8.1 Stochastic Integrals

Stochastic integrals are integrals with respect to random processes. The most basic stochastic integral is an integral against a standard Brownian motion that takes the form

$$\int_0^T X_t \, dW_t, \tag{8.1}$$

where W is a standard Brownian motion and the integrand X is a general stochastic process. There are mild conditions one needs to impose on the integrand X so as to make the integral meaningful; see Remark 8.1. Throughout the rest of the book, these conditions are implicitly assumed.

Analogous to a classical integral, a stochastic integral is defined as the limit of certain approximating sums. More precisely, the stochastic integral (8.1) is defined as

$$\int_0^T X_t \, dW_t = \lim_{n\to\infty} \sum_{i=0}^{n-1} X_{t_i}(W_{t_{i+1}} - W_{t_i}), \tag{8.2}$$

where $0 = t_0 < t_1 < \cdots < t_n = T$ is a partition of the interval $[0, T]$. It should be mentioned that we always assume

$$\max_{0\le i\le n-1} (t_{i+1} - t_i) \to 0$$

as n tends to infinity for any partition. The choice of partition does not affect the limit in (8.2). However, it is often convenient to let $t_i = iT/n$ for each i to ease computation.

We list below some of the fundamental properties of stochastic integrals. While linearity is obvious from the definition, the martingale property and Itô isometry are less so; see Remark 8.2 for a formal argument.

1. **Linearity:** Given any $a, b \in \mathbb{R}$,

$$\int_0^T (aX_t + bY_t) \, dW_t = a \int_0^T X_t \, dW_t + b \int_0^T Y_t \, dW_t.$$

2. **Martingale property**:

$$E \int_0^T X_t \, dW_t = 0.$$

3. **Itô isometry**:

$$E\left(\int_0^T X_t\, dW_t\right)^2 = \int_0^T E[X_t^2]\, dt.$$

Another useful property is that when the integrand is deterministic, the stochastic integral is normally distributed. More precisely, we have the following lemma.

Lemma 8.1. *Assume that $f(t)$ is a deterministic function. Then the stochastic integral*

$$\int_0^T f(t)\, dW_t$$

is normally distributed with mean 0 *and variance* $\int_0^T f^2(t)\, dt$.

PROOF: By definition, this stochastic integral equals the limit of the approximating sum

$$S_n = \sum_{i=1}^{n-1} f(t_i)[W_{t_{i+1}} - W_{t_i}]$$

as n tends to infinity. It is easy to see that S_n is normally distributed with mean 0 and variance

$$\mathrm{Var}[S_n] = \sum_{i=1}^{n-1} f^2(t_i)[t_{i+1} - t_i].$$

Clearly,

$$\mathrm{Var}[S_n] \to \int_0^T f^2(t)\, dt$$

as n tends to infinity. The lemma follows readily. ■

Example 8.1. The purpose of this example is to illustrate the difference between stochastic integral and classical integral by explicitly computing

$$\int_0^T 2W_t\, dW_t.$$

Note that this integral is an exception rather than the norm as stochastic integrals can rarely be explicitly evaluated.

SOLUTION: Given any n, consider the partition $0 = t_0 < t_1 < \cdots < t_n = T$ with $t_i = iT/n$. For each $i = 0, 1, \ldots, n-1$, let

$$Z_{i+1} = \frac{W_{t_{i+1}} - W_{t_i}}{\sqrt{t_{i+1} - t_i}} = \frac{W_{t_{i+1}} - W_{t_i}}{\sqrt{T/n}}.$$

Then $Z_1, \ldots, Z_n$ are iid standard normal random variables. It follows from definition (8.2) and the strong law of large numbers that

$$\begin{aligned}
&\int_0^T 2W_t \, dW_t \\
&= \lim_n \sum_{i=0}^{n-1} 2W_{t_i}(W_{t_{i+1}} - W_{t_i}) \\
&= \lim_n \sum_{i=0}^{n-1} \left[(W_{t_{i+1}} + W_{t_i}) - (W_{t_{i+1}} - W_{t_i})\right] (W_{t_{i+1}} - W_{t_i}) \\
&= W_T^2 - \lim_n \sum_{i=0}^{n-1} (W_{t_{i+1}} - W_{t_i})^2 \\
&= W_T^2 - \lim_n \frac{T}{n} \sum_{i=1}^{n} Z_i^2 \\
&= W_T^2 - T.
\end{aligned}$$

Note that if we were to calculate this stochastic integral using classical calculus, we would arrive at the *wrong* conclusion that

$$\int_0^T 2W_t \, dW_t = \int_0^T d(W_t^2) = W_T^2.$$

The difference is due to the nonzero quadratic variation of Brownian motion, that is,

$$\lim_{n\to\infty} \sum_{i=0}^{n-1} (W_{t_{i+1}} - W_{t_i})^2 = T \neq 0. \tag{8.3}$$

In a classical integral, the corresponding quadratic variation term is zero. ■

Remark 8.1. In order to ensure the stochastic integral is well defined, some conditions must be imposed on the integrand X. For one, X has to be *adapted* or *nonanticipating*. This condition basically says that the value of X_t only depends on the information up until time t. In practice, if X_t represents certain decisions an investor makes at time t, the adaptedness of X

means that the investor cannot look into the future and can only make the decision based upon the historical information available to him. Another condition that is often imposed on X requires the expected value

$$E\int_0^T X_t^2\,dt$$

to be finite. These two conditions hold automatically in nearly all applications.

Remark 8.2. The martingale property and Itô isometry can be easily understood if one observes that X_{t_i} is independent of $\Delta W_j = W_{t_{j+1}} - W_{t_j}$ for any $j \geq i$. Indeed, by the adaptedness of X, X_{t_i} only depends on the historical information up until time t_i. On the other hand, the Brownian motion increment ΔW_j is independent of the historical information up to time t_j. The independence between X_{t_i} and ΔW_j follows since $t_j \geq t_i$. Consequently,

$$E\int_0^T X_t\,dW_t = \lim_n \sum_{i=0}^{n-1} E[X_{t_i}\Delta W_i] = \lim_n \sum_{i=0}^{n-1} E[X_{t_i}]\cdot E[\Delta W_i] = 0,$$

which is the martingale property. By the same token, for every i and $j > i$

$$E[X_{t_i}\Delta W_i]^2 = E[X_{t_i}^2]\cdot E[\Delta W_i]^2 = E[X_{t_i}^2](t_{i+1}-t_i),$$

$$E[X_{t_i}\Delta W_i\cdot X_{t_j}\Delta W_j] = E[X_{t_i}X_{t_j}\Delta W_i]\cdot E[\Delta W_j] = 0.$$

Therefore,

$$\begin{aligned} E\left(\int_0^T X_t\,dW_t\right)^2 &= \lim_n E\left[\sum_{i=0}^{n-1} X_{t_i}\Delta W_i\right]^2 \\ &= \lim_n \sum_{i=0}^{n-1} E[X_{t_i}^2](t_{i+1}-t_i) \\ &= \int_0^T E[X_t^2]\,dt, \end{aligned}$$

which is exactly the Itô isometry.

Remark 8.3. The stochastic integral defined by (8.2) is said to be the *Itô integral*. The characteristic of Itô integral is that in the approximating sum the integrand X is always evaluated at the left endpoint of each of the subintervals $[t_i, t_{i+1}]$. If we replace it by the right endpoint $X_{t_{i+1}}$ or the average $(X_{t_i} + X_{t_{i+1}})/2$, the limit in (8.2) will be different and the resulting integral is said to be the *backward Itô integral* or the *Stratonovich integral*, respectively. This is very different from the classical Riemann integral, where the integrand can be evaluated at any point of the subintervals without affecting the limit. The reason for this difference is again that the Brownian motion has nonzero quadratic variation (8.3). In financial engineering, stochastic integrals are predominantly Itô integrals.

8.2 Itô Formula

The most significant result in stochastic calculus is without any doubt the celebrated *Itô formula*. Analogous to the chain rule in classical calculus, it characterizes the dynamics of functions of continuous time stochastic processes such as Brownian motions and diffusions. Itô formula plays a very important role in financial engineering. For instance, since the value of a financial derivative is in general a smooth function of the underlying asset price, Itô formula is applicable and eventually connects the value of the derivative with the solution to a partial differential equation. We will briefly touch upon this topic at the end of this chapter.

8.2.1 The Basic Itô Formula

The basic Itô formula states that a smooth function of a standard Brownian motion is the summation of a stochastic integral and an ordinary integral. It will be subsumed by the more general Itô formulas that will appear later. However, it is probably the simplest setting to illustrate the key difference between stochastic calculus and classical calculus.

Theorem 8.2. *Suppose that $f : \mathbb{R} \to \mathbb{R}$ is a twice continuously differentiable function and W is a standard Brownian motion. Then*

$$f(W_T) = f(W_0) + \int_0^T f'(W_t)\, dW_t + \frac{1}{2}\int_0^T f''(W_t)\, dt.$$

More general versions of Itô formula usually require less stringent regularity conditions on f and can accommodate a wider class of process models other than the standard Brownian motion. However, the characteristics remain the same. That is, compared with the classical chain rule, there should be an extra term analogous to

$$\frac{1}{2}\int_0^T f''(W_t)\,dt.$$

The intrinsic reason for the presence of this term is that the quadratic variation of a Brownian motion sample path is T over any time interval $[0,T]$; see equations (8.3) and (8.4).

"PROOF" OF THEOREM 8.2. In order to understand the idea behind Itô formula, we should present a formal and sketchy argument. For clarity, we will further assume that f'' is bounded by some constant M. This assumption is nonessential and can be removed by a more detailed analysis. Let $t_i = iT/n, i = 0, 1, \ldots, n,$ be a partition of the time interval $[0,T]$. We should eventually send n to infinity. Write

$$f(W_T) = f(W_0) + \sum_{i=0}^{n-1}\left[f(W_{t_{i+1}}) - f(W_{t_i})\right].$$

Denote $\Delta W_i = W_{t_{i+1}} - W_{t_i}$ and $\Delta t_i = t_{i+1} - t_i$. It follows from the Taylor expansion that

$$f(W_{t_{i+1}}) - f(W_{t_i}) = f'(W_{t_i})\Delta W_i + \frac{1}{2}f''(\xi_i)(\Delta W_i)^2,$$

where ξ_i is a random variable taking values between W_{t_i} and $W_{t_{i+1}}$. By the definition (8.2) of stochastic integral,

$$\sum_{i=0}^{n-1} f'(W_{t_i})\Delta W_i \to \int_0^T f'(W_t)\,dW_t.$$

Furthermore, thanks to equation (8.3) and the continuity of f'' and W,

$$\sum_{i=0}^{n-1}|f''(\xi_i) - f''(W_{t_i})|(\Delta W_i)^2 \le \max_i |f''(\xi_i) - f''(W_{t_i})|\sum_{i=0}^{n-1}(\Delta W_i)^2 \to 0.$$

Therefore, it suffices to show

$$\sum_{i=0}^{n-1} f''(W_{t_i})(\Delta W_i)^2 \to \int_0^T f''(W_t)\,dt. \tag{8.4}$$

The left-hand-side equals

$$\sum_{i=0}^{n-1} f''(W_{t_i})\Delta t_i + \sum_{i=0}^{n-1} f''(W_{t_i})[(\Delta W_i)^2 - \Delta t_i] = (\mathrm{I}) + (\mathrm{II}).$$

It is trivial that (I) converges to the right-hand-side of (8.4). Therefore, we only need to show (II) $\to 0$. Note that W_{t_i} and ΔW_j are independent if $j \geq i$, therefore

$$
\begin{aligned}
E(\mathrm{II}) &= \sum_{i=0}^{n-1} E[f''(W_{t_i})] \cdot E[(\Delta W_i)^2 - \Delta t_i] \;=\; 0, \\
\mathrm{Var}(\mathrm{II}) &= \sum_{i=0}^{n-1} E\left\{f''(W_{t_i})[(\Delta W_i)^2 - \Delta t_i]\right\}^2 \\
&\leq M^2 \cdot \sum_{i=0}^{n-1} E[(\Delta W_i)^2 - \Delta t_i]^2 \\
&= M^2 \cdot \sum_{i=0}^{n-1} (\Delta t_i)^2 E(Z_i^2 - 1)^2 \\
&= M^2 \cdot E(Z^2 - 1)^2 \cdot \frac{T^2}{n} \;\to\; 0,
\end{aligned}
$$

where Z and Z_i's are iid standard normal random variables. It follows that (II) converges to zero. We complete the argument. ■

Remark 8.4. Loosely speaking, since $(\Delta W_i)^2 = Z_i^2 \Delta t_i$ for some standard normal random variable Z_i, ΔW_i is of order $\sqrt{\Delta t_i}$. This leads to the non-vanishing limit (8.4), which is precisely the reason for the extra term in Itô formula. It also explains why we have applied the Taylor expansion up until order 2, that is, the higher order terms in the Taylor expansion are of orders higher than Δt_i.

8.2.2 The Differential Notation

It is notationally cumbersome to express Itô formula in its integral form. In literature, one often adopts the so-called *differential notation*. More precisely, the differential notation for a process X that satisfies

$$X_t = X_0 + \int_0^t Y_s \, ds + \int_0^t Z_s \, dW_s$$

is simply

$$dX_t = Y_t \, dt + Z_t \, dW_t. \tag{8.5}$$

Even though the differential notation has been used extensively, one should keep in mind that it is only a notational convenience. The true meaning always lies with the integral form.

In stochastic calculus, it is also convenient to define the *product* of two differentials in such a way that

$$(dt)^2 = 0, \;\; dt\, dW_t = dW_t dt = 0, \;\; (dW_t)^2 = dt$$

for any standard Brownian motion W. Not surprisingly, $(dW_t)^2 = dt$ reflects the fact that the quadratic variation of a standard Brownian motion sample path is T over any time interval $[0, T]$. From these definitions one can easily calculate the product of general differentials. For example, if

$$\begin{aligned} dX_t &= Y_t\, dt + Z_t\, dW_t \\ dM_t &= R_t\, dt + Q_t\, dW_t, \end{aligned}$$

then

$$\begin{aligned} dX_t dM_t &= Y_t R_t (dt)^2 + (Y_t Q_t + Z_t R_t)(dt dW_t) + Z_t Q_t (dW_t)^2 \\ &= Z_t Q_t\, dt. \end{aligned}$$

When W and B are two independent standard Brownian motions, the product of their differentials is defined to be

$$dW_t dB_t = 0.$$

See also Exercise 8.10.

With this notation, Theorem 8.2 can be written in a much more succinct fashion

$$df(W_t) = f'(W_t)\, dW_t + \frac{1}{2} f''(W_t)\, (dW_t)^2,$$

which is very easy to memorize because of its resemblance to the Taylor expansion.

8.2.3 General Itô Formulas and Product Rule

We state a general Itô formula below. The proof is beyond the scope of this book and thus omitted. However, the main idea is exactly the same as in Theorem 8.2. Throughout the section, all relevant processes are assumed to be of the form (8.5).

Theorem 8.3. *Suppose that $f(t, x) : [0, \infty) \times \mathbb{R}^d \to \mathbb{R}$ is continuously differentiable with respect to t and twice continuously differentiable with respect to x.*

Then for a d-dimensional process $X_t = (X_t^{(1)}, \ldots, X_t^{(d)})$,

$$\begin{aligned} df(t, X_t) &= \frac{\partial f}{\partial t}(t, X_t)\, dt + \sum_{i=1}^{d} \frac{\partial f}{\partial x_i}(t, X_t)\, dX_t^{(i)} \\ &\quad + \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d} \frac{\partial^2 f}{\partial x_i \partial x_j}(t, X_t) dX_t^{(i)} dX_t^{(j)}. \end{aligned}$$

We also want to mention two very useful corollaries. Letting $d = 1$ and f be a function of x alone, we arrive at Corollary 8.4, of which the basic Itô formula in Theorem 8.2 is a special case with $X = W$. On the other hand, if we let $d = 2$ and $f(t, x, y) = xy$, then the product rule follows.

Corollary 8.4. *Suppose that $f : \mathbb{R} \to \mathbb{R}$ is twice continuously differentiable. Then*

$$df(X_t) = f'(X_t)\, dX_t + \frac{1}{2} f''(X_t)(dX_t)^2.$$

Corollary 8.5. Product Rule. $d(X_t Y_t) = X_t\, dY_t + Y_t\, dX_t + dX_t\, dY_t.$

Example 8.2. Let W be a standard Brownian motion. Consider a geometric Brownian motion

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

Compute the differential dS_t.

SOLUTION: This is a straightforward application of Itô formula. Observe that $S_t = f(t, W_t)$, where

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

It follows from Theorem 8.3 and direct calculation (we omit the details) that

$$\begin{aligned} dS_t &= \frac{\partial f}{\partial t}(t, W_t)dt + \frac{\partial f}{\partial x}(t, W_t)dW_t + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(t, W_t)(dW_t)^2 \\ &= rS_t\, dt + \sigma S_t\, dW_t. \end{aligned}$$

This is indeed the formula often seen in the mathematical finance literature for asset prices that are geometric Brownian motions. ■

Example 8.3. Suppose that $dS_t = Y_t\,dt + Z_t\,dW_t$ and f is a twice continuously differentiable function. Define

$$X_t = e^{-rt} f(S_t).$$

Compute the differential dX_t.

SOLUTION: Let $R_t = e^{-rt}$ and $H_t = f(S_t)$. Then it follows from Corollary 8.4 that

$$\begin{aligned} dH_t &= f'(S_t)\,dS_t + \frac{1}{2} f''(S_t)\,(dS_t)^2 \\ &= f'(S_t)(Y_t\,dt + Z_t\,dW_t) + \frac{1}{2} f''(S_t) Z_t^2\,dt. \end{aligned}$$

It is trivial that $dR_t = -rR_t\,dt$. By the product rule (note that $dR_t\,dH_t = 0$),

$$\begin{aligned} dX_t &= R_t\,dH_t + H_t\,dR_t + dR_t\,dH_t \\ &= e^{-rt}\left[-rf(S_t) + f'(S_t)Y_t + \frac{1}{2} f''(S_t) Z_t^2\right] dt + e^{-rt} f'(S_t) Z_t\,dW_t. \end{aligned}$$

Another way to calculate dX_t is to write $X_t = g(t, S_t)$ where $g(t,x) = e^{-rt} f(x)$ and then apply the general Itô formula in Theorem 8.3. It leads to, of course, exactly the same result. ■

Example 8.4. Let W be a standard Brownian motion and Y an arbitrary process. Define

$$\begin{aligned} M_t &= \exp\left\{\int_0^t Y_s\,dW_s - \frac{1}{2}\int_0^t Y_s^2\,ds\right\}, \\ X_t &= W_t - \int_0^t Y_s\,ds. \end{aligned}$$

Compute the differentials dM_t and $d(X_t M_t)$.

SOLUTION: It is often helpful to introduce some intermediate processes to facilitate the calculation of differentials. Let

$$Q_t = \int_0^t Y_s\,dW_s - \frac{1}{2}\int_0^t Y_s^2\,ds.$$

Then $M_t = e^{Q_t}$ and

$$dQ_t = Y_t\,dW_t - \frac{1}{2} Y_t^2\,dt, \quad (dQ_t)^2 = Y_t^2\,dt.$$

It follows from Corollary 8.4 that

$$dM_t = e^{Q_t} dQ_t + \frac{1}{2} e^{Q_t} (dQ_t)^2 = M_t \left[dQ_t + \frac{1}{2} (dQ_t)^2 \right] = M_t Y_t \, dW_t.$$

To calculate $d(X_t M_t)$, one can just use the product rule:

$$\begin{aligned} d(X_t M_t) &= X_t \, dM_t + M_t \, dX_t + dX_t \, dM_t \\ &= X_t M_t Y_t \, dW_t + M_t (dW_t - Y_t \, dt) + M_t Y_t dt \\ &= (X_t M_t Y_t + M_t) \, dW_t. \end{aligned}$$

The explicit calculation of the differentials has some interesting implications. For example, the form of dM_t implies that

$$M_t = M_0 + \int_0^t M_s Y_s \, dW_s = 1 + \int_0^t M_s Y_s \, dW_s.$$

Thanks to the martingale property of stochastic integrals, it follows that $E[M_t] = 1$ (this is not entirely accurate because some conditions on Y are needed to ensure the martingale property). Further analysis will eventually lead to the famous Girsanov's Theorem [18]. ■

8.3 Stochastic Differential Equations

In many financial models, the price of the underlying asset is assumed to be a *diffusion process*, that is, the solution to a *stochastic differential equation* of the following form:

$$dX_t = b(t, X_t) \, dt + \sigma(t, X_t) \, dW_t, \tag{8.6}$$

where b (drift) and σ (volatility) are some given function. We say X is a solution to the stochastic differential equation (8.6) if it satisfies

$$X_T = X_0 + \int_0^T b(t, X_t) \, dt + \int_0^T \sigma(t, X_t) \, dW_t$$

for every T and some standard Brownian motion W; see Remark 8.5. Very few stochastic differential equations admit explicit solutions. However, they are not difficult to approximate by Monte Carlo simulation.

Example 8.5. Let r be a constant and $\theta(t)$ an arbitrarily given deterministic function. Solve the stochastic differential equation

$$dX_t = rX_t\,dt + \theta(t)\,dW_t, \quad X_0 = x.$$

SOLUTION: Consider the process $Y_t = e^{-rt}X_t$. Then $Y_0 = X_0 = x$. It follows from the product rule that

$$dY_t = -re^{-rt}X_t\,dt + e^{-rt}\,dX_t = e^{-rt}\theta(t)\,dW_t.$$

Therefore,

$$X_T = e^{rT}Y_T = e^{rT}\left(x + \int_0^T e^{-rt}\theta(t)\,dW_t\right).$$

Even though this is not exactly an explicit formula because of the stochastic integral involved, the distribution of X_T is

$$N\left(e^{rT}x,\ e^{2rT}\int_0^T e^{-2rt}\theta^2(t)\,dt\right)$$

from Lemma 8.1. This is all one needs to simulate samples of X_T. ■

Example 8.6. Let a, b, and σ be arbitrary constants. Solve the stochastic differential equation

$$dX_t = a(b - X_t)\,dt + \sigma\,dW_t, \quad X_0 = x.$$

SOLUTION: Analogous to the previous example, we define $Y_t = e^{at}X_t$ and use the product rule to obtain

$$dY_t = ae^{at}X_t\,dt + e^{at}\,dX_t = abe^{at}\,dt + \sigma e^{at}\,dW_t,$$

which implies that

$$\begin{aligned} X_T &= e^{-aT}Y_T \\ &= e^{-aT}\left(x + \int_0^T abe^{at}\,dt + \int_0^T \sigma e^{at}\,dW_t\right) \\ &= e^{-aT}x + b(1-e^{-aT}) + e^{-aT}\int_0^T \sigma e^{at}\,dW_t. \end{aligned}$$

Thanks to Lemma 8.1, X_T is normally distributed as

$$N\left(e^{-aT}x + b(1-e^{-aT}),\ \frac{\sigma^2}{2a}(1-e^{-2aT})\right).$$

It is particularly interesting when a is positive. In this case, X is said to be a *mean-reverting Ornstein–Uhlenbeck* process. It is mean-reverting because if $X_t > b$, then the drift $a(b - X_t)$ is negative and pushes X_t down; on the other hand, if $X_t < b$, then the drift is positive and drives the process up. Therefore, X oscillates around b. Coincidentally, the limit distribution of X_T as $T \to \infty$ is

$$N\left(b, \frac{\sigma^2}{2a}\right),$$

whose mean is exactly b. ■

Example 8.7. Let r be a constant and $\theta(t)$ an arbitrarily given deterministic function. Solve the stochastic differential equation

$$dX_t = rX_t\,dt + \theta(t)X_t\,dW_t, \quad X_0 = x > 0.$$

SOLUTION: Define the process $Y_t = \log X_t$. Then $Y_0 = \log X_0 = \log x$. It follows from Corollary 8.4 that

$$dY_t = \frac{1}{X_t}\,dX_t - \frac{1}{2}\frac{1}{X_t^2}(dX_t)^2 = r\,dt + \theta(t)\,dW_t - \frac{1}{2}\theta^2(t)\,dt.$$

Therefore,

$$X_T = e^{Y_T} = x \cdot \exp\left\{\int_0^T \left(r - \frac{1}{2}\theta^2(t)\right)\,dt + \int_0^T \theta(t)\,dW_t\right\}.$$

Letting $\theta(t)$ be a constant, say σ, we recover the classical geometric Brownian motion. ■

Continuous Time Financial Models: For the purpose of demonstration, we list a small collection of financial models that involve diffusion processes. Most of these models do not admit explicit solutions.

a. Black–Scholes model: The risk-free interest rate is assumed to be a constant r. The stock price is a geometric Brownian motion with drift μ and volatility σ:

$$\frac{dS_t}{S_t} = \mu\,dt + \sigma\,dW_t.$$

b. CEV model: The stock price under the *constant elasticity of variance* model is assumed to satisfy

$$dS_t = \mu S_t\,dt + \sigma S_t^\gamma\,dW_t,$$

where $\gamma > 0$ is a constant.

c. Stochastic volatility models: The volatility of the underlying asset is a diffusion process itself. One of the most widely used stochastic volatility models is the Heston model [15]:

$$
\begin{aligned}
\frac{dS_t}{S_t} &= \mu\,dt + \sqrt{\theta_t}\,dW_t,\\
d\theta_t &= a(b-\theta_t)\,dt + \sigma\sqrt{\theta_t}\,dB_t,
\end{aligned}
$$

where $a, b > 0$ and (W, B) is a two-dimensional Brownian motion with covariance matrix

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

d. Interest rate models: In the context of pricing interest rate derivatives, modeling the *instantaneous interest rate* or *short rate* is an important issue. Below are a couple of such models. The short rate at time t is denoted by r_t.

$$
\begin{aligned}
&\text{(Vasicek)} & dr_t &= a(b-r_t)\,dt + \sigma\,dW_t.\\
&\text{(Cox–Ingersoll–Ross)} & dr_t &= a(b-r_t)\,dt + \sigma\sqrt{r_t}\,dW_t.
\end{aligned}
$$

Remark 8.5. Rigorously speaking, there should be two types of solutions to a stochastic differential equation: *strong solutions* and *weak solutions*. The key difference between these two notions is whether the driving Brownian motion W is designated as a given input (strong solution) or viewed as part of the solution itself (weak solution). The concept of weak solutions suffices in nearly all financial applications for the following reasons. (i) In financial modeling, what really matters is the distributional properties of the relevant processes. To identify the driving Brownian motion W is both difficult and unnecessary. (ii) The condition for the existence of a strong solution is much more stringent than that of a weak solution.

