

MONTE CARLO SIMULATION IN FINANCIAL ENGINEERING

Nan Chen

L. Jeff Hong

Dept. of Systems Engineering & Engineering Management
The Chinese University of Hong Kong
Shatin, N.T., HONG KONG

Dept. of Industrial Engineering & Logistics Management
The Hong Kong University of Science and Technology
Clear Water Bay, Kowloon, HONG KONG

ABSTRACT

This paper reviews the use of Monte Carlo simulation in the field of financial engineering. It focuses on several interesting topics and introduces their recent development, including path generation, pricing American-style derivatives, evaluating Greeks and estimating value-at-risk. The paper is not intended to be a comprehensive survey of the research literature.

1 INTRODUCTION

Many problems in financial engineering focus on estimating a certain value, e.g., pricing derivative securities, computing price sensitivities, evaluating portfolio risks. The value can often be written as or transformed to an expectation of a complicated random variable whose behavior is modeled as a stochastic process. Monte Carlo simulation is a method that is often used to estimate expectations. Compared to other numerical methods, Monte Carlo simulation has several advantages. First, it is easy to use. In most situations, if the sample paths from the stochastic process model can be simulated, then the value can be estimated. Second, its rate of convergence typically does not depend on the dimensionality of the problem. Therefore, it is often attractive to apply Monte Carlo simulation to problems with high dimensions.

To apply Monte Carlo simulation to estimate a financial value, there are typically three steps: generating sample paths, evaluating the payoff along each path, and calculating an average to obtain estimation. In this paper, we will discuss the recent development of these steps and their applications. Before that, we will first provide a financial background for readers who are not familiar with financial engineering.

In Section 3 we discuss path generation. Since simulation can only generate sample paths in discrete times, how to control discretization error becomes the central issue in path generation. In this section, we first introduce Euler and Milstein discretization schemes, and compare their rates of convergence. We further introduce the recent development

on exact simulation, which aims to generate sample paths that have no discretization error.

The evaluation of payoffs along sample paths for derivatives is often straight-forward except for American-style derivatives. For the evaluation step for derivatives that are not American style, the research has focused mainly on improving the efficiency of simulation. A good review of those methods can be found in Staum (2002), which reviews the development of Monte Carlo method in financial engineering by 2002. In this paper, we focus on the pricing of American-style derivatives, and introduce some recent work, e.g., stochastic mesh method and dual method, in Section 4.

Besides pricing of derivative securities, we also introduce some applications of Monte Carlo simulation in risk management. For the practical purpose of risk management, people always want to know price sensitivities (*Greeks*) and *Value-at-Risk* (*VaR*) of an investment portfolio. We discuss the estimation of Greek and VaR estimation in Section 5.

2 FINANCIAL BACKGROUND

One topic at the core of the field of financial engineering is how to evaluate derivative securities “fairly”. Derivatives are financial instruments whose payoffs are derived from underlying market variables such as stock prices, commodity prices, market indices and interest rates, etc. A standard example of such derivatives is European options contingent on an underlying asset. The option entitles the holder a right to buy (*call*) or sell (*put*) a certain amount of underlying assets from or to the option issuer on the option maturity for a pre-specified price (*exercise price* or *strike price*).

The payoff of a derivative usually depends on the future prices of the underlying. Consider a European call option as an example. Denote K to be its strike price. When the option matures, the holder will exercise the right when the spot price of the underlying $S_T > K$ and will not do so if $S_T \leq K$. Therefore the payoff of the European option is given by $\max\{S_T - K, 0\}$. Consequently, the pricing problem now boils down to find a way to derive the present value of such

future payoffs, which are dependent on the future prices of the underlying, from the current underlying information.

As a very first step towards the target, it is indispensable to establish an accurate model to describe the underlying asset movements. In finance, the following system of SDEs is widely used:

$$\frac{dS_t^i}{S_t^i} = \mu_i(t, S_t)dt + \sum_{j=1}^k \sigma_{ij}(t, S_t)dW_t^j, \quad 1 \leq i \leq d, \quad (1)$$

where $S_t = (S_t^1, \dots, S_t^d)$ is the values of the underlyings at time t , $W = (W_t^1, \dots, W_t^k)$ is a standard k -dimensional Brownian motion to capture the random fluctuation of the underlyings, and each of μ_i and σ_{ij} are scalar-valued functions. In addition, a risk free *money market* account is often introduced, whose dynamic is given by $dS_t^0/S_t^0 = r_t dt$, where r_t is the instantaneous risk free interest rate at time t .

Based on such models, starting from Black and Scholes (1973) and Merton (1973), an elegant and remarkably practical mathematical theory of derivative pricing has been developed. Detailed treatment of the theory is obviously not suitable for a tutorial paper like this. In the rest of the section we would like to highlight some principles of the theory, especially focusing on those bridging the connection between the theory and Monte Carlo simulation, and refer readers to Björk (1998) and Duffie (2001) and the references therein for further background.

Two major approaches exist in the literature to derive financial derivative prices: one is through *replication* argument and the other through *risk neutral probability*. In details, suppose that model (1) holds and we have a derivative with payoff function $\Phi(S_T)$. Denote the derivative value at time t to be $V(t, S_t^0, S_t)$ when the money market account is S_t^0 and the underlying price is S_t . Then the former approach will show that V must satisfy the following PDE:

$$\frac{\partial V}{\partial t} + \frac{1}{2} \sum_{i,j=1}^d \Sigma_{ij} S_t^i S_t^j \frac{\partial^2 V}{\partial S_i \partial S_j} = r_t S_t^0 \frac{\partial V}{\partial S^0}. \quad (2)$$

for $0 \leq t \leq T$ and the boundary condition $V(T, S_T^0, S_T) = \Phi(S_T)$, where $\Sigma_{ij} = \sum_{l=1}^k \sigma_{il} \sigma_{lj}$.

In theory, we may call for numerical methods to solve the above PDE (2) for the derivative price. But several features limit the feasibility of the approach. First, when the dynamic (1) is complicated, the solution to the PDE (2) may be very difficult to obtain or even fail to exist; second, high dimensional assets (e.g., $d \geq 3$) will make the numerical solution of the PDE impractical; third, it is not easy to derive the corresponding PDE for such derivatives as Asian options, whose payoffs depend on the whole sample path of the underlying assets' historical price.

The introduction of *risk neutral pricing* overcomes the barriers that the PDE approach encounters. According

to the Feynman-Kac formula (Karatzas and Shreve 1991, Chapter 4), the solution to (2) has a very nice probabilistic representation:

$$V(t, s) = \tilde{E}[e^{-\int_t^T r_s ds} \Phi(S_T) | S_t = s], \quad (3)$$

where \tilde{E} is the expectation under a new probability measure \tilde{P} . Under it, the dynamic of underlying price S is given by

$$\frac{dS_t^i}{S_t^i} = r_t dt + \sum_{j=1}^k \sigma_{ij}(t, S_t) d\tilde{W}_t^j, \quad 1 \leq i \leq d, \quad (4)$$

where \tilde{W} is the standard Brownian motion under new probability \tilde{P} . Notice that in (4), the expected returns of underlying assets are always the same as the risk free interest rate r . That is why people call the new measure "risk neutral".

The relation (3) demonstrates the applicability of Monte Carlo simulation to the field of derivative pricing. Now what we need to do is simply to estimate the expectation of some functions of sample paths of a diffusion process. Notice that the difference of (1) and (4) in their financial interpretation plays no role from the view point of Monte Carlo simulation. Thus, from now on, we always consider the following general model

$$dS_t^i = a_i(t, S_t)dt + \sum_{j=1}^k b_{ij}(t, S_t)dW_t^j, \quad 1 \leq i \leq d \quad (5)$$

and how to evaluate the expectation $E[\Phi(S)]$ efficiently, skipping all irrelevant detailed financial interpretation.

In contrast to the PDE approach, Monte Carlo simulation has its own attractions. First, we bypass the technical obstacle to verifying the existence of the solution to (4), which could be extremely hard in some cases; second, Monte Carlo simulation is much easier to implement than the PDE approach, especially for high dimensional problems; third, the probabilistic representation of derivative prices (3) actually is valid for a very general class of underlying asset dynamics, such as Lévy processes, and payoff functions which may be dependent on the whole sample path.

