Proceedings of the 2003 Winter Simulation Conference
S. Chick, P. J. Sánchez, D. Ferrin, and D. J. Morrice, eds.

EFFICIENT SIMULATIONS FOR OPTION PRICING

Jeremy Staum

Industrial Engineering & Management Sciences
Northwestern University
Evanston, IL 60208-3119, U.S.A.

ABSTRACT

This paper presents an overview of techniques for improving the efficiency of option pricing simulations, including quasi-Monte Carlo methods, variance reduction, and methods for dealing with discretization error.

1 INTRODUCTION

Simulation is a valuable tool for pricing options, as Boyle (1977) pointed out. It is easy to price most options by simulation under most models, even those that are complicated or high-dimensional. (American options are a notable exception.) Simulation tends to perform better than many other numerical techniques on high-dimensional problems, for instance, those that involve many underlying securities or their prices at many times. In particular, the rate of convergence of a Monte Carlo estimate does not depend on the dimension of the problem. Another attraction of simulation is the confidence interval that it provides for the Monte Carlo estimate. A survey of the field is Boyle, Broadie and Glasserman (1997). Recent textbooks are Glasserman (2003) and Herzog and Lord (2003).

These textbooks cover the application of simulation to financial engineering in general, including other problems such as risk management. The present paper restricts itself to option pricing, broadly construed in the sense of pricing any derivative security, for instance, mortgage-backed securities and swaps as well as the many kinds of options. It is a revised version of Staum (2002), which has more background on option pricing and extra sections on estimating Greeks and on pricing American options, but less material on variance reduction and quasi-Monte Carlo. In this paper, the focus is on the efficiency of option pricing simulations: avoiding (Section 2) or mitigating (Section 3) bias from discretization, variance reduction (Section 4), and quasi-Monte Carlo (Section 5). Section 2 begins with a brief review of how to price options by simulation.

2 HOW TO PRICE BY SIMULATION

The theory of financial engineering states that (in a complete market) pricing an option is evaluating the expectation of its discounted payoff, under a specified measure. The canonical example is the European call option under the Black-Scholes model. The European call option's payoff is $\max\{S_T - K, 0\}$, where S_T is the price of a stock at time T, and K is a prespecified amount called the strike price. Under the Black-Scholes model, the stock price follows the stochastic differential equation (SDE)

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

where W is a Wiener process (Brownian motion) under the risk-neutral probability measure $\mathbf{Q}$. Applying Itô's formula and integrating,

$$\ln S_t - \ln S_0 = (r - \sigma^2/2)t + \sigma W_t.$$

Here S_0 is the initial stock price, r is the instantaneous interest rate on a riskless money market account, and σ is the volatility. Because W_t is normally distributed with mean 0 and variance t, the terminal log stock price $\ln S_T$ is normal with mean $\ln S_0 + (r - \sigma^2/2)T$ and variance $\sigma^2 T$.

Pricing the European call option under the Black-Scholes model therefore requires the generation of one standard normal random variate per path. The simulated value of S_T on the ith path is

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

and the estimated option value is

$$\frac{1}{n}\sum_{i=1}^{n} e^{-rT} \max\left\{S_T^{(i)} - K, 0\right\}.$$

In this model, the situation is not appreciably more difficult when pricing a path-dependent option whose payoff

depends on the value of the state vector at many times. For instance, a discretely monitored Asian call option has the payoff $\max\{\bar{S}_T - K, 0\}$ where $\bar{S}_T = \sum_{k=1}^{m} S_{t_k}/m$ is the average price. Now the simulation must generate the entire path $S_{t_1}, S_{t_2}, \ldots, S_{t_m}$. Assume $t_k = Tk/m = kh$. The way to simulate the whole path is to generate m independent standard normal random variables $Z_1^{(i)}, \ldots, Z_m^{(i)}$ for the ith path and set

$$S_{(k+1)h}^{(i)} = S_{kh}^{(i)} \exp\left(\left(r - \sigma^2/2\right)h + \sigma\sqrt{h}Z_k^{(i)}\right).$$

This provides the correct multivariate distribution for $(S_{t_1}, \ldots, S_{t_m})$ and hence the correct distribution for $\bar{S}_T$.

Another challenge in path generation is continuous path-dependence. While the payoff of the European call option depends only on the terminal value of the state vector, and the payoff of the discretely monitored Asian call option depends only on a finite set of observations of the state vector, some derivatives have payoffs that depend on the entire continuous-time path. An example is a down-and-in option that pays off only if a stock price goes below some barrier, or equivalently, if the minimum stock price is below the barrier. Suppose the stock price obeys the Black-Scholes model. Because

$$\min_{k=1,\ldots,m} S_{t_k} > \min_{t\in[0,T]} S_t$$

almost surely, the former is not an acceptable substitute for the latter. It is necessary to introduce a new component $M_t = \min_{u\in[0,t]} S_u$ into the state vector; this can be simulated since the joint distribution of S_t and M_t is known (Karatzas and Shreve 1991).

