbs sampling is probably one of the most commonly used MCMC methods. It is simple, intuitive, easily implemented, and designed to handle multidimensional problems. The basic limit theorem of Markov chain serves as the theoretical building block to guarantee that draws from Gibbs sampling agree with the posterior asymptotically.

Although conjugate priors are useful in Bayesian inference, it is difficult to construct a joint conjugate prior for several parameters. For a normal distribution with both mean and variance unknown, deriving the corresponding conjugate prior can be challenging. But conditional conjugate priors can be obtained relatively easily, see, for example, Gilks, Richardson and Spiegelhalter (1995). Conditioning on other parameters, a conditional conjugate prior is one dimension and has the same distributional structure as the conditional posterior.

Gibbs sampling takes advantage of this fact and offers a way to reduce a multidimensional problem to an iteration of low-dimensional problems. Specifically, let $x = (x_1, \ldots, x_n)$ be the data and let the distribution of each x_i be governed by r parameters, $\theta = (\theta_1, \theta_2, \ldots, \theta_r)$. For each $j = 1, \ldots, r$, specify the one-dimensional conditional conjugate prior $p(\theta_j)$ and construct the conditional posterior by means of the Baye's theorem. Then iterate the Gibbs procedure as follows.

Set an initial parameter vector $(\theta_2^0, \ldots, \theta_r^0)$. Update parameters by the following procedure:

- Sample $\theta_1^{(1)} \sim p(\theta_1 | \theta_2^0, \ldots, \theta_r^0, x)$;
- Sample $\theta_2^{(1)} \sim p(\theta_2 | \theta_1^{(1)}, \theta_3^0, \ldots, \theta_r^0, x)$;

  $\vdots \qquad\qquad \vdots$
- Sample $\theta_r^{(1)} \sim p(\theta_r | \theta_1^1, \theta_2^1, \ldots, \theta_{r-1}^1, x)$.

This completes one Gibbs iteration and the parameters are updated to $(\theta_1^1, \ldots, \theta_r^1)$. Using these new parameters as starting values, repeat the iteration again and obtain a new set of parameters $(\theta_1^2, \ldots, \theta_r^2)$. Repeating these iterations M times, we get a sequence of parameter vectors $\theta^{(1)}, \ldots, \theta^{(M)}$, where $\theta^{(i)} = (\theta_1^i, \ldots, \theta_r^i)$, for $i = 1, \ldots, M$. By virtue of the basic limit theorem of Markov chain, it can be shown that the Markov chain $\{\theta^{(M)}\}$ has a limiting distribution converging to the joint posterior $p(\theta_1, \theta_2, \ldots, \theta_r | x)$ when M is sufficiently large, see Tierney (1994). The number M is called the burn-in period. After simulating $\{\theta^{(M+1)}, \theta^{(M+2)}, \ldots, \theta^{(M+n)}\}$ from the Gibbs sampling, Bayesian inference can be conducted easily. For example, to compute the posterior mean, we evaluate

$$\hat{\theta}_i = \frac{1}{n} \sum_{i=1}^{n} \theta_i^{(M+i)}.$$

To acquire a clearer understanding of Gibbs sampler, consider the following example:

Example 10.2 *Let $x_1, \ldots, x_n$ be independent $\mathrm{N}(\mu, \sigma^2)$ random variables with both μ and σ^2 unknown. Estimate μ and σ^2 via Gibbs sampling.*

Recall that the conjugate prior of μ is normal for a given σ^2 and that the conjugate prior of σ^2 is inverse gamma for a given μ. Let $\mu_0 \sim \mathrm{N}(m_0, \tau_0^2)$ and $\sigma_0^2 \sim IG(\alpha_0, \beta_0)$ be random variables drawn from the initial priors. Define μ_i and σ_i^2 to be random variables generated in the i-th iteration of the Gibbs sampling procedure. The conditional posterior for μ_i can be obtained by mimicking Example 10.1. We have

$$\mu_i \sim \mathrm{N}(m_i, \tau_i^2),$$

where

$$m_i = \frac{\tau_{i-1}^2 \overline{x} + m_{i-1}\sigma_{i-1}^2/n}{\tau_{i-1}^2 + \sigma_{i-1}^2/n} \quad \text{and} \quad \tau_i^2 = \frac{\tau_{i-1}^2 \sigma_{i-1}^2}{n\tau_{i-1}^2 + \sigma_{i-1}^2}. \tag{10.5}$$

In Question 1 at the end of this chapter, the conditional posterior for σ_i^2 is found to be $\sigma_i^2 \sim IG(\alpha_i, \beta_i)$ where

$$\alpha_i = n/2 + \alpha_{i-1} \quad \text{and} \quad \beta_i = \frac{\beta_{i-1}}{4} + \frac{1}{2}\sum_{j=1}^{n}(x_j - \mu_i)^2. \tag{10.6}$$

Hence, Gibbs sampling is implemented as follows:

1. Set $i = 0$ and initial values of m_0, τ_0^2, α_0, β_0, and σ_0^2;
2. Sample $\mu_i \sim \mathrm{N}(m_i, \tau_i^2)$ and update α_{i+1} and β_{i+1} by (10.6);
3. Sample $\sigma_{i+1}^2 \sim IG(\alpha_{i+1}, \beta_{i+1})$ and update m_{i+1} and τ_{i+1}^2 by (10.5);
4. Set $i = i + 1$;
5. Go to Step 2 until i equals a prespecified integer $M + k$.

After that, we keep the last k pairs of random variables for indicies $M + 1$ to $M + k$. Estimation is achieved by taking sample means:

$$\begin{aligned} \widehat{\mu} &= \frac{1}{k}\sum_{j=1}^{k}\mu_{M+j}, \\ \widehat{\sigma^2} &= \frac{1}{k}\sum_{j=1}^{k}\sigma_{M+j}^2. \end{aligned}$$

10.4.2 Case Study: The Impact of Jumps on Dow Jones

To appreciate the usefulness of Gibbs sampling, we use it to estimate parameters of a jump-diffusion model and examine the impact of jumps in major financial indices. Note that maximum likelihood estimation does not work for this model (see Redner and Walker, 1984).

In the jump-diffusion model of Merton (1976), the dynamics of asset returns are assumed to be

$$d \log S = \mu\, dt + \sigma\, dW_t + Y\, dN_t, \tag{10.7}$$

where S is the equity price, W_t is the standard Brownian motion, N_t follows a Poisson process with an intensity λ, and Y is a normal random variable with mean k and variance s^2. We assume that dW_t, dN_t, and Y are independent random variables at each time point t. This model requires estimation of μ, σ, λ, k, and s based on observations $\{S_1, \ldots, S_n, S_{n+1}\}$, where S_i represents the equity price observed at time t_i. These prices produce n independent log-returns, which are denoted by $\{X_1, \ldots, X_n\}$ where $X_i = \log S_i - \log S_{i-1}$. With a fixed Δt, a discrete approximation to the dynamics (10.7) is

$$\Delta \log S = \mu\, \Delta t + \sigma\, \Delta W_t + Y\, \Delta N_t. \tag{10.8}$$

When Δt is sufficiently small, ΔN_t is either 1, with probability $\lambda \Delta t$, or 0, with probability $1 - \lambda \Delta t$.

Example 10.3 *Simulate 100 sample paths from the asset price dynamics of (10.7) with parameters: $\mu = 0.08, \sigma = 0.4, \lambda = 3.5, s = 0.3$, and $k = 0$. Each sample path replicates daily log-returns of a stock over a one-year horizon. Based on these 100 paths, estimate the values of μ, σ, λ, s, and k with the Gibbs sampling. Compare the results with input values.*

Simulating paths
Sample paths are simulated by assuming $n = 250$ trading days a year and so the discretization (10.8) has $\Delta t = 1/250$. On each path, l og-asset price at each time point is generated as follows,

$$\log S_{i+1} - \log S_i \;=\; \begin{cases} \mu\, \Delta t + \sigma \sqrt{\Delta t}\epsilon, & \text{if} \quad U > \lambda \Delta t \\ \mu\, \Delta t + k + \sqrt{\sigma^2 \Delta t + s^2}\epsilon, & \text{if} \quad U \le \lambda \Delta t \end{cases},$$

where $\epsilon \sim \mathrm{N}(0,1)$ and $U \sim U(0,1)$ are independent random variables. To simplify notations, we denote $x_i = \log S_{i+1} - \log S_i$. The corresponding SPLUS code and a graph of three sample paths are given below.

```
### generate observation of Y ####
MUY <- 0.08;
SIGMAY <- 0.40;
MUJ <- 0;
SIGMAJ <- 0.3;
LAMBDA <- 3.5;
m <- 100;
n <- 250;
dt<- 1/250;
```

```
Y <- matrix(100,m,n+1);

for (i in 1:n) {
JUMP     <- ifelse(runif(m)<LAMBDA*dt,1,0);
JumpSize <- JUMP*rnorm(m,MUJ,SIGMAJ);
Y[,i+1]  <- Y[,i] + rnorm(m,MUY*dt ,SIGMAY*sqrt(dt)) + JumpSize;
}

plot(Y[1,], type='l',xlab='time',ylab='stock price')
for (k in 2:100) {
plot(Y[k,], type='l',xlab='time',ylab='stock price')
}
```

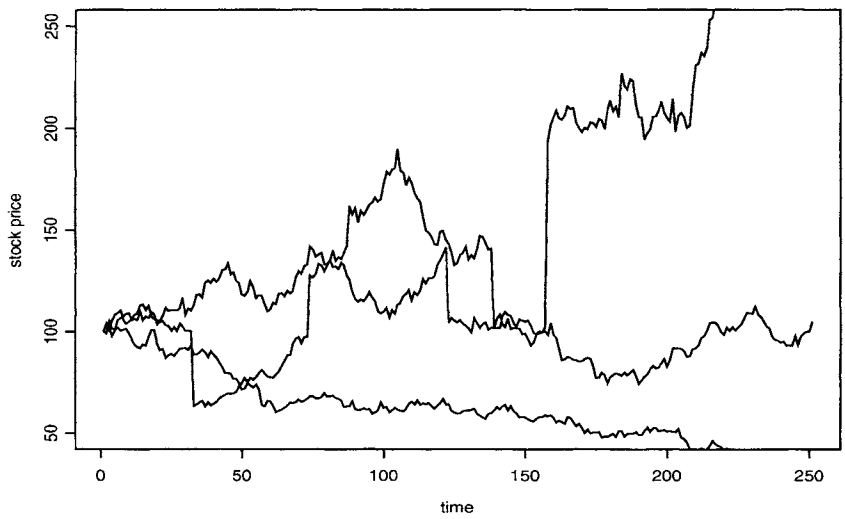


Fig. 10.1 A sample path of the jump-diffusion model.

Gibbs sampling

There are five parameters in the model so that we have to develop five conditional conjugate priors from their conditional likelihood functions. Let us proceed step by step.

1. Conditional prior and posterior for μ:
   Other things being fixed, the likelihood function of μ happens to be

proportional to a normal density. Specifically,

$$\begin{aligned} L(\mu) &\propto \prod_{i=1}^{n} \exp\left[-\frac{(x_i - \mu\Delta t - Y_i\Delta N_i)^2}{2\sigma^2\Delta t}\right] \\ &\propto \exp\left\{-\frac{1}{2\sigma^2}\left[\mu - \sum_{i=1}^{n}(x_i - Y_i\Delta N_i)\right]^2\right\}. \end{aligned}$$

Therefore, a normal distribution $\mathrm{N}(m, \tau^2)$ is suitable for μ as a conditional conjugate prior. The posterior distribution can be immediately obtained as

$$\mathrm{N}\left(\frac{\tau^2\sum_{i=1}^{n}(x_i - Y_i\Delta N_i) + m\sigma^2/n}{\tau^2 + \sigma^2/n}, \frac{\tau^2\sigma^2}{n\tau^2 + \sigma^2}\right). \tag{10.9}$$

2. Conditional prior and posterior for σ^2:
The conditional likelihood function of σ^2

$$L(\sigma^2) \propto \left(\sigma^2\right)^{-n/2} \exp\left[-\frac{1}{2\sigma^2\Delta t}\sum_{i=1}^{n}\left(x_i - \mu\Delta t - Y_i\Delta N_i\right)^2\right].$$

We select $IG(\alpha, \beta)$ as the conditional prior for σ^2. Then, the posterior distribution becomes

$$IG\left(\alpha + n/2, \beta + \frac{\sum_{i=1}^{n}\left(x_i - \mu\Delta t - Y_i\Delta N_i\right)^2}{2\Delta t}\right). \tag{10.10}$$

3. Conditional prior and posterior for λ:
The conditional likelihood of λ

$$L(\lambda) \propto (\lambda\Delta t)^N(1 - \lambda\Delta t)^{n-N},$$

where N is the total number of jumps in the horizon. From Table 10.1, we find that the appropriate conjugate prior is $Be(a, b)$. Simple computation shows that the posterior distribution is

$$Be\left(a + N, b + n - N\right). \tag{10.11}$$

4. Conditional prior and posterior for k:
Since k is the mean of the normal jump size, its prior and posterior are obtained in the same manner as μ. We state the result without proof. The prior is $\mathrm{N}(m_Y, \tau_Y^2)$ and the posterior is given by

$$\mathrm{N}\left(\frac{\tau_Y^2\sum_{j=1}^{N} Y_j/N + m_Y s^2/N}{\tau_Y^2 + s^2/n}, \frac{\tau_Y^2 s^2}{N\tau_Y^2 + s^2}\right). \tag{10.12}$$

5. Conditional prior and posterior for s^2:
   Since s^2 is the variance of the normal jump size, its prior and posterior are obtained in the same manner as σ^2. The prior is $IG(\alpha_Y, \beta_Y)$ and the posterior is given by

$$IG\left(\alpha_Y + N/2, \beta_Y + \frac{\sum_{i=1}^{N}(Y_i - k)^2}{2}\right). \tag{10.13}$$

The above priors and posteriors are distributions conditional on values of Y_i and ΔN_i. This complicates the Gibbs sampling procedure since only x_i is observable for all i. Therefore, at each time point t_i, Y_i and ΔN_i should be simulated from distributions conditional on the observed value of x_i before substituting them into the priors / posteriors. We need the following facts:

$$\begin{aligned} x_i|\Delta N_i = 0 &\sim \mathrm{N}(\mu\Delta t, \sigma^2\Delta t); \\ x_i|\Delta N_i = 1 &\sim \mathrm{N}(\mu\Delta t + k, \sigma^2\Delta t + s^2), \end{aligned}$$

which together with Baye's theorem show that

$$\begin{aligned} P(\Delta N_i = 1|x_i) &= \frac{P(x_i|\Delta N_i = 1)\lambda\Delta t}{P(x_i|\Delta N_i = 1)\lambda\Delta t + P(x_i|\Delta N_i = 0)(1 - \lambda\Delta t)}, \\ P(\Delta N_i = 0|x_i) &= 1 - P(\Delta N_i = 1|x_i). \end{aligned} \tag{10.14}$$