3 PATH GENERATION

In the section we overview the methods to generate sample paths. To avoid obscuring the main idea by unnecessarily complicated notations, we start from the case of $d = 1$ and $k = 1$ in (5) first.

The most straightforward scheme is known as the *Euler-Maruyama discretization*, which is named after the work of Maruyama (1955). The idea is to approximate the solution to SDE (5) by a finite difference recursion. Given an interval $[0, T]$ and a fixed time step h such that $h = T/N$ for a positive integer N , the approximation of S_{ih} is given by a recursion

as follows:

$$\hat{S}_i = \hat{S}_{i-1} + a((i-1)h, \hat{S}_{i-1})h + b((i-1)h, \hat{S}_{i-1})\sqrt{h}Z_i, \quad (6)$$

for $1 \leq i \leq N$ with Z a sequence of i.i.d. standard normal random variables and $\hat{S}_0 = S_0$.

To assess the quality of different schemes, we need to know how much discretization error is introduced. There are two main categories of criteria used commonly in the literature: *strong convergence criteria* and *weak convergence criteria*. We say that the approximation *converges strongly with order* $\gamma > 0$ if for all sufficiently small time steps h ,

$$E[\|\hat{S} - S\|] \leq Ch^\gamma$$

for some constant C and some norm $\|\cdot\|$. Typical choices of norms include the L^p norm $\|\hat{S}_N - S_T\|^p$ and the max norm $\sup_{0 \leq t \leq T} \|\hat{S}_{[t/h]} - S_t\|$. Under some smoothness conditions on the drift and volatility functions a and b , one can show that the Euler scheme typically has a strong order of $1/2$ (Theorem 10.2.2, Kloeden and Platen 1992).

The weak criteria are more relevant with applications in pricing derivatives than the strong ones since they characterize how close the expectations of a function computed from \hat{S} are to that computed from S . A typical weak error criterion has the form

$$|E[f(\hat{S}_N)] - E[f(S_T)]|.$$

We say that a discretization scheme has *weak order of convergence* γ if there exists some uniform constant C such that the above difference is less than Ch^γ for all sufficiently small h and all functions f whose derivatives of order $0, 1, \dots, 2\gamma + 2$ exist and are polynomially bounded. The Euler scheme turns out to have a weak order of 1 under some smoothness conditions on a and b (Theorem 14.1.5, Kloeden and Platen 1992).

Obviously, we prefer to achieve a larger γ because it implies smaller discretization error for a fixed discretization step h , which is usually associated with computational effort. One line of research is devoted to looking for schemes with higher convergence orders. To motivate the following refinements, let us rewrite

$$E[f(S_T)] = E[f(S_0)] + E\left[\sum_{i=1}^N E[f(S_{(i+1)h}) - f(S_{ih})|S_{ih}]\right].$$

If f is smooth enough, applying the Taylor expansion to f ,

$$E[f(S_{(i+1)h}) - f(S_{ih})|S_{ih}] \approx f'(S_{ih}) \cdot E[S_{(i+1)h} - S_{ih}|S_{ih}].$$

Therefore, we need accurate estimation of the conditional expectations $E[S_{(i+1)h} - S_{ih}|S_{ih}]$ to achieve a discretization scheme with a high order of weak convergence. Notice that

the difference $S_{(i+1)h} - S_{ih}$ is equal to

$$\int_{ih}^{(i+1)h} a(t, S_t)dt + \int_{ih}^{(i+1)h} b(t, S_t)dW_t. \quad (7)$$

We can see that a good approximation to the above integrals is necessary. The Euler scheme uses $a(ih, S_{ih})h$ and $b(ih, S_{ih})(W_{(i+1)h} - W_{ih})$ to approximate the two integrals respectively.

Applying Ito's formula to a , for $ih \leq t < (i+1)h$,

$$a(t, S_t) = a(ih, S_{ih}) + \int_{ih}^t \mathcal{L}^0 a(u, S_u)du + \int_{ih}^t \mathcal{L}^1 a(u, S_u)dW_u$$

where \mathcal{L}^0 and \mathcal{L}^1 are two functional operators defined as follows:

$$\begin{aligned} \mathcal{L}^0 &:= \frac{\partial}{\partial t} + a(t, S) \frac{\partial}{\partial S} + \frac{1}{2} b^2(t, S) \frac{\partial^2}{\partial S^2} \\ \mathcal{L}^1 &:= b(t, S) \frac{\partial}{\partial S}. \end{aligned}$$

Making a substitution in $\int_{ih}^{(i+1)h} a(t, S_t)dt$, we have

$$\begin{aligned} \int_{ih}^{(i+1)h} a(t, S_t)dt &= a(ih, S_{ih})h + \\ &\int_{ih}^{(i+1)h} \int_{ih}^t \mathcal{L}^0 a(u, S_u)dudt + \int_{ih}^{(i+1)h} \int_{ih}^t \mathcal{L}^1 a(u, S_u)dW_u dt. \end{aligned}$$

Finally, we approximate the above multiple integrals by the following:

$$\begin{aligned} \mathcal{L}^0 a(ih, S_{ih}) \int_{ih}^{(i+1)h} \int_{ih}^t dudt &+ \\ &+ \mathcal{L}^1 a(ih, S_{ih}) \int_{ih}^{(i+1)h} \int_{ih}^t dW_u dt \\ &\approx \mathcal{L}^0 a(ih, S_{ih}) \cdot \frac{1}{2} h^2 + \mathcal{L}^1 a(ih, S_{ih}) \Delta I_i \end{aligned}$$

where $\Delta I_i = \int_{ih}^{(i+1)h} (W_t - W_{ih})dt$.

Similarly, the integral $\int_{ih}^{(i+1)h} b(t, S_t)dW_t$ can be approximated by

$$\mathcal{L}^0 b(ih, S_{ih}) \cdot (h\Delta W_i - \Delta I_i) + \mathcal{L}^1 b(ih, S_{ih}) \cdot \frac{1}{2} ((\Delta W_i)^2 - h)$$

where $\Delta W_i = W_{(i+1)h} - W_{ih}$.

Consequently, we end up at another discretization scheme, known as the *Milstein scheme* in the literature,

which is given by the recursion:

$$\hat{S}_{i+1} = \hat{S}_i + \mathcal{L}^0 a(ih, S_{ih}) \cdot \frac{1}{2} h^2 + \mathcal{L}^1 a(ih, S_{ih}) \cdot \Delta I_i + \mathcal{L}^0 b(ih, S_{ih}) \cdot (h\Delta W_i - \Delta I_i) + \mathcal{L}^1 b(ih, S_{ih}) \cdot \frac{1}{2} ((\Delta W_i)^2 - h).$$

Talay (1984) shows that the Milstein scheme can achieve a weak order of 2 under some smoothness conditions on the coefficient functions a and b . The implementation of the scheme involves simulating two correlated normal random variables ΔW_i and ΔI_i with the joint distribution:

$$(\Delta W_i, \Delta I_i)^T \sim N\left(0, \begin{pmatrix} h & \frac{1}{2}h^2 \\ \frac{1}{2}h^2 & \frac{1}{3}h^3 \end{pmatrix}\right).$$

In theory, if the coefficients a and b are sufficiently smooth, one may apply Ito's formula on the functions $\mathcal{L}^0 a$, $\mathcal{L}^1 a$, $\mathcal{L}^0 b$ and $\mathcal{L}^1 b$ repeatedly and obtain higher-order multiple integrals representation for (7). Kloeden and Platen (1992) show that the schemes obtained through approximating those higher-order multiple integrals actually are able to produce any arbitrarily high weak or strong order (Kloeden and Platen 1992, Section 14.5). But these high order schemes are too cumbersome to be implemented in practice. We refer readers who are interested in the topic to Kloeden and Platen (1992) and Glasserman (2003), Chapter 6.

As for high dimensional S (i.e., $d > 1$ or $k > 1$ in (5)), we also can use a derivation parallel to the scalar case to derive the corresponding high-order schemes. But one difficulty of such high dimensional schemes lies in how to simulate the following mixed integral:

$$\int_t^{t+h} [W_u^k - W_u^j] dW_u^j, \quad k \neq j.$$

Several approaches are suggested in the literature to tackle the difficulty, simulating directly from its distribution, or replacing it by simpler random variables, or imposing an artificial assumption on the volatility function b . Readers may also refer to Glasserman (2003), Chapter 6 for the details.