A slightly subtler example occurs in the Hull-White model of stochastic interest rates. The SDE governing the instantaneous interest rate r_t is

$$dr_t = \alpha(\bar{r} - r_t)dt + \sigma\, dW_t$$

where $\bar{r}$ is the long-term mean interest rate, α is the strength of mean reversion, and σ is the interest rate's volatility. Integration of this SDE yields the distribution of r_t, which is normal. Then the simulated path $r_{t_1}, \ldots, r_{t_m}$ is adequate for evaluating payoffs that depend only on these interest rates, but not for evaluating the discount factor $D_T = \int_0^T r_u\, du$; the discrete approximation $h\sum_{k=1}^{m} r_{kh}$ does not have the right distribution. Instead one must add D_t to the state vector and simulate using its joint distribution with r_t, which is easily computable.

3 DISCRETIZATION ERROR

Some financial models feature SDEs that are not easily integrable, as the Black-Scholes and Hull-White models' are. An example is the Cox-Ingersoll-Ross model, in which the SDE is

$$dr_t = \alpha(\bar{r} - r_t)dt + \sigma\sqrt{r_t}\, dW_t.$$

This model's principal advantage over Hull-White is that the instantaneous interest rate must remain nonnegative. However, there is no useful expression for the distribution of r_t given r_0. A simulation of this model must rely on an approximate discretization $\hat{r}$ of the stochastic process r. Because the laws of these processes are not the same, the Monte Carlo estimate based on $\hat{r}$ may be biased for the true price based on r. This bias is known as discretization error.

Kloeden and Platen (1992) have written a major reference on the rather involved topic of discretizing SDEs, whose surface this paper barely scratches. Faced with an SDE of the generic form

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t$$

one simulates a discretized process $\hat{X}_{t_1}, \ldots, \hat{X}_{t_m}$. Even if the only quantity of interest is the terminal value X_T, it is necessary to simulate intermediate steps in order to reduce discretization error. The question is how to choose the scheme for producing the discretized process $\hat{X}$ and the number of steps m.

The most obvious method of discretizing is the Euler scheme

$$\hat{X}_{(k+1)h} = \hat{X}_{kh} + \mu\left(\hat{X}_{kh}\right)h + \sigma\left(\hat{X}_{kh}\right)\sqrt{h}Z_{k+1}$$

where $Z_1, \ldots, Z_m$ are independent standard normal random variates. The idea is simply to pretend that the drift μ and volatility σ of X remain constant over the period $[kh, (k+1)h]$ even though X itself changes. Is there a better scheme than this, and what would it mean for one discretization scheme to be better than another?

There are two types of criteria for judging discretized processes. Strong criteria evaluate the difference between the paths of the discretized and original processes produced on the same element ω of the probability space. For example, the strong criterion $\mathbf{E}[\max_k \|\hat{X}_{t_k} - X_{t_k}\|]$ measures the maximum discrepancy between the path $\hat{X}(\omega)$ and the path $X(\omega)$ over all times, then weights the elements ω with the probability measure $\mathbf{P}$. On the other hand, weak criteria evaluate the difference between the laws of the discretized and original processes: an example is $\sup_x |\mathbf{P}[\hat{X}_T < x] - \mathbf{P}[X_T < x]|$, measuring the maximum discrepancy between the cumulative distribution functions of the terminal values of $\hat{X}$ and X. Weak criteria are of greater interest in derivative pricing because the bias of the Monte Carlo estimator $f(\hat{X}_{t_1}, \ldots, \hat{X}_{t_m})$ of the true price

$\mathbf{E}[f(X_{t_1},\ldots,X_{t_m})]$, where f is the payoff, depends only on the distribution of $(\hat{X}_{t_1},\ldots,\hat{X}_{t_m})$.

Given a choice of weak criterion, a discretization scheme has weak order of convergence γ if the error is of order $m^{-\gamma}$ as the number of steps m goes to infinity. Under some technical conditions on the stochastic process X and the exact nature of the weak criterion, the weak order of the Euler scheme is 1, and a scheme with weak order 2 is

$$\begin{aligned}\hat{X}_{(k+1)h} &= \hat{X}_{kh} + \sigma Z_{k+1} h^{1/2} \\ &\quad + \left(\mu + \frac{1}{2}\sigma\sigma'\left(Z_{k+1}^2 - 1\right)\right)h \\ &\quad + \frac{1}{2}\left(\mu'\sigma + \mu\sigma' + \frac{1}{2}\sigma^2\sigma''\right) Z_{k+1} h^{3/2} \\ &\quad + \frac{1}{2}\left(\mu\mu' + \frac{1}{2}\mu''\sigma^2\right)h^2\end{aligned}$$

where μ, σ, and their derivatives are evaluated at $\hat{X}_{kh}$. This is known as the Milstein scheme, but so are some other schemes. This scheme comes from the expansion of the integral $\int_{kh}^{(k+1)h} dX_t$ to second order in h using the rules of stochastic calculus.