The jump size Y_i is necessary only when $\Delta N_i = 1$. Under such a situation, we recognize that the condition density function of Y_i is

$$f(Y_i|x_i) = f(x_i|Y_i)p(Y_i) \propto \exp\left[-\frac{(x_i - Y_i - \mu\Delta t)^2}{2\sigma^2\Delta t}\right] \exp\left[-\frac{(Y_i - k)^2}{2s^2}\right],$$

which implies

$$Y_i|x_i \sim \mathrm{N}\left(\frac{(x_i - \mu\Delta t)/\sigma^2\Delta t + k/s^2}{1/\sigma^2\Delta t + 1/s^2}, \frac{1}{1/\sigma^2\Delta t + 1/s^2}\right). \tag{10.15}$$

With all the ingredients ready, the Gibbs sampling starts by choosing initial values of μ_0, σ_0^2, k_0, λ_0, and s_0^2. We also need initial values of $Y_i^{(0)}$ and $\Delta N_i^{(0)}$, both of which can be obtained by a simulation with the initial parameters. The Gibbs sampling runs as follows:

1. Sample $\mu_j \sim p(\mu_j|\sigma_{j-1}^2, k_{j-1}, s_{j-1}^2, \lambda_{j-1})$ as given in (10.9);
2. Sample $\sigma_j \sim p(\sigma_j^2|\mu_j, k_{j-1}, s_{j-1}^2, \lambda_{j-1})$ as given in (10.10);
3. Sample $\lambda_j \sim p(\lambda_j|\mu_j, \sigma_j^2, k_{j-1}, s_j^2)$ as given in (10.11);
4. Sample $k_j \sim p(k_j|\mu_j, \sigma_j^2, s_{j-1}^2, \lambda_j)$ as given in (10.12);
5. Sample $s_j^2 \sim p(s_j^2|\mu_j, \sigma_j^2, k_j, \lambda_j)$ as given in (10.13);

6. Sample $\Delta N_i^{(j)} \sim p(\Delta N_i^{(j)}|\mu_j, \sigma_j^2, k_j, s_j^2)$ as given in (10.14) for all $i = 1, 2, \ldots, n$;

7. Sample $Y_i^{(j)} \sim p(Y_i^{(j)}|\mu_j, \sigma_j^2, k_j, s_j^2)$ as given in (10.15) at the time point t_i that $\Delta N_i = 1$;

8. Set $j = j + 1$ and go to step 1. Repeat until $j = M' + M$.

Inference is drawn by taking sample means of the values of the last M simulated parameters. The SPLUS code is given as follows:

```
### Input the stock price ###
Y <- read.table("table.txt")
Y <- log(Y)
### initial value for the Markov chain ###
dY <- diff(Y);
lambda <- 6;
mu <- 0;
sigma <- 1;
k <- 0;
s  <- 1;
jump<- c(rep(1,n))
jumpsize <- dY/2

### assign prior distribution ###
MEANmu<-mean(dY)/dt/n;
VARmu <- 1;
MEANk <- 0.5;
VARk  <- 1;
ALPHAsigma <- 2.5;
BETAsigma <- 1/25;
ALPHAs <- 2.5;
BETAs <- 1/25;
ALPHAlambda <- 2;
BETAlambda <- 60;

sample<-c(rep(0,n+5))
result<-c(rep(0,5))

### assign prior distribution ###
MEANmu<-mean(dY)/dt/n;
VARmu <- 1;
MEANk <- 0.5;
VARk <- 1;
ALPHAsigma <- 2.5;
BETAsigma <- 1/25;
```

```
ALPHAs <- 2.5;
BETAs <- 1/25;
ALPHAlambda <- 2;
BETAlambda <- 60;

sample<-c(rep(0,n+5))
result<-c(rep(0,5))

for (i in 1:m) {
### calculate the parameters for posterior distributions ###
Vmu <- sigma^2/n/dt; #normal
Mmu <- sum(dY-jumpsize)/n/dt;
Vmu2<- 1/(1/Vmu + 1/VARmu);
Mmu2<- (Mmu/Vmu+MEANmu/VARmu)/(1/Vmu+1/VARmu);
mu  <- rnorm(1,Mmu2,sqrt(Vmu2))

Asigma <- n/2+1 #inverted gamma
Bsigma <- sum((dY-mu*dt-jumpsize)^2)/2/dt;
Asigma2<- Asigma + ALPHAsigma+1;
Bsigma2<- Bsigma + BETAsigma;
sigma  <- 1/sqrt(rgamma(1,Asigma2,Bsigma2));

J <- jumpsize[jump==1]
j <- sum(jump)
Alambda <- j+1 ; #BETA
Blambda <- n-j+1;
Alambda2<- Alambda + ALPHAlambda-1;
Blambda2<- Blambda + BETAlambda-1;
lambda <- rbeta(1,Alambda2,Blambda2)/dt;

    if  (j>1) {
Mk <- mean(J); #normal
Vk <- s^2/j;
Vk2<- 1/(1/Vk + 1/VARk);
Mk2<- (Mk/Vk+MEANk/VARk)/(1/Vk+1/VARk);
k  <- rnorm(1,Mk2,sqrt(Vk2));
As <- j/2+1 #inverted gamma
Bs <- sum((J-k)^2)/2;
As2<- As + ALPHAs+1;
Bs2<- Bs + BETAs;
s  <- 1/sqrt(rgamma(1,As2,Bs2));
}

### find the probabilities of jump and distribution of jump ###
varjump  <- 1/(1/s^2+1/sigma^2/dt);
```

```
meanjump <- ( (dY-mu*dt)/sigma^2/dt + k/s^2 ) *varjump;
jumpsize <- jump*(rnorm(n)*sqrt(varjump) + meanjump);

ratio1 <- (1-lambda*dt)/(lambda*dt);
ratio2 <- sqrt((sigma^2*dt+s^2)/(sigma^2*dt));
ratio3 <- - (dY-mu*dt)^2/sigma^2/2/dt + (dY-mu*dt-k)^2/(sigma^2*dt+s^2)/2 ;
pjumps <- 1/( 1 + ratio1*ratio2*exp(ratio3));
jump <- ifelse(runif(n)<pjumps,1,0);

s <-c(mu,sigma,lambda,k,s);
sample <-sample+c(jump,s);
result <- c(result, s)
print(c(s,sum(jump)))
}

result <- matrix(result, 5)
plot(result[1,1:m], type="l", xlab="no. of step", ylab="drift")
plot(result[2,1:m], type="l", xlab="no. of step", ylab="volatility")
plot(result[3,1:m], type="l", xlab="no. of step", ylab="intensity")
plot(result[4,1:m], type="l", xlab="no. of step", ylab="mean of jump")
plot(result[5,1:m], type="l", xlab="no. of step", ylab="S.D of jump")

### the probabilities that time point has jump ###
print(sample[1:n]/m)
### the estimated parameter ###
print(sample[(n+1):(n+5)]/m)
### the true parameter ###
print(c(MUY,SIGMAY,LAMBDA,MUJ,SIGMAJ,sum(JUMP)))
### the sample mean & volatility of jump ###
JS<-JumpSize[JUMP==1]
print(c(sum(JUMP),mean(JS),sqrt(var(JS))))
```

Results and comparisons

Table 10.2 shows our estimation results. We report the averaged posterior means over the 100 sample paths and the variances. It is seen that the estimates are close to true values and variances are small. Gibbs sampling does a good job in estimating parameters for jump-diffusion models.

□

Example 10.3 shows the usefulness of Gibbs sampling in estimating the jump-diffusion model. In practice, this application can be crucial for a risk manager to assess how much risks are due to jumps. To examine the jump risk empirically, we estimate the impact of jumps on the Dow Jones Industrial

	μ	σ^2	λ	k	s^2
True value	0.08	0.4	3.5	0	0.3
Mean	0.0769	0.3986	3.8600	0.0163	0.2868
Variance	0.0233	6.5×10^{-5}	0.8895	0.0039	0.0015

Table 10.2 Performance of the Gibbs sampling.

Index. Our estimation is based on daily closing prices over the period 1995-2004. Parameters are estimated on an annual basis.

Year	μ	σ^2	λ	k	s^2
1995	0.2871	0.0901	1.9035	0.0627	0.2608
1996	0.2483	0.1172	2.818	-0.0337	0.235
1997	0.2384	0.1684	3.6587	-0.0256	0.2087
1998	0.1776	0.1752	5.5127	-0.0123	0.1782
1999	0.2177	0.1624	1.7968	-0.0176	0.2627
2000	-0.0162	0.1971	3.3364	-0.0235	0.2157
2001	0.015	0.1951	4.1797	-0.0383	0.2008
2002	-0.2188	0.2484	2.7072	0.0106	0.239
2003	0.1891	0.1626	2.0479	0.0661	0.2463
2004	0.0351	0.1111	1.7561	0.0004	0.2788

Table 10.3 Jump-diffusion estimation for Dow Jones.

In Table 10.3, the number of jumps per year, λ, range from 1.75 to 5.5. Therefore, we can have 5-6 jumps in a particular year. The impact of jumps is significant as almost all s^2 are bigger than 0.2. The variances σ^2 associated to the Brownian motion part of the model are around 0.2, but should be divided by 250 to produce a daily variance. When a jump arrives, an extra daily variance of 0.2 is added to the index return variance as $\sigma^2/250 + s^2$. The additional variance due to a jump is a relatively large quantity. Jump risk cannot be ignored! This information is useful for risk managers to construct scenarios for stress testing.

10.5 METROPOLIS-HASTINGS ALGORITHM

In this section, we explain why random draws using Gibbs sampling approximate the posterior distribution. To obtain a general result, we first introduce the Metropolis-Hastings algorithm in which the Gibbs sampling is a special case. We then show that Metropolis-Hastings algorithm constructs a Markov chain with limiting distribution following the posterior distribution. Further details are given in Casella and George (1992), Chib and Greenberg (1995), and Lee (2004).

Consider a Markov chain $\{\theta^{(n)}\}$ with a finite state space $\{1, 2, \cdots, m\}$ and transition probabilities p_{ij}. Given the transition probabilities, the limiting distribution of the chain can be found by solving the following equation:

$$\pi(j) = \sum_{i=1}^{m} \pi(i) p_{ij}.$$

When the state space is continuous, the sum is replaced by an integral (see, for example, Bhattacharya and Waymire, 1992).

In MCMC, we work with a reverse problem. Given a posterior distribution $\pi(j)$, we want to construct a Markov chain whose transition probabilities converge to the posterior distribution. If the transition probabilities satisfy the time reversibility with respect to $\pi(j)$, then its limiting distribution is guaranteed to be equal to $\pi(j)$. To explain time reversibility, write the transition probabilities p_{ij} as

$$p_{ij} = p^*_{ij} + r_i \delta_{ij},$$

$$\delta_{ii} = 1 \quad \text{and} \quad \delta_{ij} = 0 \quad \text{for} \quad i \neq j.$$

where $p^*_{ii} = 0$, $p^*_{ij} = p_{ij}$ for $i \neq j$, and $r_i = p_{ii}$.

If the equation

$$\pi(i) p^*_{ij} = \pi(j) p^*_{ji} \tag{10.16}$$

is satisfied for all i, then the probabilities p_{ij} are time reversible. This condition asserts that the probability of starting at i and ending at j when the initial probability is given by $\pi(i)$ is the same as that of starting at j and ending at i. By simple computation, we check that

$$\begin{aligned} \sum_i \pi(i) p_{ij} &= \sum_i \pi(i) p^*_{ij} + \sum_i \pi(i) r_i \delta_{ij} \\ &= \sum_i \pi(j) p^*_{ji} + \pi(j) r_j \\ &= \pi(j)(1 - r_j) + \pi(j) r_j \\ &= \pi(j). \end{aligned}$$

Therefore, $\pi(j)$ is the limiting distribution of the chain.

In other words, a Markov chain whose limiting distribution is the posterior distribution can be constructed by finding a time-reversible Markov chain. We start this process by specifying transition probabilities q_{ij}. If probabilities q_{ij} have already satisfied the time reversibility, then the corresponding Markov chain is the one we want. Otherwise, suppose that

$$\pi(i)q_{ij} > \pi(j)q_{ji}.$$

Then, it has a higher probability to move from i to j than from j to i. Therefore, we introduce a probability α_{ij} to reduce the moves from i to j. We would like to have

$$\pi(i)q_{ij}\alpha_{ij} = \pi(j)q_{ji} \tag{10.17}$$

so that

$$\alpha_{ij} = \frac{\pi(j)q_{ji}}{\pi(i)q_{ij}}.$$

Since we do not want to reduce the likelihood of moving from j to i, we set $\alpha_{ji} = 1$. Therefore, the general formula is

$$\alpha_{ij} = \min\left[\frac{\pi(j)q_{ji}}{\pi(i)q_{ij}}, 1\right]. \tag{10.18}$$

From (10.17) and (10.18), we see that the transition probabilities

$$\begin{aligned} p_{ij} &= q_{ij}\alpha_{ij}, \quad \text{for } i \neq j, \\ p_{ii} &= 1 - \sum_j q_{ij}\alpha_{ij}, \end{aligned} \tag{10.19}$$

are time reversible with respect to $\pi(i)$ and hence define a Markov chain whose limiting distribution is the required one. This method is called the Metropolis-Hasting algorithm.

Example 10.4 *Consider a random walk Markov chain:*

$$\boxed{A} \rightleftharpoons \boxed{B} \rightleftharpoons \boxed{C} \rightleftharpoons \boxed{D}$$

All transition probabilities are 0.5, except that the transitions 'from A to B' and 'from D to C' are 1. The transition matrix of the chain is given by

	A	B	C	D
A	0	1	0	0
B	0.5	0	0.5	0
C	0	0.5	0	0.5
D	0	0	1	0

Based on the Metropolis-Hasting algorithm, construct a Markov chain whose limiting distribution is $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$.