8.4 Risk-Neutral Pricing

The option pricing formula for the classical Black–Scholes model can be generalized to models with time varying (deterministic or stochastic) interest rate and complex volatility structure. To be more concrete, denote by r_t the instantaneous interest rate or short rate at time t. Then \$1 at time 0 is worth

$$
\exp\left\{\int_0^T r_t\,dt\right\}
$$

at time T. It is not surprising that the price of an option with maturity T and payoff X should be

$$v = E\left[\exp\left\{-\int_0^T r_t\,dt\right\}X\right], \tag{8.7}$$

where the expected value is taken with respect to the risk-neutral probability measure. The underlying asset price can be diffusion processes other than the geometric Brownian motion. However, the drift must equal the short rate $r_t$under the risk-neutral probability measure. An example of such a price process is

$$\frac{dS_t}{S_t} = r_t\,dt + \theta(t, S_t)\,dW_t,$$

where θ is some given function and W is a standard Brownian motion. In the special case where $r_t \equiv r$ and $\theta(t,x) \equiv \sigma$, we recover the classical pricing formula for the Black–Scholes model.

Example 8.8. Assume that the risk-free interest rate r is a constant and the price of the underlying stock satisfies the stochastic differential equation

$$\frac{dS_t}{S_t} = r\,dt + \theta(t)\,dW_t$$

for some deterministic function $\theta(t)$ under the risk-neutral probability measure. Find the price of a call option with maturity T and strike price K.

SOLUTION: The price of the option is $v = E[e^{-rT}(S_T - K)^+]$. Thanks to Example 8.7 and Lemma 8.1,

$$\begin{aligned} S_T &= S_0 \exp\left\{\int_0^T \left(r - \frac{1}{2}\theta^2(t)\right)dt + \int_0^T \theta(t)\,dW_t\right\} \\ &= S_0 \exp\left\{\int_0^T \left(r - \frac{1}{2}\sigma^2\right)dt + \sigma\sqrt{T}Z\right\}, \end{aligned}$$

where Z is a standard normal random variable and

$$\sigma = \sqrt{\frac{1}{T}\int_0^T \theta^2(t)\,dt}.$$

In other words, S_T has the same distribution as the terminal stock price in the classical Black–Scholes model with drift r and volatility σ. It follows that $v = \text{BLS\_Call}(S_0, K, T, r, \sigma)$. ■

Example 8.9. Consider the Ho–Lee model where the short rate r_t satisfies the stochastic differential equation

$$dr_t = \theta(t)\,dt + \sigma\,dW_t$$

under the risk-neutral probability measure. Assuming that $\theta(t)$ is a deterministic function, compute the price of a zero-coupon bond with payoff \$1 and maturity T.

SOLUTION: The price of this zero-coupon bond is given by the pricing formula (8.7) with $X = 1$, that is,

$$v = E\left[\exp\left\{-\int_0^T r_t\,dt\right\}\right].$$

Observe that

$$\int_0^T r_t\,dt = \int_0^T \left[r_0 + \int_0^t \theta(s)\,ds + \sigma W_t\right]dt = \mu + \sigma Y,$$

where

$$\mu = r_0 T + \int_0^T\int_0^t \theta(s)\,ds\,dt, \quad Y = \int_0^T W_t\,dt.$$

In order to determine the distribution of Y, consider an arbitrary partition $0 = t_0 < t_1 < \cdots < t_n = T$. Letting $\Delta t_i = t_{i+1} - t_i$, Y is the limit of

$$S_n = \sum_{i=0}^{n-1} W_{t_i}\Delta t_i.$$

For each n, S_n is normally distributed with mean 0. It follows from Lemma 1.7 and Exercise 2.6 that

$$\begin{aligned}
\mathrm{Var}[S_n] &= \sum_{i=0}^{n-1} \mathrm{Var}[W_{t_i}](\Delta t_i)^2 + 2\sum_{i<j}\mathrm{Cov}[W_{t_i}, W_{t_j}]\Delta t_i \Delta t_j \\
&= \sum_{i=0}^{n-1} t_i(\Delta t_i)^2 + 2\sum_{i<j} t_i \Delta t_i \Delta t_j \\
&= \sum_{i=0}^{n-1} t_i(\Delta t_i)^2 + 2\sum_{i=0}^{n-1}(T - t_{i+1})t_i\Delta t_i \\
&= -\sum_{i=0}^{n-1} t_i(\Delta t_i)^2 + 2\sum_{i=0}^{n-1}(T - t_i)t_i\Delta t_i.
\end{aligned}$$

Therefore, Y is normally distributed itself with mean 0 and variance

$$\beta^2 = \mathrm{Var}[Y] = \lim_{n\to\infty} \mathrm{Var}[S_n] = 0 + 2\int_0^T (T-t)t\,dt = \frac{1}{3}T^3,$$

and the bond price is

$$v = E\left[e^{-\mu-\sigma Y}\right] = e^{-\mu+\frac{1}{2}\sigma^2\beta^2}.$$

See Exercise 8.8 for a different approach to calculate $\mathrm{Var}[Y]$. ■

8.5 Black–Scholes Equation

The most famous application of Itô formula to the option pricing theory is probably about the connection between option prices and a family of second order partial differential equations. This leads to an entirely different approach to option pricing by means of partial differential equations. Even though the focus of the book is about Monte Carlo simulation, it is worthwhile to briefly discuss this connection.

Consider a call option with strike price K and maturity T in the classical Black–Scholes model, where the stock price is a geometric Brownian motion with drift r and volatility σ under the risk-neutral probability measure:

$$dS_t = rS_t\,dt + \sigma S_t\,dW_t.$$

Define for every $t \in [0, T]$ and $x > 0$

$$v(t,x) = E[e^{-r(T-t)}(S_T - K)^+ | S_t = x].$$

That is, $v(t,x)$ is the price of the call option evaluated at time t if the stock price at time t is x. By definition, the call option price at time 0 is $v(0, S_0)$. Moreover, for any $\theta \in [0, T]$ it follows from the tower property that

$$v(0,S_0) = E[e^{-rT}(S_T - K)^+] = E[e^{-r\theta}v(\theta, S_\theta)].$$

Assuming that v is nice and smooth, Itô formula implies that

$$e^{-r\theta}v(\theta,S_\theta) = v(0,S_0) + \int_0^\theta e^{-rt}\mathbb{L}v(t,S_t)\,dt + \int_0^\theta e^{-rt}\frac{\partial v}{\partial x}(t,S_t)\,\sigma S_t\,dW_t,$$

where

$$\mathbb{L}v = -rv + \frac{\partial v}{\partial t} + \frac{\partial v}{\partial x}rx + \frac{1}{2}\sigma^2 x^2\frac{\partial^2 v}{\partial x^2}.$$

Taking expected value on both sides and observing that by the martingale property the expected value of the stochastic integral is zero, we have

$$E[e^{-r\theta}v(\theta, S_\theta)] = v(0, S_0) + \int_0^\theta e^{-rt} E[\mathbb{L}v(t, S_t)]\, dt.$$

Therefore,

$$\int_0^\theta e^{-rt} E[\mathbb{L}v(t, S_t)]\, dt = 0$$

for every θ, which in turn implies that $E[\mathbb{L}v(t, S_t)] = 0$ for every t. Now letting $t \to 0$, we arrive at $\mathbb{L}v(0, S_0) = 0$. However, in the above derivation, the initial time and initial stock price can be *arbitrary*, which leads to $\mathbb{L}v(t, x) = 0$ for every $t \in [0, T)$ and $x > 0$. In other words, v satisfies the partial differential equation

$$-rv + \frac{\partial v}{\partial t} + \frac{\partial v}{\partial x} rx + \frac{1}{2}\sigma^2 x^2 \frac{\partial^2 v}{\partial x^2} = 0, \quad 0 \le t < T,\ x > 0, \tag{8.8}$$

with terminal condition (which is trivial from the definition of v)

$$v(T, x) = (x - K)^+, \quad x > 0. \tag{8.9}$$

Equation (8.8) is called the *Black–Scholes equation*. Solving (8.8)–(8.9) yields the classical Black–Scholes formula of call option price.

More generally, if the option payoff is a function of the terminal stock price, say $h(S_T)$, then the option price satisfies the same Black–Scholes equation (8.8) but with a different terminal condition

$$v(T, x) = h(x).$$

This can be easily verified by repeating the preceding derivation. The situation is also very similar if the stock price is other than a geometric Brownian motion; see Exercise 8.16. More discussions on option pricing by means of partial differential equations can be found in the introductory textbook [30] or the more advanced [16, 19].

Exercises

8.1 Recover the result in Example 8.1 using the basic Itô formula with $f(x) = x^2$.

8.2 Use Itô formula to compute dX_t.

(a) $X_t = W_t^3$.

(b) $X_t = e^t W_t$.

(c) $X_t = W_t \int_0^t W_s\, ds$.

(d) $X_t = W_t \int_0^t s\, dW_s$.

8.3 Suppose that $dX_t = W_t\, dt + dW_t$. Let $Y_t = X_t^2$. Compute dY_t and $d(t^2 Y_t)$.

8.4 Define $X_t = W_t \exp\{-rt + \int_0^T Y_t\, dW_t\}$, where Y is a stochastic process and r is a constant. Compute dX_t.

8.5 Suppose that $W = (W^{(1)}, \dots, W^{(d)})$ is a d-dimensional standard Brownian motion with $d \geq 2$. Let

$$R_t = \sqrt{(W_t^{(1)})^2 + \cdots + (W_t^{(d)})^2}.$$

Compute dR_t. R is said to be a *Bessel Process*.

8.6 Find the distribution of $\exp\{\int_0^T t\, dW_t\}$.

8.7 Find the distribution of X_T.

(a) $dX_t = dt + t\, dW_t$ and $X_0 = x$.

(b) $dX_t = -aX_t\, dt + dW_t$ and $X_0 = x$.

8.8 Here is another approach to determine the variance of Y in Example 8.9. Show that

$$\text{Var}[Y] = E[Y^2] = E\left[\int_0^T W_s\, ds \int_0^T W_t\, dt\right] = \int_0^T \int_0^T E[W_s W_t]\, ds\, dt.$$

Compute the double integral to obtain Var$[Y]$. Use this approach to find the distribution of $\int_0^T \theta(t)\, W_t\, dt$ for a given deterministic function $\theta(t)$.

8.9 Let $\theta(t)$ and $\sigma(t)$ be two deterministic functions. Show that the random vector

$$\left(\int_0^T \theta(t)\, dW_t,\ \int_0^T \sigma(t)\, dW_t\right)$$

is jointly normal with mean zero and covariance matrix $\Sigma = [\Sigma_{ij}]$ where

$$\Sigma_{11} = \int_0^T \theta^2(t)\, dt, \quad \Sigma_{12} = \Sigma_{21} = \int_0^T \theta(t)\sigma(t)\, dt, \quad \Sigma_{22} = \int_0^T \sigma^2(t)\, dt.$$

Hint: Given $0 = t_0 < t_1 < \cdots < t_m = T$, consider the approximating sums

$$X_n = \sum_{i=0}^{n-1} \theta(t_i)(W_{t_{i+1}} - W_{t_i}), \quad Y_n = \sum_{i=0}^{n-1} \sigma(t_i)(W_{t_{i+1}} - W_{t_i}).$$

Show that X_n and Y_n are jointly normal and compute their covariance.

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

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

Show that $dW_t dB_t = \rho dt$. *Hint:* Cholesky factorization.

8.11 Let $\theta(t)$ be a deterministic function. Solve the stochastic differential equation

$$dX_t = \theta(t)\, dt + \sigma X_t\, dW_t, \quad X_0 = x.$$

Hint: Consider the process $X_t Y_t$ where $Y_t = \exp\left\{\frac{1}{2}\sigma^2 t - \sigma W_t\right\}$.

8.12 Let a, b, and σ be constants. Solve the stochastic differential equation

$$dX_t = a(b - X_t)\, dt + \sigma X_t\, dW_t, \quad X_0 = x.$$

Hint: Let $Y_t = e^{at} X_t$. Derive the equation for Y.

8.13 Given two processes θ and σ, show that the solution to the stochastic differential equation $dX_t = \theta_t X_t\, dt + \sigma_t X_t\, dW_t$ is

$$X_T = X_0 \exp\left\{ \int_0^T \left(\theta_t - \frac{1}{2}\sigma_t^2\right) dt + \int_0^T \sigma_t\, dW_t \right\}.$$

8.14 Assume that under the risk-neutral probability measure, the stock price satisfies the stochastic differential equation

$$dS_t = rS_t\, dt + \sigma(t, S_t)\, dW_t,$$

where r is the risk-free interest rate and $\sigma(t, x)$ is some function. Show that the *put-call parity* still holds. That is

$$v_c - v_p = S_0 - e^{-rT} K,$$

where v_c is the price of a call option with maturity T and strike price K, and v_p is that of a put option with the same maturity and strike price.

8.15 Suppose that $u(t, x) : [0, \infty) \times \mathbb{R} \to \mathbb{R}$ is a solution to the *heat equation*

$$\frac{\partial u}{\partial t} = \frac{1}{2}\frac{\partial^2 u}{\partial x^2}, \quad t > 0,\ x \in \mathbb{R}$$

with initial condition

$$u(0, x) = f(x), \quad x \in \mathbb{R}.$$

(a) Fix arbitrarily $T > 0$ and $x \in \mathbb{R}$. Define $X_t = x + W_t$ and $Y_t = u(T - t, X_t)$ for $t \in [0, T]$. Compute dY_t.

(b) Use the martingale property to argue that $E[Y_T] = E[Y_0] = u(T, x)$.

(c) Show that $E[Y_T] = E[f(X_T)] = E[f(x + W_T)]$. Therefore

$$u(T, x) = \int_{\mathbb{R}} f(x + y) \frac{1}{\sqrt{2\pi T}} e^{-y^2/(2T)} \, dy.$$

This type of argument is said to be a *verification argument*. The representation of the solution to a heat equation in terms of a Brownian motion is a special case of the so called *Feynman–Kac formula* [18].

8.16 Consider a call option with strike price K and maturity T. Assume that the risk-free interest rate r is a constant and the stock price S satisfies the stochastic differential equation

$$dS_t = rS_t \, dt + \theta(S_t) \, dW_t$$

under the risk-neutral probability measure. Let $v(t, x)$ be the value of the option at time t given that the stock price at time t is x. Argue that v satisfies the partial differential equation

$$-rv + \frac{\partial v}{\partial t} + rx\frac{\partial v}{\partial x} + \frac{1}{2}\theta^2(x)\frac{\partial^2 v}{\partial x^2} = 0, \quad 0 \leq t < T, \ x > 0$$

with terminal condition

$$v(T, x) = (x - K)^+.$$

What should the partial differential equation be if the payoff of the option is $h(S_T)$ for some function h?

Chapter 9

Simulation of Diffusions

In general, diffusions or solutions to stochastic differential equations do not admit explicit formulas. To obtain quantitative information, one often resorts to Monte Carlo simulation. The idea is to discretize time and approximate the value of the diffusion at discrete time steps. Except for special cases, this operation will introduce *discretization error*, and thus the resulting estimates are usually biased. The variance reduction techniques we have discussed so far can also be brought into play. However, they will only reduce the variance of the estimate and will not affect the bias.

In this chapter, we state some of the most basic discretization schemes for stochastic differential equations and show how they can be combined with variance reduction techniques. Even though there are well-defined performance criteria to classify various discretization schemes, it is not our intention to do so here. The interested reader may want to consult the classical textbook [20] for further investigation.

9.1 Euler Scheme

Euler Scheme is the most intuitive and straightforward approximation to stochastic differential equations. To fix ideas, let $b(t,x)$ and $\sigma(t,x)$ be two continuous functions and consider the stochastic differential equation

$$dX_t = b(t, X_t)\, dt + \sigma(t, X_t)\, dW_t.$$

In order to simulate the sample paths of X on the time interval $[0, T]$, define a time discretization $0 = t_0 < t_1 < \cdots < t_m = T$. The approximating process $\hat{X}$ is defined recursively by

$$\begin{aligned}
\hat{X}_0 &= X_0, && (9.1)\\
\hat{X}_{t_{i+1}} &= \hat{X}_{t_i} + b(t_i, \hat{X}_{t_i})(t_{i+1} - t_i) + \sigma(t_i, \hat{X}_{t_i})(W_{t_{i+1}} - W_{t_i}) \\
&= \hat{X}_{t_i} + b(t_i, \hat{X}_{t_i})(t_{i+1} - t_i) + \sigma(t_i, \hat{X}_{t_i})\sqrt{t_{i+1} - t_i}Z_{i+1}, && (9.2)
\end{aligned}$$

for $i = 0, \ldots, m-1$, where $\{Z_1, \ldots, Z_m\}$ are iid standard normal random variables. In other words, in Euler scheme both the drift and the volatility are approximated by their values at the left-hand endpoint on each time interval $[t_i, t_{i+1})$.

Consider the problem of estimating $v = E[h(X_T)]$ for some function h. The Euler scheme will generate n iid sample paths and form the estimate

$$\hat{v} = \frac{1}{n}\sum_{i=1}^{n} h(\hat{X}_{T,i}),$$

where $\hat{X}_{T,1}, \ldots, \hat{X}_{T,n}$ denote the terminal values of these n sample paths. In general, since the distribution of $\hat{X}_T$ is different from that of X_T due to the time discretization,

$$E[\hat{v}] = E[h(\hat{X}_T)] \neq E[h(X_T)].$$

That is, the estimate $\hat{v}$ is *biased*. Moreover, this bias or discretization error cannot be eliminated by merely increasing the sample size n. Indeed, the strong law of large numbers implies that as n goes to infinity

$$\hat{v} - v \to E[h(\hat{X}_T)] - E[h(X_T)] \neq 0.$$

Note that variance reduction techniques will only accelerate the convergence of the estimate $\hat{v}$ to $E[h(\hat{X}_T)]$. They have no effect on the bias. On the other hand, it can be shown that under mild conditions the discretization error goes to zero as the time discretization becomes finer and finer, namely, as m tends to infinity. Given a computational budget, there is a trade off between increasing m for a smaller bias and increasing n for a smaller variance. Such analysis is often carried out under the principle of minimizing the mean square error of the estimate [8, 11].

Remark 9.1. Even though we have only described the Euler scheme for the one-dimensional diffusion processes, the extension to higher dimensions is trivial. Assume that X is a k-dimensional process and W is a d-dimensional Brownian motion with covariance matrix Σ. The Euler scheme (9.1)–(9.2) remains the same except that $\{Z_1, \ldots, Z_m\}$ become iid d-dimensional jointly normal random vectors with distribution $N(0, \Sigma)$.

9.2 Eliminating Discretization Error

The discretization error cannot be eliminated in general. However, for some special models it is possible. For example, consider a diffusion process X that satisfies the stochastic differential equation

$$dX_t = r\,dt + \theta(t)\,dW_t$$

for some constant r and deterministic function $\theta(t)$. Thanks to Lemma 8.1,

$$\begin{aligned} X_{t_{i+1}} - X_{t_i} &= r(t_{i+1} - t_i) + \int_{t_i}^{t_{i+1}} \theta(s)\,dW_s \\ &= r(t_{i+1} - t_i) + \sqrt{\int_{t_i}^{t_{i+1}} \theta^2(s)\,ds} \cdot Z_{i+1}, \end{aligned}$$

where Z_{i+1} is a standard normal random variable. It leads to the following algorithm

$$\hat{X}_0 = X_0, \quad \hat{X}_{t_{i+1}} = \hat{X}_{t_i} + r(t_{i+1} - t_i) + \sqrt{\int_{t_i}^{t_{i+1}} \theta^2(s)\,ds} \cdot Z_{i+1},$$

where $\{Z_1, \ldots, Z_m\}$ are iid standard normal random variables. It is not difficult to see that $\hat{X}_{t_i}$ has the same distribution as X_{t_i} for every i, and thus there is no discretization error.

Actually, the time discretization is not really necessary if the quantity of interest only involves X_T since one can directly simulate X_T by

$$X_T = X_0 + rT + \int_0^T \theta(t)\,dW_t = X_0 + rT + \sqrt{\int_0^T \theta^2(t)\,dt} \cdot Z,$$

where Z is a standard normal random variable. However, time discretization is still needed in case one is interested in estimating an expected value of the more general form

$$E[h(X_0, X_{t_1}, \ldots, X_{t_{m-1}}, X_{t_m})],$$

or when X is brought into play as an auxiliary process (e.g., as a control variate) and the discretization is forced upon X.

It is worth noting that even if it is possible to simulate X_{t_i} exactly, discretization error may still exist if the quantity of interest can only be approximated. For instance, options that depend on the maximum or minimum of the underlying asset price such as lookback options, or the average of the underlying asset price such as Asian options, can only be approximated in general.

9.3 Refinements of Euler Scheme

There are various refinements of the plain Euler scheme. The goal of these refined schemes is to improve the approximation and reduce the discretization error. They are *not* variance reduction techniques. We will discuss one of such refinements under the assumption that X is a diffusion process of the form

$$dX_t = b(X_t)\,dt + \sigma(X_t)\,dW_t,$$

where b and σ are twice continuously differentiable functions. Consider a time discretization $0 = t_0 < t_1 < \cdots < t_m = T$. For $i = 0, 1, \ldots, m-1$,

$$X_{t_{i+1}} = X_{t_i} + \int_{t_i}^{t_{i+1}} b(X_t)\,dt + \int_{t_i}^{t_{i+1}} \sigma(X_t)\,dW_t.$$

The Euler scheme essentially approximates $b(X_t)$ and $\sigma(X_t)$ by $b(X_{t_i})$ and $\sigma(X_{t_i})$, respectively, for every $t \in [t_i, t_{i+1})$. To improve on the Euler scheme, we should examine the coefficients $b(X_t)$ and $\sigma(X_t)$ more carefully through Itô formula.

For a twice continuously differentiable function f, it follows from Itô formula that

$$f(X_t) = f(X_{t_i}) + \int_{t_i}^{t} \mathbb{L}^0 f(X_s)\,ds + \int_{t_i}^{t} \mathbb{L}^1 f(X_s)\,dW_s,$$

where

$$\mathbb{L}^0 f(x) = f'(x)b(x) + \frac{1}{2}f''(x)\sigma^2(x), \;\; \mathbb{L}^1 f(x) = f'(x)\sigma(x).$$

Therefore,

$$X_{t_{i+1}} = X_{t_i} + b(X_{t_i})(t_{i+1} - t_i) + \sigma(X_{t_i})(W_{t_{i+1}} - W_{t_i}) + R,$$

where

$$\begin{aligned} R \;=\; & \int_{t_i}^{t_{i+1}} \left[\int_{t_i}^{t} \mathbb{L}^0 b(X_s)\,ds + \int_{t_i}^{t} \mathbb{L}^1 b(X_s)\,dW_s \right] dt \\ & + \int_{t_i}^{t_{i+1}} \left[\int_{t_i}^{t} \mathbb{L}^0 \sigma(X_s)\,ds + \int_{t_i}^{t} \mathbb{L}^1 \sigma(X_s)\,dW_s \right] dW_t \end{aligned}$$

is the remainder term. Clearly, the Euler scheme ignores the remainder term R and keeps the two leading terms: (a) $\sigma(X_{t_i})(W_{t_{i+1}} - W_{t_i})$, which is of order $\sqrt{t_{i+1} - t_i}$; (b) $b(X_{t_i})(t_{i+1} - t_i)$, which is of order $t_{i+1} - t_i$.

