Hui Wang

Monte Carlo Simulation with Applications to Finance

Chapman & Hall/CRC FINANCIAL MATHEMATICS SERIES

Monte Carlo Simulation with Applications to Finance

CHAPMAN & HALL/CRC
Financial Mathematics Series

Aims and scope:
The field of financial mathematics forms an ever-expanding slice of the financial sector. This series aims to capture new developments and summarize what is known over the whole spectrum of this field. It will include a broad range of textbooks, reference works and handbooks that are meant to appeal to both academics and practitioners. The inclusion of numerical code and concrete real-world examples is highly encouraged.

Series Editors

Published Titles
American-Style Derivatives; Valuation and Computation, *Jerome Detemple*
Analysis, Geometry, and Modeling in Finance: Advanced Methods in Option Pricing, *Pierre Henry-Labordère*
An Introduction to Exotic Option Pricing, *Peter Buchen*
Credit Risk: Models, Derivatives, and Management, *Niklas Wagner*
Engineering BGM, *Alan Brace*
Financial Modelling with Jump Processes, *Rama Cont and Peter Tankov*
Interest Rate Modeling: Theory and Practice, *Lixin Wu*
Introduction to Credit Risk Modeling, Second Edition, *Christian Bluhm, Ludger Overbeck, and Christoph Wagner*
Introduction to Stochastic Calculus Applied to Finance, Second Edition, *Damien Lamberton and Bernard Lapeyre*
Monte Carlo Methods and Models in Finance and Insurance, *Ralf Korn, Elke Korn, and Gerald Kroisandt*
Monte Carlo Simulation with Applications to Finance, *Hui Wang*
Numerical Methods for Finance, *John A. D. Appleby, David C. Edelman, and John J. H. Miller*
Option Valuation: A First Course in Financial Mathematics, *Hugo D. Junghenn*
Portfolio Optimization and Performance Analysis, *Jean-Luc Prigent*
Quantitative Fund Management, *M. A. H. Dempster, Georg Pflug, and Gautam Mitra*
Risk Analysis in Finance and Insurance, Second Edition, *Alexander Melnikov*
Robust Libor Modelling and Pricing of Derivative Products, *John Schoenmakers*
Stochastic Finance: A Numeraire Approach, *Jan Vecer*
Stochastic Financial Models, *Douglas Kennedy*
Structured Credit Portfolio Analysis, Baskets & CDOs, *Christian Bluhm and Ludger Overbeck*
Understanding Risk: The Theory and Practice of Financial Risk Management, *David Murphy*
Unravelling the Credit Crunch, *David Murphy*

Proposals for the series should be submitted to one of the series editors above or directly to:
CRC Press, Taylor & Francis Group
4th, Floor, Albert House
1-4 Singer Street
London EC2A 4BQ
UK

Chapman & Hall/CRC FINANCIAL MATHEMATICS SERIES

Monte Carlo Simulation with Applications to Finance

Hui Wang

Brown University
Providence, Rhode Island, USA

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business
A CHAPMAN & HALL BOOK

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

Version Date: 20120518

International Standard Book Number-13: 978-1-4665-6690-3 (eBook - PDF)

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

Preface

This book can serve as the text for a one-semester course on Monte Carlo simulation. The intended audience is advanced undergraduate students or students in master's programs who wish to learn the basics of this exciting topic and its applications to finance.

The book is largely self-contained. The only prerequisite is some experience with probability and statistics. Prior knowledge on option pricing is helpful but not essential. As in any study of Monte Carlo simulation, coding is an integral part and cannot be ignored. The book contains a large number of MATLAB® coding exercises. They are designed in a progressive manner so that no prior experience with MATLAB is required.

Much of the mathematics in the book is informal. For example, random variables are simply defined to be functions on the sample space, even though they should be measurable with respect to appropriate σ-algebras; exchanging the order of integrations is carried out liberally, even though it should be justified by the Tonelli–Fubini Theorem. The motivation for doing so is to avoid the technical measure theoretic jargon, which is of little concern in practice and does not help much to further the understanding of the topic.

The book is an extension of the lecture notes that I have developed for an undergraduate course on Monte Carlo simulation at Brown University. I would like to thank the students who have taken the course, as well as the Division of Applied Mathematics at Brown, for their support.

Hui Wang
Providence, Rhode Island
January, 2012

Contents

This page intentionally left blank

Chapter 1

Review of Probability

Probability theory is the essential mathematical tool for the design and analysis of Monte Carlo simulation schemes. It is assumed that the reader is somewhat familiar with the elementary probability concepts such as random variables and multivariate probability distributions. However, for the sake of completeness, we use this chapter to collect a number of basic results from probability theory that will be used repeatedly in the rest of the book.

1.1 Probability Space

In probability theory, *sample space* is the collection of all possible outcomes. Throughout the book, the sample space will be denoted by Ω. A generic element of the sample space represents a possible outcome and is called a *sample point*. A subset of the sample space is called an *event*.

1. The empty set is denoted by $\emptyset$.

2. The complement of an event A is denoted by A^c.

3. The intersection of events A and B is denoted by $A \cap B$ or simply AB.

4. The union of events A and B is denoted by $A \cup B$.

A *probability measure* $\mathbb{P}$ on Ω is a mapping from the events of Ω to the real line $\mathbb{R}$ that satisfies the following three axioms:

(i) $\mathbb{P}(\Omega) = 1$.

(ii) $0 \le \mathbb{P}(A) \le 1$ for every event A.

(iii) For every sequence of *mutually exclusive* events $\{A_1, A_2, \ldots\}$, that is, $A_i \cap A_j = \emptyset$ for all $i \neq j$,

$$\mathbb{P}\left(\cup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathbb{P}\left(A_n\right).$$

Lemma 1.1. *Let $\mathbb{P}$ be a probability measure. Then the following statements hold.*

1. $\mathbb{P}(A) + \mathbb{P}(A^c) = 1$ *for any event A.*

2. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(AB)$ *for any events A and B. More generally,*

$$\begin{aligned}\mathbb{P}(A_1 \cup \cdots \cup A_n) &= \sum_i \mathbb{P}(A_i) - \sum_{i<j} \mathbb{P}(A_i A_j) + \sum_{i<j<k} \mathbb{P}(A_i A_j A_k) \\ &\quad + \cdots + (-1)^{n+1} \mathbb{P}(A_1 \cdots A_n)\end{aligned}$$

for an arbitrary collection of events $A_1, \ldots, A_n$.

This lemma follows immediately from the three axioms. We leave the proof to the reader as an exercise.

1.2 Independence and Conditional Probability

Two events A and B are said to be *independent* if $\mathbb{P}(AB) = \mathbb{P}(A) \cdot \mathbb{P}(B)$. More generally, a collection of events $A_1, A_2, \cdots, A_n$ are said to be *independent* if

$$\mathbb{P}(A_{k_1} A_{k_2} \cdots A_{k_m}) = \mathbb{P}(A_{k_1}) \cdot \mathbb{P}(A_{k_2}) \cdot \cdots \cdot \mathbb{P}(A_{k_m})$$

for any $1 \le k_1 < k_2 < \ldots < k_m \le n$.

Lemma 1.2. *Suppose that events A and B are independent. Then so are events A^c and B, A and B^c, and A^c and B^c. Similar results hold for an arbitrary collection of independent events.*

PROOF. Consider the events A^c and B. Since $(A^c B)$ and (AB) are disjoint and $(A^c B) \cup (AB) = B$, it follows that

$$\mathbb{P}(A^c B) + \mathbb{P}(AB) = \mathbb{P}(B).$$

By the independence of A and B, $\mathbb{P}(AB) = \mathbb{P}(A)\mathbb{P}(B)$. Therefore,

$$\begin{aligned}\mathbb{P}(A^cB) &= \mathbb{P}(B) - \mathbb{P}(AB) \\ &= \mathbb{P}(B) - \mathbb{P}(A)\mathbb{P}(B) \\ &= \mathbb{P}(B)[1 - \mathbb{P}(A)] \\ &= \mathbb{P}(B)\mathbb{P}(A^c).\end{aligned}$$

In other words, A^c and B are independent. The proof for other cases is similar and thus omitted. ■

Consider two events A and B with $\mathbb{P}(B) > 0$. The *conditional probability* of A given B is defined to be

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(AB)}{\mathbb{P}(B)}. \tag{1.1}$$

When $\mathbb{P}(B) = 0$, the conditional probability $\mathbb{P}(A|B)$ is undefined. However, it is always true that

$$\mathbb{P}(AB) = \mathbb{P}(A|B)\mathbb{P}(B),$$

where the right-hand-side is defined to be 0 as long as $\mathbb{P}(B) = 0$.

Lemma 1.3. *Given any event B with $\mathbb{P}(B) > 0$, the following statements hold.*

1. *An event A is independent of B if and only if $\mathbb{P}(A|B) = \mathbb{P}(A)$.*
2. *For any disjoint events A and C,*

$$\mathbb{P}(A \cup C|B) = \mathbb{P}(A|B) + \mathbb{P}(C|B).$$

3. *For any events A_1 and A_2,*

$$\mathbb{P}(A_1A_2|B) = \mathbb{P}(A_1|B) \cdot \mathbb{P}(A_2|A_1B).$$

PROOF. All these claims follow directly from the definition (1.1). We should only give the proof of (3). The right-hand-side equals

$$\frac{\mathbb{P}(A_1B)}{\mathbb{P}(B)} \cdot \frac{\mathbb{P}(A_1A_2B)}{\mathbb{P}(A_1B)} = \frac{\mathbb{P}(A_1A_2B)}{\mathbb{P}(B)},$$

which equals the left-hand-side. We complete the proof. ■

Theorem 1.4. (Law of Total Probability). *Suppose that $\{B_n\}$ is a partition of the sample space, that is, $\{B_n\}$ are mutually exclusive and $\cup_n B_n = \Omega$. Then for any event A,*

$$\mathbb{P}(A) = \sum_n \mathbb{P}(A|B_n)\mathbb{P}(B_n).$$

PROOF. Observe that $\{AB_n\}$ are disjoint events and $\cup_n AB_n = A$. It follows that

$$\mathbb{P}(A) = \sum_n \mathbb{P}(AB_n) = \sum_n \mathbb{P}(A|B_n)\mathbb{P}(B_n).$$

We complete the proof. ■

Example 1.1. An investor has purchased bonds from five S&P AAA-rated banks and three S&P A-rated banks.

S&P rating	AAA	AA	A	BBB	BB	B	CCC
Probability	1	4	12	50	300	1100	2800

Annual default probability in basis points, 100 basis points = 1%

Assuming that all these banks are independent, what is the probability that

(a) at least one of the banks default?

(b) exactly one bank defaults?

SOLUTION: The probability that at least one of the banks default equals

$$\begin{aligned} 1 - \mathbb{P}(\text{none of the banks default}) &= 1 - (1 - 0.0001)^5 \cdot (1 - 0.0012)^3 \\ &= 1 - 0.9999^5 \cdot 0.9988^3 \\ &= 40.94\,\text{bps}. \end{aligned}$$

Observe that there are five equally likely ways that exactly one AAA-rated bank defaults and three equally likely ways that exactly one A-rated bank defaults. Hence, the probability that exactly one bank defaults equals

$$5 \cdot 0.9999^4 \cdot 0.0001 \cdot 0.9988^3 + 3 \cdot 0.9999^5 \cdot 0.9988^2 \cdot 0.0012 = 40.88\,\text{bps}.$$

The answers to (a) and (b) are nearly identical because the probability that more than one bank will default is negligible. ■

Example 1.2. A technical analyst has developed a simple model that uses the data from previous two days to predict the stock price movement of the following day. Let "$+$" and "$-$" denote the stock price movement in a trading day:

$$\begin{aligned} \text{“}+\text{”} &= \text{stock price moves up or remains unchanged,} \\ \text{“}-\text{”} &= \text{stock price moves down.} \end{aligned}$$

Below is the probability distribution.

	Tomorrow	
(Yesterday, today)	$+$	$-$
$(+,+)$	0.2	0.8
$(-,+)$	0.4	0.6
$(+,-)$	0.7	0.3
$(-,-)$	0.5	0.5

Assume that the stock price movements yesterday and today are $(-,+)$. Compute the probability that the stock price movement will be "$+$" for

(a) tomorrow,

(b) the day after tomorrow.

SOLUTION: Define the following events:

$$\begin{aligned} A_1 &= \text{stock price movement tomorrow is “}+\text{”,} \\ A_2 &= \text{stock price movement the day after tomorrow is “}+\text{”,} \\ B &= \text{stock price movements yesterday and today are } (-,+). \end{aligned}$$

Then the probability that the stock price movement tomorrow will be "$+$" is

$$\mathbb{P}(A_1|B) = \mathbb{P}(+|-,+) = 0.4.$$

By Lemma 1.3, the probability that the stock price movement the day after tomorrow will be "$+$" is

$$\begin{aligned} \mathbb{P}(A_2|B) &= \mathbb{P}(A_1A_2|B) + \mathbb{P}(A_1^cA_2|B) \\ &= \mathbb{P}(A_1|B) \cdot \mathbb{P}(A_2|A_1B) + \mathbb{P}(A_1^c|B) \cdot \mathbb{P}(A_2|A_1^cB) \\ &= \mathbb{P}(+|-,+) \cdot \mathbb{P}(+|+,+) + \mathbb{P}(-|-,+) \cdot \mathbb{P}(+|+,-) \\ &= 0.4 \times 0.2 + 0.6 \times 0.7 \\ &= 0.5. \end{aligned}$$

1.3 Random Variables

A *random variable* is a mapping from the sample space to the real line $\mathbb{R}$. The *cumulative distribution function* (cdf) of a random variable X is defined by

$$F(x) = \mathbb{P}(X \leq x)$$

for every $x \in \mathbb{R}$. It is always nondecreasing and continuous from the right. Furthermore,

$$\lim_{x \to -\infty} F(x) = 0, \qquad \lim_{x \to +\infty} F(x) = 1.$$

1.3.1 Discrete Random Variables

A random variable is said to be *discrete* if it can assume at most countably many possible values. Suppose that $\{x_1, x_2, \cdots\}$ is the set of all possible values of a random variable X. The function

$$p(x_i) = \mathbb{P}(X = x_i), \quad i = 1, 2, \cdots$$

is called the *probability mass function* of X. The *expected value* (or *expectation*, *mean*) of X is defined to be

$$E[X] = \sum_i x_i p(x_i).$$

More generally, given any function $h : \mathbb{R} \to \mathbb{R}$, the expected value of the random variable $h(X)$ is given by

$$E[h(X)] = \sum_i h(x_i) p(x_i).$$

While the expected value measures the average of a random variable, the most common measure of the variability of a random variable is the *variance*, which is defined by

$$\mathrm{Var}[X] = E[(X - E[X])^2] = E[X^2] - (E[X])^2.$$

The *standard deviation* of X is just the square root of the variance:

$$\mathrm{Std}[X] = \sqrt{\mathrm{Var}[X]}.$$

Among the most frequently used discrete random variables are Bernoulli random variables, binomial random variables, and Poisson random variables.

1. **Bernoulli with parameter p.** A random variable X that takes values in $\{0,1\}$ and
$$\mathbb{P}(X=1)=p, \quad \mathbb{P}(X=0)=1-p.$$
$$E[X]=p, \quad \mathrm{Var}[X]=p(1-p).$$

2. **Binomial with parameters (n,p).** A random variable X that takes values in $\{0,1,\ldots,n\}$ and
$$\mathbb{P}(X=k)=\binom{n}{k} p^k(1-p)^{n-k}.$$
$$E[X]=np, \quad \mathrm{Var}[X]=np(1-p).$$

3. **Poisson with parameter λ.** A random variable X that takes values in $\{0,1,\ldots\}$ and
$$\mathbb{P}(X=k)=e^{-\lambda}\frac{\lambda^k}{k!}.$$
$$E[X]=\lambda, \quad \mathrm{Var}[X]=\lambda.$$

Example 1.3. Compare the following two scenarios. The default probability of an S&P B-rated bank is assumed to be 1100 basis points.

(a) A company invests 10 million dollars in the 1-year bonds issued by an S&P B-rated bank, with an annual interest rate of 10%. Compute the expectation and standard deviation of the value of these bonds at maturity.

(b) A company diversifies its portfolio by dividing 10 million dollars equally between the 1-year bonds issued by two different S&P B-rated banks. Assume that the two banks are independent and offer the same annual interest rate of 10%. Compute the expectation and standard variation of the value of these bonds at maturity.

SOLUTION: Let X (in millions) be the value of these bonds at maturity.

(a) It is easy to see that the distribution of X is given by the following table:

Value of X	11	0
Probability	0.89	0.11

Therefore,

$$E[X] = 11 \times 0.89 + 0 \times 0.11 = 9.79,$$

$$\text{Std}[X] = \sqrt{\text{Var}[X]} = \sqrt{0.89 \times 1.21^2 + 0.11 \times (-9.79)^2} = 3.442.$$

(b) If both banks default then $X = 0$. If one of the banks defaults then $X = 5.5$. If none of the banks default then $X = 11$. The respective probabilities are given by the following table:

Value of X	11	5.5	0
Probability	0.7921	0.1958	0.0121

For example, there are two equally likely ways that one of the banks defaults, and thus the probability of $X = 5.5$ is

$$2 \times 0.11 \times (1 - 0.11) = 0.1958$$

by independence. The expected value and standard deviation of X can be similarly calculated.

$$E[X] = 11 \times 0.7921 + 5.5 \times 0.1958 + 0 \times 0.0121 = 9.79,$$

$$\text{Std}[X] = \sqrt{\text{Var}[X]} = 2.434.$$

These two investment strategies have the same expected return. However, the standard deviation or variance of the second strategy is significantly smaller than that of the first strategy. If one defines variance as the *measure of risk*, then the second strategy has the same expected return but less risk. In other words, diversification reduces risk. ■

Example 1.4. Consider the following model for stock price. Denote by S_i the price at the i-th time step. If the current price is S, then at the next time step the price either moves up to uS with probability p or moves down to dS with probability $1 - p$. Here $d < 1 < u$ are given positive constants. Given $S_0 = x$, find the distribution of S_n.

SOLUTION: Suppose that among the first n time steps there are k steps at which the stock price moves up. Then there are $(n - k)$ steps at which the stock price moves down and

$$S_n = u^k d^{n-k} S_0 = u^k d^{n-k} x.$$

Since the number of time steps at which the stock price moves up is a binomial random variable with parameters n and p, the distribution of S_n is given by

$$\mathbb{P}(S_n = u^k d^{n-k} x) = \left(\begin{array}{c} n \\ k \end{array} \right) p^k (1-p)^{n-k},$$

for $k = 0, 1, \ldots, n$. ■

1.3.2 Continuous Random Variables

A random variable X is said to be *continuous* if there exists a nonnegative function $f(x)$ such that

$$\mathbb{P}(X \in B) = \int_B f(x)\, dx$$

for any subset $B \subseteq \mathbb{R}$. The function f is said to be the *density* of X and must satisfy the equality

$$\int_{-\infty}^{\infty} f(x)\, dx = 1.$$

The relation between the cumulative distribution function F and the density f is given by

$$f(x) = F'(x), \quad F(x) = \int_{-\infty}^{x} f(t) dt.$$

The *expected value* (or *expectation*, *mean*) of X is defined to be

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

More generally, for any function $h : \mathbb{R} \to \mathbb{R}$, the expected value of $h(X)$ is given by

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

As in the discrete random variable case, the *variance* and *standard deviation* of X are defined as follows:

$$\mathrm{Var}[X] = E[(X - E[X])^2] = E[X^2] - (E[X])^2,$$

$$\mathrm{Std}[X] = \sqrt{\mathrm{Var}[X]}.$$

We should discuss some of the most widely used continuous random variables in finance: uniform random variables, exponential random variables, normal random variables, and lognormal random variables.

1. **Uniform on $[a,b]$.** A random variable X with density

$$f(x) = \begin{cases} (b-a)^{-1} & \text{if } x \in [a,b], \\ 0 & \text{otherwise.} \end{cases}$$

$$E[X] = \frac{b+a}{2}, \quad \text{Var}[X] = \frac{(b-a)^2}{12}.$$

2. **Exponential with rate λ.** A random variable X with density

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

$$E[X] = \frac{1}{\lambda}, \quad \text{Var}[X] = \frac{1}{\lambda^2}.$$

3. **Normal with mean μ and variance σ^2 : $N(\mu, \sigma^2)$.** A random variable X with density

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}, \quad x \in \mathbb{R}.$$

$$E[X] = \mu, \quad \text{Var}[X] = \sigma^2.$$

The special case where $\mu = 0$ and $\sigma = 1$ is referred to as the **standard normal**.