Simple calculation shows that the limiting distribution of the original Markov chain is $(\frac{1}{6}, \frac{1}{3}, \frac{1}{3}, \frac{1}{6})$. To construct the desired Markov chain, we have to compute probabilities α_{ij}. For instance,

$$\alpha_{AB} = \min\left(1, \frac{\pi(B)\ P(A|B)}{\pi(A)\ P(B|A)}\right) = \min\left(1, \frac{(\frac{1}{4})(\frac{1}{2})}{(\frac{1}{4})(1)}\right) = \frac{1}{2}.$$

This means the transition probability 'from A to B' is reduced from 1 to $1 \times \frac{1}{2} = \frac{1}{2}$. For node 'A', the remaining transition probabilities correspond to the event that no transition occurs. Transition probabilities for the other nodes are obtained in the same manner. The final transition matrix becomes:

	A	B	C	D
A	0.5	0.5	0	0
B	0.5	0	0.5	0
C	0	0.5	0	0.5
D	0	0	0.5	0.5

It is easy to verify that the limiting distribution of this Markov chain is $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$. □

Let us connect the Gibbs sampling with the Metropolis-Hasting algorithm. Suppose we have r parameters, i.e., $\theta = (\theta_1, \cdots, \theta_r)$, in the model. We want to generate $\theta \sim \pi(\theta)$. Let $\theta^{(n)}$ be the state of θ at time n and $\theta^{(n+1)}$ be the state of θ at time $n+1$. The sequence of vectors $\{\theta^{(n)}\}$ forms a Markov chain by virtue of the Gibbs sampling. Therefore, $\theta^{(n)}$ jumps to $\theta^{(n+1)}$ and they differ only in one component. To fix ideas, we assume the differed component to be the first component. Since we sample $\theta_1^{(n+1)}$ from the density function $p(\theta_1^{(n+1)}|\theta_2^{(n)}, \cdots, \theta_r^{(n)})$, the probability density function for the candidate Markov chain is given by

$$q\left(\theta^{(n+1)} \middle| \theta^{(n)}\right) = P\left\{\theta_1^{(n+1)} \middle| \theta_i, i \neq 1\right\}.$$

To construct a Markov chain whose limiting probability is the target one, Metropolis-Hasting algorithm multiplies q by a probability density function α, defined in (10.18). As θ is a vector of continuous random variables, the definition of α in (10.18) is modified to be

$$\alpha\left(\theta^{(n+1)} \middle| \theta^{(n)}\right) = \min\left\{\frac{\pi\left(\theta^{(n+1)}\right) q\left(\theta^{(n)} \middle| \theta^{(n+1)}\right)}{\pi\left(\theta^{(n)}\right) q\left(\theta^{(n+1)} \middle| \theta^{(n)}\right)}, 1\right\}.$$

Consider the following identities:

$$\begin{aligned} P\left(\theta^{(n+1)}\right) &= P\left(\theta_1^{(n+1)} \middle| \theta_i^{(n+1)}, i \neq 1\right) P\left(\theta_2^{(n+1)}, \cdots, \theta_r^{(n+1)}\right) \\ \pi\left(\theta^{(n+1)}\right) &= P\left(\theta_1^{(n+1)} \middle| \theta_i^{(n)}, i \neq 1\right) P\left(\theta_2^{(n)}, \cdots, \theta_r^{(n)}\right) \end{aligned}$$

$$= q\left(\theta^{(n+1)} \middle| \theta^{(n)}\right) P\left(\theta_2^{(n)}, \cdots, \theta_r^{(n)}\right); \tag{10.20}$$

and

$$\begin{aligned} P\left(\theta^{(n)}\right) &= P\left(\theta_1^{(n)} \middle| \theta_i^{(n)}, i \neq 1\right) P\left(\theta_2^{(n)}, \cdots, \theta_r^{(n)}\right) \\ \pi\left(\theta^{(n)}\right) &= P\left(\theta_1^{(n)} \middle| \theta_i^{(n+1)}, i \neq 1\right) P\left(\theta_2^{(n)}, \cdots, \theta_r^{(n)}\right) \\ &= q\left(\theta^{(n)} \middle| \theta^{(n+1)}\right) P\left(\theta_2^{(n)}, \cdots, \theta_r^{(n)}\right). \end{aligned} \tag{10.21}$$

Therefore, we deduce that

$$P\left(\theta_2^{(n)}, \cdots, \theta_r^{(n)}\right) = \frac{\pi\left(\theta^{(n+1)}\right)}{q\left(\theta^{(n+1)} \middle| \theta^{(n)}\right)} = \frac{\pi\left(\theta^{(n)}\right)}{q\left(\theta^{(n)} \middle| \theta^{(n+1)}\right)},$$

which implies that $\alpha = 1$. Hence, Gibbs sampling is a special case of the Metropolis-Hastings algorithm in which every jump is accepted.

10.6 EXERCISES

1. Suppose $X_1, \ldots, X_n$ are independent observations that follow $\mathrm{N}(\mu, \sigma^2)$, where μ is a known quantity

   (a) Show that the likelihood function $L(\sigma^2)$ satisfies

   $$L(\sigma^2) \propto (\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(X_i - \mu)^2\right\}.$$

   (b) Suppose further that $\sigma^2 \sim IG(\alpha, \beta)$. What is the conditional distribution of $\sigma^2 | X_1, \ldots, X_n$?
   Hint: Denote $p(\phi)$ as the density of the inverse Gamma distribution. Then we have

   $$p(\phi) \propto \phi^{-\alpha+1} e^{-\beta/\phi}.$$

2. A density function with a single parameter, $p(x|\theta)$, is said to be of exponential family if it takes the form

   $$p(x|\theta) = g(x)h(\theta) \exp\left[\sum t(x)\psi(\theta)\right].$$

   Show that normal mean with a known variance, normal variance with a know mean, Poisson distribution, and binomial distribution are of the exponential family.

3. Show that if the likelihood function comes from the exponential family and the prior distribution is from the exponential family, then the posterior distribution also belongs to the exponential family.

4. Simulate the daily jump-diffusion VaR of Dow Jones Industrial Index based on the data used in Section 5.2.3. Compare your number with the GED-VaR defined in Chapter 5.

5. Suppose that $x|p \sim Bin(n,p)$ and $p|x \sim Be(x+\alpha, n-x+\beta)$ where n is a Poisson variable of mean λ. Use the Gibbs sampling to find the unconditional distribution of n where $\lambda = 16$, $\alpha = 2$ and $\beta = 4$.

6. Consider the normal distribution with an unknown mean μ and a known variance.

   (a) Assume that the prior of μ is a discrete mixture of two normal densities. Show that this prior is still conjugate.

   (b) Assume that the prior of μ is a discrete mixture of k normal densities. Is the prior still conjugate?

7. Consider the following transition matrix of a Markov chain:

	1	2	3	4
1	1/6	0	1/2	1/3
2	0	1/3	1/3	1/3
3	0	1/2	0	1/2
4	1/4	1/4	1/4	1/4

   Use the Metropolis-Hastings algorithm to construct a Markov chain whose limiting distribution is $(1/6, 1/6, 1/3, 1/3)$ based on the above matrix.

8. Consider the transition matrix of another Markov chain:

	1	2	3	4
1	1/2	0	1/2	0
2	2/3	0	1/6	1/6
3	0	1/3	0	2/3
4	1/4	1/4	1/4	1/4

   Use the Metropolis-Hastings algorithm to construct a Markov chain whose limiting distribution is $(1/10, 2/10, 3/10, 4/10)$ based on the above matrix.

Simulation Techniques in Financial Risk Management
by Ngai Hang Chan and Hoi Ying Wong

11

Answers to Selected Exercises

11.1 CHAPTER 1

1.
$$\begin{aligned}\widehat{\mathrm{Var}}_g\left(\frac{f(x)}{g(x)}\right) &= \widehat{\mathrm{E}}\left(\frac{f(x)}{g(x)}\right)^2 - \left(\widehat{\mathrm{E}}(\frac{f(x)}{g(x)})\right)^2 \\ &= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{f(x_i)}{g(x_i)}\right)^2 - \widehat{I}^2. \\ \widehat{\mathrm{Var}}(\widehat{I}) &= \frac{1}{n}\widehat{\mathrm{Var}}_g\left(\frac{f(x)}{g(x)}\right) \\ &= \frac{1}{n^2}\sum_{i=1}^{n}\left(\frac{f(x_i)}{g(x_i)}\right)^2 - \frac{1}{n}\widehat{I}^2.\end{aligned}$$

4. (a) $m_s = \mathrm{E}(R_S) = \mathrm{E}(\frac{S_1-S}{S}) = \frac{\mathrm{E}(S_1)-S}{S} = \frac{[pS_u+(1-p)S_d]-S}{S}$.

(b) $\mathrm{Var}(R_S) = \mathrm{Var}(\frac{S_1-S}{S}) = \frac{\mathrm{Var}(S_1)}{S^2} = \frac{p(1-p)(S_u-S_d)^2}{S^2}$.
$v_S = \sqrt{\mathrm{Var}(R_S)} = \sqrt{p(1-p)}\frac{(S_u-S_d)}{S}$.

(c) Similarly, $\mathrm{E}(R_C) = \mathrm{E}(\frac{C_1-C}{C}) = \frac{\mathrm{E}(C_1)-C}{C} = \frac{[pC_u+(1-p)C_d]-C}{C}$.

(d) $\mathrm{Var}(R_C) = \mathrm{Var}(\frac{C_1-C}{C}) = \frac{\mathrm{Var}(C_1)}{C^2} = \frac{p(1-p)(C_u-C_d)^2}{C^2}$.
$v_C = \sqrt{\mathrm{Var}(R_C)} = \sqrt{p(1-p)}\frac{(C_u-C_d)}{C}$.

(e) Right-hand side
$= \Omega v_S$
$= \frac{C_u - C_d}{C} / \frac{S_u - S_d}{S} \sqrt{p(1-p)} \frac{(S_u - S_d)}{S}$
$= \sqrt{p(1-p)} \frac{C_u - C_d}{C}$
$= v_C$
= Left-hand side.

11.2 CHAPTER 2

1. (a)
$$\begin{aligned} \mathrm{E}X_t &= \sigma_1 \mathrm{E}(W_t - W_{t_2}) - \sigma_2 \mathrm{E}(W_{t_1} - W_{t_0}) \\ &= \sigma_1(0) - \sigma_2(0) = 0. \\ \mathrm{Var}X_t &= \sigma_1^2 \mathrm{Var}(W_t - W_{t_2}) + (-\sigma_2)^2 \mathrm{Var}(W_{t_1} - W_{t_0}) \\ &= \sigma_1^2(t - t_2) + \sigma_2^2(t_1 - t_0). \end{aligned}$$

(b)
$$\begin{aligned} X_t &= \sigma_1(W_t - W_{t_2}) - \sigma_2(W_{t_1} - W_{t_0}) \\ &= \sigma_1(W_t - W_{t_1}) + (\sigma_1 - \sigma_2)(W_{t_1} - W_{t_2}) + \sigma_2(W_{t_2} - W_{t_0}). \\ \mathrm{E}X_t &= \sigma_1 \mathrm{E}(W_t - W_{t_1}) + (\sigma_1 - \sigma_2)\mathrm{E}(W_{t_1} - W_{t_2}) \\ &\quad + \sigma_2 \mathrm{E}(W_{t_2} - W_{T_0}) = 0. \\ \mathrm{Var}X_t &= \sigma_1^2 \mathrm{Var}(W_t - W_{t_1}) + (\sigma_1 - \sigma_2)^2 \mathrm{Var}(W_{t_1} - W_{t_2}) \\ &\quad + \sigma_2^2 \mathrm{Var}(W_{t_2} - W_{t_0}) \\ &= \sigma_1^2(t - t_1) + (\sigma_1 - \sigma_2)^2(t_1 - t_2) + \sigma_2^2(t_2 - t_0). \end{aligned}$$

(c)
$$\begin{aligned} \mathrm{E}X_t &= \mathrm{E}\sum_{j=1}^{n} f(W_{t_{j-1}}, t_{j-1})(W_{t_j} - W_{t_{j-1}}) \\ &= \sum_{j=1}^{n} f(W_{t_{j-1}}, t_{j-1})\mathrm{E}(W_{t_j} - W_{t_{j-1}}) = 0. \\ \mathrm{Var}X_t &= \mathrm{Var}\sum_{j=1}^{n} f(W_{t_{j-1}}, t_{j-1})(W_{t_j} - W_{t_{j-1}}) \\ &= \sum_{j=1}^{n} f(W_{t_{j-1}}, t_{j-1})^2 \mathrm{Var}(W_{t_j} - W_{t_{j-1}}) \\ &= \sum_{j=1}^{n} f(W_{t_{j-1}}, t_{j-1})^2 (t_j - t_{j-1}). \end{aligned}$$

(d) Let $0 = t_0 < t_1 < t_2 < \cdots < t_n = t$,

$$\begin{aligned}
&\mathrm{E}\left[\int_0^t f(W_\tau, \tau)\, dW_\tau\right] \\
&= \mathrm{E}\left[\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} f(W_{t_{j-1}}, t_{j-1})(W_{t_j} - W_{t_{j-1}})\right] \\
&= \lim_{n\to\infty} \sum_{j=1}^{n} f(W_{t_{j-1}}, t_{j-1})\mathrm{E}(W_{t_j} - W_{t_{j-1}}) \\
&= \lim_{n\to\infty} \sum_{j=1}^{n} f(W_{t_{j-1}}, t_{j-1})(0) = 0.
\end{aligned}$$

$$\begin{aligned}
&\mathrm{E}\left[\int_0^t f(W_\tau, \tau)\, dW_\tau\right]^2 \\
&= \mathrm{E}\left[\lim_{n\to\infty} \sum_{j=1}^{n} f(W_{t_{j-1}}, t_{j-1})(W_{t_j} - W_{t_{j-1}})\right]^2 \\
&= \mathrm{E}\left[\lim_{n\to\infty} \sum_{i=1}^{n}\sum_{j=1}^{n} f(W_{t_{i-1}}, t_{i-1})(W_{t_i} - W_{t_{i-1}}) f(W_{t_{j-1}}, t_{j-1})(W_{t_j} - W_{t_{j-1}})\right] \\
&= \lim_{n\to\infty} \sum_{i=1}^{n}\sum_{j=1}^{n} \mathrm{E}\left[f(W_{t_{i-1}}, t_{i-1})(W_{t_i} - W_{t_{i-1}}) f(W_{t_{j-1}}, t_{j-1})(W_{t_j} - W_{t_{j-1}})\right].
\end{aligned}$$

Suppose $i < j$,

$$\begin{aligned}
&\mathrm{E}\left[f(W_{t_{i-1}}, t_{i-1}) f(W_{t_{j-1}}, t_{j-1})(W_{t_i} - W_{t_{i-1}})(W_{t_j} - W_{t_{j-1}})\right] \\
&= \mathrm{E}\left[f(W_{t_{i-1}}, t_{i-1}) f(W_{t_{j-1}}, t_{j-1})(W_{t_i} - W_{t_{i-1}})\right] \mathrm{E}(W_{t_j} - W_{t_{j-1}}) \\
&= \mathrm{E}\left[f(W_{t_{i-1}}, t_{i-1}) f(W_{t_{j-1}}, t_{j-1})(W_{t_i} - W_{t_{i-1}})\right] (0) = 0.
\end{aligned}$$

Similarly, for $i > j$, the expectation is also zero.

$$\begin{aligned}
\mathrm{E}\left[\int_0^t f(W_\tau, \tau)\, dW_\tau\right]^2 &= \lim_{n\to\infty} \sum_{j=1}^{n} \mathrm{E}\left[f(W_{t_{j-1}}, t_{j-1})^2 (W_{t_j} - W_{t_{j-1}})^2\right] \\
&= \lim_{n\to\infty} \sum_{j=1}^{n} \mathrm{E} f(W_{t_{j-1}}, t_{j-1})^2 \mathrm{E}(W_{t_j} - W_{t_{j-1}})^2 \\
&= \lim_{n\to\infty} \sum_{j=1}^{n} \mathrm{E} f(W_{t_{j-1}}, t_{j-1})^2 (t_j - t_{j-1}) \\
&= \int_0^t \mathrm{E} f(W_\tau, \tau)^2\, d\tau.
\end{aligned}$$

2. (a) Apply Itô's lemma,

$$\begin{aligned} dS_t &= \left((-\frac{1}{3})\frac{\partial S}{\partial X} + \frac{1}{2}(\frac{1}{2})^2\frac{\partial^2 S}{\partial X^2} + \frac{\partial S}{\partial t}\right) dt + (\frac{1}{2})\frac{\partial S}{\partial X}\, dW \\ &= \left(-\frac{1}{3}S + \frac{1}{8}S + 0\right) dt + \frac{1}{2}S\, dW \\ &= -\frac{5}{24}S\, dt + \frac{1}{2}S\, dW. \end{aligned}$$

(c) As X_t tends to negative infinity, S_t should tend to zero.