Refinements of the Euler scheme retain some of the higher order terms in R in order to improve the approximation. Note that of all the terms in the remainder, the double stochastic integral is the leading term with order $t_{i+1} - t_i$ [other terms are of order $(t_{i+1} - t_i)^{3/2}$ or $(t_{i+1} - t_i)^2$]. If we approximate the double stochastic integral by

$$\begin{aligned}\mathbb{L}^1\sigma(X_{t_i}) \int_{t_i}^{t_{i+1}}\int_{t_i}^{t} dW_s dW_t &= \mathbb{L}^1\sigma(X_{t_i}) \int_{t_i}^{t_{i+1}} (W_t - W_{t_i})\, dW_t \\ &= \frac{1}{2}\mathbb{L}^1\sigma(X_{t_i})[(W_{t_{i+1}} - W_{t_i})^2 - (t_{i+1} - t_i)],\end{aligned}$$

and add it to the Euler scheme while ignoring all other higher order terms, we obtain one of the *Milstein schemes*

$$\begin{aligned}\hat{X}_{t_{i+1}} &= \hat{X}_{t_i} + b(\hat{X}_{t_i})(t_{i+1} - t_i) + \sigma(\hat{X}_{t_i})\sqrt{t_{i+1} - t_i}Z_{i+1} \\ &\quad + \frac{1}{2}\sigma'(\hat{X}_{t_i})\sigma(\hat{X}_{t_i})(t_{i+1} - t_i)(Z_{i+1}^2 - 1),\end{aligned} \tag{9.3}$$

where $\{Z_1, \ldots, Z_m\}$ are iid standard normal random variables.

Remark 9.2. It is possible to define performance measures for numerical schemes [20]. Furthermore, it can be shown that the performance of the Milstein scheme is better than that of the Euler scheme, albeit not by much. It can also be shown that one can arrive at even better numerical schemes by retaining more terms from the remainder R.

9.4 The Lamperti Transform

The Milstein scheme (9.3) coincides with the plain Euler scheme if $\sigma'(x) = 0$, or equivalently, if $\sigma(x)$ is a constant function. This observation leads to the general understanding that the Euler scheme is more effective when the volatility is a constant. For a diffusion process with nonconstant volatility, the idea is to transform the process into a diffusion with constant volatility, apply the Euler scheme, and then revert back to the original process.

To be more concrete, consider a diffusion process X that satisfies the stochastic differential equation

$$dX_t = b(X_t)\, dt + \sigma(X_t)\, dW_t,$$

where σ is continuously differentiable and strictly positive. Define a function F and a process Y by

$$Y_t = F(X_t), \quad F(x) = \int \frac{1}{\sigma(x)}\, dx. \tag{9.4}$$

Since $\sigma(x)$ is strictly positive, F is well defined and strictly increasing. In particular, F^{-1} exists and $X_t = F^{-1}(Y_t)$. The process Y is said to be the *Lamperti transform* of X. By the definition of F,

$$F'(x) = \frac{1}{\sigma(x)}, \quad F''(x) = -\frac{\sigma'(x)}{\sigma^2(x)}.$$

It follows from Itô formula that

$$dY_t = F'(X_t)\, dX_t + \frac{1}{2}F''(X_t)(dX_t)^2 = \left[\frac{b(X_t)}{\sigma(X_t)} - \frac{\sigma'(X_t)}{2}\right] dt + dW_t.$$

Substituting $F^{-1}(Y_t)$ for X_t, we arrive at $dY_t = a(Y_t)\, dt + dW_t$, where

$$a(y) = \frac{b(F^{-1}(y))}{\sigma(F^{-1}(y))} - \frac{\sigma'(F^{-1}(y))}{2}.$$

In other words, Y is a diffusion with constant volatility one. Below is the pseudocode for generating *one* discrete time sample path of $\{X_t\}$ through the Lamperti transform, given a time discretization $0 = t_0 < t_1 < \cdots < t_m = T$.

Pseudocode for Euler scheme through the Lamperti transform:

set $Y_0 = F(X_0)$
for $i = 0, 1, \ldots, m-1$
  generate Z_{i+1} from $N(0,1)$
  set $\hat{Y}_{t_{i+1}} = \hat{Y}_{t_i} + a(\hat{Y}_{t_i})(t_{i+1} - t_i) + \sqrt{t_{i+1} - t_i}\, Z_{i+1}$
  set $\hat{X}_{t_{i+1}} = F^{-1}(\hat{Y}_{t_{i+1}})$

Example 9.1. Determine the Lamperti transform for the geometric Brownian motion

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

where r and σ are positive constants.

SOLUTION: In this case,

$$F(x) = \int \frac{1}{\sigma x}\, dx = \frac{1}{\sigma}\log x, \quad Y_t = F(X_t) = \frac{1}{\sigma}\log X_t,$$

and

$$X_t = F^{-1}(Y_t) = \exp\{\sigma Y_t\}.$$

By Itô formula,

$$dY_t = \left(\frac{r}{\sigma} - \frac{\sigma}{2}\right) dt + dW_t.$$

It is not difficult to check that the Euler scheme through the Lamperti transform leads to a scheme for X without discretization error. ■

Example 9.2. Determine the Lamperti transform for the Cox–Ingersoll–Ross process

$$dX_t = a(b - X_t)\,dt + \sigma\sqrt{X_t}\,dW_t,$$

where a, b, σ are positive constants and $2ab \geq \sigma^2$.

SOLUTION: In this case,

$$F(x) = \int \frac{1}{\sigma\sqrt{x}}\,dx = \frac{2}{\sigma}\sqrt{x}, \quad Y_t = F(X_t) = \frac{2}{\sigma}\sqrt{X_t},$$

and

$$X_t = F^{-1}(Y_t) = \frac{\sigma^2}{4} Y_t^2.$$

By Itô formula,

$$\begin{aligned} dY_t &= \left(\frac{ab}{\sigma\sqrt{X_t}} - \frac{a\sqrt{X_t}}{\sigma} - \frac{\sigma}{4\sqrt{X_t}}\right) dt + dW_t \\ &= \left(\frac{4ab - \sigma^2}{2\sigma^2}\frac{1}{Y_t} - \frac{aY_t}{2}\right) dt + dW_t. \end{aligned}$$

Finally, we should remark that $2ab \geq \sigma^2$ is the necessary and sufficient condition for X_t to be strictly positive with probability one for all t; see for example, [18, Chapter 5.5.C]. If this condition fails, the Lamperti transform is no longer valid because the function $\sqrt{x}$ is not differentiable at $x = 0$. ■

9.5 Numerical Examples

We present a few examples in this section to illustrate various aspects of numerical approximations to stochastic differential equations and explain how they can be combined with variance reduction techniques. Unless otherwise specified, W always stands for a standard Brownian motion.

Example 9.3. Comparison of Euler scheme and Milstein scheme. Suppose that X is a geometric Brownian motion with drift r and volatility σ. That is,

$$dX_t = rX_t\,dt + \sigma X_t\,dW_t.$$

Use the Euler scheme and Milstein scheme to approximate X_T. Compare the performance.

SOLUTION: Assume that the time discretization is $t_i = ih$ for $i = 0, 1, \ldots, m$ and $h = T/m$. For the Euler scheme, the approximation is given by

$$\hat{X}_0 = X_0, \;\; \hat{X}_{t_{i+1}} = \hat{X}_{t_i} + r\hat{X}_{t_i}h + \sigma\hat{X}_{t_i}\sqrt{h}Z_{i+1},$$

while for the Milstien scheme,

$$\bar{X}_0 = X_0, \;\; \bar{X}_{t_{i+1}} = \bar{X}_{t_i} + r\bar{X}_{t_i}h + \sigma\bar{X}_{t_i}\sqrt{h}Z_{i+1} + \frac{1}{2}\sigma^2\bar{X}_{t_i}h(Z_{i+1}^2 - 1),$$

where $\{Z_1, \ldots, Z_m\}$ are iid standard normal random variables. As the benchmark, we need a scheme without discretization error, which is simply

$$X_{t_{i+1}} = X_{t_i}\exp\left\{\left(r - \frac{1}{2}\sigma^2\right)h + \sigma\sqrt{h}Z_{i+1}\right\}.$$

For comparison, we estimate the errors $E|\hat{X}_T - X_T|$ and $E|\bar{X}_T - X_T|$. Such quantities are often used in the literature as a performance measure. The numerical results are reported in Table 9.1 with parameters

$$r = 0.12, \; \sigma = 0.2, \; T = 1, \; X_0 = 50.$$

For each value of m, we use a large sample size of $n = 10,000,000$ to get an accurate estimate of the error.

Table 9.1: Numerical approximation: Euler versus Milstein

	$m = 5$	$m = 10$	$m = 20$	$m = 40$	$m = 60$	$m = 80$	$m = 100$
Euler	0.5615	0.3994	0.2832	0.2007	0.1640	0.1421	0.1271
Milstein	0.2208	0.1121	0.0565	0.0284	0.0189	0.0142	0.0114

From Figure 9.1, it is clear that the error from the Milstein scheme is of order h and the error from the Euler scheme is of order $\sqrt{h}$ (the slope of the line in the log-log scale is roughly $1/2$). Those rates of convergence can be proved in much greater generality [20]. ■

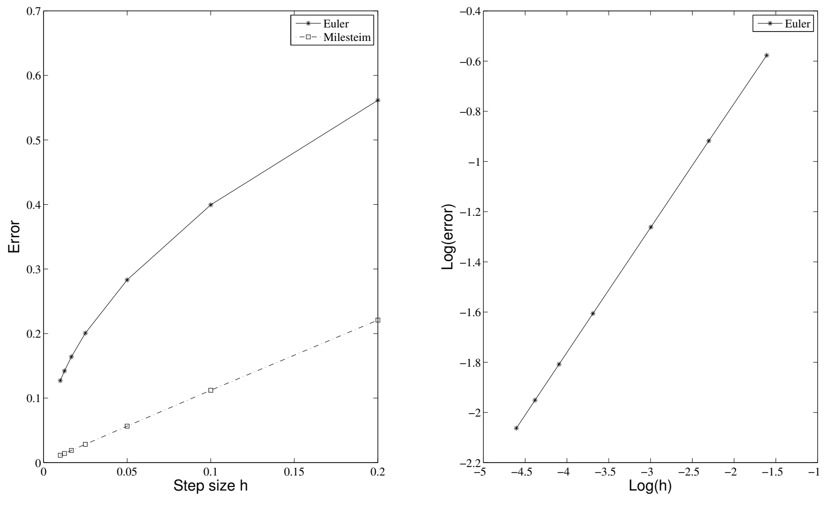

Figure 9.1: Numerical approximation: Euler versus Milstein.

Example 9.4. No discretization error. Consider a discretely monitored lookback call option with maturity T and floating strike price. The option payoff is

$$X = S_T - \min_{i=1,\dots,m} S_{t_i}$$

where $0 = t_0 < t_1 < \cdots < t_m = T$ are given dates. Assume that the risk-free interest rate r is a constant and the stock price S satisfies the stochastic differential equation

$$dS_t = rS_t\,dt + \theta(t)S_t\,dW_t$$

for some deterministic function $\theta(t)$ under the risk-neutral probability measure. Design a discretization scheme without discretization error and estimate the option price.

SOLUTION: Define $Y_t = \log S_t$. Then $Y_0 = \log S_0$. It follows from Itô formula that

$$dY_t = \frac{1}{S_t}\,dS_t - \frac{1}{2S_t^2}(dS_t)^2 = \left(r - \frac{1}{2}\theta^2(t)\right)\,dt + \theta(t)\,dW_t.$$

Therefore,

$$
\begin{aligned}
Y_{t_{i+1}} &= Y_{t_i} + \int_{t_i}^{t_{i+1}} \left(r - \frac{1}{2}\theta^2(t) \right) dt + \int_{t_i}^{t_{i+1}} \theta(t)\, dW_t \\
&= Y_{t_i} + \int_{t_i}^{t_{i+1}} \left(r - \frac{1}{2}\theta^2(t) \right) dt + \sqrt{\int_{t_i}^{t_{i+1}} \theta^2(t)\, dt} \cdot Z_{i+1} \\
&= Y_{t_i} + r(t_{i+1} - t_i) - \frac{1}{2}\sigma_i^2 + \sigma_i Z_{i+1},
\end{aligned}
$$

where

$$\sigma_i^2 = \int_{t_i}^{t_{i+1}} \theta^2(t)\, dt,$$

and $\{Z_1, \ldots, Z_m\}$ are iid standard normal random variables. Letting $S_{t_i} = \exp\{Y_{t_i}\}$, we obtain a discretization scheme for S without discretization error. Below is the pseudocode, in which Y_i and S_i stand for Y_{t_i} and S_{t_i}, respectively.

Pseudocode for lookback call option without discretization error:

for $k = 1, 2, \ldots, n$
  set $Y_0 = \log S_0$
  for $i = 0, 1, \ldots, m-1$
    generate Z_{i+1} from $N(0,1)$
    set $Y_{i+1} = Y_i + r(t_{i+1} - t_i) - \sigma_i^2/2 + \sigma_i Z_{i+1}$
    set $S_{i+1} = \exp\{Y_{i+1}\}$
  set $H_k = e^{-rT}[S_m - \min(S_1, \ldots, S_m)]$
compute the estimate $\hat{v} = \dfrac{1}{n}\sum_{i=1}^{n} H_i$
compute the standard error S.E. $= \sqrt{\dfrac{1}{n(n-1)}\left(\sum_{i=1}^{n} H_i^2 - n\hat{v}^2\right)}$.

For numerical simulation, we let $\theta(t) = \sqrt{a^2 + b^2 \sin(2\pi t/T)}$ for some positive constants a and b such that $a > b$. In this case,

$$\sigma_i^2 = a^2(t_{i+1} - t_i) + \frac{b^2 T}{2\pi}[\cos(2\pi t_i/T) - \cos(2\pi t_{i+1}/T)].$$

The simulation results are reported in Table 9.2 for

$$S_0 = 20,\ r = 0.05,\ T = 1,\ \ a = 0.4,\ b = 0.3.$$

We compare with the corresponding plain Euler scheme for $m = 5, 10, 20$, respectively (for each m, the time discretization in the Euler scheme is defined to be $0 = t_0 < t_1 < \cdots < t_m = T$). The sample size n is set to be $n = 1,000,000$ for accurate comparison.

Table 9.2: Scheme without discretization error versus Euler scheme

	$m = 5$		$m = 10$		$m = 20$	
	Euler	Exact	Euler	Exact	Euler	Exact
Estimate	4.2801	4.1409	4.9888	4.9529	5.4373	5.4239
S.E.	0.0055	0.0053	0.0059	0.0061	0.0064	0.0062

The standard errors associated with both schemes are nearly identical. However, when m is small the estimates differ because of the large bias of the Euler scheme. It is not surprising that as m increases, the bias of the Euler scheme diminishes and the two schemes become indistinguishable.

If one were to consider the discretely monitored lookback option as an approximation to the continuous time lookback call option with maturity T and payoff

$$S_T - \min_{0 \le t \le T} S_t,$$

then the scheme would be an example where the underlying process is simulated without discretization error but the estimate is still biased because the payoff function is only approximated. ■

Example 9.5. Simulating a process with sign constraints. The majority of the diffusion processes that are used for modeling stock prices, interest rate, volatility, and other financial entities are nonnegative. This property may fail for the time discretized approximating processes, which sometimes results in necessary modifications of the discretization scheme. To be more concrete, consider the Cox–Ingersoll–Ross interest rate model [6] where the short rate $\{r_t\}$ satisfies the stochastic differential equation

$$dr_t = a(b - r_t)\, dt + \sigma\sqrt{r_t}\, dW_t$$

for some positive constants a, b, and σ such that $2ab \ge \sigma^2$. Estimate the price of a zero-coupon bond with maturity T and payoff \$1.

SOLUTION: The price of the zero-coupon bond is given by the pricing formula (8.7) with $X = 1$:

$$v = E\left[\exp\left\{-\int_0^T r_t\, dt\right\}\right]. \tag{9.5}$$

Example 9.6. Method of conditioning. Suppose that under the risk-neutral probability measure, the stock price satisfies the stochastic differential equation

$$\frac{dS_t}{S_t} = r_t\,dt + \sigma\,dW_t.$$

The short rate $\{r_t\}$ is assumed to be mean-reverting and satisfy the stochastic differential equation

$$dr_t = a(b - r_t)\,dt + \theta r_t dB_t,$$

for some positive constants a, b, and θ. Assume that (W, B) is a two-dimensional Brownian motion with covariance matrix

$$\Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}.$$

Estimate the price of a call option with maturity T and strike price K.

SOLUTION: It is not difficult to verify via Itô formula that the terminal stock price is (see also Exercise 8.13)

$$\begin{aligned} S_T &= S_0 \exp\left\{\int_0^T r_t\,dt - \frac{1}{2}\sigma^2 T + \sigma W_T\right\} \\ &= S_0 \exp\left\{\left(\bar{r} - \frac{1}{2}\sigma^2\right) T + \sigma W_T\right\}, \end{aligned}$$

where

$$\bar{r} = \frac{1}{T}\int_0^T r_t\,dt$$

is the average interest rate (which a random variable). The price of the call option is $v = E\left[e^{-\bar{r}T}(S_T - K)^+\right]$, where the expected value is taken under the risk-neutral probability measure.

The Euler scheme of the two-dimensional process (S_t, r_t) is straightforward. Let $0 = t_0 < t_1 < \cdots < t_m = T$ be a time discretization. Let $\hat{S}_0 = S_0$ and $\hat{r}_0 = r_0$. Then for $i = 0, 1, \ldots, m-1$

$$\begin{aligned} \hat{S}_{t_{i+1}} &= \hat{S}_{t_i} + \hat{r}_{t_i}\hat{S}_{t_i}(t_{i+1} - t_i) + \sigma\hat{S}_{t_i}\sqrt{t_{i+1} - t_i}Z_{i+1}, \\ \hat{r}_{t_{i+1}} &= \hat{r}_{t_i} + a(b - \hat{r}_{t_i})(t_{i+1} - t_i) + \theta\hat{r}_{t_i}\sqrt{t_{i+1} - t_i}Y_{i+1}, \end{aligned}$$

where (Z_i, Y_i) are iid jointly normal random vectors with mean 0 and covariance matrix Σ. For each sample path, $\bar{r}$ can be approximated by

$$\bar{R} = \frac{1}{T}\sum_{i=0}^{m-1} \hat{r}_{t_i}(t_{i+1} - t_i).$$

The corresponding plain Monte Carlo estimate is just the sample average of iid copies of $e^{-\bar{R}T}(\hat{S}_T - K)^+$.

A more efficient algorithm is to combine the Euler scheme with the method of conditioning. The key observation is that the conditional expectation

$$h(\bar{r}, B_T) = E[e^{-\bar{r}T}(S_T - K)^+ | \bar{r}, B_T]$$

is explicitly solvable. Indeed, define

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

It is not difficult to show that (Q, B) is a two-dimensional standard Brownian motion; see Exercise 2.16. In other words, Q is a standard Brownian motion and is independent of B. Substituting $\rho B_t + \sqrt{1-\rho^2}Q_t$ for W_t, we have

$$\begin{aligned} S_T &= S_0 \exp\left\{\left(\bar{r} - \frac{1}{2}\sigma^2\right)T + \sigma\rho B_T + \sigma\sqrt{1-\rho^2}\,Q_T\right\} \\ &= X_0 \exp\left\{\left(\bar{r} - \frac{1}{2}\sigma^2(1-\rho^2)\right)T + \sigma\sqrt{1-\rho^2}\,Q_T\right\}, \end{aligned}$$

where

$$X_0 = S_0 \exp\left\{\rho\sigma B_T - \frac{1}{2}\rho^2\sigma^2 T\right\}.$$

Therefore, given $\bar{r}$ and B_T, S_T is the terminal value of a geometric Brownian motion with initial value X_0, drift $\bar{r}$, and volatility $\sigma\sqrt{1-\rho^2}$. Consequently, $h(\bar{r}, B_T)$ is the corresponding price of a call option with strike price K and maturity T. That is,

$$h(\bar{r}, B_T) = \text{BLS\_Call}(X_0, K, T, \bar{r}, \sigma\sqrt{1-\rho^2}), \tag{9.6}$$

where BLS_Call denotes the price of a call option under the classical Black–Scholes model; see Example 2.1. By the tower property, the option price is

$$v = E[h(\bar{r}, B_T)].$$

This leads to a very simple algorithm in which the stock price S does not need to be simulated at all. Only the process $\{r_t\}$ will be approximated via (say) the Euler scheme. Sampling B_T does not incur extra computational cost—it is just a by-product of the Euler scheme for $\{r_t\}$. Below is the pseudocode. Abusing notation, we use $\hat{r}_i$ to denote $\hat{r}_{t_i}$.

Pseudocode for Euler scheme and method of conditioning:

for $k = 1, 2, \ldots, n$
  set $\hat{r}_0 = r_0$ and $L = 0$
  for $i = 0, 1, \ldots, m-1$
    generate sample Y_{i+1} from $N(0,1)$
    set $\hat{r}_{i+1} = \hat{r}_i + a(b - \hat{r}_i)(t_{i+1} - t_i) + \theta \hat{r}_i \sqrt{t_{i+1} - t_i}\, Y_{i+1}$
    set $L = L + \sqrt{t_{i+1} - t_i}\, Y_{i+1}$
  set $B_T = L$
  set $R = [(t_1 - t_0)\hat{r}_0 + \cdots + (t_m - t_{m-1})\hat{r}_{m-1}]/T$
  set $X_0 = S_0 \exp\{\rho\sigma B_T - \rho^2\sigma^2 T/2\}$
  set $H_k = C(X_0, K, T, R, \sigma\sqrt{1-\rho^2})$

compute the estimate $\hat{v} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} H_i$

compute the standard error S.E. $= \sqrt{\dfrac{1}{n(n-1)}\left(\displaystyle\sum_{i=1}^{n} H_i^2 - n\hat{v}^2\right)}$.

The numerical results are reported in Table 9.4 for both the plain Monte Carlo scheme and the hybrid algorithm combining the Euler scheme and the method of conditioning. The parameters are given by

$$S_0 = 50,\ r_0 = 0.12,\ b = 0.10,\ a = 2,\ \theta = \sigma = 0.2,\ T = 1,\ \rho = 0.3.$$

The time interval is divided into $m = 50$ subintervals of equal length and the sample size is $n = 10000$.

Table 9.4: Method of conditioning with stochastic short rate

	$K = 45$		$K = 50$		$K = 55$	
	Plain	Hybrid	Plain	Hybrid	Plain	Hybrid
Estimate	10.3599	10.2925	6.8799	6.8584	4.3659	4.3207
S.E.	0.0925	0.0285	0.0822	0.0240	0.0685	0.0188

Clearly, the method of conditioning is more efficient than the plain Monte Carlo scheme when it comes to variance reduction. It is also possible to replace the Euler scheme for $\{r_t\}$ by the Milstein scheme to reduce the bias. But given the relatively small sample size, the performance is indistinguishable from the Euler scheme's. ■

Example 9.7. Control variate method by means of artificial dynamics. Consider a stochastic volatility model (the volatility is denoted by θ) where the risk-free interest rate is a constant r and under the risk-neutral probability measure,

$$\begin{aligned} dS_t &= rS_t\,dt + \theta_t S_t\,dW_t, \\ d\theta_t &= a(\Theta - \theta_t)\,dt + \beta\,dB_t, \end{aligned}$$

for some positive constants a, Θ, and β. Here (W, B) is a two-dimensional Brownian motion with covariance matrix

$$\Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}.$$

Estimate the price of a call option with strike price K and maturity T.