Recently, another line of research appears aiming to circumvent the discretization error of traditional discrete approximation schemes (e.g. Beskos and Roberts 2005, Beskos et al. 2006a, Beskos, Papaspiliopoulos, and Roberts 2006b, DiCesare and McLeish 2006). It involves acceptance-rejection sampling and returns exact draws from any finite dimensional distribution of the solutions of SDE.

Acceptance-rejection sampling (ARS) is a widely used Monte Carlo simulation technique. Let (Ω, \mathcal{F}) be a measurable space and we wish to simulate some random elements

according to a probability measure μ on Ω . Assume ν is a probability measure on Ω from which we know how to generate samples and it has the property that the Radon-Nykodym derivative of μ with respect to ν is uniformly bounded, i.e., there exists some constant $\varepsilon > 0$ such that

$$\frac{d\mu}{d\nu}(\omega) \leq \frac{1}{\varepsilon}, \text{ for all } \omega \in \Omega.$$

Let $f := \varepsilon \cdot (d\mu/d\nu)$ and we have $0 \leq f \leq 1$. The ARS technique can help us generate samples from μ through ν . The procedure is to sample X from its distribution under ν and accept it with probability $f(X)$.

Formally, at first we generate an i.i.d. sequence of $(X_n, I_n), n \geq 1$, X_n sampled from ν and a binary indicator I_n with conditional probability $P[I = 1|X = x] = f(x)$ for all $x \in \Omega$. We continue to reject X_n until we meet the first n such that $I_n = 1$. Suppose that $\tau = \inf\{n : I_n = 1\}$. Then we claim that X_τ is a sample with the distribution μ . To verify it, for any $A \in \mathcal{F}$,

$$P[X_\tau \in A] = P[X \in A|I = 1] = \frac{P[X \in A, I = 1]}{P[I = 1]}.$$

As we know,

$$\begin{aligned} P[I = 1] &= \int P[I = 1|X = x]d\nu(x) \\ &= \int f(x)d\nu(x) = \varepsilon \int d\mu(x) = \varepsilon \end{aligned}$$

and

$$\begin{aligned} P[X \in A, I = 1] &= \int_A P[I = 1|X = x]d\nu(x) \\ &= \varepsilon \int_A d\mu(x) = \varepsilon \mu(A). \end{aligned}$$

Thus, $P[X_\tau \in A] = \mu(A)$, i.e., X_τ follows the distribution μ . In summary, the ARS algorithm is valid if we can find a function $f, 0 \leq f \leq 1$, which is proportional to $d\mu/d\nu$, to generate the decision variable I whose distribution is defined by f .

Go back to the discussion of simulation of (5). Once again, we start from one dimensional case to illustrate the main idea. Our target is to generate samples of S_T . Let us introduce a function $F(t, y) = \int_0^y 1/b(t, u)du$. Under the condition that the volatility coefficient b is positive definite, F is strictly increasing in y and therefore the inverse of F exists. Denoted it by $F^{-1}(t, \cdot)$. Let $Y_t := F(t, S_t)$. Applying Ito's lemma to Y , we immediately get that

$$dY_t = \tilde{a}(t, Y_t)dt + dW_t, \tag{8}$$

where $\tilde{a}(t,y)$ is given by

$$\tilde{a}(t,y) = \left[\frac{a(t,u)}{b(t,u)} - \frac{1}{2} \frac{\partial b}{\partial S}(t,u) - \int_0^u \frac{\partial b / \partial t}{b^2}(t,v) dv \right],$$

where u is evaluated at $u = F^{-1}(t,y)$. Once we know how to simulate Y_T , we can recover S_T from the relation that $S_T = F^{-1}(T, Y_T)$.

Consider a function space $\mathcal{C}[0, T]$ which is the set of all continuous functions on $[0, T]$. Denote \mathbf{Q} and \mathbf{Z} to be the measures on $\mathcal{C}[0, T]$ induced by the process Y and the standard Brownian motion. Combining Girsanov's theorem and Ito's lemma, one can show that, for any $\omega \in \mathcal{C}[0, T]$

$$\frac{d\mathbf{Q}}{d\mathbf{Z}}(\omega) = \exp\{A(\omega_T)\} \cdot \exp\left\{-\int_0^T \phi(t, \omega_t) dt\right\}, \quad (9)$$

where $A(u) = \int_0^u \tilde{a}(y) dy$ and $\phi(t, \omega_t) = (\tilde{a}^2(t, \omega_t) + \partial \tilde{a}(t, \omega_t) / \partial y) / 2$.

In light of (9), we can do the simulation in the following steps to draw a sample of Y_T . First, simulate ω_T according to the measure \mathbf{Z} , which is a candidate to be accepted or rejected. This is easy to do because $\omega_T \sim N(0, T)$ under \mathbf{Z} . Given such ω_T , $\exp\{A(\omega_T)\}$ becomes a constant and therefore $d\mathbf{Q}/d\mathbf{Z}$ is proportional to

$$\exp\left\{-\int_0^T \phi(t, \omega_t) dt\right\}. \quad (10)$$

Now we need to simulate a decision variable I to decide whether to accept ω_T or to reject it. The probability $P[I = 1 | \omega]$ should be proportional to (10). Surprisingly, it is not necessary to know the whole path ω for the purpose of simulating I if we notice that (10) is very similar to the probability that a time-inhomogeneous Poisson process with intensity $\phi(t, \omega_t)$ at instant t has no event points occurring in $[0, T]$.

Suppose that there exists some constant k such that $0 \leq \phi(y) \leq k$ for all y . Under such assumption, we may use *stochastic thinning* to generate a Poisson process with intensity $\phi(t, \omega_t)$. Simulate a path of a homogeneous Poisson process with intensity k . This results in some event points on $[0, T]$: $0 \leq \tau_1 < \dots < \tau_n \leq T$. Generate n independent uniformly distributed random variables U_1, \dots, U_n and then accept the event i if $U_i > \phi(\tau_i, \omega_{\tau_i})/k$ and reject it otherwise. Finally, I is set to be 0 if there is any event being accepted or to be 1 otherwise. One can easily show that the probability that $I = 1$ is proportional to (10).

From the above algorithm, what we need to know is the values of ω at all τ_i 's given ω_T . In other words, all ω_{τ_i} should be sampled from a Brownian bridge with starting point 0 and ending point ω_T . One can refer to Glasserman (2003), pp. 82-86 for detailed procedure of simulating a Brownian bridge.

Another way to construct I is through Taylor expansion of the right hand side of (9), suggested by Beskos and Roberts (2005). This way decomposes the event of $I = 1$ into the union of a series of sets. Beskos, Papaspiliopoulos, and Roberts (2006b) explore how to relax the bounded assumption of ϕ to a case that ϕ is bounded below only.

As for high dimensional process S , not all processes can be transformed into the form of (8). Ait-Sahalia and Mykland (2004) show that such transform exists if and only if volatility function b satisfies the following commutativity condition: for any triplet (i, m, l)

$$\sum_{j=1}^k \frac{\partial b_{im}(t,s)}{\partial s_j} b_{jl}(t,s) = \sum_{j=1}^k \frac{\partial b_{il}(t,s)}{\partial s_j} b_{jm}(t,s).$$

Given such b , DiCesare and McLeish (2006) discuss how to implement exact simulation for high dimensional process S .

4 AMERICAN-STYLE DERIVATIVE PRICING

In this section we concentrate on American-style derivative pricing. Whereas a European option can be exercised only at a fixed date, an American-style derivative can be exercised any time up to its maturity. Thus finding its value entails finding the optimal exercise policy, which poses a significant challenge for simulation.

Suppose that we have an American-style derivative whose payoff function is given by Φ . Then its value is represented by

$$\sup_{\tau \in \mathcal{T}} E[e^{-\int_t^\tau r_s ds} \Phi(S_\tau)], \quad (11)$$

where \mathcal{T} is a class of admissible stopping times with values in $[0, T]$. In the following, to avoid cumbersome notations, we suppress the discount factor $\exp(-\int_t^\tau r_s ds)$ which may be done by appropriately redefining Φ and S . In addition, we restrict ourselves to those derivatives that can be exercised only at fixed discrete times: $0 \leq t_1 < \dots < t_K \leq T$, and consider \mathcal{T} consisting only of the stopping times valued at $t_i, 1 \leq i \leq K$. Such restriction justifies itself when we take K to be sufficiently large to approximate the value of continuously exercisable American-style derivatives.