(e) $\bar{S}_{1000} \to \mathrm{E}e^{X_{1000}} = e^{-\frac{1000}{3}+\frac{1}{2}(\frac{1}{2})^2 1000} = e^{-\frac{625}{3}}$.

3. Apply Itô's lemma,

$$\begin{aligned} dG_t &= \left(aS\frac{\partial G}{\partial S} + \frac{1}{2}(bS)^2\frac{\partial^2 G}{\partial S^2} + \frac{\partial G}{\partial t}\right) dt + bS\frac{\partial G}{\partial S}\, dW \\ &= \left(aS\frac{1}{2}S^{-\frac{1}{2}} + \frac{1}{2}(bS)^2(-\frac{1}{4})S^{-\frac{3}{2}} + 0\right) dt + bS(\frac{1}{2}S^{-\frac{1}{2}})\, dW \\ &= (\frac{1}{2}a - \frac{1}{8}b^2)G\, dt + \frac{1}{2}b\, G\, dW. \end{aligned}$$

4. (a) Apply Itô' s lemma,

$$d\log S = (0.1 - \frac{1}{2}0.3^2)\, dt + 0.3\, dW.$$

Since $\log S(0) = 0$, we have,

$$\log S(t) = 0.055t + 0.3\sqrt{t}Z, \quad \text{where} \quad Z \sim \mathrm{N}(0,1),$$

$$p = \lim_{t\to\infty}\frac{1}{t}\log S(t) = \lim_{t\to\infty}\frac{1}{t}(0.055t + 0.3\sqrt{t}Z) = 0.055 + \lim_{t\to\infty}\frac{0.3Z}{\sqrt{t}} = 0.055.$$

(b) The limit does not exist.

$$\lim_{t\to\infty}\frac{1}{t}[\log S(t) - pt]^2 = \lim_{t\to\infty}\frac{1}{t}(0.3\sqrt{t}Z)^2 = \lim_{t\to\infty}(0.09Z^2) = 0.09Z^2.$$

5. (a) The simulation result should match Itô's lemma.

(b) The difference between Itô's integral and the Stratonovich integral is $\frac{1}{2}$.

6. (a)

$$\frac{\partial Y}{\partial W} = e^W, \quad \frac{\partial^2 Y}{\partial W^2} = e^W,$$

$$dY = \frac{1}{2}e^W\, dt + e^W\, dW = \frac{1}{2}Y\, dt + Y\, dW.$$

(b) By (a)

$$\begin{aligned}\int_0^t dY &= \int_0^t \frac{1}{2} e^W \, d\tau + \int_0^t e^W \, dW. \\ \int_0^t e^W \, dW &= Y_t - Y_0 - \int_0^t \frac{1}{2} e^{W_\tau} \, d\tau \\ &= e^{W_t} - 1 - \frac{1}{2} \int_0^t e^{W_\tau} \, d\tau.\end{aligned}$$

7. (a)
$$\begin{aligned}dX &= \mu \, dt + \sigma \, dW. \\ df &= (X + \mu t) \, dt + \sigma t \, dW. \\ dg &= (X^2 + 2\mu t X + \sigma^2 t) \, dt + 2\sigma t X \, dW.\end{aligned}$$

(b) Let $0 = t_0 < t_1 < t_2 < \cdots < t_n = t$,

$$A_t = \frac{1}{t} \int_0^t X_\tau \, d\tau = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^n X_{t_j}.$$

As all X_{t_j} follow normal, A_t also follows normal.

$$\begin{aligned}\mathrm{E}A_t &= \lim_{n \to \infty} \frac{1}{n} \sum_{j=1}^n \mathrm{E}X_{t_j} \\ &= \lim_{n \to \infty} \frac{1}{n} \sum_{j=1}^n (X(0) + \mu t_j) \\ &= \int_0^t (X(0) + \mu \tau) \, d\tau \\ &= X(0) + \frac{\mu t}{2}.\end{aligned}$$

$$\begin{aligned}\mathrm{Var}A_t &= \mathrm{Var}\left[\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^n X_{t_j} \right] \\ &= \lim_{n \to \infty} \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \mathrm{Cov}(X_{t_i} X_{t_j}) \\ &= \lim_{n \to \infty} \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \sigma^2 \min(t_i, t_j) \\ &= \sigma^2 \frac{1}{t^2} \int_0^t \int_0^t \min(\rho, \tau) \, d\rho \, d\tau \\ &= \sigma^2 \frac{1}{t^2} \frac{t^3}{3} \\ &= \frac{\sigma^2 t}{3}.\end{aligned}$$

(c)
$$\begin{aligned}\mathrm{Cov}(X_t, A_t) &= \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \mathrm{Cov}(X_t, X_{t_j}) \\ &= \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \sigma^2 t_j \\ &= \sigma^2 \frac{1}{t}\int_0^t \tau\, d\tau \\ &= \frac{\sigma^2 t}{2}.\end{aligned}$$

11.3 CHAPTER 3

1. (a) Long one unit of call option and short x units of stock.
When the share price goes up: $(75-65)^+ - 75x = 10 - 75x$.
Otherwise, the share price goes down: $(50-65)^+ - 50x- = -50x$.
Therefore a riskless portfolio is found by, $10 - 75x = -50x$, $x = 0.4$.

   (b) Substitute $x = 0.4$. We consider the following cases.
When the share price goes up: $(75-65)^+ - 75(0.4) = -20$. Otherwise, the share price goes down: $(50-65)^+ - 50(0.4) = -20$.
The value of portfolio at maturity $= -20$.

   (c) Expected value $= 0.7(75-65)^+ + 0.3(50-65)^+ = 7$.

   (d) The discount factor $= \frac{60}{75(0.7)+50(0.3)} = 8/9$.
Reasonable price $= 7(8/9) = 6.222$.

2. (a)
$$\begin{aligned}c_2 &= e^{-rT}\mathrm{E}[\text{payoff}] \\ &= e^{-rT}[p^2 c_{uu} + 2p(1-p)c_{ud} + (1-p)^2 c_{dd}].\end{aligned}$$

   (b) We prove this by induction. By (a), the statement

$$c_n = e^{-rT}\sum_{j=0}^{n}\left\{{}_nC_j q^j (1-q)^{n-j}\max\left(Su^j d^{n-j} - K, 0\right)\right\}$$

is true for $n = 2$. Suppose this is true for $n = k$ and all T, i.e.,

$$c_k = e^{-rT}\sum_{j=0}^{k} {}_nC_j q^j (1-q)^{n-j}\max(Su^j d^{n-j} - K, 0).$$

For $n = k+1$, by the binomial assumption,

$$c_{k+1} = e^{-rt}[qc_k^{(u)} + (1-q)c_k^{(d)}],$$

where, $c_k^{(u)}$ and $c_k^{(d)}$ are the call price that the stock price goes up and goes down respectively, i.e.,

$$c_k^{(u)} = e^{-r(T-t)} \sum_{j=0}^{k} {}_kC_j q^j (1-q)^{k-j} \max((Su)u^j d^{k-j} - K, 0),$$

$$c_k^{(d)} = e^{-r(T-t)} \sum_{j=0}^{k} {}_kC_j q^j (1-q)^{k-j} \max((Sd)u^j d^{k-j} - K, 0).$$

$$\begin{aligned}
c_{k+1} &= e^{-rT} \sum_{j=0}^{k} {}_kC_j q^{j+1} (1-q)^{k-j} \max(Su^{j+1} d^{k-j} - K, 0) \\
&\quad + e^{-rT} \sum_{j=0}^{k} {}_kC_j q^j (1-q)^{(k+1)-j} \max(Su^j d^{(k+1)-j} - K, 0) \\
&= e^{-rT} \sum_{j=1}^{k+1} {}_kC_{j-1} q^j (1-q)^{(k+1)-j} \max(Su^j d^{(k+1)-j} - K, 0) \\
&\quad + e^{-rT} \sum_{j=0}^{k} {}_kC_j q^j (1-q)^{(k+1)-j} \max(Su^j d^{(k+1)-j} - K, 0) \\
&= e^{-rT} \sum_{j=1}^{k} ({}_kC_{j-1} + {}_kC_j) q^j (1-q)^{(k+1)-j} \max(Su^j d^{(k+1)-j} - K, 0) \\
&\quad + e^{-rT} q^{k+1} \max(Su^{k+1} - K, 0) + e^{-rT} (1-q)^{k+1} \max(Sd^{k+1} - K, 0) \\
&= e^{-rT} \sum_{j=0}^{k+1} {}_{k+1}C_j q^j (1-q)^{(k+1)-j} \max(Su^j d^{(k+1)-j} - K, 0).
\end{aligned}$$

This statement is also true for $n = k+1$, the proof is completed.

(c) Using normal approximation,

$${}_nC_x q^j (1-q)^{n-x} \to \frac{1}{\sqrt{2n\pi q(1-q)}} \exp\left\{ -\frac{(x-nq)^2}{2nq(1-q)} \right\}, \text{ as } n \to \infty.$$

$$\lim_{n\to\infty} c_k = \lim_{n\to\infty} e^{-rT} \sum_{j=0}^{n} {}_nC_j q^j (1-q)^{n-j} \max(Su^j d^{n-j} - K, 0)$$

$$= e^{-rT} \lim_{n\to\infty} \sum_{j=0}^{n} \frac{1}{\sqrt{2n\pi q(1-q)}} \exp\{ -\frac{(j-nq)^2}{2nq(1-q)} \} (Se^{\sigma\sqrt{\delta t}(2j-n)} - K)^+.$$

Let $Z_j = \frac{j-nq}{\sqrt{nq(1-q)}}$, $j = 0, 1, \cdots, n$, and $\delta Z_j = Z_j - Z_{j-1} = \frac{1}{\sqrt{nq(1-q)}}$, we have,

$$\lim_{n\to\infty} c_k = e^{-rT} \lim_{n\to\infty} \sum_{j=0}^{n} \frac{\delta Z_j}{\sqrt{2\pi}} \exp\{ -\frac{Z_j^2}{2} \} (Se^{\sigma X_n} - K)^+,$$

where $X_n = \sqrt{\delta t}[2Z_j\sqrt{nq(1-q)} + n(2q-1)]$.

$$\begin{aligned}\lim_{n\to\infty} X_n &= \lim_{n\to\infty} \sqrt{\delta t}[2Z_j\sqrt{nq(1-q)} + n(2q-1)] \\ &= \lim_{n\to\infty} [2\sqrt{n\,\delta t}Z_j\sqrt{q(1-q)} + n\sqrt{\delta t}(2q-1)] \\ &= 2\sqrt{T}Z_j \lim_{n\to\infty} \sqrt{q(1-q)} + T \lim_{n\to\infty} \frac{2q-1}{\sqrt{\delta t}}.\end{aligned}$$

Recall that $q = \frac{e^{r\delta t}-d}{u-d}$, $u = e^{\sigma\sqrt{\delta t}}$, $e^{-\sigma\sqrt{\delta t}}$, we have,

$$\lim_{n\to\infty} q = \lim_{\sqrt{\delta t}\to 0} \frac{e^{r\delta t}-d}{u-d} = \lim_{\sqrt{\delta t}\to 0} \frac{2r\sqrt{\delta t}e^{r\delta t} + \sigma d}{\sigma u + \sigma d} = 1/2,$$

$$\lim_{n\to\infty} \frac{2q-1}{\sqrt{\delta t}} = \lim_{\sqrt{\delta t}\to 0} \frac{2e^{r\delta t}-u-d}{\sqrt{\delta t}(u-d)} = \lim_{\sqrt{\delta t}\to 0} \frac{4r\sqrt{\delta t}e^{r\delta t} - \sigma u + \sigma d}{(u-d) + \sqrt{\delta t}(\sigma u + \sigma d)}$$

$$= \lim_{\sqrt{\delta t}\to 0} \frac{(8r^2\delta t + 4r)e^{r\delta t} - \sigma^2 u - \sigma^2 d}{(\sigma u + \sigma d) + (\sigma u + \sigma d) + \sqrt{\delta t}(\sigma^2 u - \sigma^2 d)} = \frac{4r - 2\sigma^2}{4\sigma}.$$

Hence, $\lim_{n\to\infty} X_n = \sqrt{T}Z_j + T(r/\sigma - \sigma/2)$.

$$\begin{aligned}&\lim_{n\to\infty} c_k \\ &= \lim_{n\to\infty} e^{-rT} \sum_{j=0}^{n} \frac{\delta Z_j}{\sqrt{2\pi}} \exp\{-\frac{Z^2}{2}\}(Se^{(r-\sigma^2/2)T+\sigma\sqrt{T}Z_j} - K)^+ \\ &= e^{-rT} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\frac{-Z^2}{2}} (Se^{(r-\sigma^2/2)T+\sigma\sqrt{T}Z} - K)^+ \, dZ \\ &= e^{-rT}\mathrm{E}\left[\max(Se^{(r-\sigma^2/2)T+\sigma\sqrt{T}Z} - K, 0)\right] \\ &= \text{call price in Black–Scholes.}\end{aligned}$$

3. Left-hand side − Right-hand side

$$\begin{aligned}&= p + S - c - Ke^{-rT} \\ &= e^{-rT}\mathrm{E}[(K-S_T)^+] + S - e^{-rT}\mathrm{E}[(S_T-K)^+] - Ke^{-rT} \\ &= e^{-rT}\mathrm{E}[(K-S_T)^+ - (S_T-K)^+] + S - Ke^{-rT} \\ &= e^{-rT}\mathrm{E}[K-S_T] + S - Ke^{-rT} \\ &= e^{-rT}K - e^{-rT}\mathrm{E}[S_T] + S - Ke^{-rT} \\ &= e^{-rT}K - S + S - Ke^{-rT} \\ &= 0.\end{aligned}$$

