*Proceedings of the 2001 Winter Simulation Conference*
*B. A. Peters, J. S. Smith, D. J. Medeiros, and M. W. Rohrer, eds.*

# SIMULATION IN FINANCIAL ENGINEERING

Jeremy Staum

School of Operations Research and Industrial Engineering
Cornell University
Ithaca, NY 14853, U.S.A.

## ABSTRACT

This paper presents an overview of the use of simulation algorithms in the field of financial engineering, assuming on the part of the reader no familiarity with finance and a modest familiarity with simulation methodology, but not its specialist research literature. The focus is on the challenges specific to financial simulations and the approaches that researchers have developed to handle them, although the paper does not constitute a comprehensive survey of the research literature. It offers to simulation researchers, professionals, and students an introduction to an application of increasing significance both within the simulation research community and among financial engineering practitioners.

## 1 INTRODUCTION

Many problems in financial engineering require numerical evaluation of an integral. Several virtues make simulation popular among practitioners as a methodology for these computations.

First, it is easy to apply to many problems. For most derivative securities and financial models, even those that are complicated or high-dimensional, it takes relatively little work to create a simulation algorithm for pricing the derivative under the model. (A notable exception, American options, occupies Section 7.) Also, pitfalls in numerical implementation of simulation algorithms are relatively rare. For the most part, a little knowledge and effort go a long way in financial simulations; with some expertise and investment of one's time, one can go further and faster.

The second virtue of simulation is its good performance on high-dimensional problems: the rate of convergence of a Monte Carlo estimate does not depend on the dimension of the problem. While other numerical integration techniques may have advantages over simulation in various situations, their rates of convergence tend to degrade as the dimension increases. The dimension of the problem is high, for instance, when dealing with models of markets that contain many fundamental sources of risk or with derivative securities that depend in a nontrivial way on prices at many times. This issue is becoming increasingly important as securities markets and financial risk management become more sophisticated.

A third attraction of simulation is the confidence interval that it provides for the Monte Carlo estimate. This information makes possible an assessment of the quality of the estimate, and of how much more computational effort might be needed in order to produce an estimate of acceptable quality.

For these reasons, simulation is a valuable tool for pricing options, as Boyle (1977) pointed out. Twenty years later, Boyle, Broadie and Glasserman (1997) surveyed this field and described research advances that had improved efficiency and broadened the domain of problems to which simulation could be profitably applied. The present paper touches on such advances in order to describe the techniques presently available to financial engineers using simulation and the challenges still confronting them, without offering a comprehensive survey of the field.

The paper continues by explaining in Section 2 the theory that underpins the use of simulation to handle financial engineering problems, and discussing in Section 3 the mechanics of generating simulated paths for this purpose. Then Section 4 deals with variance reduction, providing a philosophical perspective and examples of specific techniques and derivative securities to which they are well suited. Section 5 is a brief discussion of quasi-Monte Carlo methods. Next comes a presentation of advances that have extended the range of effective application of simulation: in Section 6, approaches to estimation of Greeks, and in Section 7 recent research into simulating American options; explanations of the technical terms "Greek" and "American" appear in those sections. The paper concludes with some thoughts about the future interplay of simulation research, and financial engineering theory and practice.

## 2 FINANCIAL BACKGROUND

Financial engineers most frequently apply simulation to derivative securities, often called simply derivatives. These are financial instruments whose payoffs derive from the values of other underlying financial variables, such as prices or interest rates. The canonical example is the European call option, whose 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. This option gives its owner the right to buy the stock at time $T$ for the strike price $K$: if $S_T > K$, the owner will exercise this right, and if not, the option expires worthless. If the future payoff of a derivative derives from the underlying, is there a way to derive the present price of the derivative from the current value of the underlying?

Under some theoretical conditions on the payoff of the derivative, the model of the stochastic process governing the underlying, and the possibilities for trading in the market, the answer is yes. If it is possible to replicate the derivative's payoff by trading in a portfolio of securities available on the open market, then the combination of executing this trading strategy and selling the derivative has no risk. This is known as hedging the sale of the derivative, and hedging strategies are of great practical interest in their own right, as well as being of theoretical interest in justifying no-arbitrage pricing. The pricing theory has this name because it postulates that there are no arbitrages, which are opportunities to make a positive amount of money with zero risk or cost. Such opportunities are supposed to disappear, should they exist, because unlimited demand for them would drive their costs above zero.