4.1 Dynamic Programming Approaches

For optimal stopping time problem (11), the solution is given by the following backward recursion:

$$V_K(s) = \Phi(s) \quad (12)$$

$$V_{i-1}(s) = \max(\Phi(s), E[V_i(S_{t_i}) | S_{t_{i-1}} = s]), \quad (13)$$

$$i = 1, \dots, K,$$

if $V_i(s)$ denotes the value of the derivative at time t_i given the underlying asset price $S_{t_i} = s$. Equation (12) states that the derivative value at the maturity must be equal to the payoff value at that time because there is no further exercise opportunity left. Equation (13) states that in intermediate steps the derivative value must be the maximum of the immediate exercise value and the expected present value of future payoffs because the derivative holder has two options at those time epochs: to stop or to continue.

The difficulty of applying Monte Carlo simulation methods in American derivative pricing is how to efficiently evaluate *continuation values* $C_{i-1}(s) := E[V_i(S_{t_i}) | S_{t_{i-1}} = s]$ for all $1 \leq i \leq K$. Once this is done, the derivative value is obtained by recursion (12) and (13) and the optimal exercise policy is determined by

$$\tau^* = \min\{t_i, 1 \leq i \leq K : \Phi(S_{t_i}) \geq C_i(S_{t_i})\}.$$

Broadie and Glasserman (1997) suggest a *random tree method* to produce two consistent estimators for the derivative value, one biased high and one biased low, and both of them converging to the true value.

Given a fixed a branching number $m > 0$, the random tree method starts by simulating a tree of sample paths. Given the initial price state S_0 , simulate m independent successors $S_{t_1}^1, \dots, S_{t_1}^m$ for possible asset prices at time t_1 ; for each $S_{t_1}^i, 1 \leq i \leq m$, simulate m independent successor $S_{t_2}^{i1}, \dots, S_{t_2}^{im}$, and so on, until generating all nodes at time t_K . Denote a generic node in the tree at time t_i by $S_{t_i}^{j_1 j_2 \dots j_i}$, where j_i is a number in $\{1, 2, \dots, m\}$. The superscript indicates that the node is reached by following the j_1 th branch of node S_0 , the j_2 th branch of node $S_{t_1}^{j_1}$, and so on.

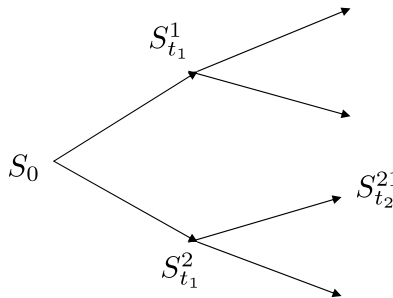


Figure 1: Random tree with $m = 2$ and $K = 2$.

Set $\tilde{V}_K^{j_1 j_2 \dots j_K} = \Phi(S_{t_K}^{j_1 j_2 \dots j_K})$ and use

$$\tilde{C}_{i-1}^{j_1 j_2 \dots j_{i-1}} = \frac{1}{m} \sum_{j=1}^m \tilde{V}_i^{j_1 j_2 \dots j_{i-1} j}, \quad 1 \leq i \leq K$$

to approximate the continuation value C_{i-1} at node $S_{t_{i-1}}^{j_1 j_2 \dots j_{i-1}}$. The approximation of the derivative value at the same node is defined as $\tilde{V}_{i-1}^{j_1 j_2 \dots j_{i-1}} = \max[\Phi(S_{t_{i-1}}^{j_1 j_2 \dots j_{i-1}}), \tilde{C}_{i-1}^{j_1 j_2 \dots j_{i-1}}]$.

A simple induction argument demonstrates that $E[\tilde{V}]$ gives an upper bound for the true value of the derivative, i.e., $E[\tilde{V}_i^{j_1 j_2 \dots j_i} | S_{t_i}^{j_1 j_2 \dots j_i}] \geq V(S_{t_i}^{j_1 j_2 \dots j_i})$ for all $0 \leq i \leq K$. First, the inequality holds obviously for $i = K$. Now if it holds for the nodes of time t_{i+1} , then

$$\begin{aligned} E[\tilde{C}_i^{j_1 j_2 \dots j_i} | S_{t_i}^{j_1 j_2 \dots j_i}] &= E[\tilde{V}_{i+1}^{j_1 j_2 \dots j_i} | S_{t_i}^{j_1 j_2 \dots j_i}] \\ &\geq E[V_{i+1}(S_{t_{i+1}}) | S_{t_i}^{j_1 j_2 \dots j_i}] = C_i(S_{t_i}^{j_1 j_2 \dots j_i}), \end{aligned}$$

where the first equality uses the fact that the m successors are independently simulated and the inequality is due to the induction hypothesis. Therefore, by Jensen's inequality,

$$\begin{aligned} E[\tilde{V}_i^{j_1 j_2 \dots j_i} | S_{t_i}^{j_1 j_2 \dots j_i}] &= E[\max(\Phi(S_{t_i}^{j_1 j_2 \dots j_i}), \tilde{C}_i^{j_1 j_2 \dots j_i}) | S_{t_i}^{j_1 j_2 \dots j_i}] \\ &\geq \max(\Phi(S_{t_i}^{j_1 j_2 \dots j_i}), E[\tilde{C}_i^{j_1 j_2 \dots j_i} | S_{t_i}^{j_1 j_2 \dots j_i}]), \end{aligned}$$

where the right hand side is exactly $V(S_{t_i}^{j_1 j_2 \dots j_i})$.

The high biased estimator \tilde{V} results from the fact that we use the same future information to estimate continuation values and to decide whether to stop or to continue at each step. Separation of successors into two disjoint subsets, one for the decision of exercise or not and the other for the estimation of continuation values, will lead to low biased estimators. More precisely, at the terminal nodes, define $\tilde{V}_K^{j_1 j_2 \dots j_K} := \Phi(S_{t_K}^{j_1 j_2 \dots j_K})$; at intermediate node $S_{t_i}^{j_1 \dots j_i}$ define

$$\tilde{V}_{i+1}^{j_1 j_2 \dots j_i} = \begin{cases} \Phi(S_{t_i}^{j_1 j_2 \dots j_i}), & \text{if } \frac{1}{m-1} \sum_{j \neq k} \tilde{V}_i^{j_1 j_2 \dots j_i j} \leq \Phi(S_{t_i}^{j_1 j_2 \dots j_i}); \\ \tilde{V}_{i+1}^{j_1 j_2 \dots j_i k}, & \text{otherwise} \end{cases}$$

for $1 \leq k \leq m$ and set the following as the estimation of the derivative value at the node: $\tilde{V}_i^{j_1 j_2 \dots j_i} = \sum_{k=1}^m \tilde{V}_{i+1}^{j_1 j_2 \dots j_i k} / m$. Broadie and Glasserman (1997) showed that $E[\tilde{V}_i^{j_1 j_2 \dots j_i} | S_{t_i}^{j_1 j_2 \dots j_i}] \leq V_i(S_{t_i}^{j_1 j_2 \dots j_i})$ and that both high biased estimator \tilde{V} and low biased estimator \tilde{v} converge in probability to the true value V as $m \rightarrow +\infty$.

The major drawback of random tree method is that the number of nodes grows exponentially as K increases, which would incur a huge computational burden when considering high dimensional problems. The *Stochastic mesh method* is introduced by Broadie and Glasserman (2004) to overcome the disadvantage. Instead of simulating a random tree in the previous method, the stochastic mesh method generates m independent sample paths starting from the initial S_0 (cf. solid arrows in Figure 2). Denote a generic node in the j th path at time t_i by $S_{t_i}^j, 1 \leq j \leq m$.