4. We know that $\log G_T \sim \mathrm{N}(\log(S_0) + (r-\sigma^2/2)T/2, \sigma^2 T/3)$. By Lemma 3.1,

$$\text{price} = e^{-rT}[\mathrm{E}(G_T)\Phi(d_1) - K\Phi(d_2)],$$

where

$$\mathrm{E}(G_T) = S_0 \exp\{(r - \sigma^2/2)T/2 + \frac{1}{2}\sigma^2 T/3\} = S_0 \exp\{(r - \sigma^2/6)T/2\},$$

$$d_1 = \frac{\log(S_0/K) + (r - \sigma^2/2)T/2 + \sigma^2 T/3}{\sigma\sqrt{T/3}} = \frac{\log(S_0/K) + (r + \sigma^2/6)T/2}{\sigma\sqrt{T/3}},$$

$$d_2 = \frac{\log(S_0/K) + (r - \sigma^2/2)T/2}{\sigma\sqrt{T/3}}.$$

5. Let X_τ be a stochastic process that follows the dynamic $dX = \mu(\tau, X)\,d\tau + \sigma(\tau, X)\,dW$, $X_t = x$. Applying Itô's lemma,

$$df = \left[\frac{\partial f}{\partial t} + \frac{1}{2}\sigma^2(t,X)\frac{\partial^2 f}{\partial X^2} + \mu(t,X)\frac{\partial f}{\partial X}\right] dt + \sigma(t,X)\frac{\partial f}{\partial X}\,dW.$$

As $\frac{\partial f}{\partial t} + \frac{1}{2}\sigma^2(t,x)\frac{\partial^2 f}{\partial x^2} + \mu(t,x)\frac{\partial f}{\partial x} + a(t,x)f = 0$,

$$\begin{aligned}
df &= -a(t,X)f\,dt + \sigma(t,X)\frac{\partial f}{\partial X}\,dW \\
df + a(t,X)f\,dt &= \sigma(t,X)\frac{\partial f}{\partial X}\,dW \\
e^{\int_t^T -a(\tau,X)\,d\tau}\, d[e^{\int_t^T a(\tau,X)\,d\tau} f] &= \sigma(t,X)\frac{\partial f}{\partial X}\,dW \\
\int_t^T d[e^{\int_t^T a(\tau,X)\,d\tau} f] &= \int_t^T e^{\int_t^T a(\tau,X)\,d\tau}\sigma(t,X)\frac{\partial f}{\partial X}\,dW.
\end{aligned}$$

The right-hand side is a Gaussian process, then the expectation is zero,

$$\begin{aligned}
\mathrm{E}\left[e^{\int_t^T a(\tau,X)\,d\tau} f(T,X_T) - e^{\int_t^t a(\tau,X)\,d\tau} f(t,X_t)\right] &= 0 \\
\mathrm{E}\left[e^{\int_t^T a(\tau,X)\,d\tau} f(T,X_T)\right] &= f(t,x),
\end{aligned}$$

where $f(T,X_T) = F(X_T)$ and $dX = \mu(\tau,X)\,d\tau + \sigma(\tau,X)\,dW$, $X_t = x$.

6. (a) Follow the proof of Theorem 3.1, and substitute r and σ by $r(t)$ and $\sigma(t)$ respectively.

   (b) Substitute $a(t,x) = -r(t)$, $\mu(t,x) = r(t)S_t$, $\sigma(t,x) = \sigma(t)S_t$, and $F(X_T) = F(S_T) = \max(S_T - K, 0)$ into Problem 5.

   (c) Let S'_T be a stochastic process following the dynamic $dS'_\tau = \bar{r}S_\tau\,d\tau + \bar{\sigma}S_\tau\,dW$, $S_t = S'_t = S$. By virtue of Itô's lemma, we have,

$$\begin{aligned}
d\log S_\tau &= \left(r(\tau) - \sigma^2(\tau)/2\right)\,d\tau + \sigma(\tau)\,dW_\tau, \\
d\log S'_\tau &= \left(\bar{r} - \bar{\sigma}^2/2\right)\,d\tau + \bar{\sigma}\,dW_\tau.
\end{aligned}$$

Both $\log S_T$ and $\log S'_T$ are normal, and

$$\begin{aligned}
\mathrm{E}[\log S_T] &= \mathrm{E}[\log S_t] + \int_t^T \left(r(\tau) - \sigma^2(\tau)/2\right) d\tau \\
&= \mathrm{E}[\log S'_t] + (T-t)\bar{r} \;=\; \mathrm{E}[\log S'_T], \\
\mathrm{Var}[\log S_T] &= \int_t^T \sigma^2(\tau)\, d\tau = (T-t)\bar{\sigma}^2 = \mathrm{Var}[\log S'_T].
\end{aligned}$$

Therefore, $\log S_T$ and $\log S'_T$ share the same distribution and S_T and S'_T have the same distribution.

$$\begin{aligned}
f(t,S) &= e^{-\int_t^T r(\tau)\,d\tau} \hat{\mathrm{E}}[\max(S_T - K, 0)] \\
&= e^{-\int_t^T r(\tau)\,d\tau} \hat{\mathrm{E}}[\max(S'_T - K, 0)] \\
&= C_{BS}(t, S_t | r = \bar{r}, \sigma = \bar{\sigma}).
\end{aligned}$$

7. (a)
$$\begin{aligned}
dS &= rS\,dt + \sigma S\,dW. \\
d\log S &= (r - \sigma^2/2)\,dt + \sigma\,dW. \\
\log S_t &\sim \mathrm{N}(\log S_\tau + (r - \sigma^2/2)(t-\tau), \sigma^2(t-\tau)).
\end{aligned}$$

$$\begin{aligned}
&\mathrm{E}^p(X(t)|X(s), s \le \tau) \\
=\;& e^{-rt}\,\mathrm{E}^p(S(t)|X(s), s \le \tau) \\
=\;& e^{-rt} \exp\{\log S_\tau + (r - \sigma^2/2)(t-\tau) + \frac{1}{2}\sigma^2(t-\tau)\} \\
=\;& S_\tau e^{-r\tau} \\
=\;& X(\tau).
\end{aligned}$$

(b)
$$\begin{aligned}
&\mathrm{E}(C(t, S_t; T)e^{r(T-t)}|X(s), s \le \tau) \\
=\;& \mathrm{E}(\mathrm{E}((S_T - K)^+|X(s), s \le t)|X(s), s \le \tau) \\
=\;& \mathrm{E}((S_T - K)^+|X(s), s \le \tau), \quad (\tau < t) \\
=\;& C(\tau, S_\tau; T)e^{r(T-\tau)}.
\end{aligned}$$

11.4 CHAPTER 4

1. (a) $F(x) = P(X \le [x]) = \sum_{j=1}^{[x]} \frac{1}{j(j+1)} = \sum_{j=1}^{[x]} \left(\frac{1}{j} - \frac{1}{j+1}\right) = 1 - \frac{1}{[x]+1}$.

Generate $Y \sim U(0,1)$, then $X = F^{-1}(Y) = \left[\frac{Y}{1-Y}\right] + 1$ has the desired distribution.

(b) Since an analytical formula of the c.d.f is not readily available, we resort to an alternative way. First compute a sufficiently long

sequence $\{f_i\}_{i=0,1,\cdots}$, where $f_0 = 0$, $f_i = F(i) = P(X \le i) = \sum_{j=0}^{i} {}_{(n+j-1)}C_j(1-p)^j p^n$, and set the last element to be 1. Then generate $y \sim U(0,1)$, and search the interval $[f_i, f_{i+1})$ in which y lies. Next set $X = i$, and X has the desired distribution.

2. (a) $X = [10U(1-U)]$, as $U \in [0,1]$, then $10U(1-U) \in [0, 2.5]$.
If $X = 0, U = \{0,1\}$,
if $X = 1, U = \{\frac{1}{2} - \sqrt{\frac{3}{20}}, \frac{1}{2} + \sqrt{\frac{3}{20}}\}$,
if $X = 2, U = \{\frac{1}{2} - \sqrt{\frac{1}{20}}, \frac{1}{2} + \sqrt{\frac{1}{20}}\}$.

$P(X=0) = P(10U(1-U) < 1) = \left[(\frac{1}{2} - \sqrt{\frac{3}{20}}) - 0\right] + \left[1 - (\frac{1}{2} + \sqrt{\frac{3}{20}})\right]$
$= 1 - \sqrt{\frac{3}{5}}$.
$P(X=1) = P(1 \le 10U(1-U) < 2)$
$= \left[(\frac{1}{2} - \sqrt{\frac{3}{20}}) - (\frac{1}{2} - \sqrt{\frac{1}{20}})\right] + \left[(\frac{1}{2} + \sqrt{\frac{1}{20}}) - (\frac{1}{2} + \sqrt{\frac{3}{20}})\right]$
$= \sqrt{\frac{3}{5}} - \sqrt{\frac{1}{5}}$.
$P(X=2) = P(10U(1-U) \ge 2) = (\frac{1}{2} + \sqrt{\frac{1}{20}}) - (\frac{1}{2} - \sqrt{\frac{1}{20}}) = \sqrt{\frac{1}{5}}$.

(b) $Y = [1/U]$, as $U \in [0,1]$, then $1/U \in [1, \infty)$.

$$\begin{aligned} P(X=k) &= P(k \le \frac{1}{U} < k+1) = P(\frac{1}{k} \ge U > \frac{1}{k+1}) \\ &= \frac{1}{k} - \frac{1}{k+1} = \frac{1}{k(k+1)}, \quad k = 1, 2, 3, \cdots. \end{aligned}$$

(c) $Z = (B-3)^2$, $Z = \{0, 1, 4, 9\}$,

$$\begin{aligned} P(Z=0) &= P(B=3) &&= {}_5C_3 0.5^5 = 10/32, \\ P(Z=1) &= P(B=2 \text{ or } 4) &&= {}_5C_2 0.5^5 + {}_5C_4 0.5^5 = 15/32, \\ P(Z=4) &= P(B=1 \text{ or } 5) &&= {}_5C_1 0.5^5 + {}_5C_5 0.5^5 = 6/32, \\ P(Z=9) &= P(B=0) &&= {}_5C_0 0.5^5 = 1/32. \end{aligned}$$

3. (a) As the functions in (a), (b) are monotonic,

$$f_X(x) = \left|\frac{\partial U}{\partial X}\right| f_U(U) = \left|(-\frac{10}{U})^{-1}\right| f_U(U) = \frac{U}{10} = \frac{1}{10} e^{-\frac{x-5}{10}}.$$

(b)

$$f_X(x) = \left|(2\pi \sec^2(\pi U))^{-1}\right| f_U(U) = \frac{1}{2\pi\sqrt{1 + \frac{(x-10)^2}{4}}}.$$

(c) Suppose $W \in [w, w+\Delta)$ and $nU \in [k, k+1)$, $k = 0, 1, \cdots, n-1$, then $W = nU - [nU] = nU - k \in [w, w+\Delta)$, $U \in [\frac{w+k}{n}, \frac{w+k+\Delta}{n})$.

$$\begin{aligned} f_W(w) &= \lim_{\Delta \to 0} \frac{1}{\Delta} P(W \in [w, w+\Delta)) \\ &= \lim_{\Delta \to 0} \sum_{k=0}^{n-1} \frac{1}{\Delta} P\left(U \in [\frac{w+k}{n}, \frac{w+k+\Delta}{n})\right) \\ &= \lim_{\Delta \to 0} \sum_{k=0}^{n-1} \frac{1}{\Delta} \frac{\Delta}{n} = 1. \end{aligned}$$

$$P(W \le w, I = i) = P\left(W \le w, U[\frac{i}{n}, \frac{i+1}{n})\right) = P\left(U \in [\frac{i}{n}, \frac{w+i}{n})\right) = \frac{w}{n}.$$

Therefore, $P(W \le w, I = i) = \frac{w}{n} = (w)\frac{1}{n} = P(W \le w)P(I = i)$.

4. (a) Generate $U \sim U(0,1)$, and accept $X = [7U] + 1$ with probability $P(X)/P(X = 6) = P(X)/0.17$.

(c) Probability of acceptance $= \sum_{i=1}^{7} \frac{P(X_i)}{0.17} \frac{1}{7} = \frac{1}{1.19}$.
Expected number of acceptance $= 1000\frac{1}{1.19} \simeq 840$.

5. (a) The c.d.f. $F(x) = \frac{x^2+x}{2}$,

$$X = F^{-1}(U) = \sqrt{2U + \frac{1}{4}} - \frac{1}{2}, \quad \text{where} \quad U \sim U(0,1).$$

(b) Generate $U \sim U(0,1)$, and accept $X = U$ with probability $\frac{2f(X)}{3}$.
Probability of acceptance $= \int_0^1 \frac{2f(X)}{3}\, dx = \frac{2}{3}$.

6. (a) $c = \max\frac{f(x)}{g(x)} = \frac{2}{\sqrt{2\pi}}\max(e^{-\frac{x^2}{2}+x}) = \frac{2}{\sqrt{2\pi}}\max(e^{1-\frac{(x-1)^2}{2}}) = \frac{2}{\sqrt{2\pi}}\, e^{1/2}$.

(b) Generate $x \sim g(x)$, and accept x with probability $\sqrt{\frac{\pi}{2}}e^{-\frac{1}{2}}\frac{f(x)}{g(x)}$.

11.5 CHAPTER 5

4. $$\mathrm{E}(\bar{X}_n) = \mathrm{E}(\frac{1}{n}\sum_{i=1}^{n} X_i) = (\frac{1}{n}\sum_{i=1}^{n} \mathrm{E}X_i) = (\frac{1}{n}\sum_{i=1}^{n} \theta) = \theta.$$

$$\mathrm{Var}(\bar{X}_n) = \mathrm{Var}(\frac{1}{n}\sum_{i=1}^{n} X_i) = (\frac{1}{n^2}\sum_{i=1}^{n} \mathrm{Var}X_i) = (\frac{1}{n}\sum_{i=1}^{n} \sigma) = \frac{\sigma}{n}.$$

5. $$\begin{aligned}
\mathrm{E}(S^2) &= \mathrm{E}\left(\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X}_n)^2\right) \\
&= \frac{1}{n-1}\mathrm{E}\sum_{i=1}^{n}[(X_i-\theta)-(\bar{X}-\theta)]^2 \\
&= \frac{1}{n-1}\mathrm{E}\sum_{i=1}^{n}[(X_i-\theta)^2-2(X_i-\theta)(\bar{X}-\theta)+(\bar{X}-\theta)^2] \\
&= \frac{1}{n-1}[\mathrm{E}\sum_{i=1}^{n}(X_i-\theta)^2-\mathrm{E}\sum_{i=1}^{n}2(X_i-\theta)(\bar{X}-\theta)+\mathrm{E}\sum_{i=1}^{n}(\bar{X}-\theta)^2] \\
&= \frac{1}{n-1}[\sum_{i=1}^{n}\mathrm{E}(X_i-\theta)^2-2\mathrm{E}n(\bar{X}-\theta)^2+n\mathrm{E}(\bar{X}-\theta)^2] \\
&= \frac{1}{n-1}[n\sigma^2-2n\frac{\sigma^2}{n}+n\frac{\sigma^2}{n}] \\
&= \sigma^2.
\end{aligned}$$

$$\begin{aligned}
& S_{j+1}^2-(1-\frac{1}{j})S_j^2 \\
= \; & \frac{1}{j}[\sum_{i=1}^{j+1}(X_i-\bar{X}_{j+1})^2-\sum_{i=1}^{j}(X_i-\bar{X}_j)^2] \\
= \; & \frac{1}{j}[\sum_{i=1}^{j}[(X_i-\bar{X}_{j+1})^2-(X_i-\bar{X}_j)^2]+\frac{1}{j}(X_{j+1}-\bar{X}_{j+1})^2 \\
= \; & \frac{1}{j}[\sum_{i=1}^{j}[\bar{X}_{j+1}^2-\bar{X}_j^2-2(\bar{X}_{j+1}-\bar{X}_j)X_i]+\frac{1}{j}(X_{j+1}-\bar{X}_{j+1})^2 \\
= \; & \bar{X}_{j+1}^2-\bar{X}_j^2-2(\bar{X}_{j+1}-\bar{X}_j)\bar{X}_j+\frac{1}{j}(X_{j+1}-\bar{X}_{j+1})^2 \\
= \; & \bar{X}_{j+1}^2-2\bar{X}_{j+1}\bar{X}_j+\bar{X}_j^2+\frac{1}{j}[j(\bar{X}_{j+1}-\bar{X}_j)]^2 \\
= \; & (j+1)(\bar{X}_{j+1}-\bar{X}_j)^2.
\end{aligned}$$

7. (a) *Gamma* measures the curvature of the portfolio.