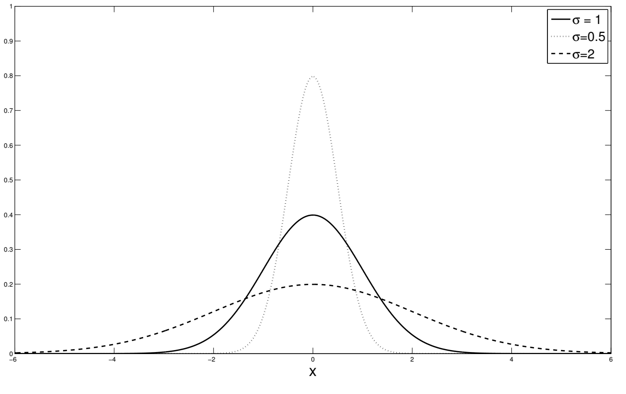


Figure 1.1: Density of $N(0, \sigma^2)$.

An important property of the normal distributions is that any linear transform of a normal random variable is still normal. More precisely, we have the following result, whose proof is left to Exercise 1.6.

Lemma 1.5. *Assume that X is $N(\mu, \sigma^2)$. Let a and b be two arbitrary constants. Then $a + bX$ is normal with mean $a + b\mu$ and variance $b^2\sigma^2$.*

An immediate corollary of the preceding lemma is that if X is $N(\mu, \sigma^2)$, then

$$Z = \frac{X - \mu}{\sigma}$$

is a standard normal random variable. Throughout the book, we will use Φ to denote the cumulative distribution function of the standard normal. That is,

$$\Phi(x) = \mathbb{P}(N(0,1) \leq x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}\, dz.$$

Due to the symmetry of the standard normal density, for every $x \in \mathbb{R}$

$$\Phi(x) + \Phi(-x) = 1.$$

4. **Lognormal with parameters μ and σ^2 : LogN(μ, σ^2).** A positive random variable X whose natural logarithm is normally distributed with mean μ and variance σ^2. In other words,

$$X = e^Y, \quad Y = N(\mu, \sigma^2).$$

$$E[X] = e^{\mu + \frac{1}{2}\sigma^2}, \quad \mathrm{Var}[X] = \left(e^{\sigma^2} - 1\right) \cdot e^{2\mu + \sigma^2}.$$

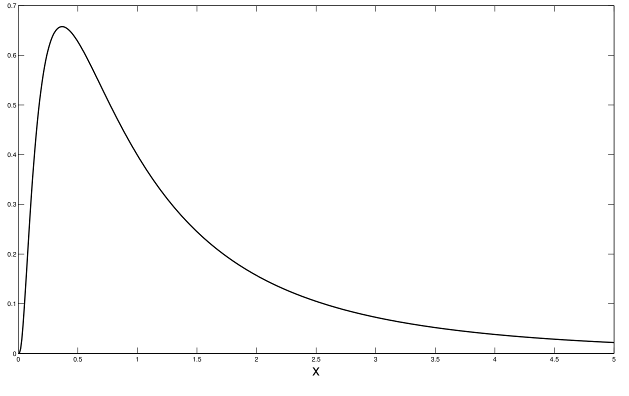


Figure 1.2: Density of LogN$(0,1)$.

Example 1.5. Value-at-risk (VaR) measures, within a confidence level, the maximum loss a portfolio could suffer. To be more precise, denote by X the change in the market value of a portfolio during a given time period. Then for a given confidence level $1-\alpha$ where $\alpha \in (0,1)$, the VaR is defined by

$$\mathbb{P}(X \leq -\mathrm{VaR}) = \alpha.$$

In other words, with probability $1-\alpha$, the maximum loss will not exceed VaR.

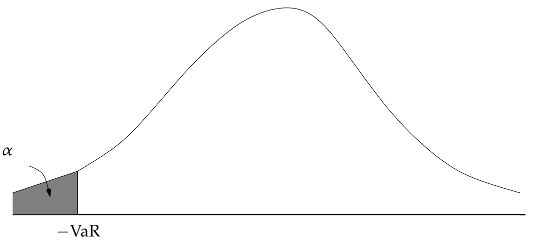


Figure 1.3: Value at Risk.

There is a simple formula for VaR when X is assumed to be normally distributed with mean μ and variance σ^2. Given a confidence level $1-\alpha$,

$$\alpha = \mathbb{P}(X \leq -\mathrm{VaR}) = \mathbb{P}\left(\frac{X-\mu}{\sigma} \leq \frac{-\mathrm{VaR}-\mu}{\sigma}\right).$$

Since $(X-\mu)/\sigma$ is a standard normal random variable, it follows that

$$\frac{-\mathrm{VaR}-\mu}{\sigma} = -z_\alpha,$$

where z_α is determined by $\Phi(-z_\alpha) = \alpha$. Therefore, $\mathrm{VaR} = z_\alpha \sigma - \mu$. ■

Example 1.6. The evaluation of call options often involves the calculation of expected values such as $E\left[(S-K)^+\right]$, where S is the price of the underlying stock, K is a given positive constant, and x^+ denotes the positive part of x, that is, $x^+ = \max\{x, 0\}$. Assuming that S is lognormally distributed with parameters μ and σ^2, compute this expected value.

SOLUTION: Since $X = \log S$ is distributed as $N(\mu, \sigma^2)$, it follows that

$$\begin{aligned} E[(S-K)^+] &= \int_{\log K}^{\infty} (e^x - K) \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}\, dx \\ &= \int_{\theta}^{\infty} \left(e^{\mu+\sigma z} - K\right) \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}\, dz, \end{aligned}$$

where in the second equality we have used the change of variable $x = \mu + \sigma z$ and let

$$\theta = \frac{\log K - \mu}{\sigma}.$$

Direct calculation yields that

$$\begin{aligned}
\int_\theta^\infty e^{\mu+\sigma z} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz &= e^{\mu+\frac{1}{2}\sigma^2} \int_\theta^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z-\sigma)^2} dz \\
&= e^{\mu+\frac{1}{2}\sigma^2} \int_{\theta-\sigma}^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz \\
&= e^{\mu+\frac{1}{2}\sigma^2} \Phi(\sigma - \theta), \\
K \int_\theta^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz &= K\Phi(-\theta).
\end{aligned}$$

Hence,

$$E[(S-K)^+] = e^{\mu+\frac{1}{2}\sigma^2} \Phi(\sigma - \theta) - K\Phi(-\theta). \qquad \blacksquare$$

1.4 Random Vectors

A *random vector* is a collection of random variables defined on the same sample space. The components of a random vector can be discrete, continuous, or mixed. We will, for the moment, restrict ourselves to a two-dimensional random vector (X, Y). The extension to general random vectors is obvious.

1. **Discrete random vectors.** If both X and Y are discrete, it is convenient to define the *joint probability mass function* by

$$p(x_i, y_j) = \mathbb{P}(X = x_i, Y = y_j).$$

The probability mass functions for X and for Y are the same as the *marginal probability mass functions*

$$p_X(x_i) = \sum_j p(x_i, y_j), \quad p_Y(y_j) = \sum_i p(x_i, y_j),$$

respectively. For any function $h : \mathbb{R}^2 \to \mathbb{R}$, the expected value of $h(X, Y)$ is given by

$$E[h(X,Y)] = \sum_{i,j} h(x_i, y_j)\, p(x_i, y_j).$$

2. **Continuous random vectors.** X and Y are said to be *jointly continuous* if there exists a nonnegative *joint density function*, say $f(x,y)$, such that
$$\mathbb{P}(X \in A, Y \in B) = \iint_{A\times B} f(x,y)\,dxdy$$
for any $A, B \subseteq \mathbb{R}$. The density functions for X and for Y are the same as the *marginal density functions*
$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy, \quad f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx,$$
respectively. For any function $h : \mathbb{R}^2 \to \mathbb{R}$, the expected value of $h(X,Y)$ is given by
$$E[h(X,Y)] = \iint_{\mathbb{R}^2} h(x,y) f(x,y)\,dxdy.$$

Let a and b be two arbitrary constants and consider the function $h(x,y) = ax + by$. We obtain the following result immediately.

Theorem 1.6. *For any random variables X and Y, any constants a and b,*
$$E[aX + bY] = aE[X] + bE[Y].$$

1.4.1 Covariance and Correlation

The *covariance* of random variables X and Y is defined to be
$$\mathrm{Cov}(X,Y) = E\left[(X - EX)(Y - EY)\right] = E[XY] - E[X]E[Y].$$

Direct calculation yields that for any random variables X, Y, Z and any constants a, b, c, the following relations hold:
$$\begin{aligned}
\mathrm{Cov}(X,X) &= \mathrm{Var}[X],\\
\mathrm{Cov}(X,Y) &= \mathrm{Cov}(Y,X),\\
\mathrm{Cov}(a,X) &= 0,\\
\mathrm{Cov}(aX+bY,Z) &= a\mathrm{Cov}(X,Z) + b\mathrm{Cov}(Y,Z),\\
\mathrm{Cov}(X,aY+bZ) &= a\mathrm{Cov}(X,Y) + b\mathrm{Cov}(X,Z).
\end{aligned}$$

We can now state the variance formula for sums of random variables. The proof is a straightforward application of the preceding identities and thus omitted.

Lemma 1.7. *For any random variables $X_1, X_2, \ldots, X_n$,*

$$\begin{aligned}\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) &= \sum_{i=1}^{n} \mathrm{Var}[X_i] + \sum_{i\neq j} \mathrm{Cov}(X_i, X_j)\\ &= \sum_{i=1}^{n} \mathrm{Var}[X_i] + 2\sum_{i<j} \mathrm{Cov}(X_i, X_j).\end{aligned}$$

The *correlation coefficient* between random variables X and Y is defined to be

$$\beta = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}X}\sqrt{\mathrm{Var}Y}}.$$

It can be shown that $-1 \leq \beta \leq 1$ [34]. If $\beta > 0$, then X and Y are said to be *positively correlated*, and if $\beta < 0$, then X and Y are said to be *negatively correlated*. Loosely speaking, positively correlated random variables tend to increase or decrease together, while negatively correlated random variables tend to move in opposite directions. When $\beta = 0$, X and Y are said to be *uncorrelated*.

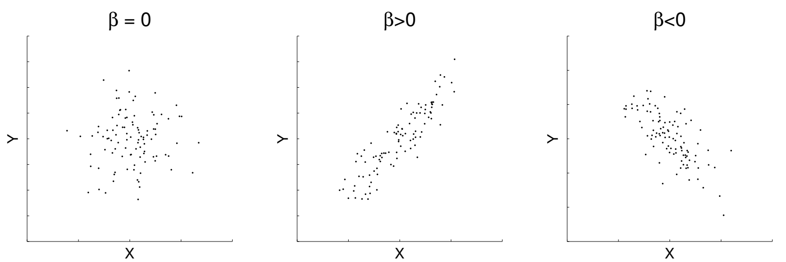


Figure 1.4: Representative samples of random vector (X, Y).

Example 1.7. A portfolio manager wishes to allocate \$1 million between two assets. Denote by X_i the return of the i-th asset, and assume $E[X_i] = r_i$ and $\mathrm{Var}[X_i] = \sigma_i^2$ for $i = 1, 2$. The correlation coefficient between X_1 and X_2 is assumed to be β. It is required that the overall expected return from the allocation be no smaller than a given level r^*. The portfolio manager would like to choose such an allocation with minimal variance. Solve this *mean variance optimization* problem under the assumption that

$$r_1 = 0.2,\ \ r_2 = 0.1,\ \ \sigma_1^2 = 0.1,\ \ \sigma_2^2 = 0.4,\ \ \beta = -0.5,\ \ r^* = 0.15.$$

SOLUTION: Suppose that the strategy is to invest $\$w_i$ million in the i-th asset. Then,

$$w_i \geq 0, \;\; w_1 + w_2 = 1.$$

The return of this strategy is $w_1X_1 + w_2X_2$. The expected return and variance are, respectively,

$$\begin{aligned} E[w_1X_1 + w_2X_2] &= w_1r_1 + w_2r_2, \\ \mathrm{Var}[w_1X_1 + w_2X_2] &= w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2\beta\sigma_1\sigma_2w_1w_2. \end{aligned}$$

Therefore, the optimization problem is to minimize

$$w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2\beta\sigma_1\sigma_2w_1w_2$$

under the constraints that

$$w_1r_1 + w_2r_2 \geq r^*, \;\; w_1 + w_2 = 1, \;\; w_i \geq 0.$$

Plugging in the given parameters and substituting $1 - w_1$ for w_2, the optimization problem reduces to minimizing

$$0.7w_1^2 - w_1 + 0.4 \;\; \text{such that} \;\; 0.5 \leq w_1 \leq 1.$$

Thus, the optimal allocation is $w_1^* = 5/7$ and $w_2^* = 1 - w_1^* = 2/7$. The expected value and standard deviation of the return from this allocation are 0.1714 and 0.2070, respectively. ■

1.4.2 Independence

Two random variables X and Y are said to be *independent* if for any $A, B \subseteq \mathbb{R}$

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A) \cdot \mathbb{P}(Y \in B).$$

If X and Y are discrete random variables, then X and Y are independent if and only if the joint probability mass function equals the product of the marginal probability mass functions:

$$p(x, y) = p_X(x)p_Y(y).$$

Similarly, if X and Y are jointly continuous, then X and Y are independent if and only if the joint density function equals the product of the marginal density functions:

$$f(x, y) = f_X(x)f_Y(y).$$

Lemma 1.8. *Assume that X and Y are independent. Then*

1. $E[XY] = E[X]E[Y]$,
2. $\mathrm{Cov}(X,Y) = 0$,
3. $\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y]$.

PROOF. Assume that X and Y are jointly continuous. Denote by $f(x,y)$ the joint density and f_X, f_Y the two marginals. It follows that

$$\begin{aligned} E[XY] &= \iint_{\mathbb{R}^2} xyf(x,y)\,dxdy \\ &= \iint_{\mathbb{R}^2} xyf_X(x)f_Y(y)\,dxdy \\ &= \int_{\mathbb{R}} xf_X(x)\,dx \int_{\mathbb{R}} yf_Y(y)\,dy \\ &= E[X]E[Y]. \end{aligned}$$

(2) follows immediately from (1), and (3) is a consequence of (2) and Lemma 1.7. The proof for the discrete case is similar. ■

A very useful result concerning normal random variables is that any linear combination of independent normal random variables is still normally distributed.

Lemma 1.9. *Assume that $\{X_1, \ldots, X_n\}$ are independent and X_i is normally distributed as $N(\mu_i, \sigma_i^2)$ for each i. Then for any constants $\{a_1, \ldots, a_n\}$,*

$$\sum_{i=1}^{n} a_i X_i = N\left(\sum_{i=1}^{n} a_i\mu_i, \sum_{i=1}^{n} a_i^2\sigma_i^2\right).$$

The proof can be found in many probability textbooks; e.g., [34, Chapter 6]. See also Exercise 1.14.

Example 1.8. Suppose that the payoff of an option depends on the prices of two stocks, say S_1 and S_2. If S_i exceeds k_i for each i, then the payoff is a fixed amount; otherwise, it is zero. The evaluation of this option will involve the expected value of $h(S_1, S_2)$, where

$$h(x_1, x_2) = \begin{cases} 1 & \text{if } x_i > k_i \text{ for } i = 1, 2, \\ 0 & \text{otherwise.} \end{cases}$$

Assume that S_1 and S_2 are independent and S_i is lognormally distributed as $\mathrm{LogN}(\mu_i, \sigma_i^2)$. Compute this expected value.

SOLUTION: Since S_1 and S_2 are independent, we can express the expected value as

$$E[h(S_1, S_2)] = \mathbb{P}(S_1 > k_1, S_2 > k_2) = \mathbb{P}(S_1 > k_1)\mathbb{P}(S_2 > k_2).$$

Note that $\log S_i$ is $N(\mu_i, \sigma_i^2)$. It follows that for each i

$$\mathbb{P}(S_i > k_i) = \mathbb{P}(\log S_i > \log k_i) = \Phi\left(-\frac{\log k_i - \mu_i}{\sigma_i}\right).$$

Therefore,

$$E[h(S_1, S_2)] = \Phi\left(-\frac{\log k_1 - \mu_1}{\sigma_1}\right) \cdot \Phi\left(-\frac{\log k_2 - \mu_2}{\sigma_2}\right). \qquad \blacksquare$$

1.5 Conditional Distributions

Consider two random variables X and Y. We are interested in the conditional distribution of X given $Y = y$. Earlier in this chapter we have defined the conditional probability of A given B, namely,

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(AB)}{\mathbb{P}(B)},$$

if $\mathbb{P}(B) > 0$. Therefore, when both X and Y are discrete, the conditional distribution of X given $Y = y$ can be defined in a straightforward manner:

$$\mathbb{P}(X = x|Y = y) = \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(Y = y)}.$$

Difficulty arises, however, when Y is a continuous random variable. In this case, the above definition fails automatically since $\mathbb{P}(Y = y) = 0$. Nonetheless, it is appropriate to replace the probability mass functions by the density functions to define analogously a conditional density function $f(x|y)$. The following theorem can be found in many introductory probability textbooks; see e.g., [34, Chapter 5].

Theorem 1.10. *Suppose that X and Y are continuous random variables with joint density $f(x, y)$. The conditional density of X given $Y = y$ is*

$$f(x|y) = \frac{f(x, y)}{f_Y(y)}.$$

Once the conditional density is obtained, it is very easy to express quantities related to the conditional distribution. For example, for every set $A \subseteq \mathbb{R}$,

$$\mathbb{P}(X \in A|Y = y) = \int_A f(x|y)\, dx,$$

and for any function $h : \mathbb{R} \to \mathbb{R}$,

$$E[h(X)|Y = y] = \int_{\mathbb{R}} h(x) f(x|y)\, dx.$$

Analogous to Theorem 1.4, one can write down another version of the law of total probability in terms of random variables. The proof is straightforward and thus omitted.

Theorem 1.11. (Law of Total Probability). *For any random variables X and Y and any subset $A \subseteq \mathbb{R}$,*

1. *Y is discrete:*

$$\mathbb{P}(X \in A) = \sum_j \mathbb{P}(X \in A|Y = y_j) P(Y = y_j).$$

2. *Y is continuous:*

$$\mathbb{P}(X \in A) = \int_{\mathbb{R}} \mathbb{P}(X \in A|Y = y) f_Y(y)\, dy.$$

Remark 1.1. In the special case where X and Y are independent, the conditional distribution of X given $Y = y$ is always the distribution of X itself, regardless of y.

Example 1.9. Consider a one-factor credit risk model where losses are due to the default of obligors on contractual payments. Suppose that there are m obligors and the i-th obligor defaults if and only if $X_i \geq x_i$ for some random variable X_i and given level x_i. The random variable X_i is assumed to take the form

$$X_i = \rho_i Z + \sqrt{1 - \rho_i^2}\, \varepsilon_i,$$

where $Z, \varepsilon_1, \ldots, \varepsilon_m$ are independent standard normal random variables, and each ρ_i is a given constant satisfying $-1 < \rho_i < 1$. Compute the probability that none of the obligors default.

SOLUTION: Note that for each i, the distribution of X_i is $N(0,1)$, thanks to Lemma 1.9. Therefore,

$$\mathbb{P}(\text{the } i\text{-th obligor does not default}) = \mathbb{P}(X_i < x_i) = \Phi(x_i).$$

Since X_i's have a common component Z, they cannot be independent unless $\rho_i = 0$ for every i. Therefore, it is in general not correct to write

$$\mathbb{P}(\text{no default}) = \prod_{i=1}^{m} \Phi(x_i).$$

However, X_i's are independent given $Z = z$. This conditional independence allows us to write

$$\mathbb{P}(\text{no default} \,|\, Z = z) = \prod_{i=1}^{m} \mathbb{P}\left(\varepsilon_i < \frac{x_i - \rho_i z}{\sqrt{1-\rho_i^2}}\right) = \prod_{i=1}^{m} \Phi\left(\frac{x_i - \rho_i z}{\sqrt{1-\rho_i^2}}\right).$$

It now follows from Theorem 1.11 that

$$\mathbb{P}(\text{no default}) = \int_{\mathbb{R}} \prod_{i=1}^{m} \Phi\left(\frac{x_i - \rho_i z}{\sqrt{1-\rho_i^2}}\right) \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}\, dz.$$

There is no closed form formula for this probability. One often resorts to approximations or Monte Carlo simulation to produce estimates for such quantities. ■

Example 1.10. Another popular measure of risk is the so-called *expected tail loss*. Let L be the loss of portfolio value and $a > 0$ a given loss threshold. The expected tail loss is defined as the conditional expectation

$$\bar{L}_a = E[L | L > a].$$

Assume that L is normally distributed as $N(\mu, \sigma^2)$. Compute $\bar{L}_a$.