The weak order of convergence remains the same if simple random variables with appropriate moments replace the standard normal random variables Z. Not only can such a substitution improve speed, but it may be necessary when the SDE involves multivariate Brownian motion, whose multiple integrals are too difficult to simulate.

It is also possible to use Richardson extrapolation in order to improve an estimate's order of convergence. For instance, let $f(\hat{X}^{(h)})$ denote the payoff simulated under the Euler scheme with step size h. The Euler scheme has weak order of convergence 1, so the leading term in the bias $\mathbf{E}[f(\hat{X}^{(h)})] - \mathbf{E}[f(X)]$ is of order h. The next term turns out to be of order h^2. Because the order h terms cancel, the bias of $2\mathbf{E}[f(\hat{X}^{(h)})] - \mathbf{E}[f(\hat{X}^{(2h)})]$ is of order h^2, and this extrapolated Euler estimate has weak order of convergence 2.

Turning to the choice of the number of steps m, one consideration is allocating computational resources between a finer discretization and a greater number of paths (Duffie and Glynn 1995). If there is a fixed computational budget C, and each simulation step costs c, then the number of paths must be $n = C/(mc)$. For a discretization scheme of weak order γ, the bias is approximately $bm^{-\gamma}$ for some constant b. Estimator variance is approximately vn^{-1} for some constant v. Therefore the mean squared error is approximately

$$b^2 m^{-2\gamma} + vn^{-1} = b^2 m^{-2\gamma} + \frac{vc}{C}m$$

which is minimized by $m \propto C^{1/(2\gamma+1)}$. With this optimal allocation, the mean squared error is proportional to $C^{-2\gamma/(2\gamma+1)}$, which is slower than the rate C^{-1} of decrease of the variance of a simulation unbiased by discretization errror. A higher order of convergence γ is associated with a coarser discretization (m smaller) and more rapid diminution of mean squared error with increased computational budget C.

4 VARIANCE REDUCTION

The standard error of a Monte Carlo estimate decreases as $1/\sqrt{C}$, where C is the computational budget. This is not an impressive rate of convergence for a numerical integration method. For simulation to be competitive for some problems, it is necessary to design an estimator that has less variance than the most obvious one. A variance reduction technique is a strategy for producing from one Monte Carlo estimator another with lower variance given the same computational budget.

A fixed computational budget is not the same as a fixed number of paths. Variance reduction techniques frequently call for more complicated estimators that involve more work per path. Where W is the expected amount of work per path, the computational budget C allows approximately $n = C/W$ paths. There is a variance per path V such that the estimator variance is approximately $V/n = VW/C$. Thus a technique achieves efficiency improvement (variance reduction given a fixed budget) if it reduces VW.

In practice, one may be concerned with human effort as well as computer time. Computing power has become so cheap that for many individual financial simulations, it is not worth anybody's time to implement variance reduction. On the other hand, some financial engineering problems are so large that variance reduction is important.

For example, it is too time-consuming to compute firmwide value at risk (VaR) for a large financial institution by simulating many future scenarios and pricing all the firm's derivatives by simulation in each scenario, so financial institutions rely on methodologies of questionable soundness for computing VaR. Lee (1998) investigates one question of efficiency for such large VaR simulations. Here variance reduction may make better answers affordable.

Another example is model calibrations that involve simulation of options' prices to compute the objective of an optimization. This takes a long time because simulations must be done at every iteration of the optimization routine. In this case, variance reduction makes possible greater responsiveness to changing market conditions.

4.1 Antithetic Variates

Because of its simplicity, the method of antithetic variates is a good introduction to variance reduction techniques,

among which it is not one of the most powerful. A quantity simulated on one path, such as a payoff, always has a representation $f(U)$ where U is uniformly distributed on $[0,1]^m$. The antithetic variate of U is $1-U = (1-U_1,\ldots,1-U_m)$. The method uses as an estimate from a pair of antithetic variates $(f(U)+f(1-U))/2$, which can be called the symmetric part of f. This is unbiased because $1-U$ is also uniformly distributed on $[0,1]^m$.

The antisymmetric part of f is $(f(U)-f(1-U))/2$. These two parts are uncorrelated and sum to $f(U)$, so the variance of $f(U)$ is the sum of the variances of the symmetric and antisymmetric parts. The estimator using antithetic variates has only the variance of the symmetric part of f, and requires at most twice as much work as the old. The variance of the antisymmetric part is eliminated, and if it is more than half the total variance of f, efficiency improves. This is true, for instance, when f is monotone, as it is in the case of the European call option in the Black-Scholes model.