   (b) Let $S=e^y$. Then

$$\begin{aligned}
\frac{\partial^2}{\partial S^2}\phi(x|\log S) &= \frac{\partial}{\partial e^y}\left[\frac{\partial}{\partial e^y}\phi(x|y)\right] \\
&= \frac{1}{e^y}\frac{\partial}{\partial y}\left[\frac{\partial}{e^y\partial y}\phi(x|y)\right] \\
&= \frac{1}{S}\frac{\partial}{\partial y}\left[e^{-y}\phi(x|y)\frac{x-y-\nu T}{\sigma^2 T}\right]
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{S}\left[-e^{-y}\phi(x|y)\frac{x-y-\nu T}{\sigma^2 T} + e^{-y}\phi(x|y)\frac{(x-y-\nu T)^2}{\sigma^4 T^2} + e^{-y}\phi(x|y)\frac{-1}{\sigma^2 T}\right] \\
&= \frac{\phi(x|y)}{S^2\sigma^2 T}\left[-(x-y-\nu T) + \frac{(x-y-\nu T)^2}{\sigma^2 T} - 1\right] \\
&= \frac{\phi(x|y)}{S^2\sigma^2 T}\left[-\sigma W_T + \frac{W_T^2}{T} - 1\right].
\end{aligned}
$$

$$
\begin{aligned}
gamma &= e^{-rT}\int_{-\infty}^{\infty} F(e^x)\frac{\partial^2}{\partial S^2}\phi(x|\log S)\,dx \\
&= e^{-rT}\int_{-\infty}^{\infty} F(S)\frac{\phi(x|y)}{S^2\sigma^2 T}\left[-\sigma W_T + \frac{W_T^2}{T} - 1\right]dx \\
&= e^{-rT}\mathrm{E}\left[\frac{F(S)}{S^2\sigma^2 T}\left(\frac{W_T^2}{T} - \sigma W_T - 1\right)\right].
\end{aligned}
$$

(c) Generate $W_T \sim \mathrm{N}(0,T)$, let $S_T = S_0 e^{(r-\sigma^2/2)+\sigma W_T}$ and $F(S_T) = (S_T - K)^+$, then take the expectation given by (b).

11.6 CHAPTER 6

1. $F(y) = P(Y \le y) = P(U \le \frac{y-a}{b-a}) = \begin{cases} 0, & y < a, \\ \int_0^{\frac{y-a}{b-a}} 1\,d\zeta = \frac{y-a}{b-a}, & a \le y < b, \\ 1, & y \ge b. \end{cases}$

$$f(y) = \frac{dF(y)}{dy} = \begin{cases} \frac{1}{b-a}, & a \le y < b, \\ 0, & \text{elsewhere.} \end{cases}$$

2.
$$
\begin{aligned}
&\mathrm{Var}(X - \hat{b}Y) \\
&= \mathrm{Var}(X) - 2\hat{b}\mathrm{Cov}(X,Y) + \hat{b}^2\mathrm{Var}(Y) \\
&= \mathrm{Var}(X) - 2\hat{b}[\mathrm{Cov}(a,Y) + \hat{b}\mathrm{Cov}(Y,Y) + \mathrm{Cov}(\epsilon,Y)] + \hat{b}^2\mathrm{Var}(Y) \\
&= \mathrm{Var}(X) - \hat{b}^2\mathrm{Var}(Y).
\end{aligned}
$$

3. As $f(u) = e^{u^2}$ is monotone on $[0,1]$, therefore e^{U^2} and $e^{(1-U)^2}$ are negative correlated, and

$$\mathrm{Var}\left[e^{U^2} + e^{(1-U)^2}\right] - \mathrm{Var}\left[e^{U_1^2} + e^{U_2^2}\right] = 2\mathrm{Cov}\left[e^{U^2}, e^{(1-U)^2}\right] < 0.$$

4. (a) $\hat{\theta} = \frac{1}{n}\sum_{i=1}^n 4X_i^3$.

(b) $\hat{\theta} = \frac{1}{2n}\sum_{i=1}^n 4((1-X_i)^3 + X_i^3)$.

(c) $\hat{\theta} = \frac{1}{kn}\sum_{i=1}^n 4(\sum_{j=0}^{k-1}(\frac{X_i+j}{k})^3)$.

(d) Let control variate be x, $\mathrm{E}[x] = 1/2$,
$\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} 4x^3 - (\hat{b}x - 1/2)$,
where $\hat{b}$ is the least square estimate of $4x^3 = a + bx + \epsilon$.

5. (a) $f(X) = (X-2)^+$.
   (b) It is known that random variable $X - 2$ conditional on $\{X \geq 2\}$ follows Exp(1), hence generate $X = -\log(U) + 2$, where $U \sim U(0,1)$.
   (c) Generate $X_k = -\log(\frac{U+k}{4}) + 2$, $k = 0, 1, 2, 3$.
   (d) $\mathrm{E}(X-2) = -1$,

$$(\hat{X}-2)^+_{CV} = (\bar{X}-2)^+ - \hat{b}((\bar{X}-2) - \mu_{(X-2)}) = (\bar{X}-2)^+ - \hat{b}(\bar{X}+1).$$

6. The theoretical price $= S_0 N(d_1) - Ke^{-rT}N(d_2) = 6.8050$.

7. $$\mathrm{E}S_T = S_0 e^{\nu T}\mathrm{E}(e^{\sigma\sqrt{T}Z}) = S_0 e^{\nu T}(e^{\sigma^2 T/2}) = S_0 e^{rT},$$
$$\mathrm{Var}[S_T] = S_0^2 e^{2\nu T}\mathrm{Var}(e^{\sigma\sqrt{T}Z}) = S_0^2 e^{2\nu T}e^{\sigma^2 T}(e^{\sigma^2 T}-1) = S_0^2 e^{2rT}(e^{\sigma^2 T}-1).$$

8. Generate $Z_i \sim \mathrm{N}(0,1)$ and let

$$S_T^{(i)} = S_0 e^{(r-\sigma^2/2)T+\sigma\sqrt{T}Z_i} \quad \text{and} \quad \tilde{S}_T^{(i)} = S_0 e^{(r-\sigma^2/2)T-\sigma\sqrt{T}Z_i}.$$

$$\hat{\mathrm{E}}(h(S_T)) = \sum_{i=1}^{n} \frac{h(S_T^i) + h(\tilde{S}_T^i)}{2n}.$$

11.7 CHAPTER 7

1. By Section 7.4, we know that $G_T = \exp\left[\frac{1}{T}\int_0^T \log S(t)\,dt\right]$, and

$$\log(G_T) \sim \mathrm{N}\left(\frac{t}{T}\log\frac{G_t}{S_t} + \log S_t + \frac{(r-\sigma^2/2)(T-t)^2}{2T}, \frac{\sigma^2(T-t)^3}{3T^2}\right).$$

Applying Lemma 3.1, we have,

$$e^{-r(T-t)}\mathrm{E}(G_T - K)^+ = e^{-r(T-t)}\mathrm{E}(G_T)\Phi(d_1) - e^{-r(T-t)}K\Phi(d_2),$$

where

$$\begin{aligned}
& e^{-r(T-t)}\mathrm{E}G_T \\
= \; & e^{-r(T-t)}\exp\left\{\frac{t}{T}\log\frac{G_t}{S} + \log S_t + \frac{(T-t)^2}{T}\left(\frac{r-\sigma^2/2}{2}\right) + \frac{1}{2}\frac{\sigma^2(T-t)^3}{3T^2}\right\} \\
= \; & S\left(\frac{G_t}{S_t}\right)^{\frac{t}{T}} e^{-r(T-t)}\exp\left\{\frac{(T-t)^2}{T}\left(\frac{r-\sigma^2/2}{2}\right) + \frac{1}{2}\frac{\sigma^2(T-t)^3}{3T^2}\right\} \\
= \; & S\left(\frac{G_t}{S_t}\right)^{\frac{t}{T}} e^{R(t;T)},
\end{aligned}$$

$$
\begin{aligned}
d_1 &= \left(\frac{\sigma^2(T-t)^3}{3T^2}\right)^{-\frac{1}{2}} \left(\log\frac{S_t}{K} + \frac{t}{T}\log\frac{G_t}{S_t} + \frac{(r-\sigma^2/2)(T-t)^2}{2T} + \frac{\sigma^2(T-t)^3}{3T^2}\right) \\
&= \left(\frac{\sigma^2(T-t)^3}{3}\right)^{-\frac{1}{2}} \left(T\log\frac{S_t}{K} + t\log\frac{G_t}{S_t} + \frac{1}{2}(r-\frac{\sigma^2}{2})(T-t)^2 + \frac{\sigma^2(T-t)^3}{3T}\right), \\
d_2 &= d1 - \sqrt{\frac{\sigma^2(T-t)^3}{3T^2}}.
\end{aligned}
$$

6. $$
\begin{aligned}
e^{-rT}\mathrm{E}(S_T - S_{t_1})^+ &= e^{-rt_1}\mathrm{E}(e^{-r(T-t_1)}\mathrm{E}((S_T - S_{t_1})^+|S_{t_1})) \\
&= e^{-rt_1}\mathrm{E}(C_{BS}(S_{t_1}, t_1; S_{t_1}, T)).
\end{aligned}
$$

Hence, we first simulate prices S_{t_1}, and compute the corresponding call option prices by Black-Scholes. Then the average would be the simulated price.

11.8 CHAPTER 8

1. Applying Taylor's series expansion, we have,

$$
\begin{aligned}
& f(t+\Delta, x+\Delta_x, y+\Delta_y) - f(t,x,y) \\
= \; & f_t\,\Delta + f_x\,\Delta_x + f_y\,\Delta_y + \frac{1}{2}f_{xx}\,\Delta_x^2 + f_{xy}\,\Delta_x\Delta_y + \frac{1}{2}f_y\,\Delta_y^2 + o(\Delta), \\
& \text{where} \begin{cases} \Delta_x &= a(t,x)\,\Delta + b(t,x)\,\Delta W_1, \\ \Delta_y &= \alpha(t,y)\,\Delta + \beta(t,y)\,\Delta W_2. \end{cases}
\end{aligned}
$$

Taking $\Delta \to 0$, we have,

$$
\begin{aligned}
dx &= a(t,x)\,dt + b(t,x)\,dW_1, \\
dx^2 &= b(t,x)^2\,dt, \\
dy &= \alpha(t,y)\,dt + \beta(t,y)\,dW_2, \\
dy^2 &= \beta(t,y)^2\,dt, \\
dxdy &= b(t,x)\beta(t,y)\lim_{\Delta\to 0}\mathrm{E}(\Delta W_1 \Delta W_2) \\
&= b(t,x)\beta(t,y)\mathrm{E}(dW_1 dW_2) = b(t,x)\beta(t,y)\rho\,dt, \\
o(\Delta) &\to 0.
\end{aligned}
$$

Substitute these into the above equation,

$$
\begin{aligned}
df &= f_t\,dt + f_x(a\,dt + b\,dW_1) + f_y(\alpha\,dt + \beta\,dW_2) \\
&\quad + \frac{1}{2}b^2 f_{xx}\,dt + \rho b\beta f_{xy}\,dt + \frac{1}{2}\beta^2 f_y\,dt \\
&= \left(f_t + af_x + \alpha f_y + \frac{b^2}{2}f_{xx} + \rho b\beta f_{xy} + \frac{\beta^2}{2}f_{yy}\right)dt \\
&\quad + bf_x\,dW_1 + \beta f_y\,dW_2.
\end{aligned}
$$

2. (a) Using the result of Problem 1, we have,

$$\begin{aligned} dX &= \left(rS_1 \frac{1}{S_2} - rS_2 \frac{S_1}{S_2^2} - \rho\sigma_1\sigma_2 S_1 S_2 \frac{1}{S_2^2} + \frac{\sigma_2^2 S_2^2}{2} 2 \frac{S_1}{S_2^3} \right) dt \\ &\quad + \sigma_1 S_1 \frac{1}{S_2} dW_1 - \sigma_2 S_2 \frac{S_1}{S_2^2} dW_2 \\ &= \frac{S_1}{S_2}(-\rho\sigma_1\sigma_2 + \sigma_2^2)\, dt + \frac{S_1}{S_2}(\sigma_1\, dW_1 - \sigma_2\, dW_2) \\ &= \frac{S_1}{S_2}\left(\sigma_1(dW_1 - \rho\sigma_2\, dt) - \sigma_2(dW_2 - \sigma_2\, dt)\right) \\ &= X(\sigma_1\, dW_1^* - \sigma_2\, dW_2^*). \end{aligned}$$

The process X is a martingale under Brown motions $W_1^*(t)$ and $W_2^*(t)$, as there is no drift term.

(b)
$$\begin{aligned} \mathrm{E}\left(\frac{dX}{X}\right) &= \mathrm{E}(\sigma_1\, dW_1 - \sigma_2\, dW_2)^2 \\ &= \sigma_1^2 \mathrm{E}(dW_1)^2 - 2\sigma_1\sigma_2 \mathrm{E}(dW_1\, dW_2) + \sigma_2^2 \mathrm{E}(dW_2)^2 \\ &= (\sigma_1^2 - 2\rho\sigma_1\sigma_2 + \sigma_2^2)\, dt \\ &= \sigma^2\, dt. \end{aligned}$$

Therefore $X(t)$ follows a geometric Brownian motion W^* with volatility σ^2.

(c) By the result of Exercise 3.7, $dX = \sigma X\, dW^*$ has a zero drift under W^*, therefore $U(t, X)$ is a martingale under Brownian motions W_1^* and W_2^*.