SOLUTION: By Itô formula, $Y_t = \log S_t$ satisfies the stochastic differential equation

$$dY_t = \left(r - \frac{1}{2}\theta_t^2\right)\,dt + \theta_t\,dW_t.$$

A very simple scheme for estimating the call option price is to apply the Euler scheme on the two-dimensional process (Y, θ). More precisely, let $0 = t_0 < t_1 < \cdots < t_m = T$ be a time discretization. Define $\hat{Y}_0 = Y_0 = \log S_0$ and $\hat{\theta}_0 = \theta_0$. For $i = 0, 1, \ldots, m-1$, recursively define

$$\begin{aligned} \hat{Y}_{t_{i+1}} &= \hat{Y}_{t_i} + \left(r - \frac{1}{2}\hat{\theta}_{t_i}^2\right)(t_{i+1} - t_i) + \hat{\theta}_{t_i}\sqrt{t_{i+1} - t_i}\,Z_{i+1}, \\ \hat{\theta}_{t_{i+1}} &= \hat{\theta}_{t_i} + a(\Theta - \hat{\theta}_{t_i})(t_{i+1} - t_i) + \beta\sqrt{t_{i+1} - t_i}\,R_{i+1}, \end{aligned}$$

where $\{(Z_1, R_1), \ldots, (Z_m, R_m)\}$ are iid jointly normal random vectors with distribution $N(0, \Sigma)$. A plain Monte Carlo estimate for the call option price is just the sample average of iid copies of $X = e^{-rT}(\exp\{\hat{Y}_T\} - K)^+$.

To improve the efficiency, one can use the control variate method by introducing an *artificial* stochastic process $\bar{Y}$ that is recursively defined by $\bar{Y}_0 = Y_0 = \log S_0$ and

$$\bar{Y}_{t_{i+1}} = \bar{Y}_{t_i} + \left(r - \frac{1}{2}\sigma^2\right)(t_{i+1} - t_i) + \sigma\sqrt{t_{i+1} - t_i}\,Z_{i+1}$$

for $i = 0, 1, \ldots, m-1$. This process $\bar{Y}$ uses the *same* sequence $\{Z_1, \ldots, Z_m\}$, and some *constant* volatility σ. One can think of $\bar{Y}$ as the logarithm of the price of a virtual stock that follows a classical geometric Brownian motion with drift r and volatility σ. Then the discounted payoff of the call option for this virtual stock, namely,

$$V = e^{-rT}(\exp(\bar{Y}_T) - K)^+$$

can be used as a control variate, whose expected value is just the classical Black–Scholes call option price $\text{BLS\_Call}(S_0, K, T, r, \sigma)$. The control variate estimate for the call option price is the sample average of iid copies of

$$X - b[V - \text{BLS\_Call}(S_0, K, T, r, \sigma)]$$

for some constant b. We opt to use $b = \hat{b}^*$, which is the sample estimate of the optimal coefficient; see formula (6.2).

The value of σ can be arbitrary in theory. However, it is usually chosen to be a typical value of θ in order to achieve a higher level of variance reduction. In our example, the process θ is a mean reverting Ornstein–Uhlenbeck process. Therefore, it is reasonable to let $\sigma = \Theta$, the long run average of θ. Below is the pseudocode. As before, we use $\hat{Y}_i$ and $\hat{\theta}_i$ to denote $\hat{Y}_{t_i}$ and $\hat{\theta}_{t_i}$, respectively.

Pseudocode for control variate method:

set $\sigma = \Theta$

for $k = 1, 2, \ldots, n$

  set $\hat{Y}_0 = \log S_0$, $\bar{Y}_0 = \log S_0$, and $\hat{\theta}_0 = \theta_0$

  for $i = 0, 1, \ldots, m-1$

    generate iid sample Z_{i+1} and U_{i+1} from $N(0,1)$

    set $R_{i+1} = \rho Z_{i+1} + \sqrt{1-\rho^2}\, U_{i+1}$

    set $\hat{Y}_{i+1} = \hat{Y}_i + (r - \hat{\theta}_i^2/2)(t_{i+1} - t_i) + \hat{\theta}_i \sqrt{t_{i+1} - t_i}\, Z_{i+1}$

    set $\bar{Y}_{i+1} = \bar{Y}_i + (r - \sigma^2/2)(t_{i+1} - t_i) + \sigma \sqrt{t_{i+1} - t_i}\, Z_{i+1}$

    set $\hat{\theta}_{i+1} = \hat{\theta}_i + a(\Theta - \hat{\theta}_i)(t_{i+1} - t_i) + \beta \sqrt{t_{i+1} - t_i}\, R_{i+1}$

  set $X_k = e^{-rT}(\exp\{\hat{Y}_m\} - K)^+$;

  set $Q_k = e^{-rT}(\exp\{\bar{Y}_m\} - K)^+ - \text{BLS\_Call}(S_0, K, T, r, \sigma)$

compute $\hat{b}^*$ from formula (6.2) [with Y replaced by Q]

for $k = 1, 2, \ldots, n$

  set $H_k = X_k - \hat{b}^* Q_k$

compute the estimate $\hat{v} = \dfrac{1}{n}\sum_{i=1}^{n} H_i$

compute the standard error S.E. $= \sqrt{\frac{1}{n(n-1)}\left(\sum_{i=1}^{n} H_i^2 - n\hat{v}^2\right)}$.

We compare the plain Euler scheme with the control variate method. The numerical results are reported in table 9.5 for

$$r = 0.05,\ S_0 = 50,\ K = 50,\ a = 3,\ \beta = 0.1,\ \theta_0 = 0.25,\ T = 1,$$

$$\Theta = 0.2,\ \rho = 0.5,\ m = 50,\ t_i = iT/m,\ n = 10000.$$

Table 9.5: Control variate method for a stochastic volatility model

	Euler	Control Variate				
		$\sigma = \Theta/5$	$\sigma = \Theta/2$	$\sigma = \Theta$	$\sigma = 2\Theta$	$\sigma = 5\Theta$
Estimate	5.5301	5.5044	5.5621	5.5688	5.5430	5.6202
S.E.	0.0867	0.0402	0.0252	0.0181	0.0171	0.0413

We have tested a wide range of σ to investigate its effect on variance reduction. The choice of $\sigma = \Theta$ has turned out to be nearly optimal. ■

Example 9.8. Importance sampling and cross-entropy method. Assume that the risk-free interest rate r is a constant and the underlying stock price is a constant elasticity of variance (CEV) process under the risk-neutral probability measure, that is,

$$dS_t = rS_t\,dt + \sigma S_t^{\gamma}\,dW_t$$

for some $0.5 \le \gamma < 1$. Estimate the price of a call option with maturity T and strike price K.

SOLUTION: Let $X_t = e^{-rt}S_t$. Then $X_0 = S_0$ and it follows from Itô formula that

$$dX_t = \sigma e^{-r(1-\gamma)t} X_t^{\gamma}\,dW_t.$$

The price of the call option is

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

In order to guarantee that the time discretized approximating process $\hat{X}$ is nonnegative, we modify the plain Euler scheme slightly. Let $0 = t_0 < t_1 < \cdots < t_m = T$ be a time discretization. Given $\hat{X}_0 = S_0$, define recursively

$$\hat{X}_{t_{i+1}} = \max\{0, \hat{X}_{t_i} + \sigma e^{-r(1-\gamma)t_i}\hat{X}_{t_i}^{\gamma}\sqrt{\Delta t_{i+1}} \cdot Z_{i+1}\}$$

for $i = 0, 1, \ldots, m-1$, where $\{Z_1, \ldots, Z_m\}$ are iid standard normal random variables and $\Delta t_{i+1} = t_{i+1} - t_i$. Note that if $\hat{X}$ ever reaches zero, it will stay at zero. This is exactly what we want because zero is an absorbing state for the original process X (loosely speaking, when X reaches zero, the volatility becomes zero and X stays at zero). The plain Monte Carlo estimate is the sample average of iid copies of

$$(\hat{X}_T - e^{-rT}K)^+,$$

which is indeed a function of $Y = (Z_1, \ldots, Z_m)$, say $h(Y)$. Note that Y is a jointly normal random vector with distribution $N(0, I_m)$.

Assuming that the alternative sampling distribution is $N(\theta, I_m)$ where $\theta = (\theta_1, \ldots, \theta_m) \in \mathbb{R}^m$ and the strike price K is not overly large, the basic cross-entropy method can be adopted to determine a nearly optimal tilting parameter $\hat{\theta}$. More precisely, by Lemma 7.1

$$\hat{\theta} = \frac{\sum_{k=1}^{N} h(Y_k) Y_k}{\sum_{k=1}^{N} h(Y_k)},$$

where Y_k's are iid jointly normal random vectors with distribution $N(0, I_m)$. Below is the pseudocode. As usual, we use $\hat{X}_i$ to denote $\hat{X}_{t_i}$.

Pseudocode for call price by the basic cross-entropy method:

generate iid pilot samples $Y_1, \ldots, Y_N$ from $N(0, I_m)$
set $\hat{\theta} = (\hat{\theta}_1, \ldots, \hat{\theta}_m) = \sum_{k=1}^{N} h(Y_k) Y_k / \sum_{k=1}^{N} h(Y_k)$
for $k = 1, 2, \ldots, n$
  for $i = 0, 1, \ldots, m-1$
    generate Z_{i+1} from $N(\hat{\theta}_{i+1}, 1)$
    set $\hat{X}_{i+1} = \max\{0, \hat{X}_i + \sigma e^{-r(1-\gamma)t_i} \hat{X}_i^{\gamma} \sqrt{\Delta t_{i+1}} \cdot Z_{i+1}\}$
  compute the discounted payoff multiplied by the likelihood ratio

$$H_k = (\hat{X}_m - e^{-rT}K)^+ \cdot \exp\left\{ -\sum_{i=1}^{m} \hat{\theta}_i Z_i + \frac{1}{2} \sum_{i=1}^{m} \hat{\theta}_i^2 \right\}$$

compute the estimate $\hat{v} = \dfrac{1}{n} \displaystyle\sum_{i=1}^{n} H_i$

compute the standard error S.E. $= \sqrt{\dfrac{1}{n(n-1)} \left(\displaystyle\sum_{i=1}^{n} H_i^2 - n\hat{v}^2 \right)}.$

We compare the plain Monte Carlo scheme with the basic cross-entropy method. The numerical results are reported in Table 9.6 for

$$S_0 = 50, \; r = 0.05, \; \sigma = 0.2, \; T = 1, \; m = 50, \; n = 10000, \; N = 2000.$$

The basic cross-entropy scheme is clearly more efficient.

Table 9.6: Basic cross-entropy method for a CEV model

$K = 50$	$\gamma = 0.5$		$\gamma = 0.7$		$\gamma = 0.9$	
	Plain	IS	Plain	IS	Plain	IS
Estimate	2.4455	2.4603	2.7788	2.7716	3.9644	4.0485
S.E.	0.0135	0.0065	0.0257	0.0127	0.0503	0.0214
$K = 55$	$\gamma = 0.5$		$\gamma = 0.7$		$\gamma = 0.9$	
	Plain	IS	Plain	IS	Plain	IS
Estimate	0.0272	0.0298	0.4131	0.4075	1.7319	1.7270
S.E.	0.0017	0.0003	0.0106	0.0035	0.0346	0.0127

Both the basic cross-entropy method and the plain Monte Carlo scheme will fail as the strike price increases to a certain level. A remedy is the general iterative cross-entropy method. However, a decent initial tilting parameter $\hat{\theta}^0$ is required for this to work. Even though it is straightforward to use the general initialization technique outlined in Section 7.2.3 to produce such an initial tilting parameter, we should present an alternative approach that does not require extra simulation. Assume that

$$\hat{\theta}^0 = x(\sqrt{\Delta t_1}, \sqrt{\Delta t_2}, \ldots, \sqrt{\Delta t_m})$$

for some constant x. Under the alternative sampling distribution $N(\hat{\theta}^0, I_m)$, the discretization scheme becomes

$$\hat{X}_{t_{i+1}} = \max\{0, \hat{X}_{t_i} + \sigma e^{-r(1-\gamma)t_i} \hat{X}_{t_i}^{\gamma} \sqrt{\Delta t_{i+1}} \cdot \bar{Z}_{i+1}\},$$

where $\bar{Y} = (\bar{Z}_1, \ldots, \bar{Z}_m)$ is a jointly normal random vector with distribution $N(\hat{\theta}^0, I_m)$. What we want is to choose x so that $E[\hat{X}_T]$ is approximately $e^{-rT}K$, and thus a reasonable fraction of samples will yield strictly positive payoffs. To this end, rewrite $\bar{Z}_{i+1} = x\sqrt{\Delta t_{i+1}} + R_{i+1}$ for each i, where $\{R_1, \ldots, R_m\}$ are iid standard normals. We arrive at

$$\hat{X}_{t_{i+1}} = \max\{0, \hat{X}_{t_i} + \sigma e^{-r(1-\gamma)t_i} \hat{X}_{t_i}^{\gamma} (x\Delta t_{i+1} + \sqrt{\Delta t_{i+1}} \cdot R_{i+1})\}.$$

It is easy to see that $\hat{X}$ is the time discretized approximation of $\bar{X}$ where

$$d\bar{X}_t = x\sigma e^{-r(1-\gamma)t} \bar{X}_t^{\gamma} \, dt + \sigma e^{-r(1-\gamma)t} \bar{X}_t^{\gamma} \, dW_t, \quad \bar{X}_0 = X_0 = S_0.$$

We compare the plain Monte Carlo scheme with the basic cross-entropy method. The numerical results are reported in Table 9.6 for

$$S_0 = 50,\ r = 0.05,\ \sigma = 0.2,\ T = 1,\ m = 50,\ n = 10000,\ N = 2000.$$

The basic cross-entropy scheme is clearly more efficient.

Table 9.6: Basic cross-entropy method for a CEV model

$K = 50$	$\gamma = 0.5$		$\gamma = 0.7$		$\gamma = 0.9$	
	Plain	IS	Plain	IS	Plain	IS
Estimate	2.4455	2.4603	2.7788	2.7716	3.9644	4.0485
S.E.	0.0135	0.0065	0.0257	0.0127	0.0503	0.0214
$K = 55$	$\gamma = 0.5$		$\gamma = 0.7$		$\gamma = 0.9$	
	Plain	IS	Plain	IS	Plain	IS
Estimate	0.0272	0.0298	0.4131	0.4075	1.7319	1.7270
S.E.	0.0017	0.0003	0.0106	0.0035	0.0346	0.0127

Both the basic cross-entropy method and the plain Monte Carlo scheme will fail as the strike price increases to a certain level. A remedy is the general iterative cross-entropy method. However, a decent initial tilting parameter $\hat{\theta}^0$ is required for this to work. Even though it is straightforward to use the general initialization technique outlined in Section 7.2.3 to produce such an initial tilting parameter, we should present an alternative approach that does not require extra simulation. Assume that

$$\hat{\theta}^0 = x(\sqrt{\Delta t_1}, \sqrt{\Delta t_2}, \ldots, \sqrt{\Delta t_m})$$

for some constant x. Under the alternative sampling distribution $N(\hat{\theta}^0, I_m)$, the discretization scheme becomes

$$\hat{X}_{t_{i+1}} = \max\{0, \hat{X}_{t_i} + \sigma e^{-r(1-\gamma)t_i} \hat{X}_{t_i}^{\gamma} \sqrt{\Delta t_{i+1}} \cdot \bar{Z}_{i+1}\},$$

where $\bar{Y} = (\bar{Z}_1, \ldots, \bar{Z}_m)$ is a jointly normal random vector with distribution $N(\hat{\theta}^0, I_m)$. What we want is to choose x so that $E[\hat{X}_T]$ is approximately $e^{-rT}K$, and thus a reasonable fraction of samples will yield strictly positive payoffs. To this end, rewrite $\bar{Z}_{i+1} = x\sqrt{\Delta t_{i+1}} + R_{i+1}$ for each i, where $\{R_1, \ldots, R_m\}$ are iid standard normals. We arrive at

$$\hat{X}_{t_{i+1}} = \max\{0, \hat{X}_{t_i} + \sigma e^{-r(1-\gamma)t_i} \hat{X}_{t_i}^{\gamma} (x\Delta t_{i+1} + \sqrt{\Delta t_{i+1}} \cdot R_{i+1})\}.$$

It is easy to see that $\hat{X}$ is the time discretized approximation of $\bar{X}$ where

$$d\bar{X}_t = x\sigma e^{-r(1-\gamma)t} \bar{X}_t^{\gamma}\, dt + \sigma e^{-r(1-\gamma)t} \bar{X}_t^{\gamma}\, dW_t, \quad \bar{X}_0 = X_0 = S_0.$$

Therefore, it suffices to choose x so that $E[\bar{X}_T]$ is approximately $e^{-rT}K$. Even though the expected value of $E[\bar{X}_T]$ is not explicitly known, we observe that by the martingale property of stochastic integrals

$$\begin{aligned} E[\bar{X}_t] &= \bar{X}_0 + x\sigma \int_0^t e^{-r(1-\gamma)s} E[\bar{X}_s^\gamma]\, ds \\ &\approx S_0 + x\sigma \int_0^t e^{-r(1-\gamma)s} (E[\bar{X}_s])^\gamma\, ds. \end{aligned} \tag{9.7}$$

Here we have approximated $E[\bar{X}_s^\gamma]$ by $(E[\bar{X}_s])^\gamma$. These two quantities would be the same if $\gamma = 1$ or $\bar{X}_s$ were a constant. This implies that the approximation is good when γ is close to 1. When γ is away from 1, the variance of $\bar{X}_s$ is often relatively small, which again justifies the validity of the approximation. Now (9.7) implies that $E[\bar{X}_t]$ is approximately $f(t)$, where $f(t)$ satisfies

$$f(t) = S_0 + x\sigma \int_0^t e^{-r(1-\gamma)s} f^\gamma(s)\, ds,$$

or equivalently,

$$\frac{1}{f^\gamma}\frac{df}{dt} = x\sigma e^{-r(1-\gamma)t}, \quad f(0) = S_0.$$

Solving this ordinary differential equation explicitly, we obtain

$$f^{1-\gamma}(t) - S_0^{1-\gamma} = \frac{x\sigma}{r}\left[1 - e^{-r(1-\gamma)t}\right].$$

Letting $f(T) = e^{-rT}K$, we arrive at

$$x = \frac{r}{\sigma}\frac{(e^{-rT}K)^{1-\gamma} - S_0^{1-\gamma}}{1 - e^{-r(1-\gamma)T}}. \tag{9.8}$$

Below is the pseudocode.

Pseudocode for the general iterative cross-entropy method:

set x by formula (9.8)
initialize $\hat{\theta}^0 = x(\sqrt{\Delta t_1}, \ldots, \sqrt{\Delta t_m})$ and set the iteration counter $j = 0$
(*) generate N iid samples $\bar{Y}_1, \ldots, \bar{Y}_N$ from $N(\hat{\theta}^j, I_m)$
set $\hat{\theta}^{j+1} = \sum_{k=1}^N h(\bar{Y}_k) e^{-\langle \hat{\theta}^j, \bar{Y}_k\rangle} \bar{Y}_k / \sum_{k=1}^N h(\bar{Y}_k) e^{-\langle \hat{\theta}^j, \bar{Y}_k\rangle}$
set the iteration counter $j = j + 1$

if $j =$ IT_NUM set $\hat{\theta} = \hat{\theta}^j$ and continue, otherwise go to step (*)

for $k = 1, 2, \ldots, n$

set $\hat{X}_0 = S_0$

for $i = 0, 1, \ldots, m-1$

generate Z_{i+1} from $N(\hat{\theta}_{i+1}, 1)$

set $\hat{X}_{i+1} = \max\{0, \hat{X}_i + \sigma e^{-r(1-\gamma)t_i} \hat{X}_i^{\gamma} \sqrt{\Delta t_{i+1}} \cdot Z_{i+1}\}$

compute the discounted payoff multiplied by the likelihood ratio

$$H_k = (\hat{X}_m - e^{-rT}K)^+ \cdot \exp\left\{ -\sum_{i=1}^{m} \hat{\theta}_i Z_i + \frac{1}{2}\sum_{i=1}^{m} \hat{\theta}_i^2 \right\}$$

compute the estimate $\hat{v} = \frac{1}{n}\sum_{i=1}^{n} H_i$

compute the standard error S.E. $= \sqrt{\frac{1}{n(n-1)}\left(\sum_{i=1}^{n} H_i^2 - n\hat{v}^2\right)}$.

Table 9.7: Iterative cross-entropy method for a CEV model

$K = 57$	$\gamma = 0.5$		$\gamma = 0.7$		$\gamma = 0.9$	
	Plain	IS	Plain	IS	Plain	IS
Estimate	0.0005	6.2068×10^{-4}	0.1416	0.1315	1.1537	1.1931
S.E.	0.0002	7.7895×10^{-6}	0.0061	0.0012	0.0282	0.0084
$K = 60$	$\gamma = 0.5$		$\gamma = 0.7$		$\gamma = 0.9$	
	Plain	IS	Plain	IS	Plain	IS
Estimate	0	1.3247×10^{-7}	0.0171	0.0157	0.6100	0.6072
S.E.	0	2.1961×10^{-9}	0.0018	0.0002	0.0202	0.0048

We present the numerical results in Table 9.7 with the same parameters, except that the strike price is larger. The sample size is $n = 10000$ and the pilot sample size is $N = 2000$ with IT_NUM $= 5$ iterations. ■

Example 9.9. Importance sampling and cross-entropy method. Consider the Heston model where the risk-free interest rate is a constant r and under the risk-neutral probability measure

$$\begin{aligned} dS_t &= rS_t\,dt + \sqrt{\theta_t} S_t\,dW_t, \\ d\theta_t &= a(b - \theta_t)\,dt + \sigma\sqrt{\theta_t}\,dB_t, \end{aligned}$$

for some positive constants a, b, and σ such that $2ab \geq \sigma^2$. Here (W, B) is a two-dimensional Brownian motion with covariance matrix

$$\Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}.$$

Estimate the price of a call option with maturity T and strike price K.

SOLUTION: The call option price is $v = E[e^{-rT}(S_T - K)^+]$. A slightly modified Euler scheme for the two-dimensional process (S, θ), which ensures that $\hat{\theta}$ stays nonnegative, is straightforward. Given a time discretization $0 = t_0 < t_1 < \ldots < t_m = T$, let $(\hat{S}_0, \hat{\theta}_0) = (S_0, \theta_0)$ and define recursively

$$\begin{aligned} \hat{S}_{t_{i+1}} &= \hat{S}_{t_i} + r\hat{S}_{t_i}\Delta t_{i+1} + \hat{S}_{t_i}\sqrt{\hat{\theta}_{t_i}\Delta t_{i+1}} \cdot Z_{i+1}, \\ \hat{\theta}_{t_{i+1}} &= \max\{0, \hat{\theta}_{t_i} + a(b - \hat{\theta}_{t_i})\Delta t_{i+1} + \sigma\sqrt{\hat{\theta}_{t_i}\Delta t_{i+1}} \cdot R_{i+1}\}, \end{aligned}$$

for $i = 0, \ldots, m-1$. Here $\Delta t_{i+1} = t_{i+1} - t_i$ and (Z_{i+1}, R_{i+1})'s are iid jointly normal random vectors with mean 0 and covariance matrix Σ. The plain Monte Carlo estimate for the call option price is just the sample average of iid copies of $e^{-rT}(\hat{S}_T - K)^+$. Write $R_{i+1} = \rho Z_{i+1} + \sqrt{1-\rho^2}Y_{i+1}$, where $\{Z_1, \ldots, Z_m, Y_1, \ldots, Y_m\}$ are iid standard normal random variables. The discounted option payoff is a function of $X = (Z_1, \ldots, Z_m, Y_1, \ldots, Y_m)$, say

$$h(X) = e^{-rT}(\hat{S}_T - K)^+. \tag{9.9}$$

As for importance sampling, let the alternative sampling distribution be $N(\mu, I_{2m})$ where $\mu = (\nu, \upsilon)$ with $\nu \in \mathbb{R}^m$ and $\upsilon \in \mathbb{R}^m$. Note that we change the sampling distribution for both stock price S and volatility θ. One can use the general iterative cross-entropy method to determine a good tilting parameter (actually for moderate strike price K the basic cross-entropy scheme should suffice, analogous to Example 9.8). To find a reasonable initial tilting parameter $\hat{\mu}^0$, we assume it takes the form

$$\hat{\mu}^0 = (\hat{\nu}^0, \hat{\upsilon}^0), \quad \hat{\nu}^0 = x(\sqrt{\Delta t_1}, \ldots, \sqrt{\Delta t_m}), \quad \hat{\upsilon}^0 = -\frac{\rho}{\sqrt{1-\rho^2}}\hat{\nu}^0 \tag{9.10}$$

for some $x \in \mathbb{R}$. To determine x, observe that the corresponding time discretized process $(\bar{S}, \bar{\theta})$ becomes

$$\begin{aligned} \bar{S}_{t_{i+1}} &= \bar{S}_{t_i} + r\bar{S}_{t_i}\Delta t_{i+1} + \bar{S}_{t_i}\sqrt{\bar{\theta}_{t_i}\Delta t_{i+1}} \cdot \bar{Z}_{i+1}, \\ \bar{\theta}_{t_{i+1}} &= \max\{0, \bar{\theta}_{t_i} + a(b - \bar{\theta}_{t_i})\Delta t_{i+1} + \sigma\sqrt{\bar{\theta}_{t_i}\Delta t_{i+1}} \cdot \bar{R}_{i+1}\}, \end{aligned}$$

where $\bar{X} = (\bar{Z}_1, \ldots, \bar{Z}_m, \bar{Y}_1, \ldots, \bar{Y}_m)$ is a sample from $N(\hat{\mu}^0, I_{2m})$, and

$$\bar{R}_{i+1} = \rho \bar{Z}_{i+1} + \sqrt{1-\rho^2}\,\bar{Y}_{i+1}.$$

Note that $\{\bar{R}_1, \ldots, \bar{R}_m\}$ are iid *standard normal* random variables. Therefore, it is easy to see that $(\bar{S}, \bar{\theta})$ approximates (abusing notation)

$$\begin{aligned} dS_t &= (r + x\sqrt{\theta_t}) S_t\, dt + \sqrt{\theta_t} S_t\, dW_t \\ d\theta_t &= a(b - \theta_t)\, dt + \sigma\sqrt{\theta_t}\, dB_t. \end{aligned}$$

We would like to pick an x so that $E[S_T]$ is approximately K. There is no explicit formula for $E[S_T]$. However, we can approximate θ_t by its long-run average b in the dynamics of S and consider instead

$$dS_t \approx (r + x\sqrt{b}) S_t\, dt + \sqrt{b} S_t\, dW_t.$$

It follows that

$$E[S_T] \approx S_0 e^{(r + x\sqrt{b})T}.$$

Letting

$$x = \frac{1}{\sqrt{b}T} \log\left(\frac{K}{S_0}\right) - \frac{r}{\sqrt{b}}, \tag{9.11}$$

we arrive at $E[S_T] \approx K$. Below is the pseudocode. As before, $\hat{S}_i$ and $\hat{\theta}_i$ stand for $\hat{S}_{t_i}$ and $\hat{\theta}_{t_i}$, respectively.