The riskless combination of a derivative minus the initial portfolio of its replicating strategy must have nonpositive cost to avoid arbitrage; assuming the same of the opposite combination, the price of the derivative must equal the cost of its initial replicating portfolio. A basic theorem of mathematical finance states that this price is the expectation of the derivative's discounted payoff under an equivalent martingale measure. This is a probability measure under which discounted asset prices are martingales, and it generally does not coincide with the original probability measure which models the real world. When discounting is done with the value of a riskless money market account, the equivalent martingale measure is known as the risk-neutral measure, because if investors had a neutral attitude toward risk, they would demand the same return on all risky assets as on a riskless asset. There are many textbook accounts of this theory, such as Björk (1998) and Duffie (1996).

Given all this, pricing a derivative is evaluating the expectation of the sum of all its discounted payoffs, under a specified measure. The discounting is crucial and allows for appropriate comparisons between cashflows, whether positive or negative, at different times. However, for brevity, henceforth "payoff" may be an abbreviation of "the sum of all discounted payoffs." Since the probability measures of financial models typically have densities, derivative pricing is evaluating the integral of the product of payoff and probability density over all possible paths of the underlying.

As an example, consider the European call option under the Black-Scholes model, for which the distribution of the log stock price $\ln S_T$ is normal with mean $\ln S_0 + (\mu - \sigma^2/2)T$ and variance $\sigma^2 T$ under a probability measure $\mathbf{P}$. Here $S_0$ is the initial stock price and $\mu$ and $\sigma$ are called respectively the drift and volatility. Under the risk-neutral measure $\mathbf{Q}$, $\ln S_T$ is normal with mean $\ln S_0 + (r - \sigma^2/2)T$ and the same variance, where $r$ is the instantaneous interest rate on a riskless money market account. The no-arbitrage price of the European call option is

$$\begin{aligned} & \mathbf{E}^{\mathbf{Q}}[e^{-rT}\max\{S_T - K, 0\}] \\ = \;& e^{-rT}\int_K^\infty (s-K)\phi\left(\frac{\ln(s/S_0) - (r-\sigma^2/2)T}{\sigma\sqrt{T}}\right) ds \\ = \;& S_0\Phi(d_1) - Ke^{-rT}\Phi(d_2) \end{aligned}$$

where

$$d_1 = \frac{\ln(S_0/K) + (r+\sigma^2/2)T}{\sigma\sqrt{T}}, \quad d_2 = d_1 - \sigma\sqrt{T}.$$

and $\Phi$ and $\phi$ are respectively the cumulative distribution and probability density functions of the standard normal. This is the famous Black-Scholes formula.

The standard Monte Carlo approach to evaluating such expectations is to simulate under the equivalent martingale measure a state vector which depends on the underlying variables, then evaluate the sample average of the derivative's payoff over all trials. This is an unbiased estimate of the derivative's price, and when the number of trials $n$ is large, the Central Limit Theorem provides a confidence interval for the estimate, based on the sample variance of the discounted payoff. The standard error is then proportional to $1/\sqrt{n}$.

The Monte Carlo approach is similar for other financial engineering problems, such as finding hedging strategies and analyzing portfolio return distributions in order to assess the risk of one's current portfolio or select a portfolio with the most attractive combination of rewards and risks. All of these rely on the same basic approach of simulating many trials, each of which is a path of underlying financial variables over a period of time, computing the values of derivatives on this path, and looking at the distribution of these values. The next section covers the generation of these paths.

## 3 PATH GENERATION

In some applications of simulation, there is no great conceptual difficulty involved in generating simulated paths, other than that of producing pseudo-random numbers with a digital computer. For instance, when estimating the steady-state mean of a random variable in a queuing system, the model specifies the transition rates from any state, and it is not theoretically difficult to sample the next state from the correct distribution. The situation in financial simulations is not so simple. The models of mathematical finance are usually specified by stochastic differential equations (SDEs) under the equivalent martingale measure used for pricing. Sometimes it is possible to integrate these SDEs and get a tractable expression for the state vector, but not always.