SOLUTION: The key step is to derive the conditional distribution of L given $L > a$. To give a general treatment, we temporarily assume that L has an arbitrary density f. Denote by φ the conditional density. It is natural that φ should take the form

$$\varphi(x) = \begin{cases} cf(x) & \text{if } x > a, \\ 0 & \text{if } x \le a, \end{cases}$$

where c is some constant to be determined. Since φ is a density, it must satisfy

$$\int_{-\infty}^{\infty} \varphi(x)\, dx = 1.$$

It follows that

$$c = \frac{1}{\int_a^\infty f(x)\, dx} = \frac{1}{\mathbb{P}(L > a)}.$$

Therefore,

$$E[L|L > a] = \int_{-\infty}^{\infty} x\varphi(x)\, dx = \frac{1}{\mathbb{P}(L > a)} \int_a^\infty x f(x)\, dx.$$

In particular, when L is normally distributed as $N(\mu, \sigma^2)$, it is not difficult to verify that $\mathbb{P}(L > a) = \Phi\,(-\theta)$, where $\theta = (a - \mu)/\sigma$, and

$$\begin{aligned}
\int_a^\infty x f(x)\, dx &= \int_a^\infty x \cdot \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}\, dx \\
&= \int_\theta^\infty (\mu + \sigma z) \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}\, dz \\
&= \mu\Phi(-\theta) + \frac{\sigma}{\sqrt{2\pi}} e^{-\frac{1}{2}\theta^2}.
\end{aligned}$$

Hence, the expected tail loss is

$$\bar{L}_a = E[L|L > a] = \mu + \frac{\sigma}{\sqrt{2\pi}\Phi(-\theta)} e^{-\frac{1}{2}\theta^2}.$$

■

1.6 Conditional Expectation

Consider two random variables X and Y. The *conditional expectation* of X given Y, denoted by $E[X|Y]$, is a function of Y and thus a random variable itself. $E[X|Y]$ can be determined as follows.

Step 1. Compute $E[X|Y = y]$ for every fixed value y. If one knows the conditional distribution of X given $Y = y$, then $E[X|Y = y]$ is just the expected value of this conditional distribution.

Step 2. Regard $E[X|Y = y]$ as a function of y and write $E[X|Y = y] = f(y)$. Replace y by Y to obtain $E[X|Y]$. That is,

$$E[X|Y] = f(Y).$$

A very important result is the following *tower property* of conditional expectations.

Theorem 1.12. (Tower Property). *For any random variables X and Y,*

$$E[E[X|Y]] = E[X].$$

PROOF. We will show for the case where X and Y have a joint density function $f(x,y)$. The proof for other cases is similar and thus omitted. By the definition of conditional expectation,

$$\begin{aligned}
E[E[X|Y]] &= \int_{\mathbb{R}} E[X|Y=y] f_Y(y)\, dy \\
&= \int_{\mathbb{R}} \left[\int_{\mathbb{R}} x f(x|y)\, dx \right] f_Y(y)\, dy \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} x f(x,y)\, dx dy \\
&= \int_{\mathbb{R}} x f_X(x)\, dx \\
&= E[X].
\end{aligned}$$

This completes the proof. ■

Lemma 1.13. *Given any random variables X, Y, Z, any constants a, b, c, and any function $h : \mathbb{R} \to \mathbb{R}$,*

1. $E[c|Y] = c$.
2. $E[h(Y)|Y] = h(Y)$.
3. $E[aX + bZ|Y] = aE[X|Y] + bE[Z|Y]$.
4. $E[h(Y)X|Y] = h(Y)E[X|Y]$.
5. $E[h(X)|Y] = E[h(X)]$, *provided that X and Y are independent.*

The proof of this lemma is left as an exercise to the reader.

Remark 1.2. Even though Y is assumed to be a random variable in our dealings with conditional distributions and conditional expectations, it can actually be much more general. For example, Y can be any random vectors. All the results that we have stated hold for the general case.

Example 1.11. Stock prices are sometimes modeled by distributions other than lognormal in order to fit the empirical data more accurately. For instance, Merton [22] introduced a jump diffusion model for stock prices. A special case of Merton's model assumes that the underlying stock price S satisfies

$$S = e^Y, \ \ Y = X_1 + \sum_{i=1}^{X_2} Z_i,$$

where X_1 is $N(\mu, \sigma^2)$, X_2 is Poisson with parameter λ, Z_i is $N(0, \nu^2)$, and $X_1, X_2, \{Z_i\}$ are all independent. The evaluation of call options involves expected values such as

$$E[(S-K)^+],$$

where K is some positive constant. Compute this expected value.

SOLUTION: For every $n \geq 0$, we can compute the conditional expected value

$$v_n = E[(S-K)^+ | X_2 = n].$$

Indeed, conditional on $X_2 = n$, Y is normally distributed as $N(\mu, \sigma^2 + n\nu^2)$, and thus S is lognormally distributed with parameters μ and $\sigma^2 + n\nu^2$. It follows from Example 1.6 that

$$v_n = e^{\mu + \frac{1}{2}(\sigma^2 + n\nu^2)} \Phi(\sqrt{\sigma^2 + n\nu^2} - \theta_n) - K\Phi(-\theta_n), \quad \theta_n = \frac{\log K - \mu}{\sqrt{\sigma^2 + n\nu^2}}.$$

By Theorem 1.12 or the tower property of conditional expectations,

$$E[(S-K)^+] = E[E[(S-K)^+ | X_2]] = \sum_{n=0}^{\infty} v_n \mathbb{P}(X_2 = n) = e^{-\lambda} \sum_{n=0}^{\infty} v_n \frac{\lambda^n}{n!}.$$

This can be evaluated numerically. ■

1.7 Classical Limit Theorems

Two of the most important limit theorems in probability theory are the strong law of large numbers and central limit theorem. They have numerous applications in both theory and practice. In particular, they provide the theoretical foundation for Monte Carlo simulation schemes.

A rigorous proof of these two theorems is beyond the scope of this book and can be found in a number of advanced probability textbooks such as [2, 5]. We will simply state the theorems without proof. In what follows, the acronym "iid" stands for "independent identically distributed."

Theorem 1.14. (Strong Law of Large Numbers). *If $X_1, X_2, \cdots$ are iid random variables with mean μ, then*

$$\mathbb{P}\left\{\frac{X_1 + X_2 + \cdots + X_n}{n} \to \mu\right\} = 1.$$

Theorem 1.15. (Central Limit Theorem). *If $X_1, X_2, \cdots$ are iid random variables with mean μ and variance σ^2, then the distribution of*

$$\frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}$$

converges to the standard normal distribution. That is, for any $a \in \mathbb{R}$,

$$\mathbb{P}\left\{\frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \le a\right\} \to \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\, dx$$

as $n \to \infty$.

Exercises

Pen-and-Paper Problems

1.1 Let A and B be two events such that $A \subseteq B$. Show that $\mathbb{P}(A) \leq \mathbb{P}(B)$.

1.2 Let A and B be two arbitrary events. Show that $\mathbb{P}(A \cup B) \leq \mathbb{P}(A) + \mathbb{P}(B)$.

1.3 Let X be a random variable and let a and b be two arbitrary constants. Show that

$$E[aX + b] = aE[X] + b, \quad \mathrm{Var}[aX + b] = a^2 \mathrm{Var}[X].$$

1.4 Let X be a random variable that is uniformly distributed on $(0, 1)$. Show that for any constants $a < b$,

$$Y = a + (b - a)X$$

is uniformly distributed on (a, b).

1.5 Show that the density function of the lognormal distribution $\mathrm{LogN}(\mu, \sigma^2)$ is given by

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(\log x - \mu)^2}, \quad x > 0.$$

1.6 Assume that X has distribution $N(\mu, \sigma^2)$. Let a and b be two arbitrary constants and $Y = a + bX$.

(a) Find the cumulative distribution function of Y in terms of Φ.

(b) Compute the density of Y and show that Y is normally distributed with mean $a + b\mu$ and variance $b^2\sigma^2$.

1.7 Assume that X is a standard normal random variable. For an arbitrary $\theta \in \mathbb{R}$, show that $E[\exp\{\theta X\}] = \exp\{\theta^2/2\}$.

1.8 Assume that X is normally distributed with mean μ and variance σ^2. For an arbitrary $\theta \in \mathbb{R}$, determine $E[\exp\{\theta X\}]$.

1.9 Suppose that X is a lognormal random variable with distribution $\mathrm{LogN}(\mu, \sigma^2)$.

(a) Given a constant $a > 0$, what is the distribution of aX?

(b) Given a constant $\alpha \in \mathbb{R}$, what is the distribution of X^α?

1.10 Suppose that S is lognormally distributed as $\mathrm{LogN}(\mu, \sigma^2)$. Compute

(a) $E[(K - S)^+]$ and $E[\max\{S, K\}]$, where K is a positive constant;

(b) $E[(S^\alpha - K)]^+$, where α and K are given positive constants.

1.11 Suppose that the joint default probability distribution of two bonds A and B is as follows.

Bond B	Bond A: Default	Bond A: No default
Default	0.05	0.10
No default	0.05	0.80

Joint default probability distribution

(a) What is the default probability of bond A?

(b) What is the default probability of bond B?

(c) Given that bond A defaults, what is the probability that bond B defaults?

(d) Are the defaults of bond A and bond B independent, positively correlated, or negatively correlated?

1.12 Let S_1 and S_2 be the prices of two assets. Assume that $X = \log S_1$ and $Y = \log S_2$ have a joint density function

$$f(x, y) = \frac{\sqrt{3}}{4\pi} e^{-\frac{1}{2}(x^2 - xy + y^2)}, \quad \text{for } x, y \in \mathbb{R}.$$

(a) What is the distribution of X?

(b) What is the distribution of Y?

(c) What is the distribution of X given $Y = y$?

(d) Determine $E[X|Y]$.

(e) Determine $E[S_1|S_2]$ and use tower property to compute $E[S_1 S_2]$.

(f) Compute $\text{Cov}(S_1, S_2)$.

The random vector (X, Y) is an example of jointly normal random vectors; see Appendix A.

1.13 Let X be a standard normal random variable. Assume that given $X = x$, Y is normally distributed as $N(x, 1)$. Find $\text{Cov}(X, Y)$. *Hint:* Use tower property to compute $E[Y]$ and $E[XY]$.

1.14 Assume that X and Y are two independent random variables. Show that

(a) if X is binomial with parameters (n, p) and Y is binomial with parameters (m, p), then $X + Y$ is binomial with parameters $(m + n, p)$;

(b) if X is Poisson with parameter λ and Y is Poisson with parameter μ, then $X + Y$ is Poisson with parameter $\lambda + \mu$;

(c) if X has cumulative distribution function F and Y has density g, then for any $z \in \mathbb{R}$

$$\mathbb{P}(X + Y \leq z) = \int_{\mathbb{R}} F(z - y) g(y)\, dy;$$

(d) if X has density f and Y has density g, then the density of $X + Y$ is given by the *convolution* $f * g$, where

$$(f * g)(z) = \int_{\mathbb{R}} f(z - y)g(y)\, dy = \int_{\mathbb{R}} g(z - x)f(x)\, dx;$$

(e) if X is normally distributed as $N(0, \sigma_1^2)$ and Y is normally distributed as $N(0, \sigma_2^2)$, then $X + Y$ is normally distributed as $N(0, \sigma_1^2 + \sigma_2^2)$;

(f) if X is exponential with rate λ and Y is exponential with rate μ, then $\min\{X, Y\}$ is exponential with rate $\lambda + \mu$.

1.15 Assume that $X_1, \ldots, X_n$ are independent random variables and that for each i, X_i is lognormal with distribution $\mathrm{LogN}(\mu_i, \sigma_i^2)$. Find the distribution of the product random variable $X_1 \cdot X_2 \cdot \cdots \cdot X_n$.

1.16 Denote by X_i the change of a portfolio's value at the i-th day. Assume that $\{X_i\}$ are independent and identically distributed normal random variables with distribution $N(\mu, \sigma^2)$. Determine the value-at-risk at the confidence level $1 - \alpha$ for the total loss over an m-day period.

1.17 Suppose that L, the loss of a portfolio's value, is lognormally distributed with parameters μ and σ^2. Given a constant $a > 0$, compute the expected tail loss

$$E[L | L > a].$$

1.18 The concept of *utility maximization* plays an important role in the study of economics and finance. The basic setup is as follows. Let X be the return from an investment. The distribution of X varies according to the investment strategy employed. The goal is to maximize the expected utility $E[U(X)]$ by judiciously picking a strategy, for some given *utility function* U that is often assumed to be concave and increasing.

Consider the following utility maximization problem with the utility function $U(x) = \log x$. An investor has a \$1 million capital and can choose to invest any portion of it. Suppose that y is the amount invested. Then with probability p the amount invested will double, and with probability $1 - p$ the amount invested will be lost. Let X be the total wealth in the end. Assuming $p > 0.5$, determine the value of y that maximizes the expected utility

$$E[\log X].$$

1.19 Assume that $X_1, \ldots, X_n$ are iid random variables with mean μ and variance σ^2. Define

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

Show that $E[\bar{X}] = \mu$ and $E[S^2] = \sigma^2$. $\bar{X}$ and S^2 are standard sample estimates for μ and σ^2, respectively.

1.20 Assume that $(X_1, Y_1), \dots, (X_n, Y_n)$ are iid random vectors. Define

$$R = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}), \qquad \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \;\; \bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i.$$

Show that $E[R] = \mathrm{Cov}(X_i, Y_i)$ for every i.

1.21 Let $X_1, \dots, X_n$ be iid random variables with common density $f_\theta(x)$, where θ is an unknown parameter. The joint density of $(X_1, \dots, X_n)$ is

$$\prod_{i=1}^{n} f_\theta(x_i).$$

The *maximum likelihood estimate* of θ is defined to be

$$\hat{\theta} = \mathrm{argmax}_\theta \prod_{i=1}^{n} f_\theta(X_i).$$

Assume that $X_1, \dots, X_n$ are iid normal random variables with distribution $N(\mu, \sigma^2)$. Show that the maximum likelihood estimate of $\theta = (\mu, \sigma^2)$ is given by

$$\hat{\mu} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

1.22 Silver Wheaton (SLW) is a mining company that generates its revenue primarily from the sales of silver. iShares Silver Trust (SLV) is a grantor trust that provides a vehicle for investors to own interests in silver. The table below contains the stock prices of SLW and SLV from Oct 24, 2011 to Nov 4, 2011.

Day i	1	2	3	4	5
SLW_i	31.24	32.21	33.40	34.46	35.97
SLV_i	30.86	32.42	32.50	32.92	34.27
Day i	6	7	8	9	10
SLW_i	34.60	34.07	34.91	36.05	36.09
SLV_i	33.44	32.33	32.23	33.61	33.17

Stock prices of SLW and SLV

Define $X_i = \log(\mathrm{SLW}_{i+1}/\mathrm{SLW}_i)$ and $Y_i = \log(\mathrm{SLV}_{i+1}/\mathrm{SLV}_i)$ for $i = 1, \dots, 9$. Assume that (X_i, Y_i) are iid random vectors. Use the results in Exercise 1.19 and Exercise 1.20 to estimate the correlation coefficient between X and Y.

1.23 A function $f : \mathbb{R} \to \mathbb{R}$ is said to be *convex* if for any $\lambda \in [0, 1]$ and $x_1, x_2 \in \mathbb{R}$,

$$f(\lambda x_1 + (1-\lambda)x_2) \le \lambda f(x_1) + (1-\lambda) f(x_2).$$

Let X be an arbitrary random variable with $\mu = E[X]$. Assume that f is differentiable at $x = \mu$. Show that for any x

$$f(x) \geq f(\mu) + f'(\mu)(x - \mu),$$

and thus

$$E[f(X)] \geq f(E[X]).$$

This inequality is called the *Jensen's inequality*.

MATLAB® Problems

The three commands, "normcdf", "norminv", and "normrnd", are MATLAB functions relevant to normal distributions. To get the detailed description of, say, "norminv", you can use the MATLAB command "help norminv."

"normcdf(x)": return $\Phi(x)$.

"norminv(α)": return the value of x such that $\Phi(x) = \alpha$.

"normrnd": generate samples from normal distributions.

1.A Compute the value-at-risk in Exercise 1.16 for $\mu = 0.1$, $\sigma = 1$, $m = 10$, and $\alpha = 0.01$.

1.B Compute the expected tail loss in Exercise 1.17 for $\mu = 0$, $\sigma = 1$, and $a = 2$.

1.C Generate 10 samples from each of the following distributions.

(a) The standard normal distribution.

(b) The lognormal distribution with parameters $\mu = 1$ and $\sigma^2 = 4$.

(c) The binomial distribution with parameters $(20, 0.5)$. Use "binornd".

(d) The uniform distribution on $(5, 7)$. Use "rand".

(e) The Poisson distribution with parameter 1. Use "poissrnd".

(f) The exponential distribution with rate 2. Use "exprnd".

1.D Use the command "plot" to plot the density functions of the exponential distributions with rate $\lambda = 1$, $\lambda = 0.5$, and $\lambda = 2$, respectively. Use a different line type for each density curve and use the command "legend" to place a legend on the picture, similar to Figure 1.1.

1.E Suppose that X is a standard normal random variable and given $X = x$, Y is normally distributed with mean x and variance one. Generate 1000 samples from the joint distribution of (X, Y) and use the command "scatter" to plot your samples. Are X and Y positively correlated or negatively correlated? See also Exercise 1.13.

1.F Repeat Exercise 1.E except that, given $X = x$, Y is normally distributed with mean $-x$ and variance one. By looking at the plot of the samples, do you expect X and Y to be positively or negatively related? Verify your answer by computing the theoretical value of the correlation coefficient between X and Y.

Chapter 2

Brownian Motion

Brownian motion was discovered in 1827 by the English botanist Robert Brown when he was studying the movement of microscopic pollen grains suspended in a drop of water. The rigorous mathematical foundation of Brownian motion was established by Norbert Wiener around 1923. For this reason, Brownian motion is also called the Wiener process. In mathematical finance, Brownian motion has been used extensively in the modeling of security prices. The celebrated Black–Scholes option pricing formula was derived upon the assumption that the underlying stock price is a geometric Brownian motion.

The purpose of this chapter is to introduce Brownian motion and its basic properties. We suggest that the reader go over Appendix A before reading this chapter since the multivariate normal distributions are indispensable for the study of Brownian motion.

2.1 Brownian Motion

A continuous time stochastic process $W = \{W_t : t \geq 0\}$ is a collection of random variables indexed by "time" t. For each fixed $\omega \in \Omega$, the mapping $t \mapsto W_t(\omega)$ is called a *sample path*. We say W is a **standard Brownian motion** if the following conditions hold:

1. Every sample path of the process W is continuous.

2. $W_0 = 0$.

3. The process has independent increments, that is, for any sequence

$0 = t_0 < t_1 < \cdots < t_n$, the increments

$$W_{t_1} - W_{t_0}, \;\; W_{t_2} - W_{t_1}, \;\; \cdots, \;\; W_{t_n} - W_{t_{n-1}}$$

are independent random variables.

4. For any $s \geq 0$ and $t > 0$, the increment $W_{s+t} - W_s$ is normally distributed with mean 0 and variance t.

It is immediate from the definition that $W_t = W_t - W_0$ is normally distributed with mean 0 and variance t.

Figure 2.1 shows some representative sample paths of a standard Brownian motion. They all exhibit a certain kind of ruggedness. Actually, it can be shown that with probability one, the Brownian motion sample paths are nowhere differentiable and nowhere monotonic [18].

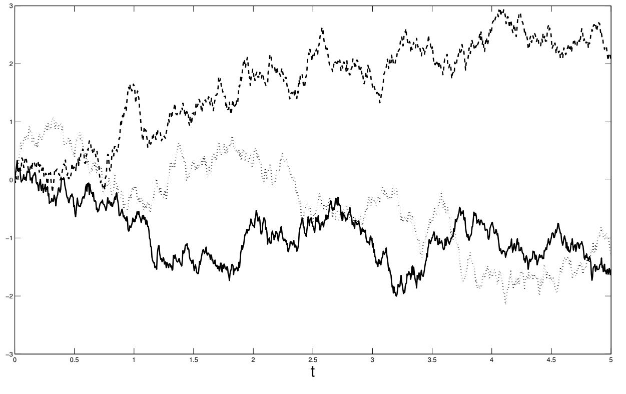


Figure 2.1: Sample paths of Brownian motion.

The next lemma follows directly from the definition of Brownian motion. We leave the proof to the reader.

Lemma 2.1. *Suppose that $W = \{W_t : t \geq 0\}$ is a standard Brownian motion. Then the following statements hold.*

1. *(Symmetry) The process $-W = \{-W_t : t \geq 0\}$ is a standard Brownian motion.*

2. *Fix an arbitrary $s > 0$ and define $B_t = W_{t+s} - W_s$ for $t \geq 0$. Then $B = \{B_t : t \geq 0\}$ is a standard Brownian motion.*

2.2 Running Maximum of Brownian Motion

The *running maximum* of a standard Brownian motion W by time t is defined to be

$$M_t = \max_{0 \leq s \leq t} W_s.$$

It is possible to derive analytically the distribution of M_t, as well as the joint distribution of (W_t, M_t), through the so-called *reflection principle* of Brownian motion.