4.2 Stratification and the Latin Hypercube

Stratification makes simulation more like numerical integration by insisting on a certain regularity of the distribution of simulated paths. This technique divides the sample space into strata and makes the fraction of simulated paths in each stratum equal to its probability in the model being simulated. Working with the representation $f(U_1,\ldots,U_m)$, one choice is to divide the sample space of U_1 into N equiprobable strata $[0,1/N],\ldots,[(N-1)/N,1]$. Then the stratified estimator is

$$\frac{1}{N}\sum_{i=1}^{N} f\left(\frac{i-1+U_1^{(i)}}{N}, U_2^{(i)},\ldots,U_m^{(i)}\right)$$

where the random variables $U_k^{(i)}$ are i.i.d. uniform on $[0,1]$. This estimator involves N paths, whose first components are chosen randomly within a predetermined stratum. Because these N paths are dependent, to get a confidence interval requires enough independent replications of this stratified estimator sufficient to make their mean approximately normally distributed.

Stratification applies in the quite general situation of sampling from a distribution that has a representation as a mixture: above, the uniform distribution on $[0,1]$ is an equiprobable mixture of N uniform distributions on intervals of size $1/N$. The general case is sampling from a distribution that is a mixture of N distributions, the ith of which has mixing probability p_i, mean μ_i, and variance σ_i^2. The mixed distribution has mean $\sum_{i=1}^N p_i\mu_i$ and variance

$$\sum_{i=1}^{N} p_i\left(\mu_i^2+\sigma_i^2\right)-\left(\sum_{i=1}^{N} p_i\mu_i\right)^2.$$

A stratified estimate has variance $\sum_{i=1}^N p_i\sigma_i^2$. The amount of variance reduction is the difference

$$\sum_{i=1}^{N} p_i\mu_i^2-\left(\sum_{i=1}^{N} p_i\mu_i\right)^2$$

which is the variance of μ_η, where η is a random variable taking on the value i with probability p_i. That is, stratification removes the variance of the conditional expectation of the outcome given the information being stratified.

This approach can be very effective when the payoff depends heavily on a single random variable, and it is possible to sample the rest of the path conditional on this random variable. For instance, if the payoff depends primarily on a terminal stock price S_T whose process S is closely linked to a Brownian motion W, then a good strategy is to stratify on W_T and simulate $W_{t_1},\ldots,W_{t_{m-1}}$ conditional on it.

Stratification in many dimensions at once poses a difficulty. Using N strata for each of d random variables results in a mixture of N^d distributions, each of which must be sampled many times if there is to be a confidence interval. If d is too large there may be no way to do this without exceeding the computational budget. Latin hypercube sampling offers a way out of this quandary.

Consider the stratification of each dimension of $[0,1]^m$ into N intervals of equal length. A Latin hypercube sample includes a point in only N of the N^d boxes formed. This sample has the property that it is stratified in each dimension separately, that is, for each stratum j and dimension k, there is exactly one point $U^{(i)}$ such that $U_k^{(i)}$ is in $[(j-1)/N,\ j/N]$. The Latin hypercube sampling algorithm illustrates:

Loop over dimension $k=1,\ldots,m$.

- Produce a permutation J of $1,\ldots,N$.
- Loop over point $i=1,\ldots,N$.
  - Choose $U_k^{(i)}$ uniformly in $[(J_i-1)/N,\ J_i/N]$.

Because points are uniformly distributed within their boxes, the marginal distributions are correct. Choosing all permutations with equal probability makes the joint distribution correct.

Because it is not full stratification, Latin hypercube sampling does not remove all the variance of the conditional expectation given the box. Writing this conditional expectation as a function $\mu(j_1,\ldots,j_m)$ where j_k is the stratum in the kth dimension, Latin hypercube sampling asymptotically removes only the variance of the additive part of this function. The additive part is the function $g(j_1,\ldots,j_m)=\sum_{k=1}^m g_k(j_k)$ that minimizes the expected squared error of its fit to the original function μ. Sometimes the fit is quite good, for instance when pricing a relatively short-term interest-rate swap in the Hull-White model. In each of a sequence of periods, the swap pays the difference

between preset interest payments and the then-prevailing interest payments. These terms are linear in the normal random variates $Z_1, \ldots, Z_m$, but for pricing must also be multiplied by nonlinear discount factors.

4.3 Importance Sampling

The intuitive way to plan a simulation to estimate the expectation of a payoff f that depends on a path $X_1, \ldots, X_m$ is to simulate paths according to the law of the process X, then compute the payoff on each path. This is a way of estimating the integral

$$\int f(x)g(x)dx = \int \left(\frac{fg}{\tilde{g}}\right)(x)\tilde{g}(x)dx$$

as long as $\tilde{g}$ is nonzero where fg is. The second integral has an interpretation as simulation of paths under a new probability measure $\tilde{\mathbf{Q}}$. The path $X_1, \ldots, X_m$ has likelihood g under $\mathbf{Q}$ and $\tilde{g}$ under $\tilde{\mathbf{Q}}$. There is also a new payoff $\tilde{f} = fg/\tilde{g}$, the product of the original payoff f and the Radon-Nikodym derivative or likelihood ratio $g/\tilde{g}$. One way in which importance sampling can arise naturally in the financial context is when $\mathbf{Q}$ and $\tilde{\mathbf{Q}}$ are both martingale measures, in which case the Radon-Nikodym derivative is the ratio of the associated numeraires' terminal values.