An example that poses no difficulties is the Black-Scholes model, which has

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

where $W$ is a Wiener process (Brownian motion) under the risk-neutral probability measure $\mathbf{Q}$. By Itô's lemma, a basic result of stochastic calculus, this is equivalent to

$$d\ln S_t = (r - \sigma^2/2)dt + \sigma\,dW_t$$

which integrates to

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

Because $W_t$ is normally distributed with mean 0 and variance $t$, the terminal log stock price $\ln S_T$ has the distribution stated previously.

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 $i$th 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 $i$th 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 $(S_{t_1}, S_{t_2}, \ldots, S_{t_m})$ 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-out option that pays off only if a stock price stays above some barrier, or equivalently, if the minimum stock price is above 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, $\alpha$ is the strength of mean reversion, and $\sigma$ 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.

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 $\mu$ and volatility $\sigma$ 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 $\omega$ 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 $\omega$ 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 $\gamma$ 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 $\mu$, $\sigma$, 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 $\gamma$, 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

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

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 $C^{-1/2}$ of decrease of the variance of a simulation unbiased by discretization errror. A higher order of convergence $\gamma$ 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 extremely important.

A large financial institution may have positions in thousands of derivative securities, involving hundreds of underlying variables. In order to manage its risks, it must assess the distribution of possible losses on its portfolio over some time horizon. One way is to compute, for instance, the one-day 5% value at risk (VaR), which is the amount $L$ such that the probability of having a loss larger than $L$ tomorrow is 5%. Despite undesirable theoretical properties, VaR is very popular and the adequacy of its computation is a matter of concern for world financial authorities. A sound way to compute this VaR would be to simulate many scenarios for tomorrow's value of the underlying variables, price all of the derivatives in each scenario, and find the level $L$ such that 5% of the scenarios have a loss larger than $L$. The difficulty is that simulation is required to price many of the derivatives, and one might need to generate, for each of one thousand scenarios, ten thousand paths of one hundred time steps and one hundred state variables, for a total of one hundred billion primitive simulation operations. Despite advances in computing technology, this is not yet affordable, and consequently financial institutions rely on methodologies of questionable soundness for computing VaR. Variance reduction makes better answers affordable.

### 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 $i$th of which has mixing probability $p_i$, mean $\mu_i$, and variance $\sigma_i^2$. 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 $\mu_\eta$, where $\eta$ 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 $k$th 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 $\mu$. 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 $g$ is. The second integral has an interpretation as simulation of paths under a new probability measure $\tilde{\mathbf{Q}}$ which is absolutely continuous with respect to the original measure $\mathbf{Q}$. 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 $g/\tilde{g}$.

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 $\mu$ is the constant of integration that makes $\tilde{g}$ a probability density. Then $\tilde{f} = \mu$ and has no variance. However, this constant $\mu$ 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.

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 (2000).

### 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 $\beta$ 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 $\beta$ is presumably unknown if $\mathbf{E}[X]$ is unknown, but the usual least squares estimate will suffice. However, using the same paths to estimate $\beta$ and evaluate the control variates estimator creates a slight bias. An alternative is to estimate $\beta$ on a small subset of the paths.

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 3. 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 Summary

The methods discussed above do not exhaust the financial engineer's repertory of variance reduction techniques, but they do 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 it is wholly 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 simulation. A quasi-Monte Carlo 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 $(\log n)^d/n$. As this result suggests, quasi-Monte Carlo methods are sometimes much more effective than Monte Carlo. Perhaps because financial instruments usually have payoff functions that are close to smooth, financial engineering is a domain that is quite favorable for quasi-Monte Carlo. Lemieux and L'Ecuyer (2001) give an overview of quasi-Monte Carlo methods for financial computations.

Here it suffices to mention along with the rewards some difficulties that beset the use of low-discrepancy sequences. The superiority of the rate of convergence to that of Monte

Carlo does not guarantee that the low-discrepancy sequence will outperform at a reasonable fixed sample size $n$. Although theory specifies this favorable rate of convergence of error bounds, in practice it is not easy to compute useful error bounds in the first place. As there is no confidence interval available, it is not simple to tell when the quality of the estimate is adequate. There is also a potential pitfall: it is possible for the sample size $n$ to be too small relative to the dimension $d$. The regularity of popular low-discrepancy sequences is such that, while the points formed from the first two coordinates $(x_1, x_2)$ may cover the unit square evenly, the points $(x_{d-1}, x_d)$ cover it very badly, with a distribution nowhere near uniform. Consequently, more care is required when using quasi-Monte Carlo than Monte Carlo.