Consider the estimation of continuation value at node $S_{t_i}^j$. We use all nodes at time t_{i+1} although $m - 1$ of them are not the successors of $S_{t_i}^j$. To correct the "abuse", we need to attach different weights to the estimated values at

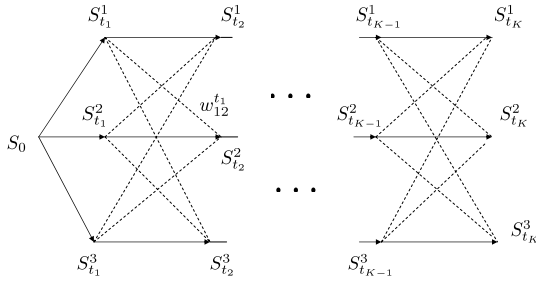


Figure 2: Stochastic mesh.

different nodes when we sum them up. Suppose that at the terminal nodes $\tilde{V}_K^j := \Phi(S_{t_K}^j)$. The estimated continuation value at $S_{t_i}^j$ is then given by

$$\tilde{C}_i^j := \frac{1}{m} \sum_{k=1}^m \tilde{V}_{i+1}^k w_{j,k}^{t_i}$$

for $1 \leq i < K$, where $w_{j,k}^{t_i}$ is the corresponding weight of $S_{t_{i+1}}^k$ used to calculate the continuation value at node $S_{t_i}^j$ (dotted lines in Figure 2). Then the estimated value of the derivative is given by $\tilde{V}_i^j := \max\{\Phi(S_{t_i}^j), \tilde{C}_i^j\}$.

We obtain high biased estimators if we choose the weights w such that $w_{j,k}^{t_i}$ is a deterministic function of $S_{t_i}^j$ and $S_{t_{i+1}}^k$ and w satisfies

$$\frac{1}{m} \sum_{k=1}^m E[V_{i+1}(S_{t_{i+1}}^k) \cdot w_{j,k}^{t_i} | S_{t_i}^j] = C_i(S_{t_i}^j). \quad (14)$$

In words, this means that if we knew the true value of the derivative of the next step, the expected weighted average at each node would be the true continuation value.

Once again we can show the estimator \tilde{V} is high biased under such selection of w by induction. Indeed, if we suppose that it holds for nodes at $i + 1$, then by Jensen's inequality,

$$\begin{aligned} E[\tilde{V}_i^j | S_{t_i}^j] &= E\left[\max\{\Phi(S_{t_i}^j), \tilde{C}_i^j\} | S_{t_i}^j\right] \\ &\geq \max\left\{\Phi(S_{t_i}^j), \frac{1}{m} \sum_{k=1}^m E[\tilde{V}_{i+1}^k w_{j,k}^{t_i} | S_{t_i}^j]\right\}. \end{aligned}$$

$\tilde{V}_{i+1}^k w_{j,k}^{t_i}$ is greater than $V_{i+1}(S_{t_{i+1}}^k) w_{j,k}^{t_i}$ by the induction hypothesis and using (14), we have $E[\tilde{V}_i^j | S_{t_i}^j] \geq V_i(S_{t_i}^j)$.

Go back to the choice of w . When $j \neq k$, $S_{t_{i+1}}^k$ is not a successor of node $S_{t_i}^j$ and thus the actual probability density function of $S_{t_{i+1}}^k$ should be given by g_{i+1} , the marginal density function of $S_{t_{i+1}}$. In (14),

$$E[V_{i+1}(S_{t_{i+1}}^k) w_{j,k}^{t_i} | S_{t_i}^j] = \int V_{i+1}(u) w_{j,k}^{t_i} g_{i+1}(u) du$$

and meanwhile, by the definition of C ,

$$C_i(S_{t_i}^j) = \int V_{i+1}(u) f_i(u; S_{t_i}^j) du,$$

where $f_i(\cdot; x)$ is the conditional probability density function of $S_{t_{i+1}}$ given $S_{t_i} = x$. Comparing the right hand sides of the above two equations, a natural choice of w should be $w_{j,k}^{t_i} = f_i(S_{t_{i+1}}^k; S_{t_i}^j) / g_{i+1}(S_{t_{i+1}}^k)$ to make them equal.

Notice that any stopping rule is suboptimal. Therefore we also can get a low biased estimator through the above estimated continuation values \tilde{C} . With the mesh held fixed, we simulate one additional independent path for the underlying asset price, $(S_{t_1}^{m+1}, \dots, S_{t_K}^{m+1})$, starting from S_0 . The derivative is exercised at a stopping time

$$\tilde{\tau} = \min\{t_i, 1 \leq i \leq K : \Phi(S_{t_i}^{m+1}) \geq \tilde{C}_i(S_{t_i}^{m+1})\}.$$

and the low biased estimator is defined as $\hat{v} = \Phi(S_{\tilde{\tau}}^{m+1})$. Broadie and Glasserman (2004) also showed that both \tilde{V} and \hat{v} are asymptotically unbiased as $m \rightarrow +\infty$.

Finally, we would like to mention the work of Cave, Donnelly, and Carriere (1996), Longstaff and Schwartz (2001) and Tsitsiklis and Van Roy (1999, 2001), which may be viewed as variations of stochastic mesh methods with different weight choices.

4.2 Duality

Rogers (2002) and Haugh and Kogan (2004) established a *dual method* to price American-style derivatives, which can be represented as the value of the following problem:

$$V_0(s) = \inf_{M \in \mathcal{M}} E \left[\max_{0 \leq i \leq K} (\Phi(S_{t_i}) - M_{t_i}) | S_0 = s \right], \quad (15)$$

where \mathcal{M} is the set of all martingales with initial value 0. The method produces upper bounds for the true derivative value if we specify a martingale in (15). Furthermore, we also can show that the infimum actually is attained by

$$M_{t_i}^* = \sum_{j=1}^i \{V_j(S_{t_j}) - E[V_j(S_{t_j}) | S_{t_{j-1}}]\}, \quad 0 \leq i \leq K. \quad (16)$$

The dual method is based on a well known fact in probability theory: the Doob-Meyer decomposition of supermartingales. We call a random variable sequence $\{X_i, 0 \leq i \leq K\}$ a *supermartingale* if $E[X_{i+1} | \mathcal{F}_i] \leq X_i$ for all i , where \mathcal{F} is the σ -algebra filtration of the sequence. For any supermartingale X , there exists a unique pair of a martingale N with $N_0 = 0$ and a decreasing process D (i.e., $D_i \geq D_{i+1}, a.s.$ for all i) such that $X_i = N_i + D_i, 0 \leq i \leq K$.

In particular, the form of N is given by

$$N_i = \sum_{j=1}^i \{X_j - E[X_j | \mathcal{F}_{j-1}]\}, \quad 0 \leq i \leq K.$$

From the recursion (13), we immediately conclude that $\{V_i(S_{t_i}), 0 \leq i \leq K\}$ is a supermartingale. Applying the Doob-Meyer decomposition to it, $V_i(S_{t_i})$ should be equal to a sum of the martingale M^* and a decreasing process D .

We may justify (15) as follows. For any martingale M with $M_0 = 0$ and any stopping time τ , by the optional stopping theorem,

$$\begin{aligned} E \left[\max_{0 \leq i \leq K} (\Phi(S_{t_i}) - M_{t_i}) | S_0 \right] &\geq E[\Phi(S_\tau) - M_\tau | S_0] \\ &= E[\Phi(S_\tau) | S_0], \end{aligned}$$

and thus

$$\begin{aligned} \inf_M E \left[\max_{0 \leq i \leq K} (\Phi(S_{t_i}) - M_{t_i}) | S_0 \right] &\geq \max_\tau E[\Phi(S_\tau) | S_0] \\ &= V_0(S_0). \end{aligned}$$

On the other hand, because $\Phi(S_{t_i}) \leq V_i(S_{t_i})$ and $V_i(S_{t_i}) = M_{t_i}^* + D_{t_i}$,

$$\begin{aligned} E \left[\max_{0 \leq i \leq K} (\Phi(S_{t_i}) - M_{t_i}^*) | S_0 \right] &\leq E \left[\max_{0 \leq i \leq K} (V_i(S_{t_i}) - M_{t_i}^*) | S_0 \right] \\ &= E \left[\max_{0 \leq i \leq K} D_{t_i} | S_0 \right] = E[D_0 | S_0], \end{aligned}$$

where the last equality uses that process D is decreasing. By the decomposition, we know that $D_0 = V_0(S_0) - M_0 = V_0(S_0)$.