The idea of importance sampling is to choose $\tilde{g}$ so that $\tilde{f}$ has less variance under $\tilde{\mathbf{Q}}$ than f does under $\mathbf{Q}$. When f is positive, the extreme choice is $\tilde{g} = fg/\mu$, where μ is the constant of integration that makes $\tilde{g}$ a probability density. Then $\tilde{f} = \mu$ and has no variance. However, this constant μ is precisely $\int f(x)g(x)dx$, the unknown quantity to be estimated. The goal is to choose $\tilde{g}$ to be a tractable density that is close to being proportional to fg. That is, one wishes to sample states x according to importance, the product of likelihood and payoff.

It is possible for importance sampling to go awry, as the following example demonstrates. Suppose $f(x) = x$ and

$$g(x) = \begin{cases} e^{-x} & \text{if } x \in [0, K] \\ \alpha x^{-4} & \text{if } x > K \end{cases}$$

where K is very large. The simulation estimates the mean of a random variable whose distribution is almost exponential, but has a power tail. The mean and variance are both finite. Suppose $\tilde{g}(x)$ is simply e^{-x} for all $x \geq 0$. As x goes to infinity, so does the likelihood ratio $g/\tilde{g}$. The new simulation variance is infinite: the new second moment is

$$\int_0^\infty \left(\frac{xg(x)}{\tilde{g}(x)}\right)^2 \tilde{g}(x)\,dx > \alpha^2 \int_K^\infty x^{-6}e^x\,dx = \infty.$$

Moreover, we are likely not to simulate any $x \gg K$, which has a large likelihood ratio, in which case the sample standard deviation will not alert us to the failure of the scheme.

The potential for mistakes aside, importance sampling has proven extremely powerful in other applications, especially in simulation of rare events, which are more common under an appropriate importance sampling measure. There have been some effective financial engineering applications in this spirit, involving the pricing of derivatives that are likely to have zero payoff. An example is an option that is deep out of the money, meaning that the underlying is currently distant from a threshold that it must cross in order to produce a positive payoff.

Importance sampling may become even more valuable in financial engineering with the advent of more sophisticated approaches to risk management. There is an increasing appreciation of the significance for risk management of extreme value theory and the heavy-tailed distributions of many financial variables. In models and applications where behavior in the tails of distributions has greater impact, importance sampling has greater potential. An example of such developments is the work of Glasserman, Heidelberger, and Shahabuddin (2002).

4.4 Control Variates

Unlike other methods that adjust the inputs to simulation, the method of control variates adjusts the outputs directly. A simulation intended to estimate an unknown integral can also produce estimates of quantities for which there are known formulas. The known errors of these estimates contain information about the unknown error of the estimate of the quantity of interest, and thus are of use in correcting it. For instance, using the risk-neutral measure, the initial stock price $S_0 = \mathbf{E}^{\mathbf{Q}}[e^{-rT}S_T]$, but the sample average $e^{-rT}\sum_{i=1}^n S_T^{(i)}/n$ will differ from S_0. If it is too large, and the payoff $f(S_T)$ has a positive correlation with S_T, then the estimate of the security price is probably also too large.

Generally, in a simulation to estimate the scalar $\mathbf{E}[X]$ which also generates a vector Y such that $\mathbf{E}[Y]$ is known, an improved estimator is $X - \beta(Y - \mathbf{E}[Y])$ where β is the multiple regression coefficient of X on Y. The variance of this estimator is the residual variance of X after regression on Y; the better the linear fit of X on the predictors Y, the less variance remains after the application of control variates. The regression coefficient β is presumably unknown if $\mathbf{E}[X]$ is unknown, but the usual least squares estimate will suffice. However, using the same paths to estimate β and evaluate the control variates estimator creates a slight bias. An alternative is to estimate β from some paths reserved for that purpose alone.

A favorite example of the great potential of control variates is the discretely monitored Asian call option in the Black-Scholes model, which appeared in Section 2. Averaging, as in the average stock price $\bar{S}_T$, is the distinguishing feature of Asian options. For economic reasons, the convention is that the averaging is arithmetic, not geometric.

For instance, an Asian option on oil futures could help a power company hedge the average cost of its planned future purchases of oil, while an option on a geometric average of prices does not have such an obvious purpose. On the other hand, the distribution of the arithmetic average of jointly lognormal random variables (such as $S_{t_1}, \ldots, S_{t_m}$) is inconvenient, while the distribution of their geometric average is again lognormal, so a geometric Asian option has a closed-form price in the Black-Scholes model. The payoffs of arithmetic and geometric Asian call options are extremely highly correlated, and therefore the geometric Asian call option makes a very effective control variate for simulation of the arithmetic Asian call option: it can reduce variance by a factor of as much as one hundred. Using this control variate, the simulation is effectively estimating only the slight difference between the arithmetic and geometric Asian options.