## 6 GREEKS

Within the theoretical framework of Section 2, the no-arbitrage price $V$ of a derivative security, or a portfolio thereof, is a function of the initial value and parameters $\psi$ of the stochastic process that models the underlying financial variables: $V = V(\psi)$. The derivatives (in the sense of differential calculus) of the price with respect to initial values and parameters are called Greeks because capital Greek letters symbolize several of the most common. For an accessible introduction, see Hull (1999).

The Greeks are important in quantifying and reducing risk. Financial institutions that sell derivative securities usually hedge these sales, often by adding securities to an existing portfolio in order to reduce its Greeks. A lesser sensitivity to changes in the environment is supposed to lead to less risk of significant loss.

Having simulated an estimate $\hat{V}(\psi)$ of $V(\psi)$, how can one estimate a derivative of the form $(\partial V/\partial\psi_1)(\psi)$? For simplicity, write the price as $V(\psi_1)$, suppressing every other component of $\psi$. An obvious way is to simulate another estimate $\hat{V}(\psi_1+\epsilon)$ of the portfolio value using a slightly different value of the parameter. Then $(V(\psi_1+\epsilon)-V(\psi_1))/\epsilon$ is the forward finite-difference approximation to the derivative evaluated at $\psi$, and $(\hat{V}(\psi_1+\epsilon)-\hat{V}(\psi_1))/\epsilon$ is an estimate of it. Somewhat better is $(\hat{V}(\psi_1+\epsilon)-\hat{V}(\psi_1-\epsilon))/(2\epsilon)$, based on the central finite-difference approximation, but this requires three rather than two simulations to estimate the price and derivative.

These estimates have biases directly related to $\epsilon$, because a finite-difference approximation is not the same as a derivative. Their variances are inversely related to $\epsilon$ because they involve division by $\epsilon$. Thus there is an optimal $\epsilon$ for the sample size $n$. Even using the optimal $\epsilon$, these finite-difference estimates perform very poorly in that, for typical problems, their root mean squared errors converge to zero at the rates $n^{-1/4}$ and $n^{-1/3}$ respectively, more slowly than the usual Monte Carlo rate of $n^{-1/2}$. Using the same random numbers in the simulations with $\psi_1$ and $\psi_1+\epsilon$ can help a great deal by making $\hat{V}(\psi_1)$ and $\hat{V}(\psi_1+\epsilon)$ positively correlated, thus reducing the variance of their difference. Even then, finite-difference estimates are still poor for the Greeks of securities such as barrier options because of their discontinuous payoffs.

Frequently, better methods are applicable. Broadie and Glasserman (1996) describe methods based on differentiating inside the expectation in the risk-neutral pricing equation

$$V(\psi) = \int f(x;\psi)g(x;\psi)dx$$

where $f(x;\psi)$ is the payoff on path $x$ and $g(x;\psi)$ is its likelihood. The freedom one has in factoring the product $fg$ is important here.

For example, for the European call option in the Black-Scholes model, the parameter vector is $\psi = (S_0, \sigma, r, T)$. One may write

$$\begin{aligned} f(x;\psi) &= e^{-rT}\max\left\{S_0 e^{(r-\sigma^2/2)T+\sigma\sqrt{T}x} - K, 0\right\} \\ g(x;\psi) &= \phi(x) \end{aligned} \tag{1}$$

so that the payoff is a function of a standard normal random variable $X$ whose density $g$ has no dependence on the parameters. Equally well one could write

$$\begin{aligned} f(x;\psi) &= e^{-rT}\max\{x-K, 0\} \\ g(x;\psi) &= \phi\left(\frac{\ln(x/S_0)-(r-\sigma^2/2)T}{\sigma\sqrt{T}}\right) \end{aligned} \tag{2}$$

so that the parameters $S_0$ and $\sigma$ appear only in the density $g$ of $S_T$, the terminal stock price.