3.
$$\begin{aligned} V_{ex} &= S_2(t) U\left(t, \frac{S_1(t)}{S_2(t)}\right) \\ &= S_2(t) \mathrm{E} U\left(T, \frac{S_1(T)}{S_2(T)}\right) \\ &= S_2(t) \mathrm{E}\left(\frac{1}{S_2(T)} V(T, S_1, S_2)\right) \\ &= S_2(t) \mathrm{E}\left(\frac{1}{S_2(T)} (S_1(T) - S_2(T))^+\right) \\ &= S_2(t) \mathrm{E}\left(\left(\frac{S_1(T)}{S_2(T)} - 1\right)^+\right) \\ &= S_2(t)\left(\frac{S_1(t)}{S_2(t)} \Phi(d_1^*) - \Phi(d_2^*)\right) \\ &= S_1(t)\Phi(d_1^*) - S_2(t)\Phi(d_2^*), \end{aligned}$$

where

$$\begin{aligned} d_1^* &= \frac{\log \frac{S_1(t)}{S_2(t)} + \sigma^2(T-t)/2}{\sigma\sqrt{T-t}}, \quad d_2^* = d_1^* - \sigma\sqrt{T-t}, \\ \sigma^2 &= \sigma_1^2 - 2\rho\sigma_1\sigma_2 + \sigma_2^2. \end{aligned}$$

5. (a) We know that arithmetic average is greater than geometric average, therefore,

$$\begin{aligned} \left(\prod_{i=1}^n S_i(T)\right)^{1/n} &\leq \frac{1}{n}\sum_{i=1}^n S_i(T) \\ \left(\prod_{i=1}^n S_i(T)^{1/n} - K\right)^+ &\leq \left(\frac{1}{n}\sum_{i=1}^n S_i(T) - K\right)^+ \\ \mathrm{E}\left(\prod_{i=1}^n S_i(T)^{1/n} - K\right)^+ &\leq \mathrm{E}\left(\frac{1}{n}\sum_{i=1}^n S_i(T) - K\right)^+. \end{aligned}$$

(b) Suppose $dS_i = rS_i\,dt + \sigma_i S_i\,dW_i$, with $\mathrm{E}(dW_i dW_j) = \rho_{ij}\,dt$, for $i, j = 1, 2, \cdots, n$. Then

$$\begin{aligned} d\log S_i &= (r - \sigma_i^2/2)\,dt + \sigma_i\,dW_i \\ \frac{1}{n}\sum_{i=1}^n d\log S_i &= (r - \sum_{i=1}^n \frac{\sigma_i^2}{2n})\,dt + \frac{1}{n}\sum_{i=1}^n \sigma_i\,dW_i \\ &= (r - \frac{1}{2n}\sum_{i=1}^n \sigma_i^2)\,dt + \sigma\,dW, \\ \text{where} \quad \sigma^2 &= \frac{1}{n^2}\sum_{i,j=1}^n \rho_{ij}\sigma_i\sigma_j. \end{aligned}$$

Let $G(t) = \prod_{i=1}^n S_i(t)^{1/n}$, then we have,

$$e^{-rT}\mathrm{E}\,(G(T) - K)^+ = e^{-rT}\mathrm{E}(G(T))\Phi(d_1) - Ke^{-rT}\Phi(d_2),$$

where

$$\begin{aligned} \mathrm{E}(G(T)) &= \exp\left\{G(t) + (r - \frac{1}{2n}\sum_{i=1}^n \sigma_i^2)(T-t) + \frac{1}{2}\sigma^2(T-t)\right\}, \\ d_1 &= \frac{-\log K + G(t) + (r - \frac{1}{2n}\sum_{i=1}^n \sigma_i^2)(T-t) + \sigma^2(T-t)}{\sqrt{\sigma^2(T-t)}} \\ &= \frac{\log \frac{G(t)}{K} + (r + \sigma^2 - \frac{1}{2n}\sum_{i=1}^n \sigma_i^2)(T-t)}{\sigma\sqrt{T-t}}, \\ d_2 &= d_1 - \sigma\sqrt{T-t}. \end{aligned}$$

11.9 CHAPTER 9

1. From Section 9.3, we have,

$$B(0,T) = \lim_{n\to\infty} \mathrm{E}\left\{\exp\left(-\Delta t \sum_{i=1}^{n} r_i\right)\right\}, \quad \text{where}$$

$$\sum_{i=1}^{n} r_i = (r_0-b)\frac{(1-a\Delta t)-(1-a\Delta t)^{n+1}}{a\Delta t} + bn + \sigma\sqrt{\Delta t}\sum_{i=1}^{n}\epsilon_{n-i+1}\frac{1-(1-a\Delta t)^i}{a\Delta t}.$$

$$\begin{aligned}
\lim_{n\to\infty} \mathrm{E}[-\Delta t\sum_{i=1}^{n} r_i] &= \lim_{n\to\infty} -(r_0-b)\frac{(1-a\Delta t)-(1-a\Delta t)^{n+1}}{a} - bT \\
&= -(r_0-b)\frac{\lim_{n\to\infty}(1-a\Delta t) - \lim_{n\to\infty}(1-a\Delta t)^{n+1}}{a} - bT \\
&= -(r_0-b)\frac{1-e^{-aT}}{a} - bT.
\end{aligned}$$

$$\begin{aligned}
\lim_{n\to\infty} \mathrm{Var}[-\Delta t\sum_{i=1}^{n} r_i] &= \lim_{n\to\infty} \frac{\sigma^2\Delta t}{a^2}\left[n - 2(1-a\Delta t)\frac{1-(1-a\Delta t)^n}{a\Delta t}\right. \\
&\quad \left. + (1-a\Delta t)^2\frac{1-(1-a\Delta t)^{2n}}{1-(1-a\Delta t)^2}\right] \\
&= \frac{\sigma^2}{a^2}\left[\lim_{n\to\infty}\Delta t\, n \lim_{n\to\infty} 2\frac{1-a\Delta t}{a}(1-\lim_{n\to\infty}(1-a\Delta t)^n)\right. \\
&\quad \left. + \lim_{n\to\infty}\frac{\Delta t(1-a\Delta t)^2}{1-(1-a\Delta t)^2}(1-\lim_{n\to\infty}(1-a\Delta t)^{2n})\right] \\
&= \frac{\sigma^2}{a^3}\left[aT - 2(1-e^{-aT}) + \frac{1}{2}(1-e^{-aT})\right].
\end{aligned}$$

$$\begin{aligned}
B(0,T) &= \lim_{n\to\infty}\exp\left[\mathrm{E}[-\Delta t\sum_{i=1}^{n} r_i] + \frac{1}{2}\mathrm{Var}[-\Delta t\sum_{i=1}^{n} r_i]\right] \\
&= \exp\left[\lim_{n\to\infty}\mathrm{E}[-\Delta t\sum_{i=1}^{n} r_i] + \frac{1}{2}\lim_{n\to\infty}\mathrm{Var}[-\Delta t\sum_{i=1}^{n} r_i]\right] \\
&= \exp\left[-bT - (r_0-b)\frac{1-e^{-aT}}{a} + \frac{1}{2}\frac{\sigma^2}{a^3}\left[aT - 2(1-e^{-aT}) + \frac{1}{2}(1-e^{-aT})\right]\right] \\
&= \exp\left[-bT - (r_0-b)\frac{1-e^{-aT}}{a} + \frac{\sigma^2}{4a^3}[aT + 4e^{-aT} - 2e^{-aT} - 3]\right].
\end{aligned}$$

2. (a) After disceretization, we have,

$$r_{i+1} = r_i + 0.1(0.05 - r_i)\,\Delta t + 0.3\sqrt{r_i\,\Delta t}\,\epsilon_i, \text{ for } i = 1, 2, \cdots.$$

Using Monle Carlo, we have sample paths of interest rate. For each scenario, we compute $\exp\{-\sum r_i \Delta t\}$, and the average will be the simulated value of discount factor.

(b) After disceretization, we have,

$$r_{i+1}^{CV} = r_i^{CV} + 0.1(0.05 - r_i^{CV})\,\Delta t + \sigma_{Vasicek}\sqrt{\Delta t}\,\epsilon_i, \text{ for } i = 1, 2, \cdots,$$

where $r_0^{CV} = 0.052$ and $\sigma_{Vasicek} = \sigma_{CIR}\sqrt{r_0} = 0.3\sqrt{0.052}$. Compute r_n^{CV} and $\exp\{-\sum r_i^{CV} \Delta t\}$ to be the control variate. Using the least squares estimation, we obtain the control variate expectation.

3. (a) Simulate scenarios of interest rate r, where

$$r_{i+1} = r_i + (a + e^{-bt_i})\,\Delta t + \sigma\sqrt{\Delta t}\,\epsilon_i, \text{ for } i = 1, 2, \cdots.$$

Take $\hat{B}(0,t)$ to be the average of $\exp\{-\sum r_i \Delta t\}$ across different scenarios.

(c)
$$\begin{aligned} r_t - r_0 &= \lim_{n\to\infty} \sum_{i=1}^{n} [(a + e^{-bt_i})\,\Delta t + \sigma\,\Delta W_i] \\ &= \int_0^t (a + e^{-b\tau})\,d\tau + \sigma \int_0^t dW \\ &= \left[a\tau - \frac{e^{-b\tau}}{b}\right]_0^t + \sigma(W_t - W_0) \\ &= at + \frac{1 - e^{-bt}}{b} + \epsilon\sigma\sqrt{t}. \end{aligned}$$

(d) Similar to Exercise 1, we have the formula of $B(0,T)$.

$$\begin{aligned} \int_0^T r_\tau\,d\tau &= \int_0^T \left(r_\tau + a\tau + \frac{1 - e^{-b\tau}}{b} + W_\tau\right) d\tau \\ &= r_0 T + \frac{aT^2}{2} + \frac{T}{b} + \frac{e^{-bT} - 1}{b^2} + \int_0^T W_\tau\,d\tau, \end{aligned}$$

which is a Gaussian process.

$$\begin{aligned} \mathrm{E}\left[-\int_t^T r_\tau\,d\tau\right] &= -\mathrm{E}\left[r_0 T + \frac{aT^2}{2} + \frac{T}{b} + \frac{e^{-bT} - 1}{b^2} + \int_0^T W_\tau\,d\tau\right] \\ &= -r_0 T - \frac{aT^2}{2} - \frac{T}{b} - \frac{e^{-bT} - 1}{b^2}. \\ \mathrm{Var}\left[-\int_t^T r_\tau\,d\tau\right] &= \mathrm{Var}\left[-\int_0^T W_\tau\,d\tau\right] = \frac{\sigma^2 T^3}{3}. \end{aligned}$$

$$\begin{aligned} B(0,T) &= \exp\left\{ \mathrm{E}\left[-\int_t^T r_\tau \, d\tau \right] + \frac{1}{2}\mathrm{Var}\left[-\int_t^T r_\tau \, d\tau \right] \right\} \\ &= \exp\left\{ -r_0 T - \frac{aT^2}{2} - \frac{T}{b} - \frac{e^{-bT}-1}{b^2} + \frac{\sigma^2 T^3}{6} \right\}. \end{aligned}$$

11.10 CHAPTER 10

1. (a)
$$\begin{aligned} L(\sigma^2) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{(X_i-\mu)^2}{2\sigma^2}} \\ &\propto \prod_{i=1}^n \frac{1}{\sigma} e^{-\frac{(X_i-\mu)^2}{2\sigma^2}} \\ &= (\sigma^2)^{-\frac{n}{2}} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (X_i-\mu)^2 \right\}. \end{aligned}$$

   (b) Conditional distribution

$$\begin{aligned} &\propto L(\sigma^2) p(\sigma^2) \\ &\propto (\sigma^2)^{-\frac{n}{2}} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (X_i-\mu)^2 \right\} (\sigma^2)^{-\alpha+1} e^{-\frac{\beta}{\sigma^2}} \\ &\propto (\sigma^2)^{-\frac{n}{2}-\alpha+1} \exp\left\{ -\frac{1}{\sigma^2} [\frac{1}{2} \sum_{i=1}^n (X_i-\mu)^2 + \beta] \right\}. \end{aligned}$$

   The posterior is $IG(\alpha + \frac{n}{2}, \beta + \frac{1}{2}\sum_{i=1}^n (X_i-\mu)^2)$.

2. (a) Normal with know variance:

$$\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2 \frac{1}{2\sigma^2}}.$$

   (b) Normal with know mean:

$$\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} = \frac{e^{-\frac{\mu^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma} e^{-x^2 \frac{1}{2\sigma^2} + x \frac{\mu+\mu}{\sigma^2}}.$$

   (c) Poisson:

$$\frac{e^{-\lambda}\lambda^x}{x!} = \frac{1}{x!} e^{-\lambda} e^{x \log \lambda}.$$

   (d) Binomial:

$${}_nC_r p^x (1-p)^{(n-x)} = {}_nC_r \, e^{x \log p + (n-x) \log(1-p)}.$$

3. Suppose that,

$$L(\theta) = g_L(x)h_L(\theta)\exp\left[\sum t_{L,i}(x)\psi_{L,i}(\theta)\right]$$

$$p(\theta) = g_p(x)h_p(\theta)\exp\left[\sum t_{p,i}(x)\psi_{p,i}(\theta)\right].$$

Then,

$$\begin{aligned}\pi(\theta) &\propto g_L(x)h_L(\theta)\exp\left[\sum t_{L,i}(x)\psi_{L,i}(\theta)\right] g_p(x)h_p(\theta)\exp\left[\sum t_{p,i}(x)\psi_{p,i}(\theta)\right] \\ &\propto (g_L(x)g_p(x))(h_L(\theta)h_p(\theta))\exp\left[\sum t_{L,i}(x)\psi_{L,i} + t_{p,i}(x)\psi_{p,i}(\theta)\right].\end{aligned}$$

6. (a) Let $p(\mu) \propto Ae^{-\frac{(\mu-\mu_1)^2}{2\sigma_1^2}} + Be^{-\frac{(\mu-\mu_2)^2}{2\sigma_1^2}}$ and $L(\mu) \propto Ce^{-\frac{(x-\mu)^2}{2\sigma^2}}$.

$$\begin{aligned}\pi(\mu) &\propto Ce^{-\frac{(x-\mu)^2}{2\sigma^2}}\left(Ae^{-\frac{(\mu-\mu_1)^2}{2\sigma_1^2}} + Be^{-\frac{(\mu-\mu_2)^2}{2\sigma_1^2}}\right) \\ &= ACe^{-\frac{(x-\mu)^2}{2\sigma^2}}e^{-\frac{(\mu-\mu_1)^2}{2\sigma_1^2}} + BCe^{-\frac{(x-\mu)^2}{2\sigma^2}}e^{-\frac{(\mu-\mu_2)^2}{2\sigma_1^2}} \\ &= A'C\exp\left\{-\frac{1}{2}\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_1^2}\right)\left(\mu - \frac{\frac{x}{\sigma^2}+\frac{\mu_1}{\sigma_1^2}}{\frac{1}{\sigma^2}+\frac{1}{\sigma_1^2}}\right)^2\right\} \\ &\quad + B'C\exp\left\{-\frac{1}{2}\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_2^2}\right)\left(\mu - \frac{\frac{x}{\sigma^2}+\frac{\mu_2}{\sigma_2^2}}{\frac{1}{\sigma^2}+\frac{1}{\sigma_2^2}}\right)^2\right\}.\end{aligned}$$

Hence, this is a mixture of two normals.