To illustrate, consider a fixed level $b > 0$ and define the first passage time of the Brownian motion to the level b:

$$T_b = \inf\{t \geq 0 : W_t = b\}.$$

Note that T_b is random and

$$\mathbb{P}(M_t \geq b) = \mathbb{P}(T_b \leq t).$$

The reflection principle asserts that

$$\mathbb{P}(W_t \leq b \,|\, T_b \leq t) = \mathbb{P}(W_t \geq b \,|\, T_b \leq t) = \frac{1}{2}. \tag{2.1}$$

The intuition is as follows. Lemma 2.1(2) basically states that a Brownian

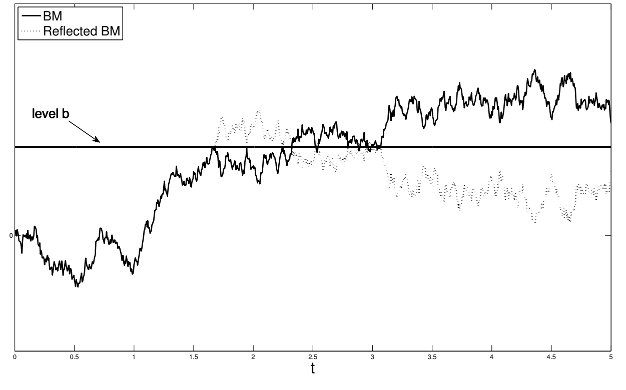


Figure 2.2: Reflected Brownian motion.

motion starts afresh at any deterministic time s. With a leap of faith, assume that it also starts afresh at the random time T_b. Therefore, given $T_b \leq t$, by the symmetry of Brownian motion, for every path that reaches a point above b at time t there is a "reflected" path that reaches a point below b at time t; see the solid path and its reflection, which is represented by the dotted path, in Figure 2.2. Since $\{W_t \geq b\} \subseteq \{T_b \leq t\}$, it follows that

$$\mathbb{P}(W_t \geq b \mid T_b \leq t) = \frac{\mathbb{P}(W_t \geq b, T_b \leq t)}{\mathbb{P}(T_b \leq t)} = \frac{\mathbb{P}(W_t \geq b)}{\mathbb{P}(T_b \leq t)}.$$

Therefore, by (2.1)

$$P(M_t \geq b) = \mathbb{P}(T_b \leq t) = 2\mathbb{P}(W_t \geq b) = 2\Phi\left(-\frac{b}{\sqrt{t}}\right).$$

Taking derivatives with respect to b on both sides, we arrive at the density of M_t:

$$f(x) = \frac{2}{\sqrt{2\pi t}} e^{-\frac{1}{2t}x^2} \tag{2.2}$$

for $x \geq 0$.

As for the joint density function of (W_t, M_t), fix arbitrarily $a \leq b$ and $b > 0$. Analogous to (2.1), we have

$$\mathbb{P}(W_t \leq a \mid T_b \leq t) = \mathbb{P}(W_t \geq 2b - a \mid T_b \leq t).$$

Therefore,

$$\begin{aligned}
\mathbb{P}(W_t \leq a, M_t \geq b) &= \mathbb{P}(W_t \leq a, T_b \leq t) \\
&= \mathbb{P}(W_t \geq 2b - a, T_b \leq t) \\
&= \mathbb{P}(W_t \geq 2b - a) \\
&= \Phi\left(-\frac{2b - a}{\sqrt{t}}\right).
\end{aligned}$$

Taking derivatives over a and b on both sides, it follows that the joint density function of (W_t, M_t) is

$$f(x, y) = \frac{2(2y - x)}{\sqrt{2\pi t^3}} e^{-\frac{1}{2t}(2y-x)^2} \tag{2.3}$$

for $x \leq y$ and $y \geq 0$.

Remark 2.1. The claim that a Brownian motion starts afresh at the first passage time T_b is a consequence of the *strong Markov property*. The proof is advanced and rather technical; see [18].

2.3 Derivatives and Black–Scholes Prices

Financial derivatives such as options derive their values from the underlying assets. For example, a call option on a prescribed stock with strike price K and maturity T grants the holder of the option the right to buy the stock at the price K at time T. Denote by S_t the stock price at time t. If at maturity the stock price S_T is above K, the holder can exercise the option, namely, buy the stock at the price K and sell it immediately at the market price S_T, to realize a profit of $S_T - K$. If the stock price S_T is at or below K, then the option expires worthless. In other words, the payoff of this call option at time T is

$$(S_T - K)^+.$$

In general, a financial derivative pays a random amount X at a given maturity date T. The form of the payoff X can be very simple and only depends on the price of the underlying asset at time T such as call options, or it can be very complicated and depend on the entire history of the asset price up until time T. The most fundamental question in the theory of financial derivatives is about evaluation: what is the fair price of a derivative with payoff X and maturity T?

To answer the question, we need a model for the price of the underlying asset. The classical Black–Scholes model assumes that the asset price is a *geometric Brownian motion*, that is,

$$S_t = S_0 \exp\left\{ \left(\mu - \frac{1}{2}\sigma^2 \right) t + \sigma W_t \right\}, \tag{2.4}$$

where $W = \{W_t : t \geq 0\}$ is a standard Brownian motion, and S_0 is the initial or current asset price. The pair of parameters μ and $\sigma > 0$ are said to be the *drift* and *volatility*, respectively. In contrast to a Brownian motion, a geometric Brownian motion is always positive. Furthermore, for every $t > 0$, S_t is lognormally distributed with distribution

$$\mathrm{LogN}(\log S_0 + (\mu - \sigma^2/2)t, \sigma^2 t).$$

Many pricing problems admit explicit solutions when the underlying asset price is modeled by a geometric Brownian motion.

In this section, we evaluate a number of financial derivatives, assuming that the price of the underlying asset is the geometric Brownian motion (2.4) with drift

$$\mu = r, \tag{2.5}$$

where r denotes the *risk-free interest rate*. Throughout the book, the interest rate is always assumed to be continuously compounded. In practice, the risk-free interest rate is usually taken to be the yield on a zero-coupon bond with similar maturity.

The price or the value of a financial derivative with payoff X and maturity T is given by

$$v = E[e^{-rT}X]. \tag{2.6}$$

Therefore, pricing a financial derivative amounts to computing an expected value. It is not difficult to understand the discounting factor e^{-rT} in the pricing formula (2.6), since one dollar at time T is worth e^{-rT} dollars at time 0. The real question is why we *equate* the drift of the underlying asset price with the risk-free interest rate r. This indeed follows from the arbitrage free principle, which is the topic of the next chapter. For the time being, we assume that the condition (2.5) and the pricing formula (2.6) are valid and use them to evaluate some derivatives. We repeat that, unless otherwise specified, the price of the underlying asset is assumed to be a geometric Brownian motion with drift r and volatility σ.

Example 2.1. (The Black–Scholes Formula). A call option with strike price K and maturity T has payoff $(S_T - K)^+$ at time T. Price this option.

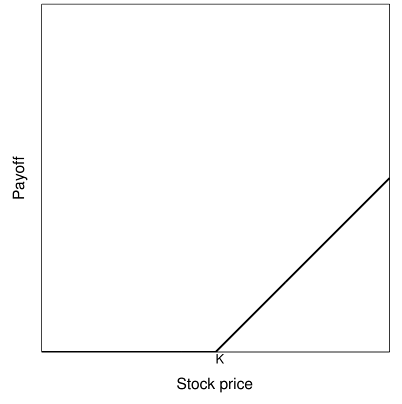


Figure 2.3: Payoff of call option with strike price K.

SOLUTION: Thanks to equation (2.6), the price of the call option is given by

$$v = E[e^{-rT}(S_T - K)^+].$$

Since S_T is lognormally distributed as $\mathrm{LogN}(\log S_0 + (r - \sigma^2/2)T, \sigma^2 T)$, it follows from Example 1.6 that

$$v = S_0 \Phi(\sigma\sqrt{T} - \theta) - Ke^{-rT}\Phi(-\theta),$$

where

$$\theta = \frac{1}{\sigma\sqrt{T}} \log \frac{K}{S_0} + \left(\frac{\sigma}{2} - \frac{r}{\sigma}\right)\sqrt{T}.$$

The call option price was first derived in the seminal paper by Fischer Black and Myron Scholes [3]. For future reference, we will denote the call option price v by

$$\text{BLS\_Call}(S_0, K, T, r, \sigma).$$

Sometimes we simply use "BLS_Call" when there is no confusion about the parameters. ■

Example 2.2. A put option with strike price K and maturity T gives the holder of the option the right to sell the stock at the price K at maturity T. Its payoff is $(K - S_T)^+$. Determine its price.

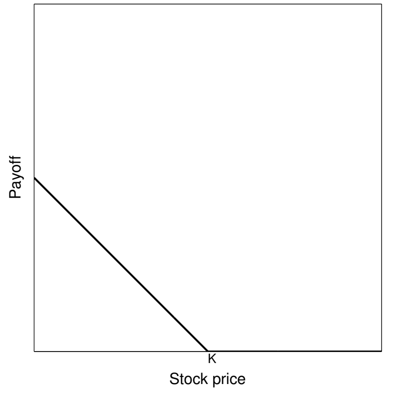


Figure 2.4: Payoff of put option with strike price K.

SOLUTION: Analogous to Example 2.1, we should denote the price of the put option by

$$\text{BLS\_Put}(S_0, K, T, r, \sigma),$$

or simply "BLS_Put" when there is no confusion about the parameters. Observe that

$$(S_T - K)^+ - (K - S_T)^+ = S_T - K.$$

Therefore, thanks to (2.6) and Exercise 2.9, we arrive at the *put-call parity*, namely,

$$\text{BLS\_Call} - \text{BLS\_Put} = E[e^{-rT}(S_T - K)] = S_0 - e^{-rT}K.$$

It follows that

$$\text{BLS\_Put} = e^{-rT}K\Phi(\theta) - S_0\Phi(\theta - \sigma\sqrt{T}),$$

where θ is as defined in Example 2.1. ■

Example 2.3. A binary call option with maturity T pays one dollar when the stock price at time T is at or above a certain level K and pays nothing otherwise. The payoff can be written in the form of an indicator function:

$$X = 1_{\{S_T \geq K\}}.$$

Compute the price of this option.

SOLUTION: It is trivial that $E[X] = \mathbb{P}(S_T \geq K)$. Therefore, the price of the option is

$$v = e^{-rT}\mathbb{P}(S_T \geq K).$$

Since $\log S_T$ is normally distributed with mean $\log S_0 + (r - \sigma^2/2)T$ and variance $\sigma^2 T$, it follows that

$$v = e^{-rT}\mathbb{P}(\log S_T \geq \log K) = e^{-rT}\Phi\left(-\frac{\log(K/S_0) - (r - \sigma^2/2)T}{\sigma\sqrt{T}}\right).$$ ■

Example 2.4. The buyer of a *future* contract is obliged to buy the underlying asset at a certain price p at a specified future time T. Unlike options, upon entering the contract, the buyer does not need to pay any premium. However, at time T the contract must be honored. Therefore, the payoff to the buyer at time T is

$$X = S_T - p.$$

The future price p is chosen so that the value of the contract at present is zero. What should be the value of p?

SOLUTION: The value of the contract to the buyer at present is

$$E[e^{-rT}X] = E[e^{-rT}S_T] - e^{-rT}p = S_0 - e^{-rT}p.$$

Hence, for the contract to have value zero one must have

$$p = e^{rT} S_0.$$

Clearly, with this choice of p the value of the contract is also zero to the seller of the contract. ■

Example 2.5. An Asian option is a path-dependent option whose payoff depends on the average of the underlying stock price over the option life. Consider a discretely monitored average price call option with payoff

$$X = (\bar{S} - K)^+$$

at maturity T, where $\bar{S}$ is the geometric average of the stock price defined by

$$\bar{S} = \left(\prod_{i=1}^{m} S_{t_i} \right)^{1/m}$$

for a given set of monitoring dates $0 \le t_1 < \cdots < t_m \le T$. Calculate the price of this option.

SOLUTION: It is not difficult to see that $\bar{S}$ is lognormally distributed. Indeed,

$$\log \bar{S} = \frac{1}{m} \sum_{i=1}^{m} \log S_{t_i} = \log S_0 + \frac{1}{m} \sum_{i=1}^{m} \left[\left(r - \frac{\sigma^2}{2} \right) t_i + \sigma W_{t_i} \right]$$

is a linear transform of the jointly normal random vector $(W_{t_1}, \ldots, W_{t_m})$ and hence normal itself (see Appendix A and Exercise 2.6). The mean and the variance of $\log \bar{S}$ are

$$\begin{aligned}
\bar{\mu} &= \log S_0 + \left(r - \frac{\sigma^2}{2} \right) \bar{t}, \quad \text{where } \bar{t} = \frac{1}{m} \sum_{i=1}^{m} t_i, \\
\bar{\sigma}^2 &= \mathrm{Var} \left(\frac{1}{m} \sum_{i=1}^{m} \sigma W_{t_i} \right) \\
&= \frac{\sigma^2}{m^2} \left[\sum_{i=1}^{m} \mathrm{Var}(W_{t_i}) + 2 \sum_{i<j} \mathrm{Cov}(W_{t_i}, W_{t_j}) \right] \\
&= \frac{\sigma^2}{m^2} \left[\sum_{i=1}^{m} t_i + 2 \sum_{i<j} t_i \right] \\
&= \frac{\sigma^2}{m^2} \sum_{i=1}^{m} (2m - 2i + 1) t_i.
\end{aligned}$$

Therefore, the price of this Asian option is, thanks to Example 1.6,

$$v = E[e^{-rT}X] = e^{-rT}\left[e^{\bar{\mu}+\frac{1}{2}\bar{\sigma}^2}\Phi(\bar{\sigma}-\theta) - K\Phi(-\theta)\right]$$

with $\theta = (\log K - \bar{\mu})/\bar{\sigma}$. ■

The next example is about pricing a lookback call option. The computation relies on Lemma 2.2, which is a preliminary version of the Girsanov's Theorem [18]. Note that we say a function is *path-dependent* if it depends on the sample paths of the relevant process. For example, if we define

$$h(W_{[0,T]}) = \max_{0\le t\le T} W_t - \min_{0\le t\le T} W_t - W_T,$$

then h is a path-dependent function and its value depends on the entire sample path $W_{[0,T]} = \{W_t : 0 \le t \le T\}$.

Lemma 2.2. *Given an arbitrary constant θ, let $B = \{B_t : t \ge 0\}$ be a Brownian motion with drift θ. That is,*

$$B_t = W_t + \theta t, \quad t \ge 0,$$

where $W = \{W_t : t \ge 0\}$ is a standard Brownian motion. Then for any $T > 0$ and path-dependent function h,

$$E\left[h(B_{[0,T]})\right] = E\left[e^{\theta W_T - \frac{1}{2}\theta^2 T} h(W_{[0,T]})\right].$$

PROOF. Since W_T is normally distributed as $N(0,T)$, it follows from the tower property that the right-hand-side equals

$$\begin{aligned}
\text{R.H.S.} &= \int_{\mathbb{R}} E\left[e^{\theta W_T - \frac{1}{2}\theta^2 T} h(W_{[0,T]})\,\Big|\, W_T = x\right]\cdot \frac{1}{\sqrt{2\pi T}}e^{-\frac{1}{2T}x^2}\,dx \\
&= \int_{\mathbb{R}} E\left[h(W_{[0,T]})\,\Big|\, W_T = x\right]\cdot f(x)\,dx, \qquad (2.7)
\end{aligned}$$

where

$$f(x) = e^{\theta x - \frac{1}{2}\theta^2 T}\cdot\frac{1}{\sqrt{2\pi T}}e^{-\frac{1}{2T}x^2} = \frac{1}{\sqrt{2\pi T}}e^{-\frac{1}{2T}(x-\theta T)^2}.$$

On the other hand, Exercise 2.12 shows that the conditional distribution of $\{B_t : 0 \le t \le T\}$ given $B_T = x$ is the same as that of $\{W_t : 0 \le t \le T\}$ given $W_T = x$. Therefore,

$$E\left[h(W_{[0,T]})\,\Big|\, W_T = x\right] = E\left[h(B_{[0,T]})\,\Big|\, B_T = x\right].$$

Since $f(x)$ is indeed the density of B_T, the tower property implies that the integral in (2.7) equals

$$E\left[E\left[h(B_{[0,T]})\middle| B_T\right]\right] = E\left[h(B_{[0,T]})\right].$$

This completes the proof. ■

Example 2.6. A lookback call option with fixed strike price K and maturity T is a path-dependent option, whose payoff is

$$X = \left(\max_{0\le t\le T} S_t - K\right)^+.$$

Assuming $K > S_0$, what is the value of this option at time zero?

SOLUTION: The value of the option is $v = E[e^{-rT}X]$. To compute this expected value, define

$$B_t = W_t + \theta t, \quad \theta = \left(\frac{r}{\sigma} - \frac{\sigma}{2}\right)t.$$

Then,

$$X = \left(\max_{0\le t\le T} S_t - K\right)^+ = \left(S_0 \exp\left\{\sigma \cdot \max_{0\le t\le T} B_t\right\} - K\right)^+.$$

By Lemma 2.2,

$$E[X] = E\left[e^{\theta W_T - \frac{1}{2}\theta^2 T}\left(S_0 e^{\sigma\cdot M_T} - K\right)^+\right], \quad M_T = \max_{0\le t\le T} W_t.$$

It follows that

$$E[X] = \int_0^\infty \int_{-\infty}^\infty e^{\theta x - \frac{1}{2}\theta^2 T}\left(S_0 e^{\sigma y} - K\right)^+ \cdot f(x,y)\,dxdy,$$

where $f(x,y)$ is the joint density of (W_T, M_T) given by (2.3) with $t = T$. The evaluation of this double integral is rather straightforward but tedious. We will only state the result and leave the details to the reader (Exercise 2.13 may prove useful for this endeavor):

$$\begin{aligned} v &= \text{BLS\_Call}(S_0, K, T, r, \sigma) \\ &+ \frac{\sigma^2}{2r} S_0 \left[\Phi(\theta_+) - e^{-rT}\left(\frac{K}{S_0}\right)^{2r/\sigma^2} \Phi(\theta_-)\right], \end{aligned}$$

where

$$\theta_{\pm} = \frac{1}{\sigma\sqrt{T}} \log \frac{S_0}{K} + \left(\frac{\sigma}{2} \pm \frac{r}{\sigma}\right) \sqrt{T}.$$

Lemma 2.2 can also be used to evaluate other similar path-dependent options such as lookback options with floating strike price and barrier options. See Exercise 2.14. ■

Example 2.7. Consider two stocks whose prices are modeled by geometric Brownian motions:

$$\begin{aligned} S_t &= S_0 \exp\left\{\left(r - \frac{\sigma_1^2}{2}\right) t + \sigma_1 W_t\right\}, \\ V_t &= V_0 \exp\left\{\left(r - \frac{\sigma_2^2}{2}\right) t + \sigma_2 B_t\right\}. \end{aligned}$$

For simplicity, we assume that W and B are two independent standard Brownian motions. The payoff of an exchange option at maturity T is

$$(S_T - V_T)^+.$$

Compute the price of this option.

SOLUTION: The price of this option is

$$v = E[e^{-rT}(S_T - V_T)^+] = E\left[e^{-rT} V_T \left(\frac{S_T}{V_T} - 1\right)^+\right].$$

Straightforward computation yields that

$$\begin{aligned} e^{-rT} V_T &= V_0 \exp\left\{-\frac{\sigma_2^2}{2} T + \sigma_2 B_T\right\}, \\ \frac{S_T}{V_T} &= \frac{S_0}{V_0} \exp\left\{\frac{\sigma_2^2 - \sigma_1^2}{2} T - \sigma_2 B_T + \sigma_1 W_T\right\}. \end{aligned}$$

Since W_T and B_T are independent $N(0, T)$ random variables, their joint density function equals

$$f(x, y) = \frac{1}{\sqrt{2\pi T}} \exp\left\{-\frac{x^2}{2T}\right\} \frac{1}{\sqrt{2\pi T}} \exp\left\{-\frac{y^2}{2T}\right\}.$$

Therefore, one can express v in terms of an integral with respect to the density f and obtain

$$v = V_0 \iint_{\mathbb{R}^2} \left[\frac{S_0}{V_0} \exp\left\{ \frac{\sigma_2^2 - \sigma_1^2}{2} T - \sigma_2 y + \sigma_1 x \right\} - 1 \right]^+ g(x,y)\, dxdy,$$

where

$$\begin{aligned} g(x,y) &= f(x,y) \exp\left\{ -\frac{\sigma_2^2}{2} T + \sigma_2 y \right\} \\ &= \frac{1}{\sqrt{2\pi T}} \exp\left\{ -\frac{x^2}{2T} \right\} \frac{1}{\sqrt{2\pi T}} \exp\left\{ -\frac{(y-\sigma_2 T)^2}{2T} \right\}. \end{aligned}$$

It is interesting to observe that $g(x,y)$ itself is the *joint density function* of two independent normal random variables, say X and Y, where X is $N(0,T)$ and Y is $N(\sigma_2 T, T)$. Therefore, the price v can be written as

$$\begin{aligned} v &= V_0 E\left[\left(\frac{S_0}{V_0} \exp\left\{ \frac{\sigma_2^2 - \sigma_1^2}{2} T - \sigma_2 Y + \sigma_1 X \right\} - 1 \right)^+ \right] \\ &= V_0 E\left[(U-1)^+ \right], \end{aligned}$$

where U stands for a lognormal random variable with distribution

$$\text{LogN}\left(\log \frac{S_0}{V_0} - \frac{\sigma_2^2 + \sigma_1^2}{2} T,\ \sigma_2^2 T + \sigma_1^2 T \right).$$

Letting $\sigma = \sqrt{\sigma_1^2 + \sigma_2^2}$, it follows from Example 1.6 that

$$v = S_0 \Phi\left(\frac{1}{\sigma\sqrt{T}} \log \frac{S_0}{V_0} + \frac{\sigma\sqrt{T}}{2} \right) - V_0 \Phi\left(\frac{1}{\sigma\sqrt{T}} \log \frac{S_0}{V_0} - \frac{\sigma\sqrt{T}}{2} \right).$$

The trick in this calculation is to value of the option in terms of the stock price V_T. This technique is called *change of numéraire*. It is very useful in the evaluation of option prices. ■

2.4 Multidimensional Brownian Motions

For the purpose of future reference, we give the definition of a multidimensional Brownian motion.