In general, M^* can not be computed exactly because it involves the value function and its conditional expectations which we do not know. Then, how can we approximate it to get a good upper bound? A general strategy is to construct an approximate martingale from either an approximate value function or stopping policy. More precisely, if we have a set of functions $\tilde{V}_i, 0 \leq i \leq K$, which approximate the true value of the derivative at time t_i , we can construct a martingale along each simulated path $S_0, S_{t_1}, \dots, S_{t_K}$ by letting

$$\tilde{M}_{t_i} = \sum_{j=1}^i \{\tilde{V}_j(S_{t_j}) - E[\tilde{V}_j | S_{t_{j-1}}]\};$$

if we have a set of stopping times $\tilde{\tau}_i, 0 \leq i \leq K$, which satisfies $\tilde{\tau}_i \geq t_i$ for each i , we can construct a martingale

$$\tilde{M}_{t_i} = \sum_{j=1}^i \{E[\Phi(S_{\tilde{\tau}_j}) | S_{t_j}] - E[\Phi(S_{\tilde{\tau}_j}) | S_{t_{j-1}}]\}.$$

The evaluation of both the martingales requires computing conditional expectations. One can do it by nested simulation (Haugh and Kogan 2004, Andersen and Broadie 2004).

The Doob-Meyer decomposition has some variations in the literature. For instance, if the supermartingale $\{X_i, 0 \leq i \leq K\}$ is positive, it admits another form of decomposition: $X_i = B_i \cdot A_i$, where B is a positive martingale with $B_0 = 1$ and A is a decreasing process. In light of such decomposition, Jamshidian (2003) provided another dual representation of the derivative value:

$$V_0(s) = \inf_{B \in \mathcal{B}} E \left[\max_{0 \leq i \leq K} \left(\frac{\Phi(S_{t_i})}{B_{t_i}} \right) \cdot B_{t_K} | S_0 = s \right],$$

where \mathcal{B} is the set of positive martingales whose initial value is 1. Furthermore, the infimum is achieved when we choose the martingale

$$B_{t_i}^* = \prod_{j=1}^i \frac{V_j(S_{t_j})}{E[V_j(S_{t_j}) | S_{t_{j-1}}]}, \quad 0 \leq i \leq K.$$

In the following, we would like to call the former dual representation by *additive dual* and the latter one by *multiplicative dual*.

One fundamental problem thus arises: how do we compare the additive and multiplicative duals since both provide upper bounds on American-style derivative prices? To answer this problem, Chen and Glasserman (2007a) showed a strong equivalence between these two in the sense that any additive dual can be improved by a multiplicative dual and vice versa. But in terms of simulation variance, they showed that the variance of the multiplicative method typically grows much faster than that of the additive method as the number of exercise dates K increases.

5 RISK MANAGEMENT

Risk management includes identifying the sources of risks, measuring them, and controlling or hedging them. Monte Carlo simulation has been implemented widely in practice to compute risk measures of the financial portfolios. However, as pointed out by Glasserman (2003), the research on ways of improving simulation applications in risk management remains limited. In this section, we are going to present two specific problems. The first is how to use Monte Carlo simulation to estimate Greeks, and the second is how to

improve the efficiency of simulation on estimating value-at-risk (VaR).

5.1 Greeks

Greeks are price sensitivities which are to quantify the risk exposure of a financial derivative investment. Each Greek measures how much the derivative value changes in response to the change of a specific parameter of the underlying asset. For instance, four popularly used Greeks are: “Delta” — the first order sensitivity with respect to the underlying; “Gamma” — the second order sensitivity to the underlying; “Vega” — the sensitivity to the underlying volatility; and “Rho”— the sensitivity to risk free interest rate.

Greeks play a vital role in risk management. They provide guiding information on how to adjust the investment portfolio to achieve the desired exposure (e.g. Delta hedging). Whereas the price of derivatives often can be observed in the market, their sensitivities can not. Therefore, an accurate and implementable calculation of sensitivities is even more important than the calculation of prices in that sense.

As we argued in Section 2, the price of a derivative is usually in the form of an expectation of the payoff function. The calculation of Greeks then can be generalized to a problem of estimating the following differentiation:

$$\frac{d}{d\theta}\alpha(\theta) := \frac{d}{d\theta}E[\Phi(S(\theta))],$$

where θ is a parameter of interest. For example, when θ is the initial underlying price, the above is the derivative’s delta; when θ is the volatility parameter, the above is the derivative’s vega.

The most naive way to do Monte Carlo simulation is by *finite difference* method. Fix a small enough ε . We simulate N independent replications $S_1(\theta), \dots, S_N(\theta)$ and N additional replications $S_1(\theta + \varepsilon), \dots, S_N(\theta + \varepsilon)$. One estimator is constructed as follows:

$$\frac{1}{\varepsilon} \left[\frac{1}{N} \sum_{i=1}^N \Phi(S_i(\theta + \varepsilon)) - \frac{1}{N} \sum_{i=1}^N \Phi(S_i(\theta)) \right]. \quad (17)$$

From the law of large numbers, both averages inside the brackets converge to $E[\Phi(S(\theta + \varepsilon))]$ and $E[\Phi(S(\theta))]$, respectively. When ε is sufficiently small, the expectation of (17) should be very closed to $\alpha'(\theta)$. But, on the other hand, the variance of the finite difference estimator usually will increase as ε decreases. Thus, we need to choose ε to balance bias and variance to reach the optimal mean square error. Related discussion can be found in Chapter 7 of Glasserman (2003) and the references therein.

The disadvantages of finite difference methods are obvious. They are biased because we use the idea of finite

difference to approximate the derivative $\alpha'(\theta)$; they are slow because we need to simulate sample paths for both parameter θ and parameter $\theta + \varepsilon$; and they also could lead to high variances.

To overcome all of these disadvantages, two categories of more sophisticated estimates are introduced in the literature. They exploit the structure about the dynamic and parameter dependence more to achieve unbiased and efficient estimators.

The first category is known as the *pathwise derivative estimates*, or more commonly *infinitesimal perturbation analysis* in the discrete event simulation literature. For the derivative price $\alpha(\theta)$, we have

$$\frac{d}{d\theta}\alpha(\theta) = E \left[\frac{d}{d\theta}\Phi(S(\theta)) \right] = E \left[\frac{d\Phi}{dS} \cdot \frac{dS(\theta)}{d\theta} \right],$$

if the interchange of the order of differentiation and expectation is valid and function Φ is smooth enough. We then end up at an unbiased estimator $d\Phi/dS \cdot dS/d\theta$ of $\alpha'(\theta)$.

The smoothness requirement on payoff function Φ turns out to be very crucial to make sure the pathwise derivative method works. One can show that one sufficient condition is that Φ is Lipschitz continuous, i.e, there exists a positive constant C such that

$$|\Phi(x) - \Phi(y)| \leq C|x - y|, \quad \text{for all } x \text{ and } y.$$

Furthermore, we have the following counterexample that the method fails if Φ is discontinuous. Suppose that the underlying asset price follows a geometric Brownian motion $dS_t/S_t = rdt + \sigma dW_t$, $S_0 = 0$ and consider the Delta of a digital option with maturity T and strike price K , whose payoff is defined as $\Phi(S_T) = 1$ if $S_T \geq K$ and $\Phi(S_T) = 0$ if $S_T < K$. It is easy to see that $d\Phi/dS = 0$, *a.e.* and therefore the pathwise derivative estimator is 0 too. But surely 0 is not unbiased to $\alpha'(S_0)$.

The second estimate, by the *likelihood ratio method*, is introduced to get rid of the dependence on the smoothness of payoff functions. Suppose that the probability density function of $S(\theta)$ is known as $g(s; \theta)$. Then, as long as we can interchange the order of differentiation and expectation,

$$\frac{d}{d\theta}\alpha(\theta) = \frac{d}{d\theta} \int \Phi(s)g(s; \theta)ds = \int \Phi(s) \frac{\partial g}{\partial \theta}(s; \theta)ds.$$

Multiplying and dividing $g(s; \theta)$ simultaneously in the integration on the right hand side,

$$\begin{aligned} \int \Phi(s) \frac{\partial g}{\partial \theta}(s; \theta)ds &= \int \Phi(s) \frac{\partial g / \partial \theta(s; \theta)}{g(s; \theta)} g(s; \theta)ds \\ &= E \left[\Phi(S(\theta)) \frac{\partial}{\partial \theta} \ln g(S(\theta), \theta) \right]. \end{aligned}$$

In other words, $\Phi(S(\theta))(\partial \ln g(S(\theta), \theta) / \partial \theta)$ is an unbiased estimator of $\alpha'(\theta)$. Notice that the estimator does not differentiate the payoff function Φ . Thus the likelihood ratio method successfully circumvents the difficulty the pathwise derivative encounters.