Using the expressions (2),

$$\frac{\partial(fg)}{\partial\sigma} = f\frac{\partial g}{\partial\sigma} = f\frac{\partial(\ln g)}{\partial\sigma}g.$$

The derivative

$$\begin{aligned} \frac{\partial V}{\partial\sigma}(\psi) &= \frac{\partial}{\partial\sigma}\int f(x;\psi)g(x;\psi)dx \\ &= \int f(x;\psi)\frac{\partial(\ln g)}{\partial\sigma}(x;\psi)g(x;\psi)dx \end{aligned}$$

because the log likelihood $\ln g$ is sufficiently smooth that it is permissible to change the order of differentiation with respect to $\sigma$ and integration with respect to $x$. The result is an expectation that simulation can estimate directly. This is called the likelihood ratio method of estimating the Greek.

Using instead the expressions (1), and writing $S_T = S_0 \exp\left(\left(r - \sigma^2/2\right) T + \sigma\sqrt{T} X\right)$,

$$\frac{\partial(fg)}{\partial\sigma} = \frac{\partial f}{\partial\sigma} g = e^{-rT}\mathbf{1}\{S_T > K\}\frac{\partial S_T}{\partial\sigma} g$$

where $\mathbf{1}\{S_T > K\}$ is the indicator function for the event that $S_T > K$. The payoff is actually not differentiable at $S_T = K$, and it is now more difficult to justify the interchange of differentiation and integration, but the result is similar: a simulation of $X$ according to the same density $g$ with $e^{-rT}\mathbf{1}\{S_T > K\}\partial S_T/\partial\sigma$ in place of the payoff gives an unbiased estimate of the derivative. This is known as the pathwise method.

These two estimators require some analytical work in performing the differentiation and checking the conditions that allow the exchange of differentiation and integration, ensuring unbiasedness in estimating the Greek. These issues are more complicated in the case of second derivatives. Still, these estimators have the great advantage that with them, a single simulation estimates the price and all desired Greeks, whereas finite difference approximations require at least one additional simulation per Greek. Both methods are faster than finite difference approximations, and the pathwise method is generally superior to the likelihood ratio method when both apply.

## 7 AMERICAN OPTIONS

An American option has the feature that the owner may decide to exercise it at any time up to a maturity date $T$, unlike a European option, which the owner may exercise only at $T$, not before. Many financial options are American, and the analysis of business investment opportunities as real options has the same feature, making this an important topic.

Whereas the risk-neutral price of a European-style security with payoff $f$ is $\mathbf{E}^{\mathbf{Q}}[f(S_T)]$, for an American-style security it is

$$\max_{\tau \leq T} \mathbf{E}^{\mathbf{Q}}[f_\tau(S_\tau)] \tag{3}$$

where $\tau$ is a stopping time that does not exceed $T$. The nominal payoff usually does not depend on time explicitly, but the discounted payoff does depend on time, requiring the notation $f_\tau$.

In this context, a stopping time is a possible policy for making the decision to exercise: the decision whether or not to exercise at time $t$ can depend on the past up to $t$, but not the future. The stopping time $\tau^*$ that attains the maximum is the optimal exercise policy, so the price is also $\mathbf{E}^{\mathbf{Q}}[f_{\tau^*}(S_{\tau^*})]$.

One minor difficulty that simulation faces in pricing an American-style security is that the optimal exercise may take place in between simulation steps. Simulation more easily prices Bermudan-style securities, for which exercise is possible only at a discrete set of times $t_1, \ldots, t_m$. The fundamental difficulty is in determining the optimal exercise policy. This is necessary for finding the price, and also for the owner to make the correct exercise decision and for the seller to hedge well. It is optimal to exercise when the payoff from doing so now is greater than the continuation value of owning the security if not exercised now, that is, when

$$f_t(S_t) > C_t(S_t) = \max_{t < \tau \leq T} \mathbf{E}^{\mathbf{Q}}[f_\tau(S_\tau) \mid S_t] \tag{4}$$

assuming the state vector process is Markov. To determine whether this is true requires knowledge of a conditional expectation whose value is not available in the simulation.

An obvious attempt at a Monte Carlo estimator is

$$\frac{1}{n}\sum_{i=1}^{n} \max_{k=1,\ldots,m} f_{t_k}\left(S_{t_k}^{(i)}\right)$$

which for each path picks the best time to have exercised, given knowledge of the entire path. This estimator is biased high, because the best time to have exercised is not a stopping time: it depends on the future and thus leads to higher average payouts than are attainable in reality. Much more useful biased estimators are possible.