Consider a continuous time process $B = \{B_t : t \geq 0\}$ where B_t is a d-dimensional random vector for each t. Let $\Sigma = [\Sigma_{ij}]$ be a $d \times d$ symmetric positive definite matrix. The process B is said to be a *d-dimensional Brownian motion with covariance matrix* Σ if the following conditions are satisfied:

1. Every sample path of the process B is continuous.

2. $B_0 = 0$.

3. The process has independent increments, that is, for any sequence $0 = t_0 < t_1 < \cdots < t_n$, the increments

$$B_{t_1} - B_{t_0}, \;\; B_{t_2} - B_{t_1}, \;\; \cdots, \;\; B_{t_n} - B_{t_{n-1}}$$

are independent random vectors.

4. For any $s \geq 0$ and $t > 0$, the increment $B_{s+t} - B_s$ is a jointly normal random vector with mean 0 and covariance matrix $t\Sigma$.

It follows immediately from the definition that B_t is a d-dimensional jointly normal random vector with distribution $N(0, t\Sigma)$.

When $\Sigma = I_d$, we say B is a *d-dimensional standard Brownian motion*. In this case, every component of B is a one-dimensional standard Brownian motion itself, and all these components are independent. Note that in general, the components of a d-dimensional Brownian motion may not be independent. Two components, say the i-th component and the j-th component, are independent if and only if $\Sigma_{ij} = 0$.

Exercises

Pen-and-Paper Problems

2.1 Assume that $X = (X_1, \ldots, X_d)'$ is a d-dimensional normal random vector with distribution $N(0, \Sigma)$. Let A be an invertible matrix such that

$$AA' = \Sigma.$$

Find the distribution of $Y = A^{-1}X$.

2.2 Suppose that a portfolio consists of two assets and the change of portfolio's value, denoted by X, can be written as

$$X = \beta X_1 + (1 - \beta)X_2,$$

where X_i denotes the change of value of the i-th asset and $0 \leq \beta \leq 1$. Assume that (X_1, X_2) is a jointly normal random vector with distribution

$$N\left(\begin{bmatrix} \mu \\ \mu \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right).$$

(a) Find the distribution of X.

(b) What is the value-at-risk at the confidence level $1 - \alpha$ for the total loss of the portfolio?

(c) For what value of β will the value-at-risk be minimized?

2.3 Suppose that a portfolio consists of two assets and the total value of the portfolio is

$$Y = \beta S_1 + (1 - \beta)S_2,$$

where β is the allocation parameter and $0 \leq \beta \leq 1$. Assume that

$$S_1 = e^{X_1}, \quad S_2 = e^{X_2},$$

where (X_1, X_2) is a jointly normal random vector with distribution

$$N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & \sigma^2 \end{bmatrix}\right).$$

(a) Compute $\mathrm{Var}[S_1]$ and $\mathrm{Var}[S_2]$.

(b) Determine the optimal allocation β that minimizes $\mathrm{Var}[Y]$.

2.4 Brownian motion can be regarded as the limit of simple random walks. Consider a sequence of iid random variable $\{X_i\}$ such that

$$\mathbb{P}(X_i = 1) = \frac{1}{2} = \mathbb{P}(X_i = -1).$$

Fix an arbitrary n. Let $t_m = m/n$ for $m = 0, 1, \ldots$ and define a discrete-time stochastic process $W^{(n)} = \{W^{(n)}_{t_m} : m = 0, 1, \ldots\}$ where

$$W^{(n)}_{t_m} = \frac{1}{\sqrt{n}} \sum_{i=1}^{m} X_i.$$

Show that $W^{(n)}$ has independent and stationary increments, that is, the increments

$$W^{(n)}_{t_1} - W^{(n)}_{t_0}, \cdots, W^{(n)}_{t_m} - W^{(n)}_{t_{m-1}}, \cdots$$

are iid. Assume that $t_m \to t + h$ and $t_k \to t$ as $n \to \infty$. Use central limit theorem to determine the limit distribution of

$$W^{(n)}_{t_m} - W^{(n)}_{t_k}$$

as $n \to \infty$. Explain intuitively that $W^{(n)}$ converges to a standard Brownian motion.

2.5 Suppose that W is a standard Brownian motion. Given an arbitrary constant $a > 0$, show that $B = \{B_t : t \geq 0\}$, where

$$B_t = \frac{1}{\sqrt{a}} W_{at},$$

is also a standard Brownian motion.

2.6 Suppose that W is a standard Brownian motion. Show that the following statements hold.

(a) The conditional distribution of W_t given $W_s = x$ is $N(x, t-s)$ for any $0 \leq s < t$.

(b) For $0 < s < t$, the conditional distribution of W_s given $W_t = x$ is $N(xs/t, s(t-s)/t)$.

(c) The covariance of W_s and W_t is $s \wedge t = \min\{s, t\}$ for any $s, t \geq 0$.

(d) For any $0 < t_1 < t_2 < \cdots < t_n$, the random vector $(W_{t_1}, \ldots, W_{t_n})'$ is jointly normal with distribution $N(0, \Sigma)$, where $\Sigma = [\sigma_{ij}]$ is an $n \times n$ matrix with $\sigma_{ij} = t_i \wedge t_j$.

(e) For any $t \geq 0$ and $\theta \in \mathbb{R}$, $E[\exp\{\theta W_t\}] = \exp\{\theta^2 t/2\}$.

2.7 Let W be a standard Brownian motion. Given any $0 = t_0 < t_1 < \cdots < t_m$ and constants $a_1, a_2, \ldots, a_m$, find the distributions of $\sum_{i=1}^{m} a_i (W_{t_i} - W_{t_{i-1}})$ and $\sum_{i=1}^{m} a_i W_{t_i}$.

2.8 Let W be a standard Brownian motion. For any $t > 0$, determine the density of the *running minimum* by time t

$$m_t = \min_{0 \le s \le t} W_s$$

and the joint density of (W_t, m_t).

2.9 Let $S = \{S_t : t \ge 0\}$ be a geometric Brownian motion with drift μ and volatility σ. Given any $T \ge 0$, show that

$$E[S_T] = e^{\mu T} S_0.$$

2.10 Show that the Black–Scholes call option price BLS_Call(S_0, K, T, r, σ) is monotonically increasing with respect to the volatility σ.

2.11 Assume that the price of an underlying asset is a geometric Brownian motion

$$S_t = S_0 \exp\left\{\left(r - \frac{1}{2}\sigma^2\right) t + \sigma W_t\right\},$$

where r is the risk-free interest rate. Determine the prices of the following derivatives with maturity T and payoff X, in terms of the Black–Scholes call option price formula "BLS_Call".

(a) Break forwards $X = \max\{S_T, S_0 e^{rT}\}$.

(b) Collar option $X = \min\{\max\{S_T, K_1\}, K_2\}$ with $0 < K_1 < K_2$.

(c) Forward-start option $X = (S_T - S_{T_0})^+$ with $T_0 < T$.

(d) Straddle

$$X = \begin{cases} K - S_T & \text{if } S_T \le K, \\ S_T - K & \text{if } S_T \ge K. \end{cases}$$

(e) Power call option $(S_T^\beta - K)^+$ for some positive constant β.

2.12 Let W be a standard Brownian motion. The process $B = \{B_t : t \ge 0\}$, where $B_t = W_t + \theta t$ for some constant θ, is said to be a *Brownian motion with drift* θ. Given $0 < t_1 < \cdots < t_n < T$, show that the conditional distribution of $(B_{t_1}, \ldots, B_{t_n})$ given $B_T = x$ does not depend on θ. In particular, letting $\theta = 0$, we conclude that the conditional distribution of $\{B_t : 0 \le t \le T\}$ given $B_T = x$ is the same as the conditional distribution of $\{W_t : 0 \le t \le T\}$ given $W_T = x$.

2.13 By convention, let Φ denote the cumulative distribution function of the standard normal distribution. Given any constants $a \ne 0$ and θ, use integration by parts to show that

$$\begin{aligned} \int_\theta^\infty e^{ax} \Phi(-x)\, dx &= -\frac{1}{a} e^{a\theta} \Phi(-\theta) + \int_\theta^\infty \frac{1}{a} e^{ax} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx \\ &= -\frac{1}{a} e^{a\theta} \Phi(-\theta) + \frac{1}{a} e^{\frac{1}{2}a^2} \Phi(-\theta + a). \end{aligned}$$

Similarly, when $a = 0$, show that

$$\begin{aligned}\int_\theta^\infty \Phi(-x)\,dx &= -\theta\Phi(-\theta) + \int_\theta^\infty x\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}\,dx \\ &= -\theta\Phi(-\theta) + \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\theta^2}.\end{aligned}$$

2.14 Use Lemma 2.2 to determine the prices of the following path-dependent options with maturity T and payoff X, assuming that the underlying stock price is a geometric Brownian motion with drift r and volatility σ.

(a) Lookback put option with floating strike price: $X = \max_{0\le t\le T} S_t - S_T$.

(b) Lookback put option with fixed strike price K: $X = (K - \min_{0\le t\le T} S_t)^+$.

(c) Down-and-out call option with strike price K and barrier b (assume $S_0 > b$):

$$X = (S_T - K)^+ \cdot 1_{\{\min_{0\le t\le T} S_t \ge b\}}.$$

(d) Up-and-in call option with strike price K and barrier b (assume $S_0 < b$):

$$X = (S_T - K)^+ \cdot 1_{\{\max_{0\le t\le T} S_t \ge b\}}.$$

2.15 Suppose that W is a d-dimensional Brownian motion with covariance matrix Σ. Let A be an $m \times d$ matrix and define

$$B_t = AW_t.$$

Show that B is an m-dimensional Brownian motion with covariance matrix $A\Sigma A'$.

2.16 Suppose that (W, B) is a two-dimensional Brownian motion with covariance matrix

$$\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$$

for some $\rho \in (-1, 1)$. Show that (Q, B) is a two-dimensional standard Brownian motion, where

$$Q_t = \frac{1}{\sqrt{1-\rho^2}}(W_t - \rho B_t).$$

In particular, Q and B are independent standard Brownian motions.

2.17 Brownian motion is the fundamental continuous time continuous process. On the other hand, the fundamental continuous time jump process is Poisson process. One way to construct a *Poisson process* $N = \{N_t : t \ge 0\}$ *with intensity* λ is as follows. Let $\{X_n\}$ be a sequence of iid exponential random variable with rate λ. Let $S_0 = 0$ and for $n \ge 1$

$$S_n = \sum_{i=1}^n X_i.$$

Define for any $t \geq 0$

$$N_t = k, \quad \text{if } S_k \leq t < S_{k+1}.$$

Show that

(a) N has independent increments, that is, for any n and $0 = t_0 < t_1 < \cdots < t_n$, the increments

$$N_{t_1} - N_{t_0}, \cdots, N_{t_n} - N_{t_{n-1}}$$

are independent;

(b) $N_{t+s} - N_s$ is a Poisson random variable with parameter λt for any $s \geq 0$ and $t > 0$. In particular, N_t is a Poisson random variable with parameter λt.

A very useful observation for analyzing Poisson processes is the *memoryless property* of exponential distributions. That is, for any $t, s \geq 0$,

$$\mathbb{P}(X > t + s | X > t) = \mathbb{P}(X > s)$$

when X is exponentially distributed.

MATLAB® Problems

2.A Write a function using the "function" command to calculate the price of a call option from the Black–Scholes formula, namely BLS_Call. The function should have input parameters

$$\begin{aligned} r &= \text{risk-free interest rate}, \\ \sigma &= \text{volatility}, \\ T &= \text{maturity}, \\ K &= \text{strike price}, \\ S_0 &= \text{initial stock price}. \end{aligned}$$

The output of the function is the Black–Scholes price of the corresponding call option. Save this function as a ".m" file.

(a) Compute the price of a call option with $S_0 = 50$, $r = 0.05$, $\sigma = 0.2$, $T = 0.5$, and strike price $K = 45, 50, 55$, respectively.

(b) Consider a call option with $S_0 = 50$, $r = 0.05$, $K = 50$, and $T = 0.5$. Plot the price of this call option with respect to the volatility σ for $0 < \sigma \leq 0.5$. Suppose that at present the market price of the call option is \$3. Find the value of the volatility from your plot such that

$$\text{BLS\_Call}(50, 50, 0.5, 0.05, \sigma) = \text{market price \$3}.$$

This volatility is said to be the *implied volatility*.

(c) Consider a call option with $S_0 = 50$, $r = 0.05$, $K = 50$, and $\sigma = 0.2$. Plot the price of this call option with respect to the maturity T for $0 \leq T \leq 1$. Is this an increasing function?

2.B Write a function using the "function" command to calculate the price of a put option from the Black–Scholes formula, namely BLS_Put. The function should have input parameters

$$\begin{aligned} r &= \text{risk-free interest rate,} \\ \sigma &= \text{volatility,} \\ T &= \text{maturity,} \\ K &= \text{strike price,} \\ S_0 &= \text{initial stock price.} \end{aligned}$$

The output of the function is the Black–Scholes price of the corresponding put option. Save this function as a ".m" file.

(a) Compute the price of a put option with $S_0 = 50$, $r = 0.05$, $\sigma = 0.2$, $T = 0.5$, and strike price $K = 45, 50, 55$, respectively.

(b) Consider a put option with $K = 50$, $r = 0.05$, $\sigma = 0.2$, and $T = 0.5$. Plot the price of this put option with respect to the initial stock price S_0 for $40 \leq S_0 \leq 60$. Is it an increasing function or decreasing function?

(c) Consider a put option with $S_0 = 50$, $r = 0.05$, $K = 50$, and $\sigma = 0.2$. Plot the price of this put option with respect to the maturity T for $0 \leq T \leq 1$. Is it an increasing function or decreasing function?

(d) Consider a put option with $S_0 = 50$, $r = 0.05$, $K = 50$, and $T = 0.5$. Plot the price of this put option with respect to the volatility σ for $0 < \sigma \leq 0.5$. Is it an increasing function or decreasing function? Suppose that at present the market price of the put option is \$2. Find the implied volatility from your plot. That is, determine the value of the volatility such that

$$\text{BLS\_Put}(50, 50, 0.5, 0.05, \sigma) = \text{market price \$2.}$$

Chapter 3

Arbitrage Free Pricing

A question from Chapter 2 remains open: why is it appropriate to equate the drift of the stock price with the risk-free interest rate r? It should be emphasized that this question is meaningful only in the context of pricing financial derivatives. It is clearly not true if we consider the real world stock price movements—in general one would expect the drift or the growth rate of a stock to be higher than the risk-free interest rate due to the risk associated with the stock.

The purpose of this chapter is to explain the key idea—arbitrage free principle—behind the pricing of financial derivatives. We will work with the binomial tree asset pricing models, not only because they are widely used in practice, but also because they provide probably the simplest setting to illustrate the mechanism of arbitrage free pricing. The answer to the open question from Chapter 2 becomes transparent once one realizes that a geometric Brownian motion can be approximated by binomial trees.

3.1 Arbitrage Free Principle

Consider a call option with strike price K and maturity T. The real world price of the underlying asset is assumed to be a geometric Brownian motion with drift μ and volatility σ:

$$S_t = S_0 \exp\left\{ \left(\mu - \frac{\sigma^2}{2} \right) t + \sigma W_t \right\},$$

where $W = \{W_t : t \geq 0\}$ is a standard Brownian motion. It is tempting to set the price of the call option as the expected value of its discounted payoff

$$E[e^{-rT}(S_T - K)^+]. \tag{3.1}$$

If this were correct, then the price of a call option would be higher if the underlying asset has a higher growth rate or drift, everything else being equal.

Unfortunately, this intuitive approach is *not* correct. As we will demonstrate later, if the option price is determined in this fashion, then one can construct portfolios that generate *arbitrage* opportunities. By arbitrage, we mean a portfolio whose value process $X = \{X_t\}$ satisfies $X_0 = 0$ and

$$\mathbb{P}(X_T \geq 0) = 1, \quad \mathbb{P}(X_T > 0) > 0.$$

The *arbitrage free principle* stipulates that there are no arbitrage opportunities or free lunch in a financial market. This principle is not far from truth. In real life, market sometimes does exhibit arbitrage. But it cannot sustain itself and will only last for a very short amount of time—as soon as it is discovered and exploited, it is removed from the market.

It turns out that under appropriate market conditions, for a financial derivative the only price that is consistent with the arbitrage free principle is the expected value of the discounted payoff, where the expected value is taken as if the drift of the underlying asset price were equal to the risk-free interest rate. That is, the right pricing formula for (say) a call option with strike price K and maturity T should still be (9.5). But instead of a geometric Brownian with drift μ, the price of the underlying asset is treated as a geometric Brownian motion with drift r:

$$S_t = S_0 \exp\left\{ \left(r - \frac{\sigma^2}{2} \right) t + \sigma W_t \right\}.$$

Therefore, the option price should *not* depend on the true drift of the underlying asset.

Since the price is derived upon the arbitrage free principle, it is called the *arbitrage free price*. But what is the motivation for arbitrage free pricing? In principle, asset prices are the results of the equilibrium between demand and supply. This equilibrium pricing approach is often used in economics to study the prices of assets *endogenously*. However, to successfully carry it out, it is necessary to characterize the preference and the attitude toward risk for each of the participating agents in the market. Clearly this is not very practical. Therefore, in financial engineering a different and more practical approach is adopted. Assuming that the prices of a collection of assets such as stocks are given *exogenously*, one tries to determine the prices of other assets such as options based on the assumption that the market is free of arbitrage. In this sense, arbitrage free pricing is a relative pricing mechanism. A thorough investigation on derivative pricing can be found in [7, 16].

Unless otherwise specified, the financial derivatives under consideration are assumed to be *European*, that is, they can only be exercised at the maturity date. On the contrary, if a derivative can be exercised at any time before or at the maturity date, it is said to be *American*. The value of an American financial derivative is obviously at least as much as that of its European counterpart. See Appendix B for the pricing of American derivatives.

Remark 3.1. The pricing mechanism is sometimes called the *risk-neutral pricing*, since the expected value is taken in an artificial world where all the risky assets have the same growth rate as the risk free saving account, or equivalently, all the investors are risk neutral.

Remark 3.2. Throughout the book, we assume that the financial market is frictionless in the sense that there is neither any restriction nor any transaction cost on buying/selling any number of financial instruments.

3.2 Asset Pricing with Binomial Trees

In a binomial tree asset pricing model, the price of the underlying asset evolves as follows. If the asset price at the current time step is S, then at the next time step the price will move up to uS with probability p and move down to dS with probability $q = 1 - p$. Here u and d are given positive constants such that $d < u$. For all binomial tree models, we use S_n to denote the asset price at the n-th time step.

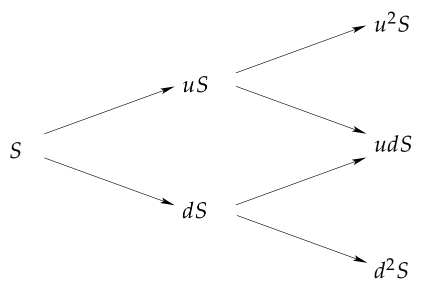


Figure 3.1: A binomial tree model.

Even though a binomial tree model seems to be too simplistic compared with the real world stock price dynamics, it proves to be a reasonable approximation in many occasions and has superior computational tractability. Figure 3.1 depicts a two-period binomial tree model.