The bottleneck of the likelihood ratio estimates is that they involve the probability density of $S(\theta)$. Unfortunately, for many diffusion process S , the explicit form of the probability densities is not available. Fournié et al. (1999, 2001) explored how to use the *Malliavin calculus* to derive unbiased estimators for several major Greeks such as Delta, Vega and Rho without any knowledge of the density. But their method calls for heavy mathematical machinery. We refer readers of interest to Nualart (2006) for a comprehensive treatment of the Malliavin calculus and its financial applications.

Different from the Malliavin calculus method, Chen and Glasserman (2007b) provide a more elementary way to reach the same result as Fournié et al. (1999) and cast more intuitive insight. The difficulty of the likelihood ratio method is that we do not know the explicit form of the density of S . Let us consider the Euler discretization of the process (5). Given $\hat{S}_i = s_i$, the transition probability density function of \hat{S}_{i+1} is

$$f(s_{i+1}; s_i) = \frac{1}{\sqrt{2\pi b(ih, s_i)}} \exp \left[-\frac{(s_{i+1} - s_i - a(ih, s_i)h)^2}{b^2(ih, s_i)} \right].$$

By the Markov property, the joint density of $(\hat{S}_1, \dots, \hat{S}_N)$ with initial value being $\hat{S}_0 = s_0$ will be

$$\hat{g}(s_1, \dots, s_N) = \prod_{i=0}^N f(s_{i+1}; s_i). \tag{18}$$

Viewing (18) as an approximation to the density of the original process S , we are able to apply the likelihood ratio method on \hat{g} to derive Greek estimators. Meanwhile the estimator will converge weakly to what the Malliavin calculus achieves as $h \rightarrow 0$.

5.2 VaR

The first step of managing market risk is to select an appropriate measure of risk. VaR is a widely used risk measure in financial applications (Jorion 2001). Let L denote the random loss of a portfolio in a certain period of time. In the subsection, we assume that L is a continuous random variable. Then the α -VaR of L , denoted as v_α , satisfies

$$P(L > v_\alpha) = 1 - \alpha, \tag{19}$$

where α often takes values 0.95 or 0.99. Note that v_α is also the α quantile of L . If we define the large losses to be the

losses in the upper $(1 - \alpha)$ -tail of the loss distribution, the α -VaR is the lower bound of the large losses. It provides information on the potential large losses that an investor may suffer.

Suppose that we can simulate the model of L , and obtain an i.i.d. sample L_1, L_2, \dots, L_n . Then v_α can be estimated by

$$\hat{v}_\alpha^n = L_{(\lceil n\alpha \rceil)},$$

where $L_{(k)}$ is the k th order statistic of L . Serfling (1980) shows that $\hat{v}_\alpha^n \rightarrow v_\alpha$ w.p.1 and

$$\sqrt{n}(\hat{v}_\alpha^n - v_\alpha) \Rightarrow \frac{\sqrt{\alpha(1-\alpha)}}{f_L(v_\alpha)} \cdot N(0, 1) \tag{20}$$

as $n \rightarrow \infty$, where $f_L(\cdot)$ is the density of L .

Since α is often 0.95 or 0.99 in financial applications, $f_L(v_\alpha)$ is often very small. Then, by Equation (20), the variance of \hat{v}_α^n is often large. To make the simulation more efficient, we need to apply variance reduction techniques. If we can find an efficient method to estimate the loss probability $P(L > x)$ with large x , then we can invert Equation (19) to find v_α . In the rest of this subsection, we demonstrate how to use importance sampling to obtain a more efficient estimator of the loss probability. The analysis and results reported here are based on Glasserman, Heidelberger, and Shahabuddin (2000).

Let $V(S, t)$ denote the portfolio value at time t and market price vector S , so the loss over time interval Δt is

$$L = -\Delta V = V(S, t) - V(S + \Delta S, t + \Delta t)$$

where ΔS is the change in S over the interval Δt . Suppose that $\Delta S \sim N(0, \Sigma_S)$, so we may write $\Delta S = CZ$ with $CC^T = \Sigma_S$ and Z being a vector of independent standard normals. To estimate $P(L > x)$ using Monte Carlo simulation, we first generate n i.i.d. observations of Z to obtain n observations of ΔS . Then we revalue portfolio and compute the loss for these n observations. Given the n i.i.d. observations of L , denoted by L_1, L_2, \dots, L_n , the loss probability $P(L > x)$ can be estimated by $\frac{1}{n} \sum_{i=1}^n \mathbf{1}\{L_i > x\}$.

Since $\{L > x\}$ is a rare event when x is large, we consider using importance sampling (IS) to improve the efficiency of the simulation. Let P denote the original probability measure of L and let \tilde{P} denote the IS measure of L . Then

$$P(L > x) = \tilde{E} \left[\left(\frac{dP}{d\tilde{P}} \right) \mathbf{1}\{L > x\} \right],$$

where the expectation is taken under the IS measure of L . Under the IS measure, the event $\{L > x\}$ may not be a rare event. Then the IS estimator of $P(L > x)$ may be more efficient. The key question is how to select an appropriate

IS measure \tilde{P} . In this subsection, we introduce a method that is based on the delta-gamma approximation of ΔV .

By the delta-gamma approximation of ΔV ,

$$\Delta V \approx \frac{\partial V}{\partial t} \Delta t + \delta^T \Delta S + \frac{1}{2} \Delta S^T \Gamma \Delta S,$$

where

$$\delta_i = \frac{\partial V}{\partial S_i}, \quad \Gamma_{ij} = \frac{\partial^2 V}{\partial S_i \partial S_j}$$

of the portfolio at time t are deterministic, and they are often available for hedging purposes. Let

$$Q = -\frac{\partial V}{\partial t} \Delta t - \delta^T \Delta S - \frac{1}{2} \Delta S^T \Gamma \Delta S.$$

It is an approximation of $L = -\Delta V$. Let $a = -\frac{\partial V}{\partial t} \Delta t$. Note that $\Delta S = CZ$. Then

$$Q = a - (C^T \delta)^T Z - \frac{1}{2} Z^T (C^T \Gamma C) Z.$$

By Glasserman, Heidelberger, and Shahabuddin (2000), we may choose C such that $-\frac{1}{2}(C^T \Gamma C) = \Lambda$, where Λ is a diagonal matrix with diagonal elements $\lambda_1, \lambda_2, \dots, \lambda_m$. Let $b = -C^T \delta$, then

$$Q = a + b^T Z + Z^T \Lambda Z = a + \sum_{j=1}^m (b_j Z_j + \lambda_j Z_j^2). \quad (21)$$

Let

$$\psi(\theta) = \log E \left(e^{\theta Q} \right)$$

be the cumulant generating function of Q . Then by Equation (21), if $\theta < 1/2$,

$$\psi(\theta) = a\theta + \frac{1}{2} \sum_{j=1}^m \left(\frac{\theta^2 b_j^2}{1 - 2\theta \lambda_j} - \log(1 - 2\theta \lambda_j) \right).$$

If our goal is to estimate $P(Q > x)$ instead of $P(L > x)$, we can apply the exponential twisting approach to find IS measures of Q . Let P_θ be a family of probability measures that satisfies

$$\frac{dP_\theta}{dP} = e^{\theta Q - \psi(\theta)}$$

with θ being any real number at which $\psi(\theta) < \infty$. Under the new measure P_θ ,

$$P(Q > x) = E_\theta \left[e^{\theta Q - \psi(\theta)} \mathbf{1}\{Q > x\} \right].$$

To minimize the variance of the new estimator, we minimize the second moment of $e^{\theta Q - \psi(\theta)} \mathbf{1}\{Q > x\}$ under P_θ . Since

$$\begin{aligned} E_\theta \left[e^{2\theta Q - 2\psi(\theta)} \mathbf{1}\{Q > x\} \right] \\ = E \left[e^{\theta Q - \psi(\theta)} \mathbf{1}\{Q > x\} \right]^2 \leq e^{-\theta x + \psi(\theta)}, \end{aligned}$$

we may choose θ that minimizes $e^{-\theta x + \psi(\theta)}$. By some algebra, we may show that the optimal θ^* satisfies $E_{\theta^*}(Q) = x$. Therefore, under the optimal IS measure P_{θ^*} , $\{Q > x\}$ is no longer a rare event, instead x is now near the center of the distribution. Glasserman, Heidelberger, and Shahabuddin (2000) show that P_{θ^*} is asymptotically optimal as $x \rightarrow \infty$.