4.5 Repricing, Matching, and Weights

As an example of a control variate, we used a stock price, which is a known expectation differing from the corresponding simulated average. Some practitioners react to such differences with dismay: when the simulation reprices market securities such as a stock incorrectly, the policy of trading at simulated prices results in arbitrage! Of course, one does not trade market securities on the basis of simulated prices, nor does one trade over-the-counter derivatives at exactly the simulated price. Rather, one establishes a bid-ask spread, accounting for model risk and profit margin. Nonetheless, the fear that errors in repricing market securities indicate arbitrage in the simulated derivative security prices may remain, leading to corrective techniques that are closely related to control variates.

Continuing with the example of a single stock, one approach is simply to change the simulated values of S_T until their sample average is indeed $e^{rT} S_0$, then computing the simulated derivative payoffs from these adjusted simulated terminal stock prices. One way to do this is to multiply S_T by $e^{rT} S_0 n / \sum_{i=1}^{n} S_T^{(i)}$. This is essentially taking the control variates concept and using it to adjust values inside the simulation, rather than to adjust the output directly. A related idea is to adjust the inputs to the simulation, the random variates. For instance, one might insist that the standard normal random variates used in the simulation have sample mean 0 and sample standard deviation 1. Affine transformation of the random variates can accomplish this.

Although such affine transformation is reminiscent of control variates, these techniques are not necessarily equivalent, because the transformation takes place at different stages in the simulation. However, like control variates, these techniques create bias. Their relative advantages vary from problem to problem.

Yet another alternative is to make the simulation estimator an unequally weighted average of the sample paths. The weights are typically chosen to minimize some measure of nonuniformity while satisfying a constraint. For example, the usual control variates estimator turns out to be of this form, where the constraint is that the control variate's sample average must equal the known mean, and the objective is the sum of squared weights. Another example is Avellaneda's (1998) use of a relative entropy criterion with the constraint that market securities' average discounted payoff must equal their market prices. This is often viewed not so much as an efficiency technique, but a corrective to the part of model risk that arises when a calibrated model does not reprice market securities exactly. For more on weighted Monte Carlo, see Glasserman and Yu (2003).

4.6 Conditional Monte Carlo

Another variance reduction technique is conditional Monte Carlo. By substituting conditional expectations when they are known, it often reduces both the work and variance per path. In derivative security pricing, this can be the simulation of the future value of the security, rather than of its payoff.

For example, the down-and-in option mentioned in Section 2 pays the same as a standard option if the underlying goes below a specified barrier, and if not, it pays nothing. Suppose there is a formula f for the standard option price. Then one may simulate the underlying path until maturity T or until the first time τ that the barrier is crossed, whichever comes first. Then the estimated option value is

$$\frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\{\tau^{(i)} \leq T\} D_{\tau^{(i)}}^{(i)} f\left(S_{\tau^{(i)}}^{(i)}\right)$$

where $\mathbf{1}$ is the indicator function. This eliminates the conditional variance of the standard option's payoff and reduces the expected number of steps per path from T to $\mathbf{E}[\tau]$.

This approach also handles knock-out options through in-out parity, and applies fairly directly to other derivatives such as forward-starting options. In a different way, conditional Monte Carlo has also been applied to stochastic volatility models in which the option price is known conditional on the volatility path.

4.7 Work Reduction

While conditional Monte Carlo should reduce not only work but also variance, as the name "variance reduction" suggests, there are methods that reduce work but not variance, or even increase variance. These might be called "work reduction" techniques. Just as a variance reduction technique that

reduces V (variance per path) while increasing W (work per path) enhances efficiency if it reduces the product VW, so an unbiased work reduction technique enhances efficiency if it reduces VW by decreasing W more than it increases V. This is reducing the simulation variance given a fixed computational budget. A work reduction technique that introduces bias can still enhance efficiency in the sense of reducing mean squared error.

One example is early stopping of some simulated paths, which can enhance efficiency if the beginning of a path contains more useful information than the end of a path. It can make sense to allocate more of the simulation resources to the steps of the path that explain more of the variance in the simulation estimator. This can be done without bias even when the decision to stop is dependent on the simulated state. See Glasserman and Staum (2002) and references therein.

A more prosaic way to reduce work, important in practice, is to code simulation programs efficiently. In part, this means simply refraining from unnecessary computation and memory access, which can be surprisingly easy to fall into. In part, this can involve more interesting techniques such as fast algorithms for numerical function evaluation and financial approximations that impart slight bias. See Staum, Ehrlichman, and Lesnevski (2003) for examples.