3.2.1 A Preliminary Example

In some sense, arbitrage free pricing is a *deterministic* pricing theory made out of probabilistic models. To be more concrete, consider a one-period binomial tree model where the current stock price is $S_0 = 10$ and $u = 1/d = 2$. The risk-free interest rate r is taken to be 0 for convenience. Consider also a call option with strike price $K = 14$ and maturity $T = 1$. The option payoff at time $T = 1$ is

$$X = (S_1 - 14)^+.$$

What should be the price or the value of this option at time 0?

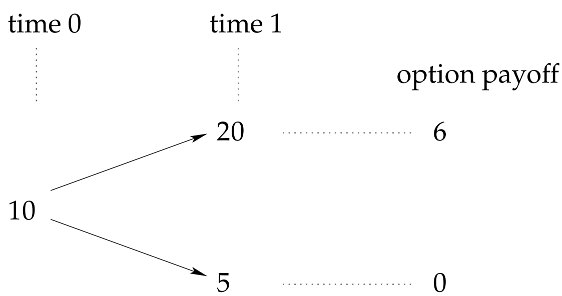


Figure 3.2: A one-period binomial pricing model.

Arbitrage Free Pricing: Consider a portfolio that consists of one share of the call option and x shares of the underlying stock. Note that the value of the portfolio at $T = 1$ is either $20x + 6$ if the stock price climbs to 20 or $5x$ if the stock price drops to 5. Therefore, if we pick x so that

$$20x + 6 = 5x \quad \text{or} \quad x = -0.4,$$

then the value of the portfolio is *fixed* at $20x + 6 = 5x = -2$ at maturity, *no matter how the stock price moves*.

Suppose that the call option's price is v at $t = 0$. Then the value of the portfolio at $t = 0$ is $10x + v = -4 + v$. Since the interest rate r equals 0, we expect that

$$-4 + v = -2 \quad \text{or} \quad v = 2.$$

That is, the option is worth \$2 at time 0. Indeed, if $v \neq 2$, one can construct portfolios that lead to arbitrage:

1. $v > 2$. In this case, the option is over valued. Starting with zero initial wealth, sell one share of the call option and buy 0.4 shares of the stock, which yields a cash position of $v - 4$. At maturity, the cash position is still $v - 4$ since $r = 0$. Now selling the stocks and honoring the call option will always yield \$2 no matter what the stock price is at maturity. Therefore, the total value of the portfolio becomes $v - 2 > 0$ at time $T = 1$. This is an arbitrage.

2. $v < 2$. In this case, the option is under valued. Starting with zero initial wealth, buy one share of the call option and sell 0.4 shares of the stock, which yields a cash position of $4 - v$. At maturity, the cash position is still $4 - v$ since $r = 0$. Now fulfilling the short position in the stock and exercising the call option will always yield $-\$2$ no matter what the stock price is at maturity. Therefore, the total value of the portfolio becomes $2 - v > 0$ at time $T = 1$. This is again an arbitrage.

3.2.2 General Pricing Formula

Now consider a general one-period binomial tree model with $S_0 = S$. We would like to price an option with maturity $T = 1$ and payoff

$$X = \begin{cases} C_u & \text{if } S_1 = uS, \\ C_d & \text{if } S_1 = dS; \end{cases}$$

see Figure 3.3. Suppose that one dollar at time $t = 0$ is worth R dollars at time $t = 1$. It is required by the arbitrage free principle (see Exercise 3.1) that

$$d < R < u.$$

Construct a portfolio with one share of the option and x shares of the underlying stock so that the portfolio will have a fixed value at maturity, regardless of the stock price. That is,

$$uS \cdot x + C_u = dS \cdot x + C_d \quad \text{or} \quad x = -\frac{C_u - C_d}{uS - dS}.$$

Suppose that the price of the option is v at time $t = 0$. Then the portfolio's value at time $t = 0$ is $v + xS$. Since one dollar at time $t = 0$ is worth R

dollars at time $t = 1$, it follows that

$$(v + xS)R = uS \cdot x + C_u,$$

which implies

$$v = \frac{1}{R}\left(\frac{R-d}{u-d} \cdot C_u + \frac{u-R}{u-d} \cdot C_d\right).$$

The reader is suggested to mimic the example in Section 3.2.1 and verify the existence of arbitrage opportunities when the option is priced differently. Note that the parameters p and q play *no* role in the pricing of the option.

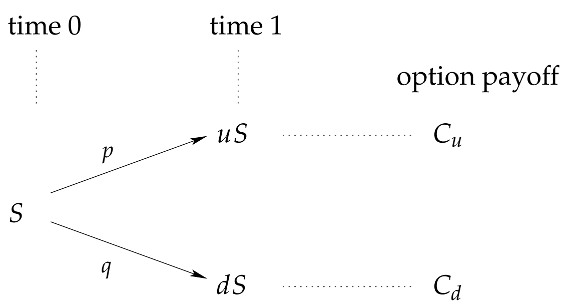


Figure 3.3: A general one-period binomial pricing model.

Risk-Neutral Probability: Observe that the price of the option can be written as the expected value of the discounted option payoff under the risk-neutral probability

$$(p^*, q^*) = \left(\frac{R-d}{u-d}, \frac{u-R}{u-d}\right). \tag{3.2}$$

That is, the option price is

$$v = E\left[\frac{1}{R}X\right] = \frac{1}{R}\left(p^* C_u + q^* C_d\right),$$

where the expected value is taken as if the stock price would move up to uS with probability p^* and move down to dS with probability $q^* = 1 - p^*$; see Figure 3.4.

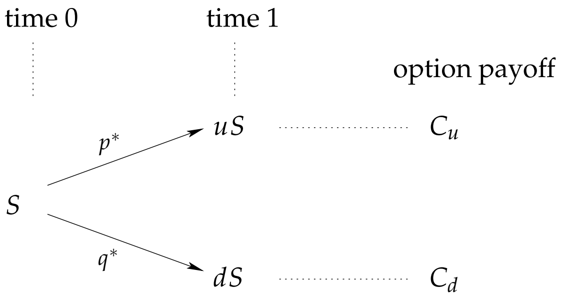


Figure 3.4: Binomial tree under the risk-neutral probability.

Summary.

1. The real world probabilities (p, q) play no role in the pricing of options.

2. The price of an option is the expected value of the discounted option payoff at maturity, under the risk-neutral probability.

3. The risk-neutral probability is determined by the structure of the binomial tree model and the interest rate, and is independent of the option payoff.

4. Under the risk-neutral probability, the growth rate of the underlying asset price equals the risk-free interest rate:

$$E[S_1] = p^* \cdot uS + q^* \cdot dS = \left(\frac{R-d}{u-d} \cdot uS + \frac{u-R}{u-d} \cdot dS \right) = RS.$$

5. **Replication:** Starting with an initial wealth that equals the price of the option, one can construct a portfolio that consists of cash and the underlying asset, and whose value at maturity completely replicates the option payoff. Indeed, starting with v dollars, one can buy

$$\delta = \frac{C_u - C_d}{uS - dS}$$

shares of the underlying asset at time $t = 0$. The cash position at time 0 will become $v - \delta S$. At time $t = 1$, the value of the portfolio is

$$R(v - \delta S) + \delta S_1 = \begin{cases} C_u & \text{if } S_1 = uS, \\ C_d & \text{if } S_1 = dS. \end{cases}$$

3.2.3 Multiperiod Binomial Tree Models

The generalization to multiperiod binomial tree models is straightforward. Consider a binomial tree with n periods and an option with maturity $T = n$ and payoff X. The parameters u, d, and R are defined as before.

The conclusion is that the price of this option at time $t = 0$ is the expected value of the discounted option payoff

$$v = E\left[\frac{1}{R^n} X\right], \tag{3.3}$$

where the expected value is taken under the risk-neutral probability measure. That is, the stock price has probability p^* to move up by a factor u and probability $q^* = 1 - p^*$ to move down by a factor d at each time step. The probabilities (p^*, q^*) are given by (3.2).

In order to verify the pricing formula (3.3), and at the same time introduce a useful recursive algorithm for computing the expected value, we specialize to a three-period binomial tree model as depicted in Figure 3.5.

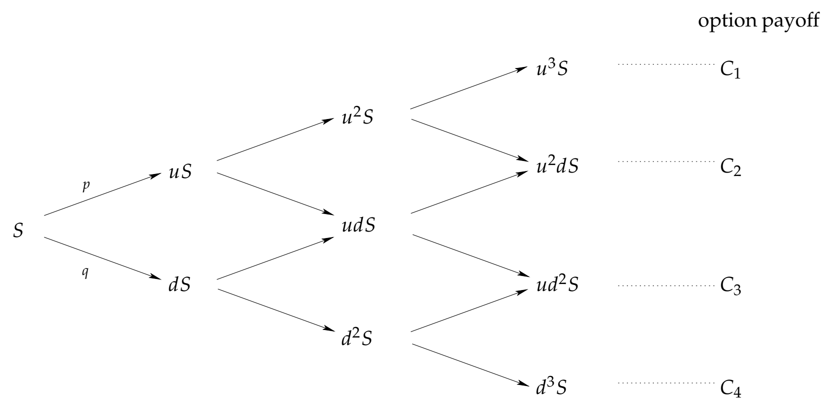


Figure 3.5: A three-period binomial tree model.

Now define V_i to be the value of the option at the i-th time step. See Figure 3.6. In general, V_i is a function of the stock price S_i and thus a random variable itself. For example, $V_2 = a$ when $S_2 = u^2S$ and $V_1 = e$ when $S_1 = dS$. The values of V_i, or $\{a, b, \cdots, f\}$, can be obtained recursively backwards in time.

1. **At time $t = 2$**:

$$a = \frac{1}{R}(p^*C_1 + q^*C_2),\ b = \frac{1}{R}(p^*C_2 + q^*C_3),\ c = \frac{1}{R}(p^*C_3 + q^*C_4).$$

2. **At time $t = 1$**:
$$d = \frac{1}{R}(p^*a + q^*b),\ e = \frac{1}{R}(p^*b + q^*c).$$

3. **At time $t = 0$**:
$$f = \frac{1}{R}(p^*d + q^*e).$$

To summarize, the option's value V satisfies the backwards recursive equation
$$V_3 = X,\ \ V_i = E\left[\frac{1}{R}V_{i+1}\,\middle|\, S_i\right],\quad i = 2, 1, 0.$$
It follows from the tower property (Theorem 1.12) that for each $i = 0, 1, 2$
$$E[V_i] = E\left[\frac{1}{R}V_{i+1}\right].$$

Therefore,
$$f = V_0 = E\left[\frac{1}{R^3}V_3\right] = E\left[\frac{1}{R^3}X\right].$$

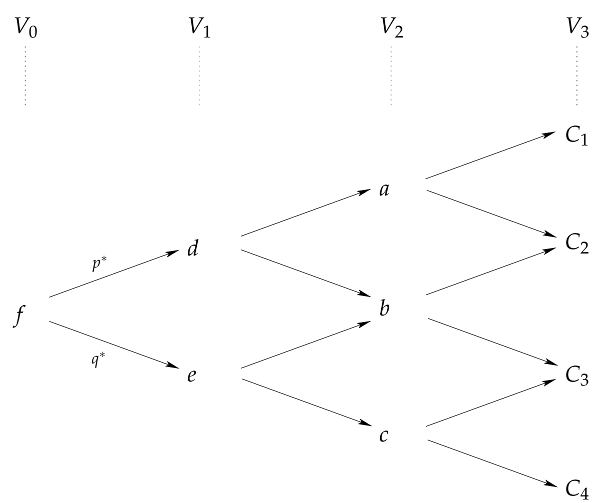


Figure 3.6: The value process of the option.

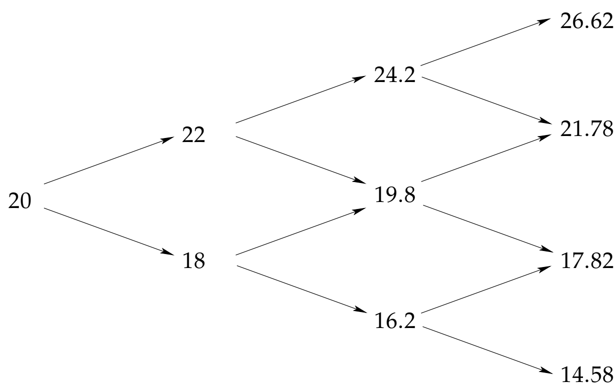
Observe that with an initial wealth of V_0, one can construct a portfolio consisting of cash and stock, and whose value at time $t = 3$ completely replicates the payoff of the option. Indeed, at time $t = 0$, it is possible to construct a portfolio whose value at $t = 1$ is exactly V_1, as described in the summary of Section 3.2.2. In a completely analogous fashion, one can adjust the portfolio at $t = 1$ (the adjustment should depend on whether $S_1 = uS$ or $S_1 = dS$) so that the value of the portfolio will replicate V_2 at $t = 2$, and so on. Now it follows immediately that V_0 is the only price of the option that is consistent with the arbitrage free principle. Indeed, if the price of the option is v and $v \neq V_0$, then one can construct arbitrage opportunities:

1. $v > V_0$: sell one share of the option and buy such a replicating portfolio.

2. $v < V_0$: buy one share of the option and sell such a replicating portfolio.

This justifies the pricing formula (3.3) for $n = 3$. The treatment for a general multiperiod binomial tree model is completely analogous.

Example 3.1. Consider a binomial tree model where each step represents 2 months in real time and $u = 1.1, d = 0.9, S_0 = 20$.



Suppose that the risk-free interest rate is 12% per annum. Compute the price of a call option with maturity $T = 6$ months and strike price $K = 21$.

SOLUTION: The risk-free interest rate for each 2-month period is $0.12/6 = 0.02$. Therefore, the discounting factor is

$$\frac{1}{R} = e^{-0.02} = 0.98,$$

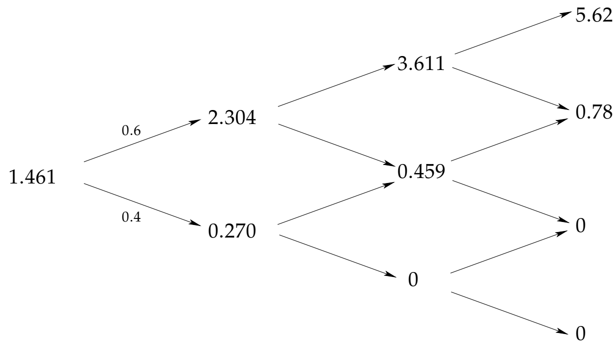


Figure 3.7: The value of the option.

and the risk-neutral probability measure is given by

$$p^* = \frac{R-d}{u-d} = 0.6, \quad q^* = \frac{u-R}{u-d} = 0.4.$$

The price of this call option at time $t = 0$ is 1.461; see Figure 3.7. The value of the option at each node is obtained backwards in time. ■

3.3 The Black–Scholes Model

Consider the Black–Scholes model where the price of the underlying stock is assumed to be a geometric Brownian motion

$$S_t = S_0 \exp\left\{ \left(\mu - \frac{\sigma^2}{2} \right) t + \sigma W_t \right\}. \tag{3.4}$$

The goal is to explain why one should replace the drift μ by the risk-free interest rate r when it comes to pricing derivatives.

The idea is to use binomial trees to approximate the geometric Brownian motion. There are many ways to achieve this approximation. We will use the following version.

1. *Binomial approximation*. Suppose that the maturity date is T. Divide the time interval $[0, T]$ into n pieces of equal length $\Delta t = T/n$. Eventually we will send n to infinity. Consider an approximating binomial tree with n periods, where each period corresponds to Δt in real time. The stock price will move up by a factor u with probability p and move down by a factor d with probability $q = 1 - p$ at each time period. The parameters p, u, and d will be determined so as to match

the distribution of the increments of the geometric Brownian motion $\{S_t\}$ over a time interval of length Δt. Even though it is impossible to make a complete match, the idea is to at least match the expected value and the variance. This leads to

$$\begin{aligned} Se^{\mu\Delta t} &= p \cdot uS + q \cdot dS, \\ S^2 e^{(2\mu+\sigma^2)\Delta t} &= p \cdot (uS)^2 + q \cdot (dS)^2. \end{aligned}$$

In order to solve for the three unknowns we impose a nonessential condition

$$u = \frac{1}{d}.$$

From these three equations we obtain (see Remark 3.3)

$$u = e^{\sigma\sqrt{\Delta t}}, \quad d = e^{-\sigma\sqrt{\Delta t}}, \quad p = \frac{e^{\mu\Delta t} - d}{u - d}. \tag{3.5}$$

It can be argued that this binomial tree approximates the geometric Brownian motion S as $n \to \infty$ [18, Invariance Principle of Donsker]. See also Exercise 3.5.

2. *Risk-neutral probability*. Since each time period in the binomial tree corresponds to Δt in real time,

$$R = e^{r\Delta t}.$$

Therefore, the risk-neutral probabilities (p^*, q^*) are given by

$$p^* = \frac{R - d}{u - d} = \frac{e^{r\Delta t} - d}{u - d}, \quad q^* = 1 - p^*.$$

The price of an option with payoff X and maturity T is

$$v = E\left[\frac{1}{R^n} X\right] = E[e^{-rT} X],$$

where the expected value is taken under the risk-neutral probability measure.

3. *The dynamics of $\{S_t\}$ under the risk-neutral probability measure*. Comparing the formula (3.5) with the formula of p^*, the only difference is that μ is replaced by r. Therefore, under the risk-neutral probability measure, the binomial tree approximates a geometric Brownian motion with drift r and volatility σ. In other words, as the limit of the binomial trees, the stock price $\{S_t\}$ is a geometric Brownian motion with drift r and volatility σ under the risk-neutral probability measure.

Summary: The price of an option with payoff X and maturity T is the expected value of the discounted option payoff, i.e.,

$$v = E[e^{-rT}X].$$

The expected value is taken with respect to the risk-neutral probability measure, under which the stock price is a geometric Brownian motion with drift r:

$$S_t = S_0 \exp\left\{\left(r - \frac{\sigma^2}{2}\right)t + \sigma W_t\right\}.$$

The above discussion also suggests a binomial tree approximation for $\{S_t\}$ under the risk-neutral probability measure: divide the time interval $[0, T]$ into n subintervals of equal length $\Delta t = T/n$ and set

$$u = e^{\sigma\sqrt{\Delta t}}, \quad d = e^{-\sigma\sqrt{\Delta t}}, \quad p^* = \frac{e^{r\Delta t} - d}{u - d}.$$

Remark 3.3. The solution is only approximate and not exact. One can check that the expected value is perfectly matched, but the variance is only matched up to order Δt. It turns out that a perfect match is not necessary, and the approximation is still valid with the choice of parameters in (3.5).

Exercises

Pen-and-Paper Problems

3.1 Show that the arbitrage free principle implies $d < R < u$ for a binomial tree asset pricing model.

3.2 Suppose that the risk-free interest rate is $r = 10\%$ per annum. Price the following financial derivatives using a two-period binomial tree with $u = 1.1$ and $d = 0.9$. The initial price of the underlying asset is assumed to be $S_0 = 50$.

(a) A call option with strike price $K = 48$ and maturity $T = 2$ months

(b) A put option with strike price $K = 50$ and maturity $T = 3$ months

(c) A vertical spread with payoff $X = (S_T - 48)^+ - (S_T - 52)^+$ and maturity $T = 6$ months

(d) A straddle with payoff $X = (50 - S_T)^+ + (S_T - 50)^+$ and maturity $T = 4$ months

3.3 Consider a put option with maturity $T = 6$ months and strike price $K = 19$. The initial price of the underlying stock is $S_0 = 20$. Assume that the risk-free interest rate is $r = 12\%$ annually. We would like to approximate the price of the put option with different binomial trees. For all the binomial trees we assume that $u = 1.1$ and $d = 0.9$.

(a) Price the put option using a one-period binomial tree.

(b) Starting with an initial wealth equal to the price you have computed in part (a), compose a portfolio that consists of cash and stock to completely replicate the payoff of the put option.

(c) Price the put option using a two-period binomial tree, each period representing 3 months in real time.

(d) Starting with an initial wealth equal to the price you have computed in part (c), compose a portfolio that consists of cash and stock to completely replicate the payoff of the put option. Determine the following.

i. The initial cash and stock positions at time $t = 0$

ii. The cash and stock positions at time $t = 1$ when the stock price goes up to uS_0

iii. The cash and stock positions at time $t = 1$ when the stock price moves down to dS_0

3.4 Consider a call option with maturity $T = 6$ months and strike price $K = 19$. The initial price of the underlying stock is $S_0 = 20$. We would like to approximate the price of this call option using a two-period binomial tree with $u = 1.1$ and $d = 0.9$. In the first period, the risk-free interest rate is $r_1 = 6\%$ per annum, but in the second period the risk-free interest rate becomes $r_2 = 8\%$ per annum. Price this option and construct a replicating portfolio consisting of cash and stock.