To implement the IS scheme, we need to know how to find $Z = (Z_1, \dots, Z_m)^T$ such that Q , calculated by Equation (21), follows the IS distribution P_{θ^*} . Glasserman, Heidelberger, and Shahabuddin (2000) show that Z_j follows a normal distribution with mean $\mu_j(\theta^*)$ and variance $\sigma_j(\theta^*)$, where

$$\mu_j(\theta^*) = \frac{\theta^* b_j}{1 - 2\lambda_j \theta^*}, \quad \sigma_j(\theta^*) = \frac{1}{1 - 2\lambda_j \theta^*}.$$

Since Q is often a good approximation of L , then the IS measure that leads to a good estimate of $P(Q > x)$ often leads to a good estimate of $P(L > x)$. Therefore, we may use P_{θ^*} as the IS measure of L . Then, we may estimate $P(L > x)$ as follows: Generate n observations of Z following the IS measure, evaluate Q from Z using Equation (21), calculate the loss $L = V(S, t) - V(S + \Delta S, t + \Delta t)$ where $\Delta S = CZ$, then estimate $P(L > x)$ by

$$\frac{1}{n} \sum_{i=1}^n e^{-\theta^* Q_i + \psi(\theta^*)} \mathbf{1}\{L_i > x\}.$$

Glasserman, Heidelberger, and Shahabuddin (2002) further extend the results to a heavy-tailed setting, where ΔS follows a multivariate t -distribution. They show that the new IS probability measure is also asymptotically optimal.

6 CONCLUSIONS

Financial engineering is a fast growing area. As the models become more and more complicated and also more and more realistic, Monte Carlo simulation often becomes the only method to evaluate the prices of derivatives and to estimate the risk measures of portfolios. There is a great need for developing simulation algorithms that are correct and efficient under these new models. In this paper we review some of the recent developments, and we believe there are still many interesting problems yet to be solved.

ACKNOWLEDGMENTS

Both of the authors acknowledge the help of the book of *Monte Carlo Methods in Financial Engineering* by Prof. Paul Glasserman. This book provides us a comprehensive overview and an excellent guideline to the whole field. The organization and several parts of this article are inspired by the book. Both authors also thank Prof. Jeremy Staum at Northwestern University and Prof. David Yao at Columbia University for their suggestions on the first draft of this article. The research of the second author was partially supported by Hong Kong Research Grants Council grant CERG 613706.

REFERENCES

- Ait-Sahalia, Y., and Per A. Mykland. 2004. Estimators of diffusions with randomly spaced discrete observation: a general theory. *Annals of Statistics* 32:2186-2222.
- Andersen, L., and M. Broadie. 2004. A primal-dual simulation algorithm for pricing multi-dimensional American options. *Management Science*, 50:1222-1234.
- Beskos, A., and G. O. Roberts. 2005. Exact simulation of diffusions. *Annals of Applied Probability* 15: 2422-2444.
- Beskos, A., O. Papaspiliopoulos, G. O. Roberts, and P. Fearnhead. 2006a. Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes (with discussion). *Journal of the Royal Statistical Society, Series B* 68: 333-383.
- Beskos, A., O. Papaspiliopoulos, G. O. Roberts. 2006b. A factorization of diffusion measure and finite sample path constructions. Working paper.
- Björk, T. 1998. *Arbitrage Pricing in Continuous Time*. Oxford University Press.
- Broadie, M., and P. Glasserman. 1996. Estimating security price derivatives using simulation. *Management Science* 42:269-285.
- Broadie, M., and P. Glasserman. 1997. Pricing American-style securities by simulation. *Journal of Economic Dynamics and Control* 21:1323-1352.
- Broadie, M., and P. Glasserman. 2004. A stochastic mesh method for pricing high-dimensional American options. *Journal of Computational Finance* 7:35-72.
- Cave, M., M. P. Donnelly and J. F. Carriere. 1996. Valuation of the early-exercise price for options using simulations and nonparametric regression. *Insurance: Mathematics and Economics* 19: 19-30.
- Chen, N., and P. Glasserman. 2007a. Additive and multiplicative duals for American option pricing. *Finance and Stochastics* 11: 153-179.
- Chen, N., and P. Glasserman. 2007b. Malliavin Greeks without Malliavin calculus. To appear in *Stochastic Processes and Their Applications*.
- DiCesare, J., and D. McLeish. 2006. Importance sampling and imputation for diffusion models. Working paper, Department of Statistics, University of Waterloo, Canada.
- Duffie, D. 2001. *Dynamic Asset Pricing Theory*, Third Edition. Princeton University Press.
- Fournié, E., J-M Lasry, J. Lebuchoux, P-L Lions, and N. Touzi. 1999. Applications of Malliavin calculus to Monte Carlo methods in finance. *Finance and Stochastics* 3: 391-412.
- Fournié, E., J-M Lasry, J. Lebuchoux, P-L Lions, and N. Touzi. 2001. Applications of Malliavin calculus to Monte Carlo methods in finance II. *Finance and Stochastics* 5: 201-236.
- Glasserman, P. 2003. *Monte Carlo Methods in Financial Engineering*. Springer-Verlag, New York.
- Glasserman, P., P. Heidelberger, and P. Shahabuddin. 2000. Variance reduction techniques for estimating value-at-risk. *Management Science*, 46: 1349-1364.
- Glasserman, P., P. Heidelberger, and P. Shahabuddin. 2002. Portfolio value-at-risk with heavy-tailed risk factors. *Mathematical Finance*, 12: 239-269.
- Haugh, M., and L. Kogan. 2004. Pricing American options: a duality approach. *Operations Research* 52: 258-270.
- Jamshidian, F. 2003. Minimax optimality of Bermudan and American claims and their Monte-Carlo upper bound approximation. Working paper.
- Kloeden, P. E., and E. Platen. 1992. *Numerical Solution of Stochastic Differential Equations*. Springer-Verlag, Berlin.
- Longstaff, F. A., and E. S. Schwartz. 2001. Valuing American options by simulation: a simple least-squares approach. *Review of Financial Studies* 14: 113-147.
- Maruyama, G. 1955. Continuous Markov processes and stochastic equations, *Rendiconti del Circolo Matematico di Palermo, Serie II* 4:48-90.
- Nualart, D. 2006. *The Malliavin Calculus and Related Topics*. Second Edition. Springer-Verlag, Berlin.
- Rogers, L. C. G. 2002. Monte Carlo valuation of American options. *Mathematical Finance* 12: 271-286.
- Staum, J. 2002. Simulation in Financial Engineering. *Proceedings of the 2002 Winter Simulation Conference*, E. Yucesan, C.-H. Chen, J. L. Snowdon, and J. M. Charnes, eds. 1482-1492.
- Talay, D. 1984. Efficient numerical schemes for the approximation of expectations of functionals of the solutions of a s.d.e., and applications. *Lecture Notes in Control and Information Sciences* 61:294-313, Springer-Verlag, Berlin.
- Tsitsiklis, J., and B. Van Roy. 1999. Optimal stopping of Markov processes: Hilbert space theory, approximation algorithms, and an application to pricing high-dimensional financial derivatives. *IEEE Transactions on Automatic Control* 44:1840-1851.

Tsitsiklis, J., and B. Van Roy. 2001. Regression methods for pricing complex American-style options. *IEEE Transactions on Neural Network* 12:694-703.

AUTHOR BIOGRAPHIES

NAN CHEN is an assistant professor in the Department of System Engineering and Engineering Management at the Chinese University of Hong Kong. He obtained his PhD degree in OR from the Department of IEOR at Columbia University in 2006. His current research interests include financial engineering, Monte Carlo simulation and applied probability. He is currently an associate editor for *Operations Research Letters*. His e-mail address is nchen@se.cuhk.edu.hk.

L. JEFF HONG is an assistant professor in industrial engineering and logistics management at the Hong Kong University of Science and Technology. His research interests include Monte-Carlo method, sensitivity analysis, simulation optimization and their applications in financial engineering. He is currently an associate editor of *Naval Research Logistics* and *ACM Transactions on Modeling and Computer Simulation*. His e-mail address is hongl@ust.hk.