Pseudocode for Heston model using the cross-entropy method:

set x by formula (9.11)

initialize $\hat{\nu}^0 = x(\sqrt{\Delta t_1}, \ldots, \sqrt{\Delta t_m})$ and $\hat{\upsilon}^0 = -\rho\hat{\nu}^0/\sqrt{1-\rho^2}$

set $\hat{\mu}^0 = (\hat{\nu}^0, \hat{\upsilon}^0)$ and set the iteration counter $j = 0$

(*) generate N iid samples $\bar{X}_1, \ldots, \bar{X}_N$ from $N(\hat{\mu}^j, I_{2m})$

set $\hat{\mu}^{j+1} = \sum_{k=1}^N h(\bar{X}_k) e^{-\langle \hat{\mu}^j, \bar{X}_k \rangle} \bar{X}_k / \sum_{k=1}^N h(\bar{X}_k) e^{-\langle \hat{\mu}^j, \bar{X}_k \rangle}$

set the iteration counter $j = j + 1$

if $j =$ IT_NUM set $\hat{\mu} = \hat{\mu}^j$ and continue, otherwise go to step (*)

write $\hat{\mu} = (\hat{\nu}, \hat{\upsilon})$

for $k = 1, 2, \ldots, n$

    set $\hat{S}_0 = S_0$ and $\hat{\theta}_0 = \theta_0$

    for $i = 0, 1, \ldots, m-1$

        generate Z_{i+1} from $N(\hat{\nu}_{i+1}, 1)$ and Y_{i+1} from $N(\hat{\upsilon}_{i+1}, 1)$

$$\text{set } R_{i+1} = \rho Z_{i+1} + \sqrt{1-\rho^2} Y_{i+1}$$
$$\text{set } \hat{S}_{i+1} = \hat{S}_i + r\hat{S}_i \Delta t_{i+1} + \hat{S}_i \sqrt{\hat{\theta}_i \Delta t_{i+1}} \cdot Z_{i+1}$$
$$\text{set } \hat{\theta}_{i+1} = \max\{0, \hat{\theta}_i + a(b-\hat{\theta}_i)\Delta t_{i+1} + \sigma\sqrt{\hat{\theta}_i \Delta t_{i+1}} \cdot R_{i+1}\}$$

compute the discounted payoff multiplied by the likelihood ratio

$$H_k = e^{-rT}(\hat{S}_m - K)^+ \cdot \exp\left\{ -\sum_{i=1}^{m} (\hat{\nu}_i Z_i + \hat{v}_i Y_i) + \frac{1}{2}\sum_{i=1}^{m} (\hat{\nu}_i^2 + \hat{v}_i^2) \right\}$$

compute the estimate $\hat{v} = \frac{1}{n}\sum_{i=1}^{n} H_i$

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 numerical results are reported in Table 9.8. The parameters are defined by

$$S_0 = 50,\ r = 0.03,\ b = 0.04,\ a = 1,\ \sigma = 0.1,\ \rho = 0.5,$$
$$\theta_0 = 0.05,\ m = 30,\ n = 10000.$$

The cross-entropy scheme uses $N = 2000$ pilot samples and IT_NUM $= 5$ iterations.

Table 9.8: Iterative cross-entropy method for a Heston model

	$K = 50$		$K = 70$		$K = 90$		$K = 110$	
	Plain	IS	Plain	IS	Plain	IS	Plain	IS
Estimate	5.0687	5.0495	0.5175	0.5521	0.0495	0.0514	0.0029	0.0054
S.E.	0.0819	0.0277	0.0281	0.0051	0.0097	0.0006	0.0014	0.0001

The iterative cross-entropy method significantly reduces the variance, especially when the strike price is higher.

Discussion on the initial tilting parameter $\hat{\mu}^0$: If one plays around with the given iterative cross-entropy scheme, it will soon become clear that the scheme will behave erratically for certain ranges of parameters. Sometimes this issue can be resolved by increasing the pilot sample size N or the number of iterations IT_NUM, but not always. The cause of this problem is that the initial tilting parameter given by (9.10) and (9.11) might be too far away from the optimal one. If we examine (9.10) and (9.11) more carefully, they represent an alternative sampling distribution where the distribution of the original volatility process remains the same and only the distribution of the stock price is changed. However, there are two ways for the stock price to hit the strike price K with nontrivial probability under an alternative sampling distribution: (1) the stock price is tilted higher toward K; (2) the volatility is tilted higher so that the stock price is more likely to reach K. If the optimal sampling distribution has significant contribution from the latter, the initial tilting parameter defined by (9.10) and (9.11) may not be a good choice.

A general solution is to adopt the initialization technique outlined in Section 7.2.3. The scheme is rather straightforward. Write the payoff function in (9.9) as

$$h(X) = e^{-rT}(\hat{S}_T - K)^+ = H(X;K)1_{\{F(X)\geq K\}}$$

where $F(X) = \hat{S}_T$ and $H(X;\alpha) = e^{-rT}(F(X) - \alpha)^+$. Below is the pseudocode for generating an initial tilting parameter.

Pseudocode for the initialization of the cross-entropy scheme:

choose ρ between 5% and 10% and set $N_0 = [N(1-\rho)]$
set $\bar{v}^0 = (0,\ldots,0)$ and $\bar{v}^0 = (0,\ldots,0)$
set $\bar{\mu}^0 = (\bar{v}^0, \bar{v}^0)$ and the iteration counter $j = 0$
(*) generate N iid samples $Y_1,\ldots,Y_N$ from $N(\bar{\mu}^j, I_{2m})$
set $V_k = F(Y_k)$ and the order statistics $V_{(1)} \leq \cdots \leq V_{(N)}$
set $\bar{\alpha}_{j+1} = V_{(N_0)}$
update $\bar{\mu}^{j+1}$ according to (7.18) with $\hat{\theta}^j$ replaced by $\bar{\mu}^j$
set the iteration counter $j = j+1$
if $\bar{\alpha}_j \geq K$ set $\hat{\mu}^0 = \bar{\mu}^j$ and stop, otherwise go to step (*).

Note that the final tilting parameter from this scheme will serve as the *initial tilting parameter* for the general iterative cross-entropy scheme.

Table 9.9: Cross-entropy method with general initialization technique

	$K = 130$	$b = 0.01,\ K = 100$	$\theta_0 = 0.01,\ K = 130$
Estimate	6.3417×10^{-4}	0.0064	2.5654×10^{-6}
S.E.	9.2801×10^{-6}	0.0001	4.6579×10^{-8}
NUM_IT	3	2	3

Additional numerical results are reported in Table 9.9. We have chosen those cases where the previous cross-entropy scheme [i.e., with the initial tilting parameter given by (9.10) and (9.11)] has difficulty with. The plain Euler scheme estimates are not reported as they are meaningless. The parameters of the model, except those indicated in the table, remain the same. With the general initialization technique, the iterative cross-entropy scheme is quite robust and constantly yields very accurate estimates. We should mention that the number of iterations for the general iterative cross-entropy scheme is still set at IT_NUM $= 5$. The entry "NUM_IT" in Table 9.9 records the number of iterations performed in the initialization program. ■

Exercises

Pen-and-Paper Problems

9.1 Suppose that X satisfies the stochastic differential equation

$$dX_t = (2re^t \sqrt{X_t} + \sigma^2)\,dt + 2\sigma\sqrt{X_t}\,dW_t,$$

where r and σ are both positive constants.

(a) Find a function F such that $Y_t = F(X_t)$ is a diffusion process with constant diffusion coefficient σ. That is, Y satisfies a stochastic differential equation of the form

$$dY_t = b(t, Y_t)\,dt + \sigma\,dW_t.$$

(b) For your choice of F, compute the drift function b.

(c) Write down a discretization scheme for X with no discretization error.

9.2 Consider a diffusion process X that satisfies the stochastic differential equation

$$dX_t = b(t, X_t)\,dt + \sigma(t, X_t)\,dW_t,$$

where $\sigma(t,x)$ is a strictly positive function. Find a function $F(t,x)$ such that $Y_t = F(t, X_t)$ satisfies

$$dY_t = \mu(t, Y_t)\,dt + dW_t$$

for some function μ. Determine F and μ. This is a generalized version of the Lamperti transform.

9.3 Find the Lamperti transform of the following diffusion processes. All the constants involved are assumed to be strictly positive.

(a) $dX_t = a(b - X_t)\,dt + \sigma\sqrt{X_t}\,dW_t$

(b) $dX_t = (r_0 + r_1 X_t)\,dt + \sigma X_t\,dW_t$

9.4 Suppose that r is a strictly positive constant. Find the Lamperti transform of the diffusion process X that satisfies

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

Is the Lamperti transform valid when r is negative?

9.5 Write down the Euler scheme and the Milstein scheme for the following diffusion processes.

(a) $dX_t = r\,dt + \sigma X_t\,dW_t.$

(b) $dX_t = rX_t\,dt + \sqrt{\sigma_0^2 + \sigma_1^2 X_t^2}\,dW_t.$

9.6 Let a, b, and β be positive constants. Consider a mean-reverting Ornstein–Uhlenbeck process X such that

$$dX_t = a(b - X_t)\,dt + \beta dW_t\,.$$

Use the results in Example 8.6 to devise a discretization scheme for X with no discretization error.

9.7 Let $r(t)$ and $\theta(t)$ be two deterministic functions. Devise a discretization scheme for X with no discretization error, where

$$dX_t = r(t)\,X_t\,dt + \theta(t)X_t\,dW_t\,.$$

MATLAB® Problems

For all the numerical schemes, divide the time interval $[0, T]$ into m subintervals of equal length. That is,

$$t_i = \frac{i}{m}T, \quad i = 0, 1, \ldots, m.$$

The sample size is always denoted by n.

9.A Estimate the error $E|\hat{X}_T - X_T|$ where $\hat{X}$ is the time discretized approximating process for the diffusion

$$dX_t = a(b - X_t)\,dt + \sigma\,dW_t$$

via the Euler scheme. You will need a discretization scheme for X without discretization error (Exercise 9.6), and it will be necessary to simulate the random vector

$$\left(W_{t_{i+1}} - W_{t_i},\ \int_{t_i}^{t_{i+1}} \sigma e^{at}\,dW_t\right),$$

for which you might want to use the results from Exercise 8.9. Report your results for

$$a = 2,\ \ b = 1,\ \ \sigma = 0.2,\ \ T = 1,\ \ X_0 = 1,$$

and $m = 5, 10, 20, 40, 60, 80, 100$, respectively. Let the sample size be $n = 1{,}000{,}000$. Describe the relationship between the error and the step size $h = T/m$.

9.B Assume that under the risk-neutral probability measure, the short rate r_t is a mean-reverting process:

$$dr_t = a(b - r_t)\,dt + \sigma r_t dW_t\,.$$

Write a function to compare the Euler scheme and the Milstein scheme in the estimation of the price of a zero-coupon bond with maturity T. Report your results for

$$r_0 = 0.12,\ a = 1,\ b = 0.10,\ \sigma = 0.2,\ T = 1,\ m = 50,\ n = 10000.$$

Is there really a difference in the performance of these two schemes?

9.C Suppose that under the risk-neutral probability measure, the stock price satisfies the stochastic differential equation

$$\frac{dS_t}{S_t} = r_t\, dt + \sigma\, dW_t$$

and the short rate $\{r_t\}$ is a Cox–Ingersoll–Ross process:

$$dr_t = a(b - r_t)\, dt + \theta\sqrt{r_t}\, dB_t,$$

where (W, B) is a two-dimensional Brownian motion with covariance matrix

$$\Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix},$$

and a, b, σ, θ are all positive constants. Estimate the price of a call option with strike price K and maturity T by the method of conditioning. Compare with the plain Monte Carlo estimate. Report your results for

$$S_0 = 50,\ \sigma = 0.2,\ r_0 = 0.04,\ a = 1,\ b = 0.06,\ \theta = 0.1,\ \rho = 0.8,\ T = 1,$$

$$m = 50,\ \ n = 10000.$$

and $K = 45, 50, 55$, respectively.

9.D Assume that under the risk-neutral probability measure, the stock price S satisfies the stochastic differential equation

$$dS_t = rS_t\, dt + \sqrt{\sigma_0^2 + \sigma_1^2 S_t^2}\, dW_t,$$

where r is the risk-free interest rate and (σ_0, σ_1) are two positive constants. Estimate the prices of the following options with the control variate method and explain how you construct the control variate. Compare with the plain Monte Carlo estimate.

(a) A call option with maturity T and strike price K.

(b) An average price call option with maturity T and strike price K, whose payoff is

$$\left(\frac{1}{m}\sum_{i=1}^{m} S_{t_i} - K\right)^+.$$

Report your results for

$$S_0 = 50,\ r = 0.03,\ K = 50,\ T = 1,\ m = 30,\ \sigma_0 = 2,\ \sigma_1 = 0.2,$$

with sample size $n = 10000$.

9.E Repeat Exercise 9.D for the Heston model where the risk-free interest rate is a constant r and under the risk-neutral probability measure

$$\begin{aligned} dS_t &= rS_t\,dt + \sqrt{\theta_t}S_t\,dW_t \\ d\theta_t &= a(b - \theta_t)\,dt + \sigma\sqrt{\theta_t}\,dB_t, \end{aligned}$$

for some positive constants a, b, and σ such that $2ab \ge \sigma^2$. Here (W, B) is a two-dimensional Brownian motion with covariance matrix

$$\Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}.$$

Report your results for

$$S_0 = 50,\ r = 0.03,\ a = 1,\ b = 0.2,\ \sigma = 0.6,\ \theta_0 = 0.3,\ T = 1,\ K = 50,$$

$$\rho = 0.6,\ m = 50, n = 10000.$$

9.F Suppose that under the risk-neutral probability measure the stock price is a CEV process:

$$dS_t = rS_t\,dt + \sigma\sqrt{S_t}\,dt.$$

(a) Write a function to estimate the price of a put option with maturity T and strike price K by the basic cross-entropy method. Report your results for

$$S_0 = 50,\ r = 0.05,\ \sigma = 0.2,\ T = 1,\ m = 50,\ n = 10000,$$

and $K = 55, 52, 50$, respectively. The pilot sample size is $N = 2000$.

(b) Write a function to estimate the put option price by the general iterative cross-entropy method when the strike price K is lower. Indicate your choice of the initial tilting parameter. Report your results for $K = 48, 45$, respectively. Use IT_NUM $= 5$ iterations. All other parameters remain the same.

9.G Suppose that under the risk-neutral probability measure the stock price is a CEV process:

$$dS_t = rS_t\,dt + \sigma S_t^{\gamma}\,dt$$

for some $0.5 \le \gamma < 1$. Write a function to estimate the price of an average price call option with maturity T and payoff

$$\left(\frac{1}{m}\sum_{i=1}^{m} S_{t_i} - K\right)^+,$$

by the general iterative cross-entropy method. Use IT_NUM $= 5$ iterations, each iteration using $N = 2000$ pilot samples. Report your results for

$$S_0 = 50,\ r = 0.05,\ \sigma = 0.2,\ T = 1,\ m = 50,\ n = 10000,$$

and (a) $\gamma = 0.5$, $K = 50, 52, 55$; (b) $\gamma = 0.7$, $K = 50, 55, 60$; (c) $\gamma = 0.9$, $K = 50, 60, 70$, respectively. Indicate your choice of the initial tilting parameter.

Chapter 10

Sensitivity Analysis

Sensitivity analysis is important for the purpose of hedging a portfolio or managing risk [16, 24]. Consider a simple scenario where an investment is made on a stock option with maturity T. Denote the value of the option at time $t \in [0, T]$ by

$$V(S, r, \sigma, t),$$

where S is the underlying stock price at time t, r is the risk-free interest rate, and σ is the volatility. Consider a strategy of hedging one share of the option with x shares of the stock. The value of this portfolio at time t is

$$V(S, r, \sigma, t) + xS.$$

Letting the derivative of this value with respect to the stock price S be zero, we obtain

$$x = -\frac{\partial V}{\partial S}(S, r, \sigma, t).$$

With this choice of x, one would expect the portfolio value to be insensitive to the movement of the stock price. The partial derivative

$$\frac{\partial V}{\partial S}$$

is said to be the *delta* of the option, which indicates how sensitive the value of the option is with respect to the movement of the underlying asset price. For example, if the delta of an option is 0.5, then the price of the option will roughly increase or decrease by \$0.5 if the stock price moves up or down by \$1, respectively. Similarly, if the delta of an option is -0.6, then the price of the option will roughly decrease or increase by \$0.6 if the stock price moves up or down by \$1, respectively.

Partial derivatives such as delta are referred to as *Greeks*. They measure the sensitivity of the value of a financial instrument with respect to the underlying parameters such as asset prices and volatilities. The name "Greeks" is used because these sensitivities are often denoted by Greek letters.

10.1 Commonly Used Greeks

Let $V = V(S, r, \sigma, t)$ denote the value of an option at time t, where S stands for the underlying stock price at time t, r is the risk-free interest rate, and σ is the volatility. Some commonly used Greeks are listed in Table 10.1.

Table 10.1: Commonly used Greeks

Greeks	Notation	Definition
Delta	Δ	$\partial V/\partial S$
Gamma	Γ	$\partial^2 V/\partial S^2$
Rho	ρ	$\partial V/\partial r$
Theta	Θ	$\partial V/\partial t$
Vega	ν	$\partial V/\partial \sigma$

Many times it is only one parameter, say S, that is of interest. In these cases, the option value is simply denoted by $V(S)$ if there is no confusion about other parameters.

Example 10.1. Suppose that under the risk-neutral probability measure, the underlying stock price is a geometric Brownian motion

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

Consider a call option with maturity T and strike price K. Calculate its delta at time $t = 0$.

SOLUTION: By the Black–Scholes formula, the price of the option is

$$V(S_0) = S_0 \Phi(\alpha) - K e^{-rT} \Phi(\alpha - \sigma\sqrt{T}),$$

where

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

It follows from the definition of delta and straightforward calculation (we omit the details) that

$$\Delta = \frac{\partial V}{\partial S_0} = \Phi(\alpha). \tag{10.1}$$

In particular, the delta of a call option is always between 0 and 1. ■

10.2 Monte Carlo Simulation of Greeks

Since explicit formulas for Greeks are not available in general, Monte Carlo simulation is often used to produce estimates. Generically, one can write the Greeks in the form of derivatives. To be more concrete, consider an option with the discounted payoff $X(\theta)$, where θ is the parameter of interest. For example, if one wishes to estimate delta, then $\theta = S$; if one wishes to estimate vega, then $\theta = \sigma$. The value of the option is

$$V(\theta) = E[X(\theta)],$$

where the expected value is taken under the risk-neutral probability measure. Then estimating the Greek amounts to estimating the derivative

$$V'(\theta) = \frac{\partial}{\partial \theta} E[X(\theta)].$$

10.2.1 Methods of Finite Difference

A popular finite difference method for estimating derivatives is based on the *central difference* approximation

$$V'(\theta) = \lim_{h \downarrow 0} \frac{V(\theta + h) - V(\theta - h)}{2h}.$$

Given a small number h, let $\hat{V}(\theta + h)$ and $\hat{V}(\theta - h)$ be estimates for $V(\theta + h)$ and $V(\theta - h)$, respectively. Then an estimate for $V'(\theta)$ is

$$\frac{\hat{V}(\theta + h) - \hat{V}(\theta - h)}{2h}. \tag{10.2}$$

To reduce the variance, common random numbers are usually used to produce estimates of $\hat{V}(\theta + h)$ and $\hat{V}(\theta - h)$.

Even if $\hat{V}(\theta \pm h)$ are both unbiased, the estimate (10.2) for $V'(\theta)$ can be biased and the magnitude of the bias follows directly from the Taylor's expansion:

$$\text{bias} = \frac{V(\theta+h) - V(\theta-h)}{2h} - V'(\theta) \approx \frac{1}{6}V'''(\theta)h^2.$$

To reduce the bias, one would like to decrease h. However, the variance of the estimate

$$\frac{1}{4h^2}\text{Var}[\hat{V}(\theta+h) - \hat{V}(\theta-h)]$$

may remain bounded away from zero or even explode as $h \to 0$. Therefore, reducing h alone will not make the estimate more accurate unless one increases the sample size accordingly.

Example 10.2. Estimate the delta at $t = 0$ of a call option with maturity T and strike price K, assuming that the underlying stock price is a geometric Brownian motion with drift r (the risk-free interest rate) and volatility σ under the risk-neutral probability measure.

SOLUTION: In this case the parameter of interest is $\theta = S_0$. Given an arbitrary h, the estimate for $V(S_0 \pm h)$ is the sample average of iid copies of

$$e^{-rT}\left[(S_0 \pm h)\exp\left\{\left(r - \frac{1}{2}\sigma^2\right)T + \sigma\sqrt{T}Z\right\} - K\right]^+.$$

The scheme is fairly straightforward. Below is the pseudocode. Note that the same samples $\{Z_i\}$ are used for estimating $V(S_0 + h)$ and $V(S_0 - h)$.

Pseudocode of the central difference method:

for $i = 1, 2, \ldots, n$

  generate Z_i from $N(0,1)$

  set $X_i = (S_0 + h) \cdot \exp\{(r - \sigma^2/2)T + \sigma\sqrt{T}Z_i\}$

  set $Y_i = (S_0 - h) \cdot \exp\{(r - \sigma^2/2)T + \sigma\sqrt{T}Z_i\}$

  set $H_i = \dfrac{1}{2h}[e^{-rT}(X_i - K)^+ - e^{-rT}(Y_i - K)^+]$

compute the estimate $\hat{v} = \dfrac{1}{n}(H_1 + H_2 + \cdots + H_n)$

compute the standard error S.E. $= \sqrt{\dfrac{1}{n(n-1)}\left(\sum_{i=1}^{n} H_i^2 - n\hat{v}^2\right)}$.

The numerical results are reported in Table 10.2. The theoretical values are calculated from formula (10.1). We let

$$S_0 = 50, \ \ r = 0.05, \ \ \sigma = 0.2, \ \ T = 1, \ \ n = 10,000,000,$$

and estimate the delta for $K = 50$ and 55, respectively.