3.5 The intuitive justification that the binomial tree defined by (3.5) approximates the geometric Brownian motion (3.4) is very similar to Exercise 2.4. The binomial tree has n periods, each period representing $\Delta t = T/n$ in real time. Denote by B_m the value of the tree at the m-th period for $0 \leq m \leq n$. Clearly $B_0 = S_0$. Let $t_m = m\Delta t$ for $m = 0, 1, \ldots, n$, and define a discrete time process $X = \{X_{t_m} = B_m\}$. Note that $\log X$ can be written in the following fashion:

$$\log X_{t_m} = \log S_0 + \sigma\sqrt{\Delta t}\sum_{i=1}^{m} Y_i$$

where Y_i are iid random variables with

$$\mathbb{P}(Y_i = +1) = p = \frac{e^{\mu\Delta t} - d}{u - d}, \quad \mathbb{P}(Y_i = -1) = 1 - p.$$

(a) Show that $\{\log X_{t_m} : m = 0, 1, \ldots, n\}$ has independent and stationary increments. That is,

$$\log X_{t_1} - \log X_{t_0}, \ldots, \log X_{t_n} - \log X_{t_{n-1}}$$

are iid random variables.

(b) Identify the limit of $E[\log X_{t_m}]$ and $\mathrm{Var}[\log X_{t_m}]$ when $t_m \to t$.

(c) In light of the central limit theorem, assume that as $t_m \to t$, X_{t_m} converges to a normal distribution. Argue that the limit distribution is the same as the distribution of $\log S_t$.

(d) Explain intuitively that $\{X_{t_m}\}$ converges to $\{S_t\}$ on interval $[0, T]$.

MATLAB® Problems

3.A Write a function using the "function" command to price call and put options with the same strike price and maturity, via the binomial tree method. Save this function as a ".m" file. The function should include the following parameters as input:

$$\begin{aligned}
u &= \text{factor by which the stock price moves up in each period,} \\
d &= \text{factor by which the stock price moves down in each period,} \\
N &= \text{number of periods,} \\
S_0 &= \text{initial stock price,} \\
r &= \text{risk-free interest rate,} \\
T &= \text{maturity,} \\
K &= \text{strike price.}
\end{aligned}$$

The output of the function should be the call option price and the put option price. Observe that in a binomial tree model, the possible values of the stock price at the N-th period are

$$S_0 u^i d^{N-i}, \quad i = N, N-1, ..., 0.$$

In the programming you may want to use matrices. There are various MATLAB commands to initialize a variable as a matrix. For example,

"zeros(m,n)": returns an $m \times n$ matrix with each entry 0,

"ones(m,n)": returns an $m \times n$ matrix with each entry 1,

"eye(n)": returns an $n \times n$ identity matrix.

3.B Suppose that the stock price S is a geometric Brownian motion under the risk-neutral probability measure:

$$S_t = S_0 \exp\left\{ \left(r - \frac{1}{2}\sigma^2 \right) t + \sigma W_t \right\}$$

with the initial price $S_0 = 20$. Use the binomial tree approximation with $N = 30, 60, 120$ periods, respectively, to price the call and put options with strike price $K = 20$ and maturity $T = 1$ year, assuming that the risk-free interest rate is $r = 8\%$ annually and $\sigma = 0.3$. Compare your results with the theoretical Black–Scholes prices, namely, BLS_Call and BLS_Put.

Chapter 4

Monte Carlo Simulation

Monte Carlo simulation is a very flexible and powerful tool for estimating integrals and expected values. Since most of the quantitative analysis in finance or risk management involves computing quantities that are indeed expected values, Monte Carlo simulation is widely used in the financial industry. This chapter aims to give a quick introduction to Monte Carlo simulation, as well as its pros and cons.

4.1 Basics of Monte Carlo Simulation

Consider the generic question of estimating the expected value of a function of some random variable X:

$$\mu = E[h(X)].$$

A plain Monte Carlo simulation scheme can be roughly divided into two steps:

1. Generate samples, or independent identically distributed (iid) random variables $X_1, X_2, \ldots, X_n$, that have the same distribution as X.

2. The estimate of the expected value μ is defined to be the sample average
$$\hat{\mu} = \frac{1}{n}\left[h(X_1) + h(X_2) + \cdots + h(X_n)\right].$$

Sometimes we simply refer to the samples $X_1, X_2, \ldots, X_n$ as *iid copies* of X. The number of samples n is the *sample size*, which is usually chosen to be a large number. It should be noted that μ, the quantity we wish to estimate, is an unknown *fixed* number, whereas the Monte Carlo estimate $\hat{\mu}$ is a *random variable*. The value of $\hat{\mu}$ will vary depending on the samples.

It is possible to design many different Monte Carlo simulation algorithms for estimating the same expected value μ. We briefly mention a couple of alternatives.

(a) *Importance Sampling:* Assuming that X admits a density f, we can write

$$\mu = \int_{\mathbb{R}} h(x)f(x)\,dx.$$

Consider an arbitrary nonzero density function $g(x)$. It follows that

$$\mu = \int_{\mathbb{R}} h(x)\frac{f(x)}{g(x)} \cdot g(x)\,dx = E\left[h(Y)\frac{f(Y)}{g(Y)}\right],$$

where Y is a random variable with density g. The corresponding Monte Carlo estimate for μ is

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} h(Y_i)\frac{f(Y_i)}{g(Y_i)},$$

where $Y_1, Y_2, \ldots, Y_n$ are iid copies of Y.

(b) *Control Variates:* Suppose that there is a random variable Y such that $E[Y] = 0$. Then one can write

$$\mu = E[h(X) + Y].$$

The corresponding Monte Carlo estimate for μ is

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n}[h(X_i) + Y_i],$$

where $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ are iid copies of (X, Y).

All these different Monte Carlo schemes can be generically described as follows. Let H be a random variable such that $\mu = E[H]$. Then the corresponding Monte Carlo estimate is given by

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} H_i,$$

where $H_1, H_2, \ldots, H_n$ are iid copies of H.

The underlying principle for Monte Carlo simulation is the strong law of large numbers. That is, as the sample size n tends to infinity,

$$\hat{\mu} = \frac{1}{n}\left(H_1 + H_2 + \cdots + H_n\right) \to E[H] = \mu$$

with probability one. Therefore, the estimate $\hat{\mu}$ is expected to be close to the true value μ when n is large.

Remark 4.1. Since a probability can be expressed as an expected value, Monte Carlo simulation is commonly used for estimating probabilities as well. For example, for any subset $A \subseteq \mathbb{R}$, one can write

$$\mathbb{P}(X \in A) = E\left[h(X)\right],$$

where h is an indicator function defined by

$$h(x) = \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{otherwise.} \end{cases}$$

In this case, the plain Monte Carlo estimate for $\mathbb{P}(X \in A)$ is just the proportion of samples that fall into set A.

4.2 Standard Error and Confidence Interval

Let μ be the unknown quantity of interest and H a random variable such that $\mu = E[H]$. A Monte Carlo estimate for μ is

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} H_i,$$

where $H_1, H_2, \ldots, H_n$ are iid copies of H. As we have mentioned, the strong law of large numbers asserts that $\hat{\mu}$ is close to μ when n is large. But how close? Since $\hat{\mu}$ is a random variable, so is the error $\hat{\mu} - \mu$. Therefore, what we are really looking for is the distribution of this error, not the error bound in the usual sense.

Denote by σ_H^2 the variance of H. It follows from the central limit theorem that as $n \to \infty$,

$$\frac{H_1 + H_2 + \cdots + H_n - n\mu}{\sigma_H\sqrt{n}} = \frac{\sqrt{n}(\hat{\mu} - \mu)}{\sigma_H}$$

converges to the standard normal distribution. That is, for every $a \in \mathbb{R}$,

$$\mathbb{P}\left\{ \frac{\sqrt{n}(\hat{\mu} - \mu)}{\sigma_H} \leq a \right\} \to \Phi(a). \tag{4.1}$$

In other words, the error $\hat{\mu} - \mu$ is approximately normally distributed with mean 0 and variance σ_H^2/n. This asymptotic analysis also produces *confidence intervals* for the Monte Carlo estimate $\hat{\mu}$. More precisely, it follows from (4.1) that the $1 - \alpha$ confidence interval for μ is approximately

$$\hat{\mu} \pm z_{\alpha/2} \frac{\sigma_H}{\sqrt{n}},$$

where $z_{\alpha/2}$ is defined by the equation $\Phi(-z_{\alpha/2}) = \alpha/2$. Confidence intervals are *random* intervals. A $1 - \alpha$ confidence interval has probability $1 - \alpha$ of covering the true value μ. Note that the width of a confidence interval decreases as the sample size increases. If one quadruples the sample size, the width is reduced by half.

In practice, the standard deviation σ_H is rarely known. Instead, the sample standard deviation

$$s_H = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (H_i - \hat{\mu})^2} = \sqrt{\frac{1}{n-1} \left(\sum_{i=1}^{n} H_i^2 - n\hat{\mu}^2 \right)}$$

is used in place of σ_H. This substitution is appropriate since s_H converges to σ_H with probability one when the sample size n tends to infinity, and hence the central limit theorem still holds if σ_H is replaced by s_H. The empirical $1 - \alpha$ confidence interval thus becomes

$$\hat{\mu} \pm z_{\alpha/2} \frac{s_H}{\sqrt{n}}.$$

The quantity $s_H/\sqrt{n}$ is often said to be the *standard error* of $\hat{\mu}$. That is,

$$\text{S.E.} = \sqrt{\frac{1}{n(n-1)} \left(\sum_{i=1}^{n} H_i^2 - n\hat{\mu}^2 \right)}.$$

Therefore, the commonly used 95% confidence interval is just the estimate $\hat{\mu}$ plus/minus twice the standard error (note that $z_{0.025} \approx 2$). In Monte Carlo simulation, it is customary to report both the estimate and the standard error.

Remark 4.2. The variance of H determines the width of a confidence interval and in some sense, the size of the error $\hat{\mu} - \mu$. Given a fixed sample size, the smaller the variance, the tighter the confidence interval, and hence the more accurate the estimate. This leads to the following criterion: *when comparing two estimates, the one with the smaller variance is more efficient.* Naturally, such a statement is a great simplification of the more scientific efficiency criteria that also take into consideration the computational effort of generating samples [11]. Nonetheless, it is a valuable guiding principle and will be used throughout the book to analyze the efficiency of various Monte Carlo schemes.

Remark 4.3. The Monte Carlo estimates that we have discussed so far are all *unbiased*, that is,

$$E[\hat{\mu}] = \mu.$$

Unbiasedness is a desirable property, but it is not always attainable. For example, consider estimating the price of a lookback call option with payoff

$$\left(\max_{0 \le t \le T} S_t - K\right)^+$$

at maturity T. Except for some rare occasions, it is impossible to exactly simulate the maximum of a continuous time sample path. Instead, one often uses the discrete time analogue

$$\max_{i=1,\dots,m} S_{t_i}, \quad t_i = iT/m$$

as an approximation. The plain Monte Carlo estimate for the price is just the sample mean of iid copies of

$$e^{-rT}\left(\max_{i=1,\cdots,m} S_{t_i} - K\right)^+$$

under the risk-neutral probability measure. This estimate has a negative bias since

$$\max_{i=1,\dots,m} S_{t_i} \le \max_{0 \le t \le T} S_t.$$

Note that the bias is very different from the random error of a Monte Carlo estimate. While the latter decreases when the sample size becomes larger, the bias can only be reduced by increasing the discretization parameter m.

4.3 Examples of Monte Carlo Simulation

We first study a few simple examples to illustrate the basic structure and techniques of Monte Carlo simulation. For each example, we report not only the estimate but also the standard error because the latter indicates how accurate the former is.

Example 4.1. Simulate W_T. Consider the problem of estimating the price of a call option under the assumption that the underlying stock price is a geometric Brownian motion.

SOLUTION: Recall that the price of a call option with strike price K and maturity T is

$$v = E[e^{-rT}(S_T - K)^+],$$

where r is the risk-free interest rate. The expected value is taken under the risk-neutral probability measure, where the stock price is a geometric Brownian motion with drift r:

$$S_T = S_0 \exp\left\{\left(r - \frac{1}{2}\sigma^2\right)T + \sigma W_T\right\}.$$

In order to generate samples of the option payoff, it suffices to generate samples of W_T. Since W_T is normally distributed with mean 0 and variance T, one can write $W_T = \sqrt{T}Z$ for some standard normal random variable Z.

Pseudocode:

for $i = 1, 2, \ldots, n$

  generate Z_i from $N(0,1)$

  set $Y_i = S_0 \exp\left\{\left(r - \frac{1}{2}\sigma^2\right)T + \sigma\sqrt{T}Z_i\right\}$

  set $X_i = e^{-rT}(Y_i - K)^+$

compute the estimate $\hat{v} = \frac{1}{n}(X_1 + X_2 + \cdots + X_n)$

compute the standard error S.E. $= \sqrt{\frac{1}{n(n-1)}\left(\sum_{i=1}^{n} X_i^2 - n\hat{v}^2\right)}.$

The simulation results are reported in Table 4.1 with the parameters given by

$$S_0 = 50, \;\; r = 0.05, \;\; \sigma = 0.2, \;\; T = 1.$$

For comparison, the theoretical values are calculated from the Black–Scholes formula BLS_Call; see Example 2.1.

Table 4.1: Monte Carlo simulation for call options

	Sample size $n = 2500$			Sample size $n = 10000$		
Strike price K	40	50	60	40	50	60
Theoretical value	12.2944	5.2253	1.6237	12.2944	5.2253	1.6237
M.C. Estimate	12.2677	5.2992	1.6355	12.3953	5.2018	1.6535
S.E.	0.1918	0.1468	0.0873	0.0964	0.0727	0.0438

Note that the standard errors of the estimates drop roughly 50% when the sample size is quadrupled. ■

Example 4.2. Simulate a Brownian Motion Sample Path. Consider a discretely monitored average price call option whose payoff at maturity T is

$$\left(\frac{1}{m} \sum_{i=1}^{m} S_{t_i} - K \right)^+ ,$$

where $0 < t_1 < \cdots < t_m = T$ are a fixed set of dates. Assume that under the risk-neutral probability measure,

$$S_t = S_0 \exp\left\{ \left(r - \frac{1}{2}\sigma^2 \right) t + \sigma W_t \right\}.$$

Estimate the price of the option.

SOLUTION: The key issue is to generate iid copies of the discrete time sample path $(S_{t_1}, S_{t_2}, \ldots, S_{t_m})$. They should be simulated sequentially since

$$S_{t_{i+1}} = S_{t_i} \exp\left\{ \left(r - \frac{1}{2}\sigma^2 \right) (t_{i+1} - t_i) + \sigma(W_{t_{i+1}} - W_{t_i}) \right\},$$

and $(W_{t_1} - W_{t_0}, \ldots, W_{t_m} - W_{t_{m-1}})$ are *independent* normal random variables. Below is the pseudocode for generating *one* sample path.

Pseudocode:

for $i = 1, \ldots, m$

generate Z_i from $N(0, 1)$

set $S_{t_i} = S_{t_{i-1}} \exp\left\{ \left(r - \frac{1}{2}\sigma^2 \right) (t_i - t_{i-1}) + \sigma\sqrt{t_i - t_{i-1}} Z_i \right\}$

compute the discounted payoff $X = e^{-rT} \left(\frac{1}{m} \sum_{i=1}^{m} S_{t_i} - K \right)^+.$

The Monte Carlo algorithm will repeat the above steps n times to obtain n sample paths and n iid copies of X, say $X_1, \ldots, X_n$. The estimate and its standard error are given by

$$\begin{aligned} \hat{v} &= \frac{1}{n}(X_1 + X_2 + \cdots + X_n), \\ \text{S.E.} &= \sqrt{\frac{1}{n(n-1)}\left(\sum_{k=1}^{n} X_k^2 - n\hat{v}^2\right)}. \end{aligned}$$

The simulation results are reported in Table 4.2. The parameters are set to be

$$S_0 = 50,\ r = 0.05,\ \sigma = 0.2,\ T = 1,\ m = 12,\ t_i = \frac{i}{12}.$$

Explicit formula for the option price is not available.

Table 4.2: Monte Carlo simulation for average price call options

	Sample size $n = 2500$			Sample size $n = 10000$		
Strike price K	40	50	60	40	50	60
M.C. Estimate	10.7487	3.0730	0.3837	10.8810	3.0697	0.3490
S.E.	0.1183	0.0846	0.0332	0.0597	0.0422	0.0152

As in the previous example, the standard errors of the estimates reduce roughly by half when the sample size increases fourfold. ■

Example 4.3. Simulate 2D Jointly Normal Random Vectors. Estimate the price of a spread call option whose payoff at maturity T is

$$(X_T - Y_T - K)^+,$$

where $\{X_t\}$ and $\{Y_t\}$ are the prices of two underlying assets. Assume that under the risk-neutral probability measure,

$$\begin{aligned} X_t &= X_0 \exp\left\{\left(r - \frac{1}{2}\sigma_1^2\right)t + \sigma_1 W_t\right\}, \\ Y_t &= Y_0 \exp\left\{\left(r - \frac{1}{2}\sigma_2^2\right)t + \sigma_2 B_t\right\}, \end{aligned}$$

where (W, B) is a two-dimensional Brownian motion with covariance matrix

$$\Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}.$$

SOLUTION: By assumption (W_T, B_T) is a jointly normal random vector with mean 0 and covariance matrix

$$T \cdot \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}.$$

If the two Brownian motions W and B are uncorrelated (that is, if $\rho = 0$), then the simulation is straightforward—one could just sample two independent standard normal random variables Z_1 and Z_2, and let

$$\begin{aligned} X_T &= X_0 \exp\left\{\left(r - \frac{1}{2}\sigma_1^2\right) T + \sigma_1 \sqrt{T} Z_1\right\}, \\ Y_T &= Y_0 \exp\left\{\left(r - \frac{1}{2}\sigma_2^2\right) T + \sigma_2 \sqrt{T} Z_2\right\}. \end{aligned}$$

When $\rho \neq 0$, the simulation is more involved. Recall that if Z_1 and Z_2 are two independent standard normal random variables, then

$$Z = \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix}$$

is a two-dimensional standard normal random vector (see Appendix A). Therefore, for any 2×2 matrix $C = [C_{ij}]$, the random vector

$$R = CZ = \begin{bmatrix} C_{11} Z_1 + C_{12} Z_2 \\ C_{21} Z_1 + C_{22} Z_2 \end{bmatrix}$$

is jointly normal with mean 0 and covariance matrix CC'. If there exists a matrix C such that

$$CC' = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}, \tag{4.2}$$

then $\sqrt{T} R$ will have the same distribution as (W_T, B_T) and we can let

$$\begin{aligned} X_T &= X_0 \exp\left\{\left(r - \frac{1}{2}\sigma_1^2\right) T + \sigma_1 \sqrt{T} R_1\right\}, \\ Y_T &= Y_0 \exp\left\{\left(r - \frac{1}{2}\sigma_2^2\right) T + \sigma_2 \sqrt{T} R_2\right\}. \end{aligned}$$

There are many choices of C that satisfy (4.2). A particularly convenient one is when C is a lower triangular matrix:

$$C = \begin{bmatrix} C_{11} & 0 \\ C_{21} & C_{22} \end{bmatrix}.$$

In this case

$$CC' = \begin{bmatrix} C_{11}^2 & C_{11}C_{21} \\ C_{21}C_{11} & C_{21}^2 + C_{22}^2 \end{bmatrix} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}.$$

Taking $C_{11} = 1$, we arrive at

$$C = \begin{bmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{bmatrix}.$$

Below is the pseudocode for estimating the price of the spread call option.

Pseudocode:

set $C_{11} = 1, C_{21} = \rho, C_{22} = \sqrt{1-\rho^2}$
for $i = 1, 2, \ldots, n$
  generate Z_1 and Z_2 from $N(0,1)$
  set $R_1 = C_{11}Z_1$ and $R_2 = C_{21}Z_1 + C_{22}Z_2$
  set $X_i = X_0 \exp\left\{\left(r - \frac{1}{2}\sigma_1^2\right)T + \sigma_1\sqrt{T}R_1\right\}$
  set $Y_i = Y_0 \exp\left\{\left(r - \frac{1}{2}\sigma_2^2\right)T + \sigma_2\sqrt{T}R_2\right\}$
  compute the discounted payoff $H_i = e^{-rT}(X_i - Y_i - K)^+$
compute the estimate $\hat{v} = \frac{1}{n}(H_1 + H_2 + \cdots + H_n)$
compute the standard error S.E. $= \sqrt{\frac{1}{n(n-1)}\left(\sum_{i=1}^{n} H_i^2 - n\hat{v}^2\right)}$.