(b) Let $p(\mu) \propto \sum_{i=1}^{K} A_i e^{-\frac{(\mu-\mu_i)^2}{2\sigma_i^2}}$ and $L(\mu) \propto Ce^{-\frac{(x-\mu)^2}{2\sigma^2}}$

$$\begin{aligned}\pi(\mu) &\propto Ce^{-\frac{(x-\mu)^2}{2\sigma^2}}\sum_{i=1}^{K} A_i e^{-\frac{(\mu-\mu_i)^2}{2\sigma_i^2}} \\ &= \sum_{i=1}^{k} A_i Ce^{-\frac{(x-\mu)^2}{2\sigma^2}}e^{-\frac{(\mu-\mu_i)^2}{2\sigma_i^2}} \\ &= \sum_{i=1}^{k} A_i' C\exp\left\{-\frac{\mu^2}{2}\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_i^2}\right) + \mu\left(\frac{x}{\sigma^2}+\frac{\mu_i}{\sigma_i^2}\right)\right\} \\ &= \sum_{i=1}^{k} A_i'' C\exp\left\{-\frac{1}{2}\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_i^2}\right)\left(\mu - \frac{\frac{x}{\sigma^2}+\frac{\mu_i}{\sigma_i^2}}{\frac{1}{\sigma^2}+\frac{1}{\sigma_i^2}}\right)^2\right\}.\end{aligned}$$

Hence, this is a mixture of normals.

7. Evaluating $\alpha_{i,j} = \min\left\{1, \frac{\pi(j)P(i|j)}{\pi(i)P(j|i)}\right\}$, we have,

$\alpha_{i,j}$	1	2	3	4
1	--	--	0	1
2	--	--	1	1
3	--	1/3	--	1/2
4	2/3	2/3	1	--

As $P(i,j) = p(i,j)\alpha_{i,j}$, for all $i \neq j$, we have

$$P = \begin{pmatrix} -- & 0 & 0 & 1/3 \\ 0 & -- & 1/3 & 1/3 \\ 0 & 1/6 & -- & 1/4 \\ 1/6 & 1/6 & 1/4 & -- \end{pmatrix} = \begin{pmatrix} 2/3 & 0 & 0 & 1/3 \\ 0 & 1/3 & 1/3 & 1/3 \\ 0 & 1/6 & 7/12 & 1/4 \\ 1/6 & 1/6 & 1/4 & 5/12 \end{pmatrix}.$$

The transition matrix can be verified by checking the equation $\pi P = \pi$.

8. Similarity,

$\alpha_{i,j}$	1	2	3	4
1	--	--	0	--
2	0	--	1	1
3	--	1/3	--	1/2
4	0	1/3	1	--

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 2/3 & 1/6 & 1/6 \\ 0 & 1/9 & 5/9 & 1/3 \\ 0 & 1/12 & 1/4 & 2/3 \end{pmatrix}.$$

Although the matrix P satisfies the equation $\pi P = \pi$, the Markov chain is reducible, it is not suitable for MCMC. This exercise shows that not all Markov chain can be reduced to a desired chain via the Metropolis-Hastings algorithm. But for most common continuous state Markov chains, the Metropolis-Hastings algorithm still works.

References

1. Alexander, C.A. (ed.) (1998). *Risk Management and Analysis, Volume 1: Mesasuring and Modeling Financial Risk.* Wiley, Chichester.

2. Alexander, C.A. (ed.) (1998). *Risk Management and Analysis, Volume 2: New Markets and Products.* Wiley, Chichester.

3. Anderson, T.W. (2003). *An Introduction to Multivariate Statistical Analysis*, 3rd ed. Wiley, New York.

4. Barnett, V. (1991). *Sample Surveys: Principles and Methods*, 2nd ed. Oxford University Press, New York.

5. Berger, J.O. *Statistical Decision Theory and Bayesian Analysis*, 2nd ed. Springer-Verlag, New York.

6. Bernardo, J.M. and Smith, A.F.M. (2000). *Bayesian Theory.* Wiley, Chichester.

7. Bhattacharya, R.N. and Waymire, E.C. (1990). *Stochastic Processes with Applications.* Wiley, New York.

8. Billingsley, P. (1999). *Convergence of Probability Measures*, 2nd ed. Wiley, New York.

9. Black, F. and Scholes, M. (1973). The pricing of option and corporate liabilities. *Journal of Political Economy* **81**, 637–659.

10. Box, G.E.P. and Tiao, G.C. (1973). *Bayesian Inference in Statistical Analysis.* Addison-Wesley, Reading, Massachusetts.

11. Boyle, P. (1977). Options: a Monte Carlo approach. *Journal of Financial Economics* **4**, 323–338.

12. Boyle, P., Broadie, M. and Glasserman, P. (1997). Monte Carlo methods for security pricing. *Journal of Economic Dynamics and Control* **21**, 1276–1321.

13. Brandimarte, P. (2002). *Numerical Methods in Finance: A MATLAB-Based Introduction.* Wiley, New York.

14. Broadie, M. and Glasserman, P. (1998). Simulation for option pricing and risk management. In Alexander, C. (ed.), *Risk Management and Analysis, Volume 1: Measuring and Modeling Financial Risk*, 173–206. Wiley, Chichester.

15. Casella, G. and Berger, R.L. (2001). *Statistical Inference*, 2nd ed. Duxbury Press, Belmont, California.

16. Casella, G. and Goerge, E. (1992). Explaining the Gibbs sampler. *The American Statistician* **46**, 167–174.

17. Chan, N.H. (2002). *Time Series: Applications to Finance.* Wiley, New York.

18. Chib, S. and Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm. *The American Statistician* **49**, 327–335.

19. Conte, S.D. and de Boor, C. (1980). *Elementary Numerical Analysis: An Algorithmic Approach*, 3rd ed. McGraw-Hill, New York.

20. Conze, A. and Viswanthan. (1991). Path-dependent options: the case of lookback option. *Journal of Finance* **46**, 1893–1907.

21. Cox, J., Ross, A. and Rubinstein, M. (1979). Option pricing: a simplified approach. *Journal of Financial Economics* **7**, 229–264.

22. Cox, J., Ingersoll, J. and Ross, A. (1985). A theory of the term structure of interest rates. *Econometrica* **53**, 385–407.

23. Crouchy, M., Galai, D. and Robert, M. (2001). *Risk Management.* McGraw-Hill, New York.

24. Dana, R.A. and Jeanblanc, M. (2002). *Financial Markets in Continuous Time.* Springer-Verlag, Berlin.

25. DeGroot, M. H. (1970). *Optimal Statistical Decisions.* McGraw-Hill, New York.

26. Dempster, M.A.H. (ed.) (2002). *Risk Management: Value at Risk and Beyond.* Cambirdge University Press, Cambridge.

27. El Babsiri, M. and Noel, G. (1988). Simulating path-dependent options: a new approach. *Journal of Derivatives* **6**, 65–83.

28. Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance.* Springer-Verlag, Berlin.

29. Finkenstädt, B. and Rootzén, H. (eds.) (2004). *Extreme Values in Finance, Telecommunications and the Environment.* Chapman and Hall, New York.

30. Fournie, E., Lasry, J., Lebuchoux, J., Lions, P. and Touzi, N. (1999). Applications of Malliavin Calculus to Monte Carlo methods in finance. *Finance and Stochastics* **3**, 391–412.

31. Fournie, E., Lasry, J., Lebuchoux, J. and Lions, P. (2000). Applications of Malliavin Calculus to Monte Carlo methods in finance II. *Finance and Stochastics* **5**, 201–236.

32. Gelman, A., Carlin, J.B., Stern, H.S., and Rubin, D. B. (eds.) (2003). *Bayesian Data Analysis*, 2nd ed. Chapman and Hall, New York.

33. Gilks, W.R., Richardson, S. and Speigelhalter, D.J. (eds.) (1995). *Markov Chain Monte Carlo in Practice.* Chapman and Hall, New York.

34. Glasserman, P. (2003). *Monte Carlo Methods in Financial Engineering.* Springer-Verlag, New York.

35. Hansen, A.T. and Jorgensen, P.L. (2000). Analytical valuation of American-style Asian options. *Management Science* **46**, 1116–1136.

36. Heath, D., Jarrow, R. and Morton, A. (1992). Bond pricing and the term structure of interest rates. *Econometrica* **60**, 77–106.

37. Hickernell, F.J., Lemieux, C. and Owen, A.B. (2005). Control variates for Quasi-Monte Carlo (with discussions). *Statistical Science* **20**, 1–31.

38. Ho, T. and Lee, S. (1986). Term structure movements and pricing interest rate contingent claims. *Journal of Finance* **41**, 1011–1029.

39. Hogg, R.V. and Tanis, E.A. (2006). *Probability and Statistical Inference*, 7th ed. Prentice Hall, Englewood Cliffs, New Jersey.

40. Hull, J. (2006). *Options, Futures and Other Derivatives*, 6th ed. Prentice Hall, Englewood Cliffs, New Jersey.

41. Hull, J. and White, A. (1988). The use of the control variate technique in option pricing. *Journal of Financial and Quantitative Analysis* **23**, 237–251.

42. Jaeckel, P. (2002). *Monte Carlo Methods in Finance.* Wiley, New York.

43. Jamshidian, F. and Zhu, Y. (1997). Scenario simulation: theory and methodology. *Finance and Stochastics* **1**, 43–67.

44. Jarrow, R.A. (2002). *Modeling Fixed-Income Securities and Interest Rate Options.* Stanford University Press, Stanford, California.

45. Jorion, P. (2000). *Value at Risk*, 2nd ed. McGraw-Hill, New York.

46. Joshi, M. (2003). *The Concepts and Practice of Mathematical Finance.* Cambridge University Press, Cambridge.

47. Karatzas, I. and Shreve, S.E. (1997). *Brownian Motion and Stochastic Calculus,* 2nd ed. Springer-Verlag, New York.

48. L'Ecuyer, P. (1994). Uniform random number generation. *Annals of Operations Research* **53**, 77–120.

49. Lee, P. (2004). *Bayesian Statistics: An Introduction*, 3rd ed. Edward Arnold, London.

50. Longstaff, F. and Schwartz, E.S. (2001). Valuing American options by simulation: a simple least-squares approach. *Review of Financial Studies* **14**, 113–147.

51. Margrabe, W. (1978). The value of an option to exchange one asset for another. *Journal of Finance* **33**, 177–186.

52. McLeish, D.L. (2005). Monle Carlo Simulation and Finance. Wiley, NewYork.

53. McNeil, A.J., Frey, R., and Embrechts, P. (2005). *Quantitative Risk Management.* Princeton University Press, Princeton, New Jersey.

54. Merton, R. (1973). Theory of rational option pricing. *Bell Journal of Economics and Management Science* **4**, 141–183.

55. Morgan, J.P. (1995). *Riskmetrics Technical Manual.* J.P. Morgan, New York.

56. Moreno, M. and Navas, J.F. (2003). On the robustness of least-squares Monte Carlo for pricing American derivatives. *Review of Derivatives Research* **6**, 107–128.

57. O'Hagan, A. (1994). *Kendall's Advanced Theory of Statistics* **2B**, *Bayesian Inference.* Edward Arnold, London.

58. Pliska, S.R. (1997). *Introduction to Mathematical Finance: Discrete Time Models.* Blackwell, Maiden, Massachusetts.

59. Redner, R. and Walker, H. (1984). Mixture densities, maximum likelihood and the EM algorithm. *SIAM Review* **26**, 195–239.

60. Robert, C.P. (2001). *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*, 2nd ed. Springer-Verlag, New York.

61. Robert, C.P. (2005). *Monte Carlo Statistical Methods.* Springer-Verlag, New York.

62. Ross, S. (2002). *Simulation,* 3rd ed. Academic Press, San Diego.

63. Stentoft, L. (2004). Assessing the least squares Monte-Carlo approach to American option valuation. *Review of Derivatives Research* **7**, 129–168.

64. Taylor, H. and Karlin, S. (1998). *An Introduction to Stochastic Modeling*, 3rd ed. Academic Press, San Diego.

65. Tierney, L. (1994). Markov chains for exploring posterior distributions, with discussions. *Annals of Statistics* **22**, 1701–1762.

66. Tsay, R.S. (2006). *Analysis of Financial Time Series*, 2nd ed. Wiley, New York.

67. Weisberg, S. (1985). *Applied Linear Regression*, 2nd ed. Wiley, New York.

68. Wong, H.Y. and Cheung, Y.L. (2004). Geometric Asian options: valuation and calibration with stochastic volatility. *Quantitative Finance* **4**, 301–314.

69. Wong, H.Y. and Kwok, Y.K. (2003). Sub-replication and replenishing premium: efficient pricing of multi-state lookbacks. *Review of Derivatives Research* **6**, 83–106.

70. Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics* **5**, 177–188.

71. Venables, W.N. and Ripley, B.D. (2002). *Modern Applied Statistics with S-Plus*, 4th ed. Springer-Verlag, New York.

Simulation Techniques in Financial Risk Management
by Ngai Hang Chan and Hoi Ying Wong

Index

Simulation Techniques in Financial Risk Management
by Ngai Hang Chan and Hoi Ying Wong

STATISTICS IN PRACTICE

Human and Biological Sciences

Brown and Prescott · Applied Mixed Models in Medicine
Ellenberg, Fleming and DeMets · Data Monitoring Committees in Clinical Trials: A Practical Perspective
Lawson, Browne and Vidal Rodeiro · Disease Mapping With WinBUGS and MLwiN
Lui · Statistical Estimation of Epidemiological Risk
*Marubini and Valsecchi · Analysing Survival Data from Clinical Trials and Observation Studies
Parmigiani · Modeling in Medical Decision Making: A Bayesian Approach
Senn · Cross-over Trials in Clinical Research, *Second Edition*
Senn · Statistical Issues in Drug Development
Spiegelhalter, Abrams and Myles · Bayesian Approaches to Clinical Trials and Health-Care Evaluation
Whitehead · Design and Analysis of Sequential Clinical Trials, *Revised Second Edition*
Whitehead · Meta-Analysis of Controlled Clinical Trials

Earth and Environmental Sciences

Buck, Cavanagh and Litton · Bayesian Approach to Interpreting Archaeological Data
Glasbey and Horgan · Image Analysis in the Biological Sciences
Helsel · Nondetects and Data Analysis: Statistics for Censored Environmental Data
McBride · Using Statistical Methods for Water Quality Management: Issues, Problems and Solutions
Webster and Oliver · Geostatistics for Environmental Scientists

Industry, Commerce and Finance

Aitken and Taroni · Statistics and the Evaluation of Evidence for Forensic Scientists, *Second Edition*
Chan and Wong · Simulation Techniques in Financial Risk Management
Lehtonen and Pahkinen · Practical Methods for Design and Analysis of Complex Surveys, *Second Edition*
Ohser and Mücklich · Statistical Analysis of Microstructures in Materials Science

*Now available in paperback.