For instance, Broadie and Glasserman (1997a) produce a low-biased and a high-biased estimator from simulated trees. In these trees, each path has $b$ branches at each of $m$ steps, so branches are conditionally independent given their most recent common ancestor, but are generally dependent.

Using dynamic programming to find the continuation value and exercise decision on these trees still results in a positive bias, but it is inversely related to the branching factor $b$. On the other hand, using only some of the branches to make the exercise decision and the rest of the branches to estimate value produces a negative bias: the stopping time is suboptimal and has no foresight on branches that evolve conditionally independently. The total cost is of order $b^m$ and thus decreasing the bias is expensive, and it is difficult to handle problems with many exercise opportunities. However, there is no trouble with high-dimensional state vectors, and a confidence interval is still available, by creating $n$ independent trees.

Broadie and Glasserman (1997b) also propose a stochastic mesh method which produces a low-biased and a high-biased estimator. This method is designed to handle large problems with a more manageable amount of work. In the stochastic mesh, again each path has $b$ branches, but the total number of nodes at each step is only $b$. The paths are drawn by connecting every node at step $k$ to every node at step $k+1$, requiring $b^2$ connections at each of $m$ steps for a cost of just $mb^2$. The success of this method depends on a good way of choosing the weights associated with these connections. Again, the high-biased estimator comes from applying dynamic programming to the mesh, and this time the low-biased estimator comes from generating entirely new paths and using the suboptimal exercise policy estimated from the mesh. It seems that to be effective, this method requires intensive application of variance reduction. Avramidis and Hyden (1999) do further work on improving stochastic mesh estimators.

Another line of research combines simulation with regression (Carrière 1996, Tsitsiklis and Van Roy 1999, Longstaff and Schwartz 2001). These papers differ in their details; what follows is an algorithm in their spirit. The basic idea is to approximate the continuation value $C_t(S_t)$ in condition (4) by regressing the simulated rewards to continuation on the state vector $S_t$.

Working backward through the possible exercise dates $t_m, \ldots, t_1$, the algorithm creates an estimated continuation value function $\hat{C}_t$ and an estimated value function $\hat{V}_t$. At the last step, $\hat{V}_{t_m}(S_{t_m}) = f_{t_m}(S_{t_m})$. At step $k$ on path $i$, the simulated reward to continuation from state $S^{(i)}_{t_k}$ is $\hat{V}_{t_{k+1}}(S^{(i)}_{t_{k+1}})$. Regression produces the estimated continuation value function $\hat{C}_{t_k}$ fit to these rewards, and then $\hat{V}_{t_k} = \max\{f_t, \hat{C}_t\}$. This approach has had success in practice because most American-style securities have a continuation value that is easy to approximate well by regression on the state vector.

Tsitsiklis and Van Roy (2000) and Clément, Lamberton, and Protter (2001) prove convergence results for such regression-based methods. They are often much faster to arrive at an acceptable approximation to the price than the two Broadie-Glasserman methods, at least when the dimension of the state vector is low, but do not provide a confidence interval with a guaranteed minimum probability of containing the price.

However, using one set of paths to produce a suboptimal stopping policy and a separate set of paths to estimate the price using this policy will result in an estimator biased low. Haugh and Kogan (2001) and Rogers (2001) offer methods of producing an estimator biased high by considering the dual of the American optimal stopping problem (3). Rogers shows that this dual approach is related to the American option seller's hedging strategy and does not depend on finding the American option buyer's exercise policy and the related low-biased estimator. Andersen and Broadie (2001) describe a primal-dual simulation algorithm that is practical for solving this important class of problems.

## 8 CONCLUSIONS

The application of simulation in financial engineering has been a great success story and occasioned much fruitful cross-pollination. Most evidently, financial simulations draw strength from financial theory. One often has theoretical knowledge that makes simulation a more effective tool, because most financial problems are close to an analytically tractable problem, or have analytically tractable elements. This is the key to successful variance reduction and the invention of methods that extend simulation's applicability to new types of problems.