4.8 Summary

The methods discussed above illustrate two major types of variance reduction. Importance sampling and control variates rely on knowledge about the structure of the problem to change the payoff or sampling distribution. Stratified and Latin hypercube sampling also benefit from a good choice of the variables to stratify. However, these methods and antithetic variates work by making Monte Carlo simulation less purely random and more like other numerical integration techniques that use regular, not random, distributions of points. Similarly, quasi-Monte Carlo simulation is a numerical integration technique that bears a resemblance to Monte Carlo, although its foundations are deterministic.

5 QUASI-MONTE CARLO

A sample from the multidimensional uniform distribution usually covers the unit hypercube inefficiently: to the eye it seems that there are clusters of sample points and voids bare of sample points. A rectangular grid of points looks more attractive, but the bound on the error of this numerical integration technique converges as $n^{-2/d}$ where n is the number of points used and d is the dimension of the hypercube. For dimension four or higher, there is no advantage compared to the order $n^{-1/2}$ convergence of the standard error of a Monte Carlo (MC) simulation. The quasi-Monte Carlo (QMC) approach, often used in financial engineering, is to generate a deterministic set of points that fills space efficiently without being unmanageably numerous in high dimension. Several authors have proposed rules for generating such sets, known as low-discrepancy sequences: see Niederreiter (1992). The name "quasi-Monte Carlo" does not indicate that these sequences are somewhat random, but rather that they look random; indeed they look more random than actual random sequences, because the human mind is predisposed to see patterns that are statistically insignificant.

The great attraction of low-discrepancy sequences is that they produce an error of integration whose bound converges as $(\log n)^d/n$, a better asymptotic rate than $n^{-1/2}$. As this result suggests, QMC is often much more efficient than MC, at least if d is not too large. If dimension d is too large relative to sample size n, two things can go wrong. First, the regularity of popular low-discrepancy sequences is such that, while coordinates 1 and 2 of points $1, \ldots, n$ in a low-discrepancy sequence may cover the unit square evenly, coordinates $d-1$ and d of these n points may cover it very badly, causing potentially large error. See, for instance, Figure 2 of Imai and Tan (2002). Second, if $(\log n)^d/n > n^{-1/2}$, it suggests that MC may be more accurate than QMC.

However, QMC is often more accurate than MC even when the dimension d is large and the sample size n is not. An explanation for this surprise is the low effective dimension of many high-dimensional financial simulation problems. Roughly, effective dimension means the number of dimensions required to explain, in the sense of analysis of variance, a large proportion of the entire variance of the integrand. For precise definitions and distinctions, see Caflisch, Morokoff, and Owen (1997). Owen (2002) demonstrates that low effective dimension is necessary for scrambled $(0, m, d)$-nets, a type of low-discrepancy sequence, to beat MC; it is an open question whether it is necessary for all QMC methods.

Such observations lead to contemplation of effective dimension reduction. If one can change the simulation scheme so that the integrand has the same integral on the unit hypercube but a lower effective dimension, then QMC may be more effective. For example, some such transformations use Brownian bridge or principal components as the basis for producing a sample path, which would ordinarily proceed by using one random variate at each time step in turn. Imai and Tan (2002) review and extend efforts in this area.

Another promising development is randomized quasi-Monte Carlo (RQMC), which randomizes a low-discrepancy sequence so that it gains desirable statistical properties while retaining its regularity properties. An RQMC algorithm produces dependent random vectors $U^{(1)}, \ldots, U^{(n)}$ each uniformly distributed on $[0,1]^m$. This makes RQMC much like MC with a variance reduction technique: the uniformity of each $U^{(i)}$ means that the estimator is unbiased, while

dependence suitable for the problem provides reduced variance. An example is the random shift. Taking $\tilde{U}^{(i)}$ from a low-discrepancy sequence and Δ uniformly distributed on $[0,1]^m$, $U^{(i)} = (\tilde{U}^{(i)} + \Delta) \mod 1$ is also uniformly distributed on $[0,1]^m$, but retains the original spacing. From repeated random draws of the shift Δ, a confidence interval is available. As with importance sampling, there is the potential for great improvement in efficiency, but a mistake can lead to increased variance. For further information, see the survey of L'Ecuyer and Lemieux (2002).

Financial engineering has proved to be a domain that is quite favorable for QMC. The combination of QMC with variance reduction techniques can be particularly powerful. For an overview of QMC methods for financial computations and further references, see Lemieux and L'Ecuyer (2001).

6 CONCLUSIONS

Many general simulation efficiency techniques apply to option pricing. However, because many of these general techniques require problem-specific knowledge to be applied to best advantage, much research has gone into their application in the financial context. The knowledgeable practitioner can use these ideas to achieve high-quality estimates despite constraints on time and computing power. This process is freeing financial engineers from a dependence on closed-form solutions and tractable but unrealistic models to simulate more realistic models, leading to better answers.

ACKNOWLEDGMENTS

This paper owes much to cited sources, especially Boyle, Broadie, and Glasserman (1997) and Glasserman (2003). The author thanks Paul Glasserman, Shane Henderson, and Pierre L'Ecuyer for comments and discussions. The views expressed are those of the author, who is solely responsible for any errors.