Table 10.2: Finite difference method: delta for call options

	$K = 50$ $(\Delta = 0.6368)$			$K = 55$ $(\Delta = 0.4496)$		
	$h = 1$	$h = 0.1$	$h = 0.01$	$h = 1$	$h = 0.1$	$h = 0.01$
Estimate	0.6364	0.6368	0.6366	0.4495	0.4497	0.4497
S.E.	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002

In this example, the variance of the estimates remains bounded away from zero. Reducing h alone will not produce more accurate estimates if the sample size is not increased accordingly. ■

10.2.2 Method of Pathwise Differentiation

The basic idea of the pathwise differentiation method is to exchange the order of expectation and differentiation to obtain

$$V'(\theta) = \frac{\partial}{\partial \theta} E\left[X(\theta)\right] = E\left[\frac{\partial}{\partial \theta} X(\theta)\right].$$

The plain Monte Carlo estimate of $V'(\theta)$ is just the sample average of iid copies of

$$\frac{\partial}{\partial \theta} X(\theta),$$

which is an unbiased estimate. All the variance reduction techniques we have discussed so far can be applied to improve the efficiency of the estimate. Note that $X(\theta)$ is a random variable depending on θ. Therefore, its derivative with respect to θ is a pathwise differentiation and a random variable itself. In general, the evaluation or approximation of this derivative is rather straightforward. Many times explicit formulas are available.

Finally, it should be pointed out that the method of pathwise differentiation does not work universally. Its validity hinges on the interchangeability of the order of differentiation and expectation. The rule of thumb is that it is justifiable when the payoff function is everywhere continuous and almost everywhere continuously differentiable with respect to the parameter. One should be especially careful when the payoff is discontinuous. For example, consider the delta at $t = 0$ for a binary option with maturity T and discounted payoff

$$X(S_0) = e^{-rT} 1_{\{S_T \geq K\}}.$$

Even though

$$\frac{\partial X}{\partial S_0} = 0$$

almost everywhere, the delta of this option is clearly not zero.

Example 10.3. Use the method of pathwise differentiation to estimate the delta at $t = 0$ for a call option with maturity T and strike price K, assuming that the underlying stock price is a geometric Brownian motion with drift r (the risk-free interest rate) and volatility σ under the risk-neutral probability measure.

SOLUTION: Note that $\theta = S_0$ for the calculation of delta. The stock price is

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

and the discounted option payoff is

$$X(S_0) = e^{-rT}(S_T - K)^+ = e^{-rT}(S_T - K) 1_{\{S_T \geq K\}}.$$

It follows that

$$\frac{\partial X}{\partial S_0} = 1_{\{S_T \geq K\}} e^{-rT} \frac{\partial S_T}{\partial S_0} = 1_{\{S_T \geq K\}} \frac{e^{-rT} S_T}{S_0}$$

and

$$\Delta = E\left[1_{\{S_T \geq K\}} \frac{e^{-rT} S_T}{S_0} \right].$$

Monte Carlo simulation of Δ is straightforward. The numerical results are reported in Table 10.3 with parameters

$$S_0 = 50, \;\; r = 0.05, \;\; \sigma = 0.2, \;\; T = 1, \;\; n = 10000.$$

The theoretical values of delta are obtained from formula (10.1). ■

Table 10.3: Pathwise differentiation: delta for call options

Strike price K	40	45	50	55	60
Theoretical value	0.9286	0.8097	0.6368	0.4496	0.2872
Estimate	0.9244	0.8064	0.6373	0.4469	0.2867
S.E.	0.0036	0.0049	0.0058	0.0059	0.0054

Example 10.4. The setup is the same as Example 10.3. Consider a discretely monitored average price call option with maturity T and payoff

$$(\bar{S} - K)^+, \qquad \bar{S} = \frac{1}{m}\sum_{i=1}^{m} S_{t_i},$$

where $0 = t_0 < t_1 < \cdots < t_m = T$ are prefixed dates. Argue that the delta of this option at $t = 0$ is

$$E\left[1_{\{\bar{S}\geq K\}} \frac{e^{-rT}\bar{S}}{S_0}\right].$$

See Exercise 10.5 for a generalization of this result.

PROOF: The discounted payoff of this option is $X(S_0) = e^{-rT}(\bar{S} - K)^+$. By the method of pathwise differentiation, its delta is $\Delta = E[\partial X/\partial S_0]$. Similar to Example 10.3,

$$\begin{aligned} \frac{\partial X}{\partial S_0} &= 1_{\{\bar{S}\geq K\}} \cdot e^{-rT} \frac{\partial \bar{S}}{\partial S_0} \\ &= 1_{\{\bar{S}\geq K\}} \cdot e^{-rT} \frac{1}{m}\sum_{i=1}^{m} \frac{\partial S_{t_i}}{\partial S_0} \\ &= 1_{\{\bar{S}\geq K\}} \cdot e^{-rT} \frac{1}{m}\sum_{i=1}^{m} \frac{S_{t_i}}{S_0} \\ &= 1_{\{\bar{S}\geq K\}} \cdot e^{-rT} \frac{\bar{S}}{S_0}. \end{aligned}$$

We complete the proof. ■

Example 10.5. Control variate method by means of artificial dynamics. Consider a stochastic volatility model where under the risk-neutral probability measure

$$\begin{aligned} dS_t &= rS_t\,dt + \theta_t S_t\,dW_t, \\ d\theta_t &= a(\Theta - \theta_t)\,dt + \beta\,dB_t. \end{aligned}$$

Here r is the risk-free interest rate, (a, Θ, β) are all positive constants, and (W, B) is a two-dimensional Brownian motion with covariance matrix

$$\Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}.$$

Estimate the delta of a call option with maturity T and strike price K.

SOLUTION: This example is very similar to Example 9.7. It follows from Itô formula that

$$S_t = S_0 \exp\{Y_t\}, \tag{10.3}$$

where Y_t satisfies

$$dY_t = \left(r - \frac{1}{2}\theta_t^2\right) dt + \theta_t \, dW_t, \quad Y_0 = 0.$$

Since the process Y does *not* depend on S_0, the delta of the call option is

$$\Delta = E\left[1_{\{S_T \geq K\}} e^{-rT} \frac{\partial S_T}{\partial S_0}\right] = E\left[1_{\{S_T \geq K\}} e^{-rT} \frac{S_T}{S_0}\right].$$

Plugging in (10.3), we have

$$\Delta = E\left[1_{\{Y_T \geq y\}} \exp\{Y_T - rT\}\right], \quad y = \log\left(\frac{K}{S_0}\right).$$

It is straightforward to apply the Euler scheme on the two-dimensional process (Y, θ) to estimate Δ. More precisely, let $0 = t_0 < t_1 < \cdots < t_m = T$ be a time discretization. Define $\hat{Y}_0 = Y_0 = 0$ and $\hat{\theta}_0 = \theta_0$. For $i = 0, 1, \ldots, m-1$, recursively define

$$\begin{aligned} \hat{Y}_{t_{i+1}} &= \hat{Y}_{t_i} + \left(r - \frac{1}{2}\hat{\theta}_{t_i}^2\right)(t_{i+1} - t_i) + \hat{\theta}_{t_i}\sqrt{t_{i+1} - t_i}Z_{i+1}, \\ \hat{\theta}_{t_{i+1}} &= \hat{\theta}_{t_i} + a(\Theta - \hat{\theta}_{t_i})(t_{i+1} - t_i) + \beta\sqrt{t_{i+1} - t_i}R_{i+1}, \end{aligned}$$

where $\{(Z_1, R_1), \ldots, (Z_m, R_m)\}$ are iid jointly normal random vectors with distribution $N(0, \Sigma)$. The plain Monte Carlo estimate for Δ is just the sample average of iid copies of

$$X = 1_{\{\hat{Y}_T \geq y\}} \exp\{\hat{Y}_T - rT\}.$$

Similar to Example 9.7, one can introduce an *artificial* stochastic process $\bar{Y}$ defined recursively by $\bar{Y}_0 = Y_0 = 0$ and

$$\bar{Y}_{t_{i+1}} = \bar{Y}_{t_i} + \left(r - \frac{1}{2}\sigma^2\right)(t_{i+1} - t_i) + \sigma\sqrt{t_{i+1} - t_i}Z_{i+1}$$

for $i = 0, 1, \ldots, m-1$. This process $\bar{Y}$ uses the *same* sequence $\{Z_1, \ldots, Z_m\}$ and some *constant* volatility σ. The control variate is defined to be

$$V = 1_{\{\bar{Y}_T \geq y\}} \exp\{\bar{Y}_T - rT\}.$$

Clearly, Y_T would equal $\bar{Y}_T$ if the volatility θ_t were the constant σ. This implies that $E[V]$ is the delta of a call option with the same maturity T and strike price K when the underlying stock price is a geometric Brownian motion with drift r and volatility σ. In particular, by formula (10.1),

$$E[V] = \Phi(\alpha), \quad \alpha = \frac{1}{\sigma\sqrt{T}} \log\frac{S_0}{K} + \left(\frac{\sigma}{2} + \frac{r}{\sigma}\right)\sqrt{T}. \tag{10.4}$$

The control variate estimate for the delta is the sample average of iid copies of

$$X - b[V - \Phi(\alpha)]$$

for some constant b. As in Example 9.7, we choose $\sigma = \Theta$ because Θ is the long run average of $\{\theta_t\}$ and use $b = \hat{b}^*$, the sample estimate for the optimal coefficient b^*; see formula (6.2).

Below is the pseudocode. As before, we use $\hat{Y}_i$, $\bar{Y}_i$, and $\hat{\theta}_i$ to denote $\hat{Y}_{t_i}$, $\bar{Y}_{t_i}$, and $\hat{\theta}_{t_i}$, respectively.

Pseudocode for control variate method:

set $\sigma = \Theta$, $y = \log(K/S_0)$, and α as in (10.4)
for $k = 1, 2, \ldots, n$
  set $\hat{Y}_0 = 0$, $\bar{Y}_0 = 0$, and $\hat{\theta}_0 = \theta_0$
  for $i = 0, 1, \ldots, m-1$
    generate iid sample Z_{i+1} and U_{i+1} from $N(0,1)$
    set $R_{i+1} = \rho Z_{i+1} + \sqrt{1-\rho^2}\, U_{i+1}$
    set $\hat{Y}_{i+1} = \hat{Y}_i + (r - \hat{\theta}_i^2/2)(t_{i+1} - t_i) + \hat{\theta}_i \sqrt{t_{i+1} - t_i}\, Z_{i+1}$
    set $\bar{Y}_{i+1} = \bar{Y}_i + (r - \sigma^2/2)(t_{i+1} - t_i) + \sigma \sqrt{t_{i+1} - t_i}\, Z_{i+1}$
    set $\hat{\theta}_{i+1} = \hat{\theta}_i + a(\Theta - \hat{\theta}_i)(t_{i+1} - t_i) + \beta \sqrt{t_{i+1} - t_i}\, R_{i+1}$
  set $X_k = 1_{\{\hat{Y}_T \geq y\}} \exp\{\hat{Y}_T - rT\}$;
  set $Q_k = 1_{\{\bar{Y}_T \geq y\}} \exp\{\bar{Y}_T - rT\} - \Phi(\alpha)$
compute $\hat{b}^*$ from formula (6.2) [with Y replaced by Q]
for $k = 1, 2, \ldots, n$
  set $H_k = X_k - \hat{b}^* Q_k$
compute the estimate $\hat{v} = \dfrac{1}{n}\sum_{i=1}^{n} H_i$

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

We compare the plain Monte Carlo scheme with the control variate method. The numerical results are reported in table 10.4 for

$$S_0 = 50,\ r = 0.03,\ T = 1,\ a = 2,\ \beta = 0.1,\ \theta_0 = 0.25,\ \Theta = 0.2,$$

$$\rho = 0.5,\ m = 50,\ t_i = iT/m,\ n = 10000.$$

Table 10.4: Estimating Δ: control variate method

	$K = 45$		$K = 50$		$K = 55$	
	Plain	Control	Plain	Control	Plain	Control
Estimate	0.7218	0.7221	0.5978	0.6030	0.4816	0.4839
S.E.	0.0062	0.0032	0.0066	0.0026	0.0067	0.0023

The control variate method does produce more accurate estimates. More extensive numerical experiments should show that the choice of $\sigma = \Theta$ is nearly optimal. ■

Example 10.6. Delta by Euler Scheme. Assume that the risk-free interest rate r is a constant and the underlying stock price S satisfies the stochastic differential equation

$$dS_t = rS_t\,dt + \sigma(S_t)\,dW_t$$

under the risk-neutral probability measure. Design an estimate for the delta at $t = 0$ of an option with maturity T and payoff $h(S_T)$.

Solution: The price of the option is $V(S_0) = E[e^{-rT}h(S_T)]$. Therefore, by the chain rule,

$$\Delta = E\left[e^{-rT}\frac{\partial}{\partial S_0}h(S_T)\right] = E\left[e^{-rT}h'(S_T)\frac{\partial S_T}{\partial S_0}\right].$$

However, unlike all the examples we have seen so far, it is not possible to explicitly calculate

$$\frac{\partial S_T}{\partial S_0} \tag{10.5}$$

in general. To approximate it, consider the Euler scheme for S. Given a time discretization $0 = t_0 < t_1 < \cdots < t_m = T$. Define recursively

$$\begin{aligned} \hat{S}_0 &= S_0, && (10.6) \\ \hat{S}_{t_{i+1}} &= \hat{S}_{t_i} + r\hat{S}_{t_i}(t_{i+1} - t_i) + \sigma(\hat{S}_{t_i})\sqrt{t_{i+1} - t_i}\, Z_{i+1} && (10.7) \end{aligned}$$

for $i = 0, 1, \ldots, m-1$, where $\{Z_1, \ldots, Z_m\}$ are iid standard normal random variables. Define

$$\hat{\Delta}_{t_i} = \frac{\partial \hat{S}_{t_i}}{\partial S_0}, \quad i = 0, 1, \ldots, m.$$

Taking derivatives over S_0 on both sides of (10.6) and (10.7), it follows from the chain rule that

$$\begin{aligned} \hat{\Delta}_0 &= 1, && (10.8) \\ \hat{\Delta}_{t_{i+1}} &= \hat{\Delta}_{t_i} + r\hat{\Delta}_{t_i}(t_{i+1} - t_i) + \sigma'(\hat{S}_{t_i})\hat{\Delta}_{t_i}\sqrt{t_{i+1} - t_i}\, Z_{i+1}. && (10.9) \end{aligned}$$

Equations (10.6)–(10.9) offer a recursive algorithm to compute the pathwise derivative

$$\hat{\Delta}_{t_m} = \hat{\Delta}_T = \frac{\partial \hat{S}_T}{\partial S_0},$$

which can be used as an approximation of (10.5). The plain Monte Carlo estimate of the delta is simply the sample average of iid copies

$$e^{-rT} h'(\hat{S}_T)\hat{\Delta}_T.$$

Note that the above calculation can be extended to path-dependent options and more general price dynamics. We leave this as an exercise to the interested reader. ■

10.2.3 Method of Score Function

Contrary to the method of pathwise differentiation, the method of score function, sometimes also referred to as the *likelihood ratio method*, does not require any smoothness condition on the payoff function.

Recall that $X(\theta)$ stands for the discounted payoff of the option and the option price $V(\theta)$ is given by

$$V(\theta) = E[X(\theta)].$$

Often, one can write $X(\theta) = H(Y)$ for some function H that has *no* dependence on θ, and some random variable or random vector Y that has a θ-dependent density $f_\theta(y)$. It follows that

$$V(\theta) = E[H(Y)] = \int H(y) f_\theta(y)\, dy.$$

Taking derivate with respect to θ and exchanging the order of integration and differentiation, we arrive at

$$V'(\theta) = \int H(y) \frac{\partial f_\theta}{\partial \theta}(y)\, dy = \int H(y) \frac{\partial \log f_\theta}{\partial \theta}(y) \cdot f_\theta(y)\, dy,$$

or equivalently,

$$V'(\theta) = E\left[H(Y)\frac{\partial \log f_\theta}{\partial \theta}(Y)\right] = E\left[X(\theta)\frac{\partial \log f_\theta}{\partial \theta}(Y)\right].$$

Therefore, an unbiased estimate for $V'(\theta)$ is just the sample mean of iid copies of

$$X(\theta)\frac{\partial \log f_\theta}{\partial \theta}(Y).$$

Note that the particular form of H is not important. The essential requirement is that the discounted payoff X can be written as a θ-independent function of some random variable or random vector Y. There are many possible ways to choose Y. But usually Y is chosen to be the building block of the sampling scheme for $X(\theta)$ or one with the simplest form of density.

Observe the difference between the pathwise differentiation method and the score function method. In the former, the differentiation is taken with respect to the payoff function, while in the latter, the differentiation is taken with respect to the density function of some underlying random variable or random vector. For the score function method to work, it requires that the density of Y be smooth with respect to θ, which is much milder than the regularity condition on $X(\theta)$ for the pathwise differentiation method. Another advantage of this method is that once the *score function*

$$\frac{\partial \log f_\theta}{\partial \theta}$$

is obtained, it can be applied to any payoff function X. However, the practical use of this method is often limited by the explicit knowledge of the density function f_θ.

We should illustrate the score function method through examples. In all these examples, we assume that the risk-free interest rate r is a constant and the underlying stock price is a geometric Brownian motion with drift r and volatility σ under the risk-neutral probability measure. All the Greeks under consideration are assumed to be at $t = 0$. For Greeks at a general time point $t < T$, it suffices to replace T by $T - t$ and S_0 by S_t.

Example 10.7. Write down an estimate for the delta of an option with maturity T and payoff $h(S_T)$.

SOLUTION: The parameter of interest is $\theta = S_0$. The stock price at maturity T is

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

where Z is a standard normal random variable. Define

$$Y = Z + \frac{1}{\sigma\sqrt{T}}\log S_0.$$

Then the option price is

$$V(S_0) = E[H(Y)]$$

where H is a function independent of $\theta = S_0$:

$$H(Y) = e^{-rT}h(S_T) = e^{-rT}h\left(\exp\left\{\left(r - \frac{1}{2}\sigma^2\right)T + \sigma\sqrt{T}Y\right\}\right)$$

The distribution of Y is normal and depends on S_0. Its density, denoted by f_{S_0}, is

$$f_{S_0}(y) = \frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\left(y - \frac{1}{\sigma\sqrt{T}}\log S_0\right)^2\right\},$$

which satisfies

$$\frac{\partial \log f_{S_0}}{\partial S_0}(y) = \frac{y}{S_0\sigma\sqrt{T}} - \frac{1}{S_0\sigma^2 T}\log S_0.$$

It follows that

$$\frac{\partial \log f_{S_0}}{\partial S_0}(Y) = \frac{Y}{S_0\sigma\sqrt{T}} - \frac{1}{S_0\sigma^2 T}\log S_0 = \frac{Z}{S_0\sigma\sqrt{T}},$$

which in turn implies that

$$\Delta = E\left[H(Y)\frac{\partial \log f_{S_0}}{\partial S_0}(Y)\right] = E\left[e^{-rT}h(S_T)\frac{Z}{S_0\sigma\sqrt{T}}\right].$$

An unbiased estimate for Δ is just the sample average of iid copies of the random variable inside expectation. In this formulation, h does not need to be continuous, for example, h can be the payoff of a binary option. ■

Example 10.8. Write down an estimate for the vega of an option with maturity T and payoff $h(S_T)$.

SOLUTION: The parameter of interest is $\theta = \sigma$. The stock price at maturity T is

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

where Z is a standard normal random variable. Define

$$Y = \log\left(\frac{S_T}{S_0}\right) - rT = \sigma\sqrt{T}Z - \frac{1}{2}\sigma^2 T. \tag{10.10}$$

Then the price of the option is

$$V(\sigma) = E[H(Y)],$$

where

$$H(Y) = e^{-rT} h(S_T) = e^{-rT} h(S_0 \cdot \exp\{Y + rT\}).$$

The distribution of Y depends on $\theta = \sigma$. Indeed, Y is normally distributed with distribution

$$N\left(-\frac{1}{2}\sigma^2 T, \sigma^2 T\right),$$

and its density, denoted by f_σ, satisfies

$$\frac{\partial \log f_\sigma}{\partial \sigma}(y) = \frac{y^2}{\sigma^3 T} - \frac{\sigma T}{4} - \frac{1}{\sigma}.$$

Therefore,

$$\begin{aligned} \frac{\partial \log f_\sigma}{\partial \sigma}(Y) &= \frac{Y^2}{\sigma^3 T} - \frac{\sigma T}{4} - \frac{1}{\sigma} \\ &= \frac{1}{\sigma^3 T}\left(\sigma\sqrt{T}Z - \frac{1}{2}\sigma^2 T\right)^2 - \frac{\sigma T}{4} - \frac{1}{\sigma} \\ &= \frac{Z^2 - 1}{\sigma} - \sqrt{T}Z. \end{aligned} \tag{10.11}$$

It follows that

$$\nu = E\left[H(Y)\frac{\partial \log f_\sigma}{\partial \sigma}(Y)\right] = E\left[e^{-rT} h(S_T)\left(\frac{Z^2 - 1}{\sigma} - \sqrt{T}Z\right)\right].$$

An unbiased estimate for vega is just the sample average of iid copies of the random variable inside expectation. We want to repeat that h does not need to be continuous. ■

Example 10.9. Write down an estimate for the vega of an average price call option with maturity T and payoff

$$(\bar{S} - K)^+, \quad \bar{S} = \frac{1}{m} \sum_{i=1}^{m} S_{t_i},$$

where $0 = t_0 < t_1 < \cdots < t_m = T$ are given dates.

SOLUTION: In this case $\theta = \sigma$. Denote $\Delta t_{i+1} = t_{i+1} - t_i$. Then we can recursively write

$$S_{t_{i+1}} = S_{t_i} \exp\left\{ \left(r - \frac{1}{2}\sigma^2 \right) \Delta t_{i+1} + \sigma \sqrt{\Delta t_{i+1}} Z_{i+1} \right\},$$

for $i = 0, 1, \ldots, m-1$. Similar to (10.10), define for each i

$$Y_{i+1} = \log\left(\frac{S_{t_{i+1}}}{S_{t_i}} \right) - r\Delta t_{i+1} = \sigma \sqrt{\Delta t_{i+1}} Z_{i+1} - \frac{1}{2}\sigma^2 \Delta t_{i+1}.$$

Since $S_{t_{i+1}} = S_{t_i} \exp\{r\Delta t_{i+1} + Y_{i+1}\}$ for each i, it follows easily that

$$e^{-rT}(\bar{S} - K)^+ = H(Y_1, \ldots, Y_m)$$

for some function H that does not depend on σ. Because Y_i's are independent, the joint density of $(Y_1, \ldots, Y_m)$, denoted by $f_\sigma(y_1, \ldots, y_m)$, satisfies

$$f_\sigma(y_1, \ldots, y_m) = \prod_{i=1}^{m} f_\sigma^{(i)}(y_i),$$

where $f_\sigma^{(i)}$ is the marginal density for Y_i. Therefore, thanks to (10.11),

$$\frac{\partial \log f_\sigma}{\partial \sigma}(Y_1, \ldots, Y_m) = \sum_{i=1}^{m} \frac{\partial \log f_\sigma^{(i)}}{\partial \sigma}(Y_i) = \sum_{i=1}^{m} \left(\frac{Z_i^2 - 1}{\sigma} - \sqrt{\Delta t_i} Z_i \right).$$

It follows that

$$\begin{aligned} \nu &= E\left[H(Y_1, \ldots, Y_m) \frac{\partial \log f_\sigma}{\partial \sigma}(Y_1, \ldots, Y_m) \right] \\ &= E\left[e^{-rT}(\bar{S} - K)^+ \sum_{i=1}^{m} \left(\frac{Z_i^2 - 1}{\sigma} - \sqrt{\Delta t_i} Z_i \right) \right]. \end{aligned}$$

In the preceding derivation, the specific nature of the average price call option is not important at all. As long as the option payoff is of the form $h(S_{t_1}, \ldots, S_{t_m})$, the formula for vega is still valid if one replaces the payoff of the average price call by h. ■

Exercises

Unless otherwise specified, we assume that the risk-free interest rate r is a constant and the underlying stock price is a geometric Brownian motion with drift r and volatility σ under the risk-neutral probability measure. All Greeks under consideration are assumed to be at $t = 0$.

Pen-and-Paper Problems

10.1 Explicitly evaluate the delta and vega of a put option with maturity T and strike price K .

10.2 Explicitly evaluate the delta of a binary option with maturity T and payoff

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

10.3 Use the method of pathwise differentiation to show that the vega of a call option with maturity T and strike price K is

$$\nu = E\left[1_{\{S_T \geq K\}} e^{-rT} S_T \left(\frac{1}{\sigma} \log \frac{S_T}{S_0} - \left(\frac{r}{\sigma} + \frac{\sigma}{2}\right) T\right)\right].$$

10.4 Use the method of pathwise differentiation to find an unbiased estimate for the delta of a discretely monitored lookback call option with maturity T and payoff

$$S_T - \min_{1 \leq i \leq m} S_{t_i},$$

where $0 = t_0 < t_1 < \cdots < t_m = T$ are given dates.