The simulation results are reported in Table 4.3. The parameters are given by

$$X_0 = 50, \ \ Y_0 = 45, \ \ r = 0.05, \ \ \sigma_1 = 0.2, \ \ \sigma_2 = 0.3, \ \ \rho = 0.5, \ \ T = 1.$$

Table 4.3: Monte Carlo simulation for spread call options

	Sample size $n = 2500$			Sample size $n = 10000$		
Strike price K	0	5	10	0	5	10
M.C. Estimate	7.9593	4.9831	2.7024	7.9019	5.0056	2.8861
S.E.	0.1680	0.1330	0.0990	0.0838	0.0683	0.0521

The matrix C is a special case of the *Cholesky factorization*. It can be generalized to simulate higher dimensional jointly normal random vectors. We will take the discussion further in Chapter 5. ■

We should give a couple of examples to demonstrate that neither does every Monte Carlo estimate take the form of sample average, nor is the construction of an efficient Monte Carlo scheme always automatic. The first example is concerned with estimating value-at-risk, which is essentially a quantile of an unknown distribution. The difficulty there is the construction of confidence intervals. The second example is about estimating the probability of a large loss in a credit risk model. Such probabilities are usual very small, which renders the plain Monte Carlo scheme quite inefficient or even infeasible.

Example 4.4. Simulate Value-at-Risk. Denote by X_i the daily return of a portfolio. Assume that $X = \{X_1, X_2, \ldots\}$ is a *Markov chain*, that is, the conditional distribution of X_{i+1} given $(X_i, X_{i-1}, \ldots, X_1)$ only depends on X_i for each i. Let X be autoregressive conditional heteroskedastic (ARCH) in the sense that given $X_i = x$, X_{i+1} is normally distributed with mean zero and variance $\beta_0 + \beta_1 x^2$ for some $\beta_0 > 0$ and $0 < \beta_1 < 1$. The total return within an m-day period is

$$S = \sum_{i=1}^{m} X_i.$$

Assuming that X_1 is a standard normal random variable, estimate the value-at-risk at the confidence level $1 - p$.

SOLUTION: The simulation of a Markov chain is done sequentially as the distribution of X_{i+1} depends on the value of X_i. Below is the pseudocode for generating one sample of S:

Pseudocode for one sample of the total return S:

generate X_1 from $N(0,1)$
for $i = 2, 3, \ldots, m$
    generate Z from $N(0,1)$
    set $X_i = \sqrt{\beta_0 + \beta_1 X_{i-1}^2} \cdot Z$
set $S = \sum_{i=1}^{m} X_i$.

The Monte Carlo scheme will repeat the above steps n times to generate n iid copies of S, say $S_1, \ldots, S_n$. Recall that the value-at-risk (VaR) at the confidence level $1 - p$ is defined by

$$\mathbb{P}(S \leq -\mathrm{VaR}) = p.$$

Thus it makes sense to estimate VaR by a number $\hat{x}$ such that the fraction of samples at or below level $-\hat{x}$, i.e.,

$$\frac{\text{number of samples among } \{S_1, \ldots, S_n\} \text{ at or below } -\hat{x}}{n},$$

is close to p. Consider the *order statistics* $\{S_{(1)}, \ldots, S_{(n)}\}$, which is a permutation of $\{S_1, \ldots, S_n\}$ in an increasing order:

$$S_{(1)} \leq S_{(2)} \leq \cdots \leq S_{(n)}.$$

A common choice for $\hat{x}$ is to let $k = [np]$ (the integer part of np) and set

$$-\hat{x} = S_{(k)}.$$

Note that $\hat{x}$ is neither the sample average nor an unbiased estimate.

The difficulty here lies in the construction of confidence intervals for this estimate. Even though there are various approaches, we will only describe a simple method based on the order statistics, which works best when the sample size n is large. Suppose that one is interested in a $1-\alpha$ confidence interval. The goal is to find integers $k_1 < k_2$ such that

$$\mathbb{P}(S_{(k_1)} \leq -\text{VaR} < S_{(k_2)}) = 1 - \alpha. \tag{4.3}$$

If this can be done exactly or approximately, then $(-S_{(k_2)}, -S_{(k_1)}]$ serves as a $1-\alpha$ confidence interval. To this end, observe that the left-hand-side of (4.3) is just $\mathbb{P}(k_1 \leq Y < k_2)$, where Y denotes the number of samples that are less than or equal to $-$VaR. Since Y is binomial with parameters (n, p), this probability equals

$$\sum_{j=k_1}^{k_2-1} \binom{n}{j} (1-p)^j p^{n-j}.$$

When n is large, a direct evaluation of this summation is difficult. However, one can use the normal approximation of binomial distributions (see Exercise 4.1) to conclude that

$$\mathbb{P}(S_{(k_1)} \leq -\text{VaR} < S_{(k_2)}) \approx 1 - \Phi\left(\frac{k_1 - np}{\sqrt{np(1-p)}}\right) - \Phi\left(-\frac{k_2 - np}{\sqrt{np(1-p)}}\right),$$

which leads to the following choice of k_1 and k_2 (taking integer part if necessary):

$$k_1 = np - \sqrt{np(1-p)}z_{\alpha/2}, \quad k_2 = np + \sqrt{np(1-p)}z_{\alpha/2}.$$

The simulation results are reported in Table 4.4, given $\beta_0 = \beta_1 = 0.5$ and $m = 10$. We estimate the value-at-risk at the confidence level $1-p$ and give a 95% confidence interval for $p = 0.05, 0.02, 0.01$, respectively. The sample size is $n = 10000$.

Table 4.4: Monte Carlo simulation for value-at-risk

p	0.05	0.02	0.01
k	500	200	100
(k_1, k_2)	(458,542)	(173, 227)	(80,120)
Estimate $\hat{x}$	4.9978	6.4978	7.7428
95% C.I.	[4.8727, 5.1618]	[6.3381,6.6835]	[7.4650, 8.0430]

The probability in (4.3) does not depend on the underlying distribution of S. For this reason, the confidence intervals constructed here are said to be *distribution free* or *nonparametric*. ■

Example 4.5. Difficulty in Estimating Small Probabilities. Consider the one-factor portfolio credit risk model in Example 1.9. Let c_k denote the loss from the default of the k-th obligor. Then the total loss is

$$L = \sum_{k=1}^{m} c_k 1_{\{X_k \geq x_k\}}.$$

Estimate the probability that L exceeds a given large threshold h.

SOLUTION: The simulation algorithm is straightforward.

Pseudocode:

for $i = 1, 2, \ldots, n$

  generate independent samples Z, $\varepsilon_1, \ldots, \varepsilon_m$ from $N(0,1)$

  compute $X_k = \rho_k Z + \sqrt{1-\rho_k^2}\,\varepsilon_k$ for $k = 1, \ldots, m$

  compute $L = \sum_{k=1}^{m} c_k 1_{\{X_k \geq x_k\}}$

  set $H_i = 1$ if $L > h$; set $H_i = 0$ otherwise

compute the estimate $\hat{v} = \frac{1}{n}(H_1 + H_2 + \cdots + H_n)$

compute the standard error S.E. $= \sqrt{\frac{1}{n(n-1)}\left(\sum_{i=1}^{n} H_i^2 - n\hat{v}^2\right)}$.

Set $m = 3, c_1 = 2, c_2 = 1, c_3 = 4, \rho_1 = 0.2, \rho = 0.5, \rho_3 = 0.8$. The levels are assumed to be $x_1 = 1$, $x_2 = 1$, $x_3 = 2$. The simulation results are reported in Table 4.5, where we have also included

$$\textit{Empirical Relative Error} = \frac{\text{Standard Error}}{\text{Estimate}}.$$

Table 4.5: Monte Carlo simulation for a credit risk model

	Sample size $n = 2500$			Sample size $n = 10000$		
Threshold h	1	2	4	1	2	4
M.C. Estimate	0.1840	0.0476	0.0136	0.1780	0.0528	0.0150
S.E.	0.0078	0.0043	0.0023	0.0038	0.0022	0.0012
R.E.	4.24%	9.03%	16.91%	2.13%	4.17%	8.00%
	Sample size $n = 10000$			Sample size $n = 40000$		
Threshold h	6	8	10	6	8	10
M.C. Estimate	0.0028	0.0000	0.0000	0.0032	0.0000	0.0000
S.E.	0.0005	0.0000	0.0000	0.0003	0.0000	0.0000
R.E.	18.87%	NaN	NaN	8.76%	NaN	NaN

An interesting observation is that as the threshold h gets larger, the probability gets smaller and the quality of the estimates deteriorates. The reason is that only for a very small fraction of samples will the total loss exceed the large threshold h. With so few hits, the estimate cannot be accurate. Therefore, the plain Monte Carlo is inefficient for estimating small probabilities; see also Exercise 4.2. Clearly, a more efficient Monte Carlo scheme is needed for estimating such small quantities. ■

4.4 Summary

Monte Carlo simulation is a very useful tool for the quantitative analysis of financial models. It is well suited for parallel computing, and its flexibility can accommodate complicated models that are otherwise inaccessible.

Monte Carlo simulation is a random algorithm. A different run of simulation will yield a different estimate. It is very different from those deterministic numerical schemes for evaluating integrals, which are usually designed for problems of low dimensions. Since many of the pricing problems in financial engineering are intrinsically problems of evaluating integrals of large or infinite dimensions, these deterministic algorithms are not well suited for this type of tasks. On the contrary, the central limit theorem asserts that the standard error of a Monte Carlo estimate decays in the order of $O(1/\sqrt{n})$ with respect to the sample size n, *regardless of the dimension*.

But the Monte Carlo method is not without shortcomings. Even though it is often possible to improve the efficiency of a given Monte Carlo scheme, little can be done to accelerate the convergence above the rate $O(1/\sqrt{n})$. A large sample size is often required in order to achieve a desirable accuracy level.

Finally, the design of Monte Carlo schemes is not as straightforward as one might think. This is especially true when the quantity of interest is associated with events of small probabilities, which is a common scenario in risk analysis. Here one has to be very cautious, since it is not uncommon that a seemingly very accurate estimate (i.e., an estimate with a very small standard error; see Exercise 4.F, for example) can be far off from the true value. Theoretical justification of a Monte Carlo scheme should be provided whenever possible.

Exercises

Pen-and-Paper Problems

4.1 Suppose that $X_1, \ldots, X_n$ are iid Bernoulli random variables with parameter p. Then

$$S_n = X_1 + \ldots + X_n$$

is a binomial random variable with parameters (n, p). Use the central limit theorem to explain that, when n is large, the binomial distribution with parameters (n, p) can be approximated by the normal distribution with mean np and variance $np(1-p)$. That is, for $x \in \mathbb{R}$,

$$\mathbb{P}\left(\frac{S_n - np}{\sqrt{np(1-p)}} \leq x\right) \approx \Phi(x).$$

4.2 Suppose that $X_1, \ldots, X_n$ are iid Bernoulli random variables with unknown parameter p. The average

$$\bar{X} = \frac{1}{n}(X_1 + \cdots + X_n)$$

can be used to estimate p.

(a) What are the expected value and standard deviation of $\bar{X}$?

(b) Write down a 95% confidence interval for p.

(c) Define the relative error as

$$\textit{Relative Error} = \frac{\text{Standard Deviation of } \bar{X}}{\text{Expected Value of } \bar{X}}.$$

How large should n be so that the relative error is at most 5%?

4.3 In a class of 100 students, each student is asked to run a simulation to estimate the price of an option and provide a 95% confidence interval, independently from others. What is the distribution of the number of confidence intervals that cover the true value of the option price? Is it likely that all the confidence intervals cover the true value of the option price?

4.4 Let X be a random variable with density $f(x)$. It is easy to see that estimating the integral

$$\int_{\mathbb{R}} h(x) f(x)\, dx$$

amounts to estimating the expected value $E[h(X)]$. Use this observation to design Monte Carlo schemes for estimating the following integrals:

$$\int_0^\infty e^{-x} \sin(x)\, dx, \quad 4\int_0^1 \sqrt{1-x^2}\, dx, \quad \int_0^\infty \frac{1}{\sqrt{x}} e^{-x^2}\, dx.$$

Write down the pseudocode (it should report both the estimate and the standard error).

4.5 Let $\hat{\mu}$ be an estimate for some unknown quantity μ. The difference $E[\hat{\mu}] - \mu$ is said to be the *bias* of $\hat{\mu}$. The *mean square error* (M.S.E.) of $\hat{\mu}$ is defined to be $E[(\hat{\mu} - \mu)^2]$. Show that

$$\text{M.S.E.} = (\text{Bias of } \hat{\mu})^2 + \text{Var}[\hat{\mu}].$$

In general, it is beneficial to allocate the computational budget in order to balance bias and variance. The rule of thumb is to make the bias and the standard deviation of the estimate roughly the same order [11].

4.6 Samples of a random vector (X, Y) can often be drawn in a sequential manner: one first samples X from its marginal distribution and then samples Y from its conditional distribution given X. Explain that it is essentially what has been done in Example 4.3 to simulate the jointly normal random vector $R = (R_1, R_2)$ with mean 0 and covariance matrix Σ.

4.7 Let $W = \{W_t : t \geq 0\}$ be a standard Brownian motion. Consider the random vector (W_t, M_t), where M_t is the running maximum of W by time t, that is,

$$M_t = \max_{0 \leq s \leq t} W_s.$$

Show that the conditional distribution of M_T given $W_T = x$ is identical to the distribution of

$$\frac{1}{2}\left(x + \sqrt{x^2 + 2TY}\right),$$

where Y is an exponential random variable with rate one. Use this result to design a scheme to draw samples from

(a) (W_T, M_T);

(b) $(W_{t_1}, M_{t_1}, W_{t_2}, M_{t_2}, \ldots, W_{t_m}, M_{t_m})$ given $0 < t_1 < t_2 < \cdots < t_m = T$;

(c) (B_T, M_T^B), where $B_t = W_t + \theta t$ is Brownian motion with drift θ and M^B is the running maximum of B, that is,

$$M_t^B = \max_{0 \leq s \leq t} B_s;$$

(d) $(B_{t_1}, M_{t_1}^B, B_{t_2}, M_{t_2}^B, \ldots, B_{t_m}, M_{t_m}^B)$ given $0 < t_1 < t_2 < \cdots < t_m = T$.

Write down the pseudocode. *Hint:* Recall Exercise 2.12 for (c) and (d).

4.8 Consider the following two Monte Carlo schemes for estimating $\mu = E[h(X) + f(X)]$. The total sample size is $2n$ in both schemes.

(a) *Scheme I: Use Same Random Numbers.* Generate $2n$ iid copies of X, say $\{X_1, \ldots, X_{2n}\}$. The estimate is

$$\hat{\mu}_1 = \frac{1}{2n} \sum_{i=1}^{2n} [h(X_i) + f(X_i)].$$

(b) *Scheme II: Use Different Random Numbers.* Write $\mu = E[h(X)] + E[f(X)]$ and estimate the two expected values separately. Generate $2n$ iid copies of X, say $\{X_1, \ldots, X_n, Y_1, \ldots, Y_n\}$. The estimate is

$$\hat{\mu}_2 = \frac{1}{n}\sum_{i=1}^{n} h(X_i) + \frac{1}{n}\sum_{i=1}^{n} f(Y_i).$$

Show that both $\hat{\mu}_1$ and $\hat{\mu}_2$ are unbiased estimates for μ, but $\hat{\mu}_1$ always has a smaller variance because

$$\mathrm{Var}[\hat{\mu}_1] - \mathrm{Var}[\hat{\mu}_2] = -\frac{1}{2n}\mathrm{Var}[h(X) - f(X)] \leq 0.$$

This exercise shows that if one wants to estimate the price of a financial instrument such as straddle (the combination of a call and a put with the same strike price), it is beneficial to use the same random numbers to estimate its price altogether rather than estimate the call and the put prices separately.

MATLAB® Problems

In Exercises 4.A – 4.C, assume that the underlying stock price is a geometric Brownian motion under the risk-neutral probability measure:

$$S_t = S_0 \exp\left\{\left(r - \frac{1}{2}\sigma^2\right)t + \sigma W_t\right\},$$

where W is a standard Brownian motion and r is the risk-free interest rate.

4.A Write a function to estimate the price of a binary put option with maturity T and payoff

$$X = 1_{\{S_T \leq K\}}.$$

The input parameters are S_0, r, σ, K, T, and the sample size n. The function should output the estimate of the price and its standard error. Report your results for

$$S_0 = 30,\ r = 0.05,\ \sigma = 0.2,\ K = 30,\ T = 0.5,\ n = 10000.$$

Compare with the theoretical value of the option price.

4.B Write a function to estimate the price of a discretely monitored down-and-out call option with maturity T and payoff

$$(S_T - K)^+ \cdot 1_{\{\min(S_{t_1}, S_{t_2}, \ldots, S_{t_m}) \geq b\}}.$$

The monitoring dates $t_1, \ldots, t_m$ are prespecified and $0 < t_1 < \cdots < t_m = T$. The function should have the following parameters as input

$$S_0,\ r,\ \sigma,\ T,\ K,\ b,\ m,\ (t_1, \ldots, t_m),\ n,$$

where n denotes the sample size. The output of the function should include the estimate for the option price and the standard error. Report your results for

$$S_0 = 50,\ r = 0.10,\ \sigma = 0.2,\ T = 1,\ K = 50,\ b = 45,$$
$$m = 12,\ t_i = iT/m,\ n = 10000.$$

4.C Consider a lookback put option with floating strike price and maturity T, whose payoff is

$$X = \max_{0\le t\le T} S_t - S_T.$$

(a) Write a function to estimate the price of this lookback option, where the maximum of the stock price is approximated by

$$\max(S_{t_0}, S_{t_1}, \ldots, S_{t_m})$$

for some $0 = t_0 < t_1 < \ldots < t_m = T$. Let the input of the function be

$$S_0,\ r,\ \sigma,\ T,\ m,\ (t_0, t_1, \ldots, t_m),\ n,$$

where n is the sample size. Is the estimate unbiased? If it is not, is the bias positive or negative?

(b) Write a function that yields an unbiased estimate for the price of the lookback put option. The input of the function should be S_0, r, T, σ, n. *Hint:* Use Exercise 4.7 (a) and Lemma 2.2, or use Exercise 4.7 (c).

(c) Report your estimates and standard errors from (a) and (b) with the parameters given by

$$S_0 = 20,\ r = 0.03,\ \sigma = 0.2,\ T = 0.5,\ n = 10000.$$

For part (a) let $t_i = iT/m$ and $m = 10, 100, 1000$, respectively.

4.D Suppose that the stock price S is a geometric Brownian motion with jumps:

$$S_t = S_0 \exp\left\{\left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma W_t + \sum_{i=1}^{N_t} Y_i\right\},$$

where W is a standard Brownian motion, $N = \{N_t : t \ge 0\}$ is a Poisson process with rate λ, and Y_i's are iid normal random variables with distribution $N(0, \nu^2)$. Assume that W, N, $\{Y_i\}$ are independent. Write a function to estimate the probability

$$\mathbb{P}\left(\max_{1\le i\le m} S_{t_i} \ge b\right),$$

where $0 < t_1 < \cdots < t_m = T$ are prespecified dates and b is a given threshold. The function should have input parameters

$$S_0,\ \mu,\ \sigma,\ \lambda,\ \nu,\ b,\ T,\ m,\ (t_1, \ldots, t_m),\ n,$$

where n is the sample size. Report your estimate and its standard error for

$$S_0 = 50,\ \mu = 0.1,\ \sigma = 0.2,\ \lambda = 2,\ \nu = 0.3,\ b = 55,\ T = 1,$$

$$m = 50,\ t_i = iT/m,\ n = 10000.$$

4.E The setup is analogous to Example 4.4. Consider the following GARCH model: for every $i \geq 1$, $X_{i+1} = \sigma_{i+1} Z_{i+1}$, where Z_{i+1} is a standard normal random variable independent of $\{X_1, \ldots, X_i\}$ and

$$\sigma_{i+1}^2 = \beta_0 + \beta_1 X_i^2 + \beta_2 \sigma_i^2.$$

Assuming that X_1 is normally distributed as $N(0, \sigma_1^2)$ for some constant σ_1, write a function to estimate the value-at-risk of the total return

$$S = \sum_{i=1}^{m} X_i$$

at the confidence level $1 - p$. The input parameters of the function should be

$$\sigma_1,\ \beta_0,\ \beta_1,\ \beta_2,\ m,\ p,\ n,$$

where n denotes the sample size. Report your estimates and 95% confidence intervals for

$$\sigma_1 = 1,\ \beta_0 = 0.5,\ \beta_1 = 0.3,\ \beta_2 = 0.5,\ m = 10,\ n = 10000,$$

and $p = 0.05, 0.01$, respectively.

4.F Consider the problem of estimating $E[\exp\{\theta Z - \theta^2/2\}]$ for some constant θ and standard normal random variable Z. Use the plain Monte Carlo scheme with sample size $n = 1,000,000$. Report your simulation results for $\theta = 6$ and $\theta = 7$, respectively. Explain why your results are inconsistent with the theoretical value, which is one (see Exercise 1.7). *Hint:* The expected value can be written as

$$\int_{\mathbb{R}} e^{\theta x - \frac{1}{2}\theta^2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\, dx.$$

For what range of x does the majority of contribution to the integral come from? Can you trust the standard errors?