REFERENCES

Avellaneda, M. 1998. Minimum-Relative-Entropy Calibration of Asset-Pricing Models. *International Journal of Theoretical and Applied Finance* 1: 447–472.

Boyle, P. 1977. Options: A Monte Carlo Approach. *Journal of Financial Economics* 4: 323–338.

Boyle, P., M. Broadie, and P. Glasserman. 1997. Monte Carlo Methods for Security Pricing. *Journal of Economic Dynamics and Control* 21: 1267–1321.

Caflisch, R. E., W. Morokoff, and A. Owen. 1997. Valuation of Mortgage Backed Securities Using Brownian Bridges to Reduce Effective Dimension. *Journal of Computational Finance* 1: 27–46.

Duffie, D., and P. Glynn. 1995. Efficient Monte Carlo Simulation of Security Prices. *Annals of Applied Probability* 5: 897–905.

Glasserman, P. 2003. *Monte Carlo Methods in Financial Engineering*. New York: Springer-Verlag.

Glasserman, P., P. Heidelberger, and P. Shahabuddin. 2002. Portfolio Value-at-Risk with Heavy-Tailed Risk Factors. *Mathematical Finance* 12: 239–270.

Glasserman, P., and J. Staum. 2001. Stopping Simulated Paths Early. In *Proceedings of the 2001 Winter Simulation Conference*, ed. B. A. Peters, J. S. Smith, D. J. Medeiros, and M. W. Rohrer, 318–324. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers. Available online via <http://www.informs-cs.org/wsc01papers/040.PDF>.

Glasserman, P., and B. Yu. 2003. Large Sample Properties of Weighted Monte Carlo Estimators. Working paper, Columbia Business School. Available online via <http://www.paulglasserman.com>.

Herzog, T. N., and G. Lord. 2002. *Applications of Monte Carlo Methods to Finance and Insurance.* Winstead, Conn.: ACTEX Publications.

Imai, J., and K. S. Tan. 2002. Enhanced Quasi-Monte Carlo Methods with Dimension Reduction. In *Proceedings of the 2002 Winter Simulation Conference*, ed. E. Yücesan, C.-H. Chen, J. L. Snowdon, and J. M. Charnes, 1502–1510. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers. Available online via <http://www.informs-cs.org/wsc02papers/205.PDF>.

Karatzas, I., and S. E. Shreve. 1991. *Brownian Motion and Stochastic Calculus.* 2nd ed. New York: Springer-Verlag.

Kloeden, P. E., and E. Platen. 1992. *Numerical Solution of Stochastic Differential Equations.* New York: Springer-Verlag.

L'Ecuyer, P., and C. Lemieux. 2002. Recent Advances in Randomized Quasi-Monte Carlo Methods. In *Modeling Uncertainty: An Examination of Stochastic Theory, Methods, and Applications*, ed. M. Dror, P. L'Ecuyer, and F. Szidarovszki, 419–474. New York: Kluwer Academic Publishers. Available online via <http://www.iro.umontreal.ca/~lecuyer/papers.html>.

Lee, S.-H., 1998. *Monte Carlo Computation of Conditional Expectation Quantiles*. Doctoral dissertation, Stanford University.

Lemieux, C., and P. L'Ecuyer. 2001. On the Use of Quasi-Monte Carlo Methods in Computational Finance. In *Computational Science–ICCS 2001*, 607–616. New York: Springer-Verlag. Available online via <http://www.iro.umontreal.ca/~lecuyer/papers.html>.

Niederreiter, H. 1992. *Random Number Generation and Quasi-Monte Carlo Methods.* Philadelphia: Society for Industrial and Applied Mathematics.

Owen, A. B. 2002. Necessity of Low Effective Dimension. Working paper, Stanford University. Available online via <http://www-stat.stanford.edu/~owen/reports/>.

Staum, J. 2002. Simulation in Financial Engineering. In *Proceedings of the 2002 Winter Simulation Conference*, ed. E. Yücesan, C.-H. Chen, J. L. Snowdon, and J. M. Charnes, 1481–1492. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers. Available online via <http://www.informs-cs.org/wsc02papers/203.PDF>.

Staum, J., S. Ehrlichman, and V. Lesnevski. 2003. Work Reduction in Financial Simulations. In *Proceedings of the 2003 Winter Simulation Conference*, ed. S. Chick, P. J. Sánchez, D. Ferrin, and D. J. Morrice. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.

AUTHOR BIOGRAPHY

JEREMY STAUM is Assistant Professor in the Department of Industrial Engineering and Management Sciences at Northwestern University. He received his Ph. D. in Management Science from Columbia University in 2001. His research interests include variance reduction techniques and financial engineering. His e-mail address is <staum@iems.northwestern.edu>, and his web page is <www.iems.northwestern.edu/~staum>.