10.5 Assume that the stock price S satisfies the stochastic differential equation

$$dS_t = r_t S_t \, dt + \theta_t S_t \, dW_t,$$

under the risk-neutral probability measure. Here r_t and θ_t are the short rate and volatility at time t, respectively. They can be deterministic or stochastic, but are assumed to have no dependence on S_0. Consider an option with maturity T and payoff $h(S_{t_0}, S_{t_1}, \ldots, S_{t_m})$, where $0 = t_0 < t_1 < \cdots < t_m = T$ are give dates. Show that the delta of this option is

$$\Delta = E\left[\sum_{i=0}^{m} \exp\left\{-\int_0^T r_t \, dt\right\} \frac{S_{t_i}}{S_0} \frac{\partial}{\partial S_{t_i}} h(S_{t_0}, S_{t_1}, \ldots, S_{t_m})\right].$$

10.6 Consider the following generalization of Example 10.6. Let X be a diffusion process that satisfies the stochastic differential equation

$$dX_t = b(X_t) \, dt + \sigma(X_t) \, dW_t.$$

Define $V(X_0) = E[h(X_{t_0}, X_{t_1}, \ldots, X_{t_m})]$, where $0 = t_0 < t_1 < \cdots < t_m = T$ are given time points. Design an Euler scheme to approximate

$$\frac{\partial V}{\partial X_0}.$$

You can assume that h is nice so that the order of differentiation and expectation can be exchanged.

10.7 Mimic Example 10.9 to write down an unbiased estimate for the delta of a path-dependent option with maturity T and payoff $h(S_{t_1}, \ldots, S_{t_m})$, where $0 = t_0 < t_1 < \cdots < t_m = T$ are given dates. Use the score function method.

10.8 Suppose that we are interested in an unbiased estimate for the delta and vega of a down-and-in barrier option with maturity T and payoff

$$(S_T - K)^+ 1_{\{\min(S_{t_1}, \ldots, S_{t_m}) \le b\}},$$

where $0 = t_0 < t_1 < \cdots < t_m = T$ are given dates. Use the score function method to design your estimates. Will the pathwise differentiation method work? Why?

MATLAB® Problems

10.A Use the central difference method to estimate the following Greeks. The sample size is always $n = 100000$, and the model parameters are

$$S_0 = 50, \quad r = 0.03, \quad \sigma = 0.2, \quad T = 1.$$

(a) Delta of a put option with maturity T and strike price K. Report your results for $K = 50$ and $h = 1, 0.1, 0.01$, respectively.

(b) Vega of a call option with maturity T and strike price K. Report you results for $K = 50$ and $h = 0.1, 0.01, 0.001$, respectively.

(c) Delta of a up-and-out barrier option with maturity T and payoff

$$(S_T - K)^+ 1_{\{\max(S_{t_1}, \ldots, S_{t_m}) \le b\}}.$$

Report your results for $K = 50$, $b = 65$, $m = 30$, $t_i = iT/m$, and $h = 10^{-k}$ with $k = 0, 1, \ldots, 6$, respectively. What is your conclusion?

10.B Consider an average price call option with maturity T and strike price K, whose payoff is

$$(\bar{S} - K)^+, \quad \bar{S} = \frac{1}{m} \sum_{i=1}^{m} S_{t_i}.$$

Write a function to estimate the delta by

(a) the method of pathwise differentiation.

(b) the method of pathwise differentiation, combined with the control variate method to reduce variance. Use the delta of the corresponding average price call option with geometric mean as the control variate. Use the sample estimate of b^*. *Hint:* For an average price call option with geometric mean, show that its delta is

$$\Delta = E\left[1_{\{\bar{S}_G \geq K\}} \frac{e^{-rT}\bar{S}_G}{S_0}\right], \quad \bar{S}_G = \left(\prod_{i=1}^{m} S_{t_i}\right)^{\frac{1}{m}},$$

and derive the theoretical delta value from the explicit pricing formula for the average price call option with geometric mean.

Report your estimates and standard errors for

$$S_0 = 50,\ r = 0.03,\ K = 50,\ \sigma = 0.2,\ T = 1,\ m = 30,\ t_i = iT/m,$$

with sample size $n = 10000$.

10.C Suppose that under the risk-neutral probability measure, the stock price satisfies the stochastic differential equation:

$$\begin{aligned} dS_t &= r_t S_t\, dt + \sigma S_t\, dW_t \\ dr_t &= a(b - r_t)\, dt + \theta r_t\, dB_t. \end{aligned}$$

Here a, b, σ, θ are all positive constants, and (W, B) is a two-dimensional Brownian motion with covariance matrix

$$\Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}.$$

Let $Y_t = \log S_t$. Consider a call option with maturity T and strike price K.

(a) Apply Exercise 10.5 to express Δ in terms of an expected value.

(b) Estimate Δ by the Euler scheme on (Y_t, r_t).

(c) Mimic Example 9.6 to estimate Δ by the method of conditioning.

Report your estimates and standard errors for

$$S_0 = 50,\ r_0 = 0.03,\ a = 1,\ b = 0.02,\ \sigma = 0.2,\ \theta = 0.1,\ T = 1,$$

$$K = 50,\ m = 50,\ t_i = iT/m,\ n = 10000,$$

and $\rho = 0.1, 0.3, 0.5, 0.7, 0.9$, respectively. Explain why the variance reduction becomes less significant when ρ gets larger.

10.D Assume that the risk-free interest rate r is a constant and under the risk-neutral probability measure the underlying stock price satisfies the stochastic differential equation

$$dS_t = rS_t\,dt + \sigma\sqrt{S_t}\,dW_t.$$

Design an Euler scheme to estimate the following Greeks.

(a) The delta of a call option with maturity T and strike price K.

(b) The delta of an average price call option with maturity T and strike price K, whose payoff is

$$(\bar{S} - K)^+, \quad \bar{S} = \frac{1}{m}\sum_{i=1}^{m} S_{t_i}.$$

You will need a slight modification on the Euler scheme to ensure that the time discretized process $\hat{S}$ stays nonnegative. Modify the Euler scheme on Δ accordingly. Report your results for

$$S_0 = 50, \;\; r = 0.03, \;\; \sigma = 0.3, \;\; K = 50, \;\; T = 1,$$

$$m = 30, \;\; t_i = iT/m, \;\; n = 10000.$$

10.E Use the score function method to estimate the delta and vega of a binary option with maturity T and payoff $1_{\{S_T \geq K\}}$. Report your results for

$$S_0 = 50, \; r = 0.03, \; K = 50, \; \sigma = 0.2, \; T = 1, \; n = 10000.$$

Compare with the theoretical values.

10.F Use the score function method to estimate the delta and vega of an up-and-out barrier option with maturity T and payoff

$$(S_T - K)^+ 1_{\{\max(S_{t_1}, \ldots, S_{t_m}) \leq b\}}.$$

Report your results for

$$S_0 = 50, \;\; r = 0.05, \;\; \sigma = 0.2, \;\; T = 1, \;\; m = 30, \;\; t_i = iT/m,$$

$$K = 50, \; b = 70, \; n = 10000.$$

Appendix A

Multivariate Normal Distributions

In this appendix, we review some of the important results concerning jointly normal random vectors and multivariate normal distributions.

The convention in this appendix is that the vectors are always in *column* form. Given a random vector $X = (X_1, \ldots, X_n)'$, its *expected value* is defined to be the vector

$$E[X] = (E[X_1], \ldots, E[X_n])'.$$

Let $Y = (Y_1, \ldots, Y_m)'$ be another random vector. The *covariance matrix* of X and Y is defined to be the $n \times m$ matrix

$$\mathrm{Cov}(X, Y) = [\mathrm{Cov}(X_i, Y_j)].$$

That is, the entry of the covariance matrix in the i-th row and j-th column is $\mathrm{Cov}(X_i, Y_j)$. In particular, the $n \times n$ symmetric matrix $\mathrm{Var}[X] = \mathrm{Cov}(X, X)$ is said to be the covariance matrix of X.

A random vector $X = (X_1, X_2, \ldots, X_n)'$ is said to be *jointly normal* and have a *multivariate normal distribution* $N(\mu, \Sigma)$, if the joint density function of X takes the form

$$f(x) = (2\pi)^{-\frac{n}{2}} |\Sigma|^{-\frac{1}{2}} e^{-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)}, \quad x \in \mathbb{R}^n.$$

Here $\mu = (\mu_1, \ldots, \mu_n)'$ is an n-dimensional vector and $\Sigma = [\Sigma_{ij}]$ is an $n \times n$ symmetric positive definite matrix whose determinant is denoted by $|\Sigma|$. In the case where $\mu = 0$ and $\Sigma = I_n$ ($n \times n$ identity matrix), we say the distribution is the *n-dimensional standard normal*.

Since a collection of continuous random variables are independent if and only if the joint density function is the product of individual marginal density functions, we have the following result immediately.

Lemma A.1. *$X = (X_1, \ldots, X_n)'$ is an n-dimensional standard normal random vector if and only if $X_1, \ldots, X_n$ are independent standard normal random variables. More generally, if $X_1, \ldots, X_n$ are independent normal variables with $E[X_i] = \mu_i$ and $Var[X_i] = \sigma_i^2$, then $X = (X_1, \ldots, X_n)'$ is a jointly normal random vector with distribution $N(\mu, \Sigma)$ where*

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n^2 \end{bmatrix}.$$

The converse is also true.

We will list some of the properties of multivariate normal distributions without proof. For a comprehensive treatment, see [32]. Assume from now on that $X = (X_1, \ldots, X_n)'$ is jointly normal with distribution $N(\mu, \Sigma)$.

a. $E[X] = \mu$ and $\text{Var}[X] = \Sigma$.

b. X_i has distribution $N(\mu_i, \Sigma_{ii})$ for each $i = 1, \ldots, n$.

c. X_i and X_j are independent if and only if $\Sigma_{ij} = \text{Cov}(X_i, X_j) = 0$.

d. For any $m \times n$ matrix C and $m \times 1$ vector b, the random vector

$$Y = CX + b$$

is jointly normal with distribution $N(C\mu + b, C\Sigma C')$.

e. Given any $k < n$, let $Y_1 = (X_1, \ldots, X_k)'$ and $Y_2 = (X_{k+1}, \ldots, X_n)'$. Consider the partition of X, μ, and Σ given by

$$X = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}, \quad \mu = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{bmatrix},$$

where

$$\theta_i = E[Y_i], \quad \Lambda_{ij} = \text{Cov}(Y_i, Y_j).$$

Then the conditional distribution of Y_1 given $Y_2 = y$ is $N(\bar{\mu}, \bar{\Sigma})$ with

$$\bar{\mu} = \theta_1 + \Lambda_{12}\Lambda_{22}^{-1}(y - \theta_2), \quad \bar{\Sigma} = \Lambda_{11} - \Lambda_{12}\Lambda_{22}^{-1}\Lambda_{21}.$$

Appendix B

American Option Pricing

Throughout the book, it has been assumed that an option can only be exercised at maturity. Such options are called *European options*. The mechanism of an *American option* is no different, except that the holder of the option can exercise the option at any time before or at maturity. Therefore, an American option is at least as valuable as its European counterpart.

In a nutshell, evaluating an American option amounts to solving a control problem—at any given time, there are two actions (i.e., controls) that are available to the option holder, namely, to exercise the option immediately or to wait. If the value of exercising is larger than that of waiting, the option holder should exercise, and vice versa. While it is trivial that the value of exercising equals the option payoff, the difficult part of this decision process is to determine the value of waiting.

A systematic approach for solving such control problems is *dynamic programming*. It is a recursive algorithm running *backwards* in time. The main goal of the appendix is to discuss this approach and use it to price American options via binomial tree approximations. Due to the complexity of evaluating American options in general, we refer the reader to [11] for an in-depth discussion on various different techniques.

B.1 The Value of an American Option

Consider an American option with maturity T and payoff X_t if the option is exercised at time $t \leq T$. For example, for an American put option with strike price K

$$X_t = (K - S_t)^+.$$

The value of this option can be expressed in terms of the solution to an optimization problem:

$$v = \max_{0 \leq \tau \leq T} E[e^{-r\tau} X_\tau], \tag{B.1}$$

where the maximum is taken over all meaningful random times τ. Not surprisingly, the expected value is taken under the risk-neutral probability measure.

We will first discuss what are "meaningful" random times. Roughly speaking, a meaningful random time is an exercise time one can realize *without* looking into the future. For instance, a strategy of exercising the option whenever the stock price reaches down to a certain level b corresponds to the random time

$$\tau = \min\{t \in [0, T] : S_t \leq b\};$$

a strategy of exercising the option at a prefixed time t_0 no matter how the stock price behaves corresponds to

$$\tau = t_0.$$

These two examples of random times or strategies are meaningful and practical. On the contrary, a strategy of exercising the option when the stock price reaches the absolute minimum on the time interval $[0, T]$ corresponds to

$$\tau = \min\{t \in [0, T] : S_t = \min_{0 \leq u \leq T} S_u\},$$

which is neither practical nor meaningful. It will require the option holder to know the *entire* path of the stock price—whether the absolute minimum is reached or not cannot be determined unless at maturity. This is clearly unrealistic since the option holder, making a decision at time t, can only use the information he has gathered up until time t. From these discussions, it is not surprising that meaningful random times are also said to be *nonanticipating*.

The pricing formula (B.1) can be interpreted as follows. If the option is to be exercised at a meaningful random time τ, the value of the option at $t = 0$ is

$$E[e^{-r\tau} X_\tau],$$

where the expected value is taken under the risk-neutral probability measure. Since the option can be exercised at any meaningful random time τ, it will be natural for the option holder to maximize the above expected value. Therefore, the value of the option at $t = 0$ should be

$$\max_{0\le\tau\le T} E[e^{-r\tau}X_\tau],$$

and the optimal strategy for the option holder is to exercise the option at a random time τ^* that solves the above optimization problem.

Remark B.1. From now on, whenever we say random times, we implicitly assume that they are meaningful or nonanticipating. Such random times are also called *stopping times* in the literature, and this type of specialized control problems (B.1) are often said to be "optimal stopping problems."

B.2 Dynamic Programming and Binomial Trees

Evaluating an American option amounts to solving a stochastic optimization problem. The principle of dynamic programming essentially breaks it down into a sequence of much simpler, static optimization problems that can be solved backwards in time. We should illustrate this approach through the binomial tree pricing models.

The setup is exactly the same as that of Section 3.2.3. Consider a binomial tree with N periods. At each time period, the stock price either moves up by a factor u or moves down by a factor d, and the discounting factor is $1/R$. Under the risk-neutral probability measure, the stock price moves up with probability p^* and moves down with probability q^*, where

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

We will index the time steps by $n = 0, 1, \ldots, N$ and let S_n denote the stock price at time n. Consider an American option with maturity N and payoff, if exercised at time n,

$$H(S_n).$$

To solve this American option pricing problem, let $V_n(x)$ be the value of the option at time n if $S_n = x$. The price of the American option at time 0 is, by definition,

$$v = V_0(S_0).$$

By introducing these unknown functions $\{V_n(x) : n = 0, 1 \ldots, N\}$, it seems that we have made the matter more complicated. However, this allows us to write down a recursive equation (i.e., *dynamic programming equation*) in the following fashion: for $n = 0, 1, \ldots, N-1$,

$$\begin{aligned} V_n(x) &= \max\{H(x),\ E[V_{n+1}(S_{n+1})|S_n = x]/R\}, && \text{(B.2)} \\ V_N(x) &= H(x). && \text{(B.3)} \end{aligned}$$

Equation (B.3) is said to be the *terminal condition*. It is self-explanatory in the sense that the value of the option equals its payoff at maturity. The recursive equation (B.2) amounts to that the value of an American option at time n (given $S_n = x$) equals the maximum of the value of exercising, which is option payoff $H(x)$, and the value of waiting, which is $E[V_{n+1}(S_{n+1})|S_n = x]/R$. In the setting of binomial trees, the value of waiting is easy to evaluate and

$$E[V_{n+1}(S_{n+1})|S_n = x] = p^* V_{n+1}(ux) + q^* V_{n+1}(dx).$$

These considerations lead to a simple recursive algorithm for computing the value of the American option at any time. It is very similar to European option pricing with binomial trees: (1) compute the value of the option at time N by the terminal condition; (2) recursively compute the value of the option at each node by the dynamic programming equation, backwards in time. This recursive algorithm also yields the optimal exercise strategy: at any node, it is optimal to exercise if the value of exercising is greater than the value of waiting, and vice versa.

Example B.1. Consider an American put option with maturity $T = 1$ year and strike price $K = 50$. The underlying stock price is assumed to be a geometric Brownian motion with initial price $S_0 = 50$ and volatility $\sigma = 0.2$. The risk-free interest is $r = 12\%$ annually. Approximate the price of this American put option with a binomial tree of three periods.

SOLUTION: There are $N = 3$ periods. Each period represents $\triangle t = T/N = 1/3$ year in real time. Therefore, the parameters of the approximating binomial tree are

$$R = e^{r\triangle t} = 1.0408, \ \ u = e^{\sigma\sqrt{\triangle t}} = 1.1224, \ \ d = e^{-\sigma\sqrt{\triangle t}} = 0.8909,$$

and the risk-neutral probabilities are

$$p^* = \frac{R-d}{u-d} = 0.6475, \quad q^* = \frac{u-R}{u-d} = 0.3525.$$

Figure B.1 includes the tree of stock price, the tree of option value, and the tree of optimal strategy. For example, consider the node where the stock price is 39.6894. For this node, the value of exercising is

$$(50 - 39.6894)^+ = 10.3106$$

and the value of waiting is

$$(p^* \cdot 5.4526 + q^* \cdot 14.6389)/R = 8.3501.$$

Therefore, the value of the option at this node is 10.3106 and the optimal strategy is to immediately exercise. The price of the option at time 0 is 2.2359. ■

Lemma B.1. *The value of an American call option equals the value of an European call option with the same maturity and strike price.*

PROOF. We only need to show that for an American call option it is always optimal to wait until maturity to exercise. Denote by $V_n(x)$ the value of the American call option at time n if $S_n = x$. It follows from the dynamic program equation (B.2) that, for $n = 0, 1, \ldots, N-1$,

$$V_n(x) = \max\{(x-K)^+, \ E[V_{n+1}(S_{n+1})|S_n = x]/R\}.$$

Using the trivial observation $V_{n+1}(x) \geq (x-K)^+$ and the inequality $x^+ + y^+ \geq (x+y)^+$ for any x and y, we arrive at

$$\begin{aligned} E[V_{n+1}(S_{n+1})|S_n = x] &= p^* V_{n+1}(ux) + q^* V_{n+1}(dx) \\ &\geq p^*(ux-K)^+ + q^*(dx-K)^+ \\ &\geq (p^* ux + q^* dx - K)^+ \\ &= (Rx-K)^+ \\ &\geq R(x-K)^+, \end{aligned}$$

which in turn implies that

$$V_n(x) = E[V_{n+1}(S_{n+1})|S_n = x]/R.$$

In other words, it is always optimal to wait at time n for $n = 0, 1, \ldots, N-1$. This completes the proof. ■

B.3 Diffusion Models: Binomial Approximation

When the underlying asset price is a diffusion, it is possible to approximate it with binomial trees. We have already seen the use of such approximation to geometric Brownian motions. This section will discuss the extension to general diffusions.

Recall the binomial tree approximation of a geometric Brownian motion S that satisfies

$$\frac{dS_t}{S_t} = r\,dt + \sigma\,dW_t.$$

It starts with a single node S_0. In each period it moves up by a factor u with probability p^* or moves down by a factor d with probability $1 - p^*$, where

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

and Δt represents the real-time increment of each period of the binomial tree. Now suppose that we take logarithm of the stock price at each node of the binomial tree. The resulting new binomial tree starts at $\log S_0$ and, in each period, moves up by $h = \log u = \sigma\sqrt{\Delta t}$ with probability p^* or moves down by $-h = \log d$ with probability $1 - p^*$, where

$$p^* = \frac{e^{r\Delta t} - d}{u - d} = \frac{e^{r\Delta t} - e^{-h}}{e^h - e^{-h}}.$$

Obviously, this tree is an approximation to $X_t = \log S_t$, which satisfies

$$dX_t = \left(r - \frac{1}{2}\sigma^2\right) dt + \sigma\,dW_t.$$

Note that the binomial tree approximation of a geometric Brownian motion is constructed by matching the first and second moments of the increments of the process. Since such a construction is purely local, it is not difficult to generalize to a binomial tree approximation of a more general diffusion process (abusing notation) X_t that satisfies

$$dX_t = \left[b(X_t) - \frac{1}{2}\sigma^2(X_t)\right] dt + \sigma(X_t)\,dW_t.$$

More precisely, the approximating binomial tree will start at X_0. Given that its value at the current node is x, at the next time step, it will move up by $h(x)$ with probability $p^*(x)$ and move down by $-h(x)$ with probability $1 - p^*(x)$, where

$$h(x) = \sigma(x)\sqrt{\Delta t}, \quad p^*(x) = \frac{e^{b(x)\triangle t} - e^{-h(x)}}{e^{h(x)} - e^{-h(x)}}. \tag{B.4}$$

This is indeed a legitimate approximation in theory under mild regularity conditions. However, it has little use in practice because the increment $h(x)$ depends on x, and hence the nodes will not collapse as in the case of a geometric Brownian motion. Indeed, for a tree with N periods, the total number of nodes can be

$$1 + 2 + \cdots + 2^N = 2^{N+1} - 1,$$

whereas the approximating tree to a geometric Brownian motion will only have

$$1 + 2 + \cdots + (N+1) = (N+1)(N+2)/2$$

nodes. A simple solution to this issue is to make $h(x)$, or equivalently, $\sigma(x)$, independent of x. This motivates the use of the Lamperti transform (see Section 9.4) to convert the original process into a diffusion with constant volatility and then apply the binomial tree approximation given by (B.4). We will illustrate this approach through an example.

Example B.2. Suppose that the risk-free interest rate r is a constant and under the risk-neutral probability measure the underlying stock price satisfies the stochastic differential equation

$$dS_t = rS_t\,dt + \sigma\sqrt{S_t}\,dW_t.$$

Use binomial tree approximation to estimate the price of an American put option with maturity T and strike price K.

SOLUTION: For this process the Lamperti transform is

$$F(x) = \int \frac{1}{\sigma\sqrt{x}}\,dx = \frac{2}{\sigma}\sqrt{x}, \quad X_t = F(S_t) = \frac{2}{\sigma}\sqrt{S_t}.$$

It follows from Itô formula that

$$dX_t = \left(\frac{rX_t}{2} - \frac{1}{2X_t}\right)dt + dW_t = \left[b(X_t) - \frac{1}{2}\right]dt + dW_t,$$

where

$$b(x) = \frac{rx}{2} - \frac{1}{2x} + \frac{1}{2}.$$

The parameters for the binomial tree approximation of X is given by (B.4), that is,

$$h = \sqrt{\Delta t}, \quad p^*(x) = \frac{e^{b(x)\triangle t} - e^{-h}}{e^h - e^{-h}}.$$

Provided that the value of X at the current node is x, it will go up to $x + h$ with probability $p^*(x)$ and drop down to $x - h$ with probability $1 - p^*(x)$.

This description of the binomial tree is not entirely accurate, however. Some slight modifications are necessary.

1. To make sure that $p^*(x)$ is a real probability, that is, $0 \leq p^*(x) \leq 1$, we actually let

$$p^*(x) = \min\left\{1, \left(\frac{e^{b(x)\triangle t} - e^{-h}}{e^h - e^{-h}}\right)^+\right\}. \tag{B.5}$$

2. Since X is nonnegative, whenever the value of a node reaches below zero, we reset it to zero.

3. When $x = 0$, $p^*(x)$ is not well defined. However, observe that $S_t = 0$ if $X_t = 0$, and that the stock price will remain at zero once it reaches zero. Therefore, it is natural to let the approximating binomial tree stay at zero once it reaches zero. In other words, there are no more bifurcations once the value of a node reaches zero.

These modifications can be easily implemented and the resulting binomial tree is still a valid approximation to the process X.

The corresponding dynamic programming equation is straightforward. Suppose that we divide the time interval $[0, T]$ into N equal-length subintervals and let $t_n = nT/N$ for $n = 0, 1, \ldots, N$. Let $\Delta t = T/N$, $h = \sqrt{\Delta t}$, and define $p^*(x)$ by (B.5) for $x > 0$. Since $S_t = \sigma^2 X_t^2/4$, we define

$$H(x) = \left(K - \frac{1}{4}\sigma^2 x^2\right)^+,$$

which is the option payoff given $X_t = x$.