Also, as financial engineering becomes increasingly important in the global economy, and the computational power needed to solve more problems by simulation becomes increasingly affordable, more researchers investigate simulation methods designed for financial problems. As these problems are typically members of some class of similarly structured problems from many domains, such research arrives at methods of general applicability. In financial engineering, as anywhere, the ideal simulation algorithm takes advantage of all available knowledge of the problem's structure to deploy computational resources as effectively as possible in reducing variance and any bias that might be present.

Finally, one interpretation of present events is that the success of simulation in financial engineering is having an impact on financial theory. Mathematical finance is unsettled because its models do not describe financial processes very well at all. Older models strove for simplicity and analytical tractability at the expense of caricaturing reality and fitting data poorly. Newer models tend to sacrifice simplicity in exchange for capturing features of reality that had been unaccounted for in the past: for instance, jumps, stochastic volatility, heavy tails, and transaction costs. The continuing success of simulation allows financial engineers to adopt methods that do not yield analytical solutions and are computationally expensive, but are more successful in describing and controlling financial risks. The result of better engineering should be more efficient markets and fewer disasters.

## ACKNOWLEDGMENTS

This paper owes a debt not only to cited sources, especially the survey of Boyle, Broadie, and Glasserman (1997), but also to unpublished lecture notes from Paul Glasserman's 1998 course on Monte Carlo methods for option pricing at Columbia Business School. The author thanks Peter Carr and Paul Glasserman for providing references.

## REFERENCES

Andersen, L., and M. Broadie. 2001. A primal-dual simulation algorithm for pricing multi-dimensional American options. Working paper, Graduate School of Business, Columbia University, New York.

Avramidis, A. N., and P. Hyden. 1999. Efficiency improvements for pricing American options with a stochastic mesh. In *Proceedings of the 1999 Winter Simulation Conference*, ed. P. A. Farrington, H. B. Nembhard, D. T.

Sturrock, and G. W. Evans, 344–350. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers. Available online via <http://www.informs-cs.org/wsc99papers/048.PDF>.

Björk, T. 1998. *Arbitrage Theory in Continuous Time.* New York: Oxford University Press.

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.

Broadie, M., and P. Glasserman. 1996. Estimating security price derivatives using simulation. *Management Science* 42: 269–825.

Broadie, M., and P. Glasserman. 1997a. Pricing American-style securities using simulation. *Journal of Economic Dynamics and Control* 21: 1323–1352.

Broadie, M., and P. Glasserman. 1997b. A stochastic mesh method for pricing high-dimensional American options. Working paper, Graduate School of Business, Columbia University, New York.

Carrière, J. F. 1996. Valuation of the early-exercise price for options using simulations and nonparametric regression. *Insurance Mathematics and Economics* 19: 19–30.

Clément, E., D. Lamberton, and P. Protter. 2001. An analysis of the Longstaff-Schwartz algorithm for American option pricing. Working paper, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, New York.

Duffie, D. 1996. *Dynamic Asset Pricing Theory.* 2nd ed. Princeton, New Jersey: Princeton University Press.

Duffie, D., and P. Glynn. 1995. Efficient Monte Carlo simulation of security prices. *Annals of Applied Probability* 5: 897–905.

Glasserman, P., P. Heidelberger, and P. Shahabuddin. 2000. Portfolio value-at-risk with heavy-tailed risk factors. Available online via <http://www.paulglasserman.com>.

Haugh, M., and L. Kogan. 2001. Pricing American options: a duality approach. Working paper, The Wharton School, University of Pennsylvania, Philadelphia.

Hull, J. C. 1999. *Options, Futures, and Other Derivatives.* 4th ed. Upper Saddle River, New Jersey: Prentice-Hall.

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.

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>.

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.

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

Rogers, L. C. G. 2001. Monte Carlo valuation of American options. Available online via <http://www.bath.ac.uk/~maslcgr/papers.html>.

Tsitsiklis, J. N., 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. N., and B. Van Roy. 2000. Regression methods for pricing complex American-style options. Forthcoming, *IEEE Transactions on Neural Networks.*

## AUTHOR BIOGRAPHY

**JEREMY STAUM** is a Visiting Assistant Professor in the School of Operations Research and Industrial Engineering at Cornell University. He received his Ph.D. from Columbia University in 2001. His research interests include variance reduction techniques and financial engineering.