Denote by $V_n(x)$ the value of the American option at time t_n when $X_{t_n} = x$. The dynamic programming equation is simply

$$
\begin{aligned}
V_N(x) &= H(x), \\
V_n(x) &= \max\left\{H(x),\, e^{-r\Delta t}[p^*(x)V_{n+1}(x+h) + q^*(x)V_{n+1}(x-h)]\right\},
\end{aligned}
$$

where $q^*(x) = 1 - p^*(x)$. As we have mentioned, (i) when $x - h < 0$, we simply replace it by zero; (ii) when $x = 0$, the recursive equation becomes

$$V_n(0) = \max\left\{H(0), e^{-r\Delta t}V_{n+1}(0)\right\},$$

since the binomial tree will stay at zero.

Table B.1 presents some numerical results for the price of an American put option with maturity T and strike price K, where $S_0 = 50$, $r = 0.05$, $\sigma = 1.0$, and $T = 0.5$.

Table B.1: American option pricing with binomial trees.

	$K = 45$			$K = 50$		
	$N = 50$	$N = 200$	$N = 500$	$N = 50$	$N = 200$	$N = 500$
Price	0.2464	0.2442	0.2444	1.5251	1.5300	1.5310

Note that the binomial tree approximation is not a Monte Carlo simulation technique. Therefore, there is no standard errors associated with these prices. ■

Remark B.2. The binomial tree approximation described here is a special case of the so-called Markov chain approximation to diffusion processes. It is a powerful numerical technique for solving general continuous time stochastic control problems. There are a myriad of ways to construct implementable approximations to diffusion processes, and the Lamperti transform is not really necessary. The interested reader may consult [21].

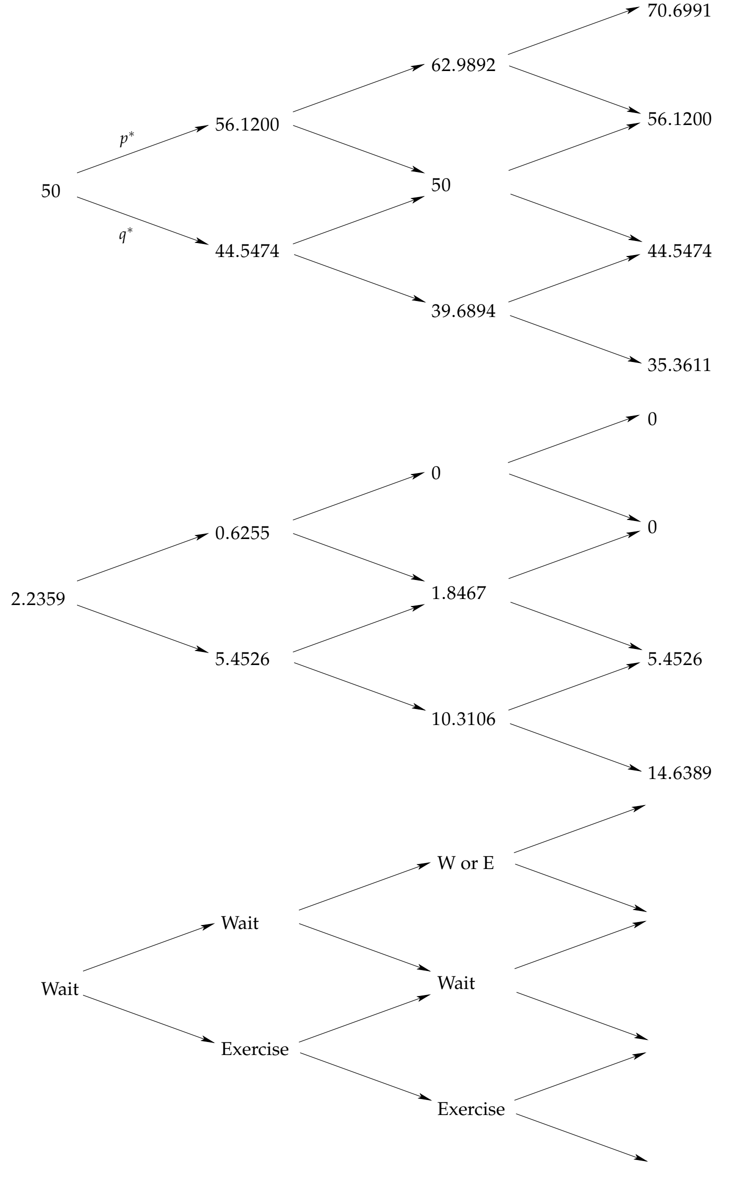

Figure B.1: American option pricing with binomial trees.

Appendix C

Option Pricing Formulas

In this appendix, we collect a number of explicit option pricing formulas. Unless otherwise specified, the underlying stock price is assumed to be 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, r is the risk-free interest rate, and σ is the volatility. All options are assumed to have maturity T. The option payoff is denoted by X.

i. **Binary call option: $X = \mathbf{1}_{\{S_T \geq K\}}$.**

$$\text{Price} = e^{-rT}\,\Phi\left(-\frac{\log(K/S_0) - (r - \sigma^2/2)T}{\sigma\sqrt{T}} \right).$$

ii. **Binary put option: $X = \mathbf{1}_{\{S_T \leq K\}}$.**

$$\text{Price} = e^{-rT}\,\Phi\left(\frac{\log(K/S_0) - (r - \sigma^2/2)T}{\sigma\sqrt{T}} \right).$$

iii. **Call option: $X = (S_T - K)^+$.**

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

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

iv. **Put option: $X = (K - S_T)^+$.**

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

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

v. **Average price call option with geometric mean: $X = (\bar{S}_G - K)^+$ with $\bar{S}_G = \left(\prod_{k=1}^{m} S_{t_k}\right)^{1/m}$.**

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

$$\bar{\mu} = \log S_0 + \frac{1}{m}\left(r - \frac{\sigma^2}{2}\right)\sum_{k=1}^{m} t_k, \quad \bar{\sigma} = \frac{\sigma}{m}\sqrt{\sum_{k=1}^{m}(2m-2k+1)t_k},$$

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

vi. **Average price put option with geometric mean: $X = (K - \bar{S}_G)^+$ with $\bar{S}_G = \left(\prod_{k=1}^{m} S_{t_k}\right)^{1/m}$.**

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

$$\bar{\mu} = \log S_0 + \frac{1}{m}\left(r - \frac{\sigma^2}{2}\right)\sum_{k=1}^{m} t_k, \quad \bar{\sigma} = \frac{\sigma}{m}\sqrt{\sum_{k=1}^{m}(2m-2k+1)t_k},$$

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

vii. **Lookback call option with floating strike: $X = S_T - \min_{0\le t\le T} S_t$.**

$$\text{Price} = S_0\Phi(\theta) - \frac{S_0\sigma^2}{2r}\Phi(-\theta) - S_0e^{-rT}\left(1 - \frac{\sigma^2}{2r}\right)\Phi(\theta - \sigma\sqrt{T}),$$

$$\theta = \left(\frac{r}{\sigma} + \frac{\sigma}{2}\right)\sqrt{T}.$$

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

$$\text{Price} = S_0 e^{-rT} \left(1 - \frac{\sigma^2}{2r}\right) \Phi(\sigma\sqrt{T} - \theta) - S_0 \Phi(-\theta) + \frac{S_0 \sigma^2}{2r} \Phi(\theta),$$

$$\theta = \left(\frac{r}{\sigma} + \frac{\sigma}{2}\right)\sqrt{T}.$$

ix. Lookback call option with fixed strike: $X = \left(\max_{0 \le t \le T} S_t - K\right)^+$.

(a) If $S_0 < K$:

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

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

(b) If $S_0 \ge K$:

$$\text{Price} = S_0 \Phi(\theta_+) \left(1 + \frac{\sigma^2}{2r}\right) + S_0 e^{-rT} \Phi(\theta_-) \left(1 - \frac{\sigma^2}{2r}\right) - K e^{-rT},$$

$$\theta_\pm = \left(\frac{\sigma}{2} \pm \frac{r}{\sigma}\right)\sqrt{T}.$$

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

(a) If $S_0 < K$:

$$\text{Price} = K e^{-rT} + S_0 e^{-rT} \Phi(\theta_-) \left(\frac{\sigma^2}{2r} - 1\right) - S_0 \Phi(\theta_+) \left(1 + \frac{\sigma^2}{2r}\right),$$

$$\theta_\pm = -\left(\frac{\sigma}{2} \pm \frac{r}{\sigma}\right)\sqrt{T}.$$

(b) If $S_0 \ge K$:

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

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

xi. Down-and-out call option: $X = (S_T - K)^+ \mathbf{1}_{\{\min_{0\le t\le T} S_t \ge b\}}$ **with** $S_0 > b$**.**

(a) If $b \le K$**:**

$$\begin{aligned}\text{Price} &= S_0\Phi(\theta) - Ke^{-rT}\Phi(\theta - \sigma\sqrt{T}) - b\left(\frac{b}{S_0}\right)^{2r/\sigma^2}\Phi(\mu) \\ &\quad + e^{-rT}\frac{KS_0}{b}\left(\frac{b}{S_0}\right)^{2r/\sigma^2}\Phi(\mu - \sigma\sqrt{T}),\end{aligned}$$

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

(b) If $b > K$**:**

$$\begin{aligned}\text{Price} &= S_0\Phi(\theta_+) - Ke^{-rT}\Phi(\theta_+ - \sigma\sqrt{T}) - b\left(\frac{b}{S_0}\right)^{2r/\sigma^2}\Phi(\theta_-) \\ &\quad + e^{-rT}\frac{KS_0}{b}\left(\frac{b}{S_0}\right)^{2r/\sigma^2}\Phi(\theta_- - \sigma\sqrt{T}),\end{aligned}$$

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

xii. Down-and-in call option: $X = (S_T - K)^+ \mathbf{1}_{\{\min_{0\le t\le T} S_t < b\}}$ **with** $S_0 > b$**.**

(a) If $b \le K$**:**

$$\text{Price} = b\left(\frac{b}{S_0}\right)^{2r/\sigma^2}\Phi(\theta) - e^{-rT}\frac{KS_0}{b}\left(\frac{b}{S_0}\right)^{2r/\sigma^2}\Phi(\theta - \sigma\sqrt{T}),$$

$$\theta = \frac{1}{\sigma\sqrt{T}}\log\frac{b^2}{KS_0} + \left(\frac{\sigma}{2} + \frac{r}{\sigma}\right)\sqrt{T}.$$

(b) If $b > K$**:**

$$\begin{aligned}\text{Price} &= S_0\left[\Phi(\mu) - \Phi(\theta_+)\right] - Ke^{-rT}\left[\Phi(\mu - \sigma\sqrt{T}) - \Phi(\theta_+ - \sigma\sqrt{T})\right] \\ &\quad + \left(\frac{b}{S_0}\right)^{2r/\sigma^2}\left[b\Phi(\theta_-) - e^{-rT}\frac{KS_0}{b}\Phi(\theta_- - \sigma\sqrt{T})\right],\end{aligned}$$

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

xiii. Up-and-out call option: $X = (S_T - K)^+ \mathbf{1}_{\{\max_{0\le t\le T} S_t \le b\}}$ with $S_0 < b$.

(a) If $b \le K$:

$$\text{Price} = 0.$$

(b) If $b > K$:

$$\begin{aligned}\text{Price} &= S_0\left[\Phi(\mu) - \Phi(\theta_+)\right] - Ke^{-rT}\left[\Phi(\mu - \sigma\sqrt{T}) - \Phi(\theta_+ - \sigma\sqrt{T})\right] \\ &\quad + b\left(\frac{b}{S_0}\right)^{2r/\sigma^2}\left[\Phi(\nu - \sigma\sqrt{T}) - \Phi(-\theta_-)\right] \\ &\quad - e^{-rT}\frac{KS_0}{b}\left(\frac{b}{S_0}\right)^{2r/\sigma^2}\left[\Phi(\nu) - \Phi(-\theta_- + \sigma\sqrt{T})\right],\end{aligned}$$

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

$$\nu = \frac{1}{\sigma\sqrt{T}}\log\frac{KS_0}{b^2} + \left(\frac{\sigma}{2} - \frac{r}{\sigma}\right)\sqrt{T}.$$

xiv. Up-and-in call option: $X = (S_T - K)^+ \mathbf{1}_{\{\max_{0\le t\le T} S_t > b\}}$ with $S_0 < b$.

(a) If $b \le K$:

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

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

(b) If $b > K$:

$$\begin{aligned}\text{Price} &= S_0\Phi(\theta_+) - Ke^{-rT}\Phi(\theta_+ - \sigma\sqrt{T}) \\ &\quad - b\left(\frac{b}{S_0}\right)^{2r/\sigma^2}\left[\Phi(\nu - \sigma\sqrt{T}) - \Phi(-\theta_-)\right] \\ &\quad + e^{-rT}\frac{KS_0}{b}\left(\frac{b}{S_0}\right)^{2r/\sigma^2}\left[\Phi(\nu) - \Phi(-\theta_- + \sigma\sqrt{T})\right],\end{aligned}$$

$$\theta_\pm = \left(\frac{r}{\sigma} + \frac{\sigma}{2}\right)\sqrt{T} \pm \frac{1}{\sigma\sqrt{T}}\log\frac{S_0}{b}, \;\; \nu = \frac{1}{\sigma\sqrt{T}}\log\frac{KS_0}{b^2} + \left(\frac{\sigma}{2} - \frac{r}{\sigma}\right)\sqrt{T}.$$

xv. Down-and-out put option: $X = (K - S_T)^+ \mathbf{1}_{\{\min_{0\le t\le T} S_t \ge b\}}$ with $S_0 > b$.

(a) If $b \le K$:

$$\begin{aligned}\text{Price} &= S_0\left[\Phi(\mu) - \Phi(\theta_+)\right] - Ke^{-rT}\left[\Phi(\mu - \sigma\sqrt{T}) - \Phi(\theta_+ - \sigma\sqrt{T})\right] \\ &\quad + b\left(\frac{b}{S_0}\right)^{2r/\sigma^2}\left[\Phi(\nu - \sigma\sqrt{T}) - \Phi(-\theta_-)\right] \\ &\quad - e^{-rT}\frac{KS_0}{b}\left(\frac{b}{S_0}\right)^{2r/\sigma^2}\left[\Phi(\nu) - \Phi(-\theta_- + \sigma\sqrt{T})\right],\end{aligned}$$

$$\theta_\pm = \left(\frac{r}{\sigma} + \frac{\sigma}{2}\right)\sqrt{T} \pm \frac{1}{\sigma\sqrt{T}}\log\frac{S_0}{b},\ \ \mu = \theta_+ + \frac{1}{\sigma\sqrt{T}}\log\frac{b}{K},$$
$$\nu = \frac{1}{\sigma\sqrt{T}}\log\frac{KS_0}{b^2} + \left(\frac{\sigma}{2} - \frac{r}{\sigma}\right)\sqrt{T}.$$

(b) If $b > K$:

$$\text{Price} = 0.$$

xvi. Down-and-in put option: $X = (K - S_T)^+ \mathbf{1}_{\{\min_{0\le t\le T} S_t < b\}}$ with $S_0 > b$.

(a) If $b \le K$:

$$\begin{aligned}\text{Price} &= -S_0\Phi(-\theta_+) + Ke^{-rT}\Phi(\sigma\sqrt{T} - \theta_+) \\ &\quad - b\left(\frac{b}{S_0}\right)^{2r/\sigma^2}\left[\Phi(\nu - \sigma\sqrt{T}) - \Phi(-\theta_-)\right] \\ &\quad + e^{-rT}\frac{KS_0}{b}\left(\frac{b}{S_0}\right)^{2r/\sigma^2}\left[\Phi(\nu) - \Phi(-\theta_- + \sigma\sqrt{T})\right],\end{aligned}$$

$$\theta_\pm = \left(\frac{r}{\sigma} + \frac{\sigma}{2}\right)\sqrt{T} \pm \frac{1}{\sigma\sqrt{T}}\log\frac{S_0}{b},\ \ \nu = \frac{1}{\sigma\sqrt{T}}\log\frac{KS_0}{b^2} + \left(\frac{\sigma}{2} - \frac{r}{\sigma}\right)\sqrt{T}.$$

(b) If $b > K$:

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

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

xvii. Up-and-out put option: $X = (K - S_T)^+ \mathbf{1}_{\{\max_{0 \le t \le T} S_t \le b\}}$ with $S_0 < b$.

(a) If $b \le K$:

$$\begin{aligned} \text{Price} &= -S_0 \Phi(-\theta_+) + K e^{-rT} \Phi(\sigma\sqrt{T} - \theta_+) + b \left(\frac{b}{S_0}\right)^{2r/\sigma^2} \Phi(-\theta_-) \\ &\quad - e^{-rT} \frac{K S_0}{b} \left(\frac{b}{S_0}\right)^{2r/\sigma^2} \Phi(\sigma\sqrt{T} - \theta_-), \end{aligned}$$

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

(b) If $b > K$:

$$\begin{aligned} \text{Price} &= -S_0 \Phi(-\theta) + K e^{-rT} \Phi(\sigma\sqrt{T} - \theta) + b \left(\frac{b}{S_0}\right)^{2r/\sigma^2} \Phi(-\mu) \\ &\quad - e^{-rT} \frac{K S_0}{b} \left(\frac{b}{S_0}\right)^{2r/\sigma^2} \Phi(\sigma\sqrt{T} - \mu), \end{aligned}$$

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

xviii. Up-and-in put option: $X = (K - S_T)^+ \mathbf{1}_{\{\max_{0 \le t \le T} S_t > b\}}$ with $S_0 < b$.

(a) If $b \le K$:

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

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

(a) If $b > K$:

$$\text{Price} = -b \left(\frac{b}{S_0}\right)^{2r/\sigma^2} \Phi(-\theta) + e^{-rT} \frac{K S_0}{b} \left(\frac{b}{S_0}\right)^{2r/\sigma^2} \Phi(\sigma\sqrt{T} - \theta),$$

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

xix. Exchange options: $X = [S_T^{(1)} - S_T^{(2)}]^+$. Assume that under the risk-neutral probability measure, the prices of the two underlying assets are

$$\begin{aligned} S_t^{(1)} &= S_0^{(1)} \exp\left\{\left(r - \frac{1}{2}\sigma_1^2\right) t + \sigma_1 W_t^{(1)}\right\}, \\ S_t^{(2)} &= S_0^{(2)} \exp\left\{\left(r - \frac{1}{2}\sigma_2^2\right) t + \sigma_2 W_t^{(2)}\right\}, \end{aligned}$$

where $(W^{(1)}, W^{(2)})$ is a two-dimensional Brownian motion with covariance matrix

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

Then

$$\text{Price} = S_0^{(1)} \Phi(\theta) - S_0^{(2)} \Phi(\theta - \sigma\sqrt{T}),$$

$$\theta = \frac{1}{\sigma\sqrt{T}} \log \frac{S_0^{(1)}}{S_0^{(2)}} + \frac{\sigma\sqrt{T}}{2}, \quad \sigma = \sqrt{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}.$$

Bibliography

[1] M. Baxter and A. Rennie. *Financial Calculus*. Cambridge University Press, Cambridge, UK, 2000.

[2] P. Billingsley. *Probability and Measure*. John Wiley & Sons, New York, 1995.

[3] F. Black and M. Scholes. The pricing of options and corporate liabilities. *Journal of Political Economy* 81:637–654, 1973.

[4] Y.S. Chow and H. Teicher. *Probability Theory: Independence, Interchangeability, Martingales*. Springer-Verlag, New York, 2003.

[5] K.L. Chung. *A Course in Probability Theory*. Academic Press, New York, 2000.

[6] J.C. Cox, J.E. Ingersoll, and S.A. Ross. An intertemporal general equilibrium model of asset prices. *Econometrica* 53:363–384, 1985.

[7] D. Duffie. *Dynamic Asset Pricing Theory*. Princeton University Press, Princeton, NJ, 1996.

[8] D. Duffie and P. Glynn. Efficient Monte Carlo simulation for security prices. *Annals of Applied Probability* 5:897–905, 1995.

[9] P. Dupuis and H. Wang. Importance sampling, large deviations, and differential games. *Stochastics and Stochastics Reports* 76:481–508, 2004.

[10] P. Dupuis and H. Wang. Subsolutions of an Isaacs equation and efficient schemes for importance sampling. *Mathematics of Operations Research* 32:723–757, 2007.

[11] P. Glasserman. *Monte Carlo Methods in Financial Engineering*. Springer-Verlag, New York, 2004.

[12] P. Glasserman, P. Heidelberger, and P. Shahabuddin. Asymptotically optimal importance sampling and stratification for path-dependent options. *Mathematical Finance* 9:117–152, 1999.

[13] P. Glasserman and Y. Wang. Counter examples in importance sampling for large deviations probabilities. *Annals of Applied Probability* 7:731–746, 1997.

[14] G. Hardy, J.E. Littlewood, and G. Pólya. *Inequalities*. Cambridge University Press, Cambridge, UK, 1952.

[15] S.L. Heston. A closed-form solution for options with stochastic volatility with applications to bonds and currency options. *The Review of Financial Studies* 6:327–343, 1993.

[16] J. Hull. *Options, Futures, and Other Derivative Securities*. Prentice-Hall, New Jersey, 2000.

[17] S. Kou. A jump diffusion model for option pricing. *Management Sciences* 48:1086–1101, 2002.

[18] I. Karatzas and S.E. Shreve. *Brownian Motion and Stochastic Calculus*. Springer-Verlag, New York, 1999.

[19] I. Karatzas and S.E. Shreve. *Methods of Mathematical Finance*. Springer-Verlag, New York, 1998.

[20] P.E. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations.* Springer-Verlag, Berlin, 1992.

[21] H.J. Kushner and P. Dupuis. *Numerical Methods for Stochastic Control Problems in Continuous Time.* Springer-Verlag, New York, 1992.

[22] R.C. Merton. Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics* 3:125–144, 1976.

[23] R.C. Merton. *Continuous-Time Finance*. Blackwell Publishers, Cambridge, MA, 1994.

[24] S. Natenberg. *Option Volatility and Pricing*. McGraw-Hill, New York, 1994.

[25] B. Oksendal. *Stochastic Differential Equations: An Introduction with Applications*. Springer, Berlin, 1998.

[26] L.C.G. Rogers and D. Williams. *Diffusions, Markov Processes, and Martingales: Volume 1, Foundations*. Cambridge University Press, Cambridge, UK, 2001.

[27] L.C.G. Rogers and D. Williams. *Diffusions, Markov Processes, and Martingales: Volume 2, Itô Calculus*. Cambridge University Press, Cambridge, UK, 2001.

[28] R.Y. Rubinstein. Optimization of computer simulation models with rare events. *European Journal of Operations Research* 99:89–112, 1997.

[29] R.Y. Rubinstein and D.P. Kroese. *Simulation and the Monte Carlo Method*. John Wiley & Sons, New Jersey, 2008.

[30] R.U. Seydel. *Tools for Computational Finance*. Springer-Verlag, Berlin, 2009.

[31] D. Siegmund. Importance sampling in the Monte Carlo study of sequential tests. *Annals of Statistics* 4:673–684, 1976.

[32] Y.L. Tong. *The Multivariate Normal Distribution*. Springer-Verlag, New York, 1989.

[33] S.R.S. Varadhan. *Large Deviations and Applications*. CBMS-NSF Regional Conference Series in Mathematics, SIAM, Philadelphia, 1984.

[34] D.D. Wackerly, W. Mendenhall, and R.L. Scheaffer. *Mathematical Statistics with Applications.* Duxbury Press, USA, 2008.

Monte Carlo Simulation with Applications to Finance

Developed from the author's course on Monte Carlo simulation at Brown University, **Monte Carlo Simulation with Applications to Finance** provides a self-contained introduction to Monte Carlo methods in financial engineering. Only requiring some familiarity with probability and statistics, the book keeps much of the mathematics at an informal level and avoids technical measure-theoretic jargon to provide a practical understanding of the basics. It includes a large number of examples as well as MATLAB® coding exercises that are designed in a progressive manner so that no prior experience with MATLAB is needed.

The author first presents the necessary mathematical tools for simulation, arbitrary free option pricing, and the basic implementation of Monte Carlo schemes. He then describes variance reduction techniques, including control variates, stratification, conditioning, importance sampling, and cross-entropy. The text concludes with stochastic calculus and the simulation of diffusion processes.

Features

- Presents common variance reduction techniques as well as the cross-entropy method
- Covers the simulation of diffusion process models
- Requires minimal background in mathematics and finance
- Contains numerous examples of option pricing, risk analysis, and sensitivity analysis
- Includes many hand-and-paper and MATLAB coding exercises at the end of every chapter

CRC Press
Taylor & Francis Group
an **informa** business
www.crcpress.com

6000 Broken Sound Parkway, NW
Suite 300, Boca Raton, FL 33487
711 Third Avenue
New York, NY 10017
2 Park Square, Milton Park
Abingdon, Oxon OX14 4RN, UK

K12713

ISBN: 978-1-4398-5824-0
90000
9 